Author/s:
Hung, Kuo-Chih

Title:
Methods for Volunteered Geographic Information Quality Assessment in Disaster Response

Date:
2019

Persistent Link:
https://hdl.handle.net/11343/234356

Methods for Volunteered Geographic Information Quality Assessment in Disaster Response

By

Kuo-Chih Hung

This thesis submitted to the University of Melbourne in fulfilment of the degree of Doctor of Philosophy

Centre for Disaster Management and Public Safety,
Department of Infrastructure Engineering, School of Engineering,
The University of Melbourne, Victoria, Australia

February 2020

THIS PAGE INTENTIONALLY LEFT BLANK

For my parents, my family, and my sister Ling-yu in heaven

Abstract

Disaster response is the most crucial phase in the cycle of disaster management. With increased information, disaster management stakeholders are possible to reduce community harm and save lives. Currently, the spatial information used in the response phase is collected from two sources, authoritative channels and Volunteered Geographic Information. Although the authoritative information is more reliable, it might be outdated or limitedly available during the response. By contrast, Volunteered Geographic Information (VGI), which is a direct geographic contribution by volunteers or affected residents, has the potential to provide the complementary information to the authoritative information. In this research, the author focused on text-based VGI that contains textual and spatial contents. Due to the potential for a large number of contributors, text-based VGI can be updated more quickly than authoritative sources and then provide valuable information for decision making.

However, during a disaster, particularly an extreme event, the use of text-based VGI leads to both benefits and challenges. The volume of a text-based VGI collection is often huge. The stakeholders often lack scalability to identify high-quality content such as credible and informative information pieces. Text-based VGI is still underutilized in disaster response. Most of the existing efforts lack efficient methods for text-based VGI credibility and text content quality assessment. Specifically, the use of geographic characteristics of VGI in the process of quality assessment is rarely considered.

To provide credible and informative content for disaster response efficiently, this thesis investigated a modelling approach for VGI quality assessment in a case study. The quality problem was treated as a classification problem. The VGI instances generated during two extreme flooding events in Brisbane were collected. Innovative data filtering processes and quality annotation processes were developed. Particularly, the correlation between quality and geographic characteristics of VGI instances was analysed and validated.

Based on the annotated data, Supervised learning techniques were further employed for the development and evaluation of the quality assessment methods; the geographic method used the correlated geographic characteristics as the predictors, and the integrated method tested a combination of the geographic predictors and text-related predictors. The established approach that considered text contents only was used as a baseline method for the comparison. The developed experiments showed that the usage of geographic predictors can support the classification of high-quality VGI content more efficiently. The geographic method had an excellent performance in credibility assessment but was limited in text content quality assessment. On the other hand, the integrated method can be used in both credibility and text content quality assessment; models based on the integrated method can have better performance in comparison with the baseline method. The knowledge gained suggests great potential for using the developed methods to harvest VGI for the information needs of disaster response.

Declaration

This is to certify that:

i. the thesis comprises only my original work towards PhD;
ii. due acknowledgement has been made in the text to all other material used;
iii. the thesis is fewer than 100,000 words in length, exclusive of references, tables;
iv. parts of this work were published in scientific journals, books, and professional magazines.

Kuo-Chih Hung

February 2020

Acknowledgements

This thesis could never have been completed without the generous support from many individuals and institutions. For these years at the University of Melbourne, I am grateful for the patience, support, and encouragement from all of my supervisors, advisory committee, examiners, and colleagues.

First of all, I wish to express my sincere appreciation to my supervisors, Associate Professor Mohsen Kalantari and Professor Abbas Rajabifard. They have convincingly guided and encouraged me to be professional and do the right thing even when the road got tough. Professor Mohsen Kalantari particularly put tremendous effort in reading every draft of this thesis and discussing my developing ideas with me throughout my writing stage. Many useful comments of my Professor Abbas Rajabifard helped to shape the final version of this thesis. Without their persistent help, the goal of this research project would not have been realized. Furthermore, I would also like to deeply extend my appreciation to my research advisory committee and the chair of examiners: Professor Stephan Winter, Professor Tuan Ngo, and Professor Nelson Lam. They were always available for me when I needed help.

The physical and technical support of the Department of Infrastructure Engineering is truly appreciated. Without the support and funding, this project could not have achieved. Besides, during the development of this research project, I received great help from many people in the Centre for Spatial Data Infrastructure and Land Administration (CSDILA) and the Centre for Disaster Management and Public Safety (CDMPS). I would also like to express my deepest appreciation to the industry advisors: David Williams, Ged Griffin, and Geoff Spring for their important insights about my research topic and research design. With their support, I was able to conduct several difficult tasks. I would also like to thank my colleagues, including Greg Ireton, Dr. Soheil Sabri, Dr. Benny (Yiqun) Chen, Dr. Sam Amirebrahimi, Dr. Davood Shojaei, Dr. Hamed Olfat, Dr. Katie Potts, Dr. Muyiwa Agunbiade, Dr. Farhad Laylavi, Dr. Behnam Atazadeh, Dr. Hosna Tashakkori, Dr. Farzad

Alamdara, Dr. Chen Chen, Dr. Shima Rahmatizadeh, Dr. Arash Kaviani, Dr. Aleksandrov Mitko, Dr. Milad Haghani, Dr. Zahra Shahhoseini, Alireza Kashian, Yan Li, and Jing Ren. Their generosity and hospitality made my days in the office very enjoyable and made this project possible.

Finally, I acknowledge the generous support from my parents, my family, and my sisters. This thesis is simply impossible without them. A special thanks to my current working organization, National Center for High-Performance Computing (Taiwan) for providing me with financial support throughout the last stage of this research.

Table of Contents

List of Figures

List of Tables

List of Publications during the PhD program

Journal articles

- Hung, K.-C., Kalantari, M., & Rajabifard, A. (2016). Methods for assessing the credibility of volunteered geographic information in flood response: A case study in Brisbane, Australia. *Applied Geography, 68*, 37-47. doi: doi:10.1016/j.apgeog.2016.01.005
- Hung, K.-C., Kalantari, M., & Rajabifard, A. (2017). An integrated method for assessing the text content quality of volunteered geographic information in disaster management. International Journal of Information Systems for Crisis Response and Management (IJISCRAM), 9(2), 1-17.

Peer-reviewed conference article

- Hung, K.-C., Kalantari, M., & Rajabifard, A. (2016). Assessing the quality of building footprints on OpenStreetMap: a case study in Taiwan. Paper presented at the GSDI 15 Conference, Taiwan.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 1

Introduction

1.1 Research background

The critical role of spatial information in decision-making processes has been repeatedly highlighted in different fields, including public administration, land suitability evaluation, and disaster management (Densham, 1991; Pereira & Duckstein, 1993; Rajabifard, Crompvoets, Kalantari, & Kok, 2010). Spatial information has usually been generated by resourceful governments or private companies with sufficient knowledge, technology, and labour for capture, analysis, and digitisation. However, recent developments in online mapping services, GPS-enabled communication technologies, and social media have changed this context. New web tools provide cost-effective means to enable ordinary citizens without professional cartographic skills to generate and disseminate spatial data, termed "Volunteered Geographic Information" (VGI), "Crowdsourced Geographic Information" or "User-Generated Geographic Content" (UGGC) (Elwood, 2008; Goodchild, 2007). VGI initiatives such as OpenStreetMap (OSM), Wikimapia, and Google My Maps are rapidly multiplying. The rise of VGI has attracted considerable interest in many different fields. Among these fields, the use of VGI is addressed in the domain of disaster management because the complementary information from the crowd often contributes to decision-making.

The most time-sensitive aspect of disaster management is the response phase, wherein the details of an event and up-to-date information are necessary to minimize loss and to deploy resources efficiently (Coburn, Spence, & Pomonis, 1992; Cova, 1999; Murphy et al., 2008; Tadokoro, 2009). In many cases, such a short time period has posed a variety of challenges in information availability (Goodchild & Glennon, 2010). Currently, most required disaster information in chaotic situations of response is collected from authorities (e.g. in-situ sensors or emergency responders). This authoritative approach is limited and sometimes information is not complete or outdated during the response. By contrast, VGI can be timely and contain crucial information. VGI has the potential to assist disaster agencies such as information managers to make informed decisions. Taking the response to Hurricane Katrina in 2005 as an example, people were faced with government failure and therefore bypassed official agencies to render assistance spontaneously.

Technology initiatives such as mapping platforms and forums were used to collect disaster-related information by the affected crowd and volunteers (e.g. Katrina PeopleFinder Project, Google Maps mashup at Scipionus.com, see: Kawasaki, Berman, & Guan, 2013; Miller, 2006). Text description with location and linked contents were reported to notify the presence and extent of major accidents, incidents, and natural or man-made disasters, (Coleman, Georgiadou, & Labonte, 2009; Starbird, 2011). These VGI initiatives provide the pivotal information in the new disaster management context, particularly for extreme events like the Haiti earthquake and massive forest fires in Russia (Meier, 2012; Zook, Graham, Shelton, & Gorman, 2010).

These VGI initiatives enable a new paradigm of using Geographic Information (GI) for disaster management. From the literature review, a large part of the existing research has predominately focused on VGI containing textual and spatial contents, which is called text-based VGI in this research. Text-based VGI can be extracted from the emergency reporting platforms (e.g. Ushahidi) or social media by the specific techniques. However, as there is a sheer volume of text-based VGI instances available in extreme events, the identification of trustworthy and informative VGI is difficult. A challenge of VGI Quality Assessment (QA) has emerged.

First, the involvement of vandalism or false information is always a threat. For example, it was found that some fake incidents were detected in the Ushahidi-Chile projects (Iacucci, 2010). False information can cause resources misallocation leading to loss of life. Since the access to accurate information is often limited during disaster response, the credibility is hard to be assessed under such information overload. According to the volunteers' experience, this became an important issue in the use of VGI (Ford, 2011); as the verification highly relied on human interpretation, if the credibility of VGI cannot be verified, volunteers were difficult to convince the government officials to endorse the voluntary information. How to improve the credibility assessment process is an important issue.

Second, text-based VGI is often insufficiently structured or documented. Particularly, the text contents extracted from social media are with different intentions and difficult to be assessed. Some text contents denoted by specific keywords or hashtags are on-topic to a disaster event, but not relevant to the actions of disaster management stakeholders. Even the available text contents are relevant, they are often not detailed enough to correspond with the needs (see: Morrow, Mock, Papendieck, & Kocmich, 2011). To enable the coordination of tasks among different stakeholders, the text contents are expected to be informative, which is identified as a "text-content quality" (i.e. informative level) problem in this research. Using a manual classification is often adopted to identify high-quality content but this method is costly and not efficient in the response to extreme events.

This research focuses on the two abovementioned issues, credibility and text content quality of text-based VGI. In the past few years, the problem of analysing text-based VGI in the context of crises has been addressed by a growing body of existing studies. Although various methods have been developed for text-based VGI QA, these methods are often limited to the text contents and metadata. The "geographic aspect" of information is often ignored in these established methods (see: Chapter 3), or the QA methods related to the geographic aspect are not flexible (e.g. Spinsanti & Ostermann, 2013). This leads to incomplete QA methods for the use of text-based VGI.

The hypothesis posted here is a geographic assumption: the credible and informative text-based VGI pieces are correlated with geographic characteristics of VGI, and the correlation can be further used in mathematical models for QA. For example, during a crisis event, a local area with many VGI contributions might be under severe conditions. Therefore, VGI contributions from this area are probably more credible to the disaster responders. Likewise, VGI from disaster-prone areas is more likely to be informative for disaster response in comparison with VGI out of the disaster-prone area.

For the validation of this geographic correlation, this research aims to leverage the existing knowledge related to statistics, spatial analysis, and text classification techniques. Based on the existing knowledge, this research addresses to develop methods to examine the

significant geographic characteristics and justify their variances on the QA process. The developed methods analysed the geographic characteristics as the predictors of text-based VGI QA models (i.e. the geographic method), or use both geographic predictors and the text-related predictors (i.e. the integrated method). The goal is to provide an efficient process for the identification of the high-quality VGI pieces in disaster response. VGI quality problem was treated as a classification problem. The developed method is expected to be a more comprehensive solution, capable of dealing with a large volume of text-based VGI collection.

1.2 Research problem

As outlined above, to efficiently assess the quality and examine the correlation between quality and geographic characteristics of text-based VGI, the statement of the research problem is formulated as follow:

In disaster response, the use of text-based VGI has the potential to assist disaster agencies to make informed decisions. However, in extreme events, text-based VGI is still underutilized due to the often-overwhelming amount of information. Most of the existing efforts lack efficient methods for VGI credibility and text content quality assessment. And the correlation between quality and geographic characteristics of text-based VGI is not validated yet.

To operationalise the concept of credibility and text content quality, this research defines the two information quality criteria as follows. Credibility in this research represents that the information correctly represents the reality (credibility-as-accuracy) or it complies with the perceived truth from the perspective of information receiver (i.e. credibility-as-perception or believability). Text content quality means that the text content meets the information need of conciseness, completeness, clarity, and relevance; the text content provides the possibility of correct interpretation of information and observable phenomena to understand the situation.

1.3 Research aim

In considering the research problem, the aim of this research is:

To develop and evaluate methods based on geographic characteristics of text-based VGI for generating credibility and text content quality labels in the disaster response phase.

This research seeks to develop the text-based VGI QA methods for the response phase with an in-depth understanding of the associated information factors. The developed methods can support the information management process of managers in an emergency response centre. They can collect mass information from local responders, disaster volunteers, and social media, and further filter credible and informative content to provide critical information and distribute aids.

The methods developed in this research aimed to classify VGI credibility and text content quality. By the integration of authoritative data, the geographic characteristics of the collected text-based VGI are tested in the models. For example, the geographic context such as the physical elevation of VGI instances can be used as the predictors. However, if the geographic correlation is too weak to develop a model of high-performance, this research will seek to use another kind of predictors. For example, most studies have shown that linguistic characteristics have proven capability in the text classification problem. The integrated method is therefore further developed and evaluated. Note that this research focused on the quality issue of text-based VGI and limited to the georeferenced VGI instances. Crowdsourced data without geospatial information (e.g. most social media feeds) were not discussed yet they might be still critical for the response tasks.

1.4 Research questions

In accordance with the research problem and aim, the state-of-the-art methods for utilizing VGI/text-based VGI have to be investigated. The research questions are:

1. What are the existing approaches and methods for VGI QA? Particularly, what methods can be used to deal with text-based VGI in disaster management? And what are the advantages and limitations of these methods?
2. What geographic characteristics are correlated to the assessment of credibility and text content quality of text-based VGI in the specific disaster events? How to use these geographic characteristics to develop the QA models?
3. If the geographic correlation is validated in a specific type of disaster and the model performs well, is such a model applicable to a different event in the same area (i.e. a future event)?

1.5 Research objectives

To achieve the aim of the research and answering the research questions, this research sets the geographic context in the response of extreme flooding events (due to the availability of VGI instances in Australia), and the objectives of this research are defined and outlined below:

1. To investigate and understand the use of VGI in the context of disaster management and the associated QA methods, including
    a. the concept of VGI and relevant cases;
    b. the existing QA methods in various forms of VGI;
    c. the advantages and limitations of these methods, particularly for text-based VGI QA.
2. To validate the correlation between the text-based VGI quality and the geographic characteristics of text-based VGI in the extreme flooding events.
    a. design and development of relevant processes for data collection and data preparation in a given context;
    b. identification of the significant characteristics impacting credibility and text content quality of VGI for the validation of geographic correlation;
3. To design, develop, and evaluate the methods for assessing text-based VGI quality in disaster response, including

a. the geographic method and the integrated method based on models of Supervised learning;
b. testing the predictive capability of the methods in a different flooding event.

1.6 Research approach

An interdisciplinary approach based on statistical and data mining techniques is adopted in a case study of the Australian extreme events. The research method combines the state-of-the-art techniques in the fields of Geographic Information Science (GIScience), Information Retrieval (IR), Natural Language Processing (NLP), Machine Learning (ML). These techniques were used to discover the geographic correlation and further used in the QA methods.

The selected methodology used in this research is "Design Science Research Methodology (DSRM)" (Hevner, 2007; Hevner, March, Park, & Ram, 2004). DSRM uses systematic knowledge to solving a problem by constructing an artefact. The constructed artefact in this research refers to the QA methods for providing quality labels of the individual VGI instances. Based on the theory and guideline of DSRM, this research consists of four phases, namely: (1) Conceptual, (2) Design, (3) Development and Evaluation (D&E) phase, and (4) Synthesis and Communication phase. Research objective 1 is conducted in the conceptual phase. Research objective 2 and objective 3 are mainly addressed in D&E phase. These phases are presented schematically as follows:

- Conceptual phase

The conceptual is in correspondence to the "awareness of the problem" phase in Design Science. The main activity is to conduct an exhaustive literature review based on the examination of relevant publications. The practical cases and current status of VGI in disaster management are investigated and the state-of-art of VGI in the context of disaster management are presented. Scientific publications used in the literature review included journal papers, conference proceedings, books, and organization reports. Different forms of VGI and various problems of information needs were identified. Specifically, the methods used for VGI retrieval and QA were reviewed and summarized. The advantages

and limitations of these methods were revealed. The results of the literature review are presented in Chapter 2 and Chapter 3. In addition to the literature review, an investigation to understand the use of text-based VGI in two Australian emergency management stakeholders was made, which is presented in Section 2.3.5.

Based on the findings in this phase, this research concludes that (1) the existing methods of text-based VGI QA have various limitations for the response of extreme events, and (2) the geographic correlation has rarely been discussed and validated in the existing methods. Therefore, the VGI QA methods using geographic characteristics are further designed and developed.

- Design phase

Following the findings in the conceptual phase, the research steps for achieving the research objective 2 and 3 were designed. First, a research flowchart was defined. Relevant statistical tools and techniques used in D&E phase were identified. Next, a plan of collecting the VGI instances for D&E phase was made. As the uncertainty of extracting VGI geographic characteristics might harm the validation of geographic correlation, only the geotagged VGI instances were used. From an initial survey of this research, Text-based VGI instances generated during the Queensland Floods in 2011 and 2013 had more volume than other recent disaster events in Australia as volunteers used Ushahidi to geotag information. Therefore, VGI instances were collected from Ushahidi and Twitter and flood response is a specific scenario to use VGI in this research. Moreover, Brisbane in the State of Queensland was one of the most affected areas in the two flood events. Brisbane is selected as a case study area.

- Development and Evaluation phase (D&E phase)

In this phase, the geographic correlation between the VGI quality and the associated characteristics was validated. Next, by using the corresponding statistical/ML tools, the text-based VGI QA methods were developed by mathematical models in two ways. The geographic method analysed the geographic characteristics of the VGI instances and tested

the geographic correlation in the models; the collected VGI instances were pre-processed, filtered and then annotated into different quality labels. On the other hand, an integrated method to explore the multiple information factors correlated to text-based VGI QA was implemented. Since text-related predictors have demonstrated good performances in the quality classification problem in the literature, this method addresses to test the performance variance by using geographic predictors. The evaluation of the developed methods addressed to compare the performances and understand the predictive capability of the models.

- Synthesis and Communication phase

In the final phase, the problem of this research, solution and its novelty, as well as the results and findings in the research activities are synthesised and presented to other researchers and relevant audience via various communication channels, such as academic journals, conference proceedings, and oral presentations.

1.7 Structure of the thesis

After the introductory chapter, the dissertation is comprised of six research chapters and a final concluding chapter.

Chapter 2, *VGI for disaster management*, provides a review of relevant literature on VGI and crisis informatics. The concept and components of VGI are discussed, and several VGI initiatives for disaster management in recent extreme events are investigated.

Chapter 3, *Quality assessment method of VGI*, a review of existing QA methods is conducted. Limitations and gaps in these methods and are identified. Chapter 2 and Chapter 3 are the background study to address the first research question.

Chapter 4, *Research design and methodology*, outlines the research design and the research methodology to achieve the research aim and objectives. This chapter first describes a conceptual design framework of this research. The concept and the employment of the adopted methodology, DSRM is described. Then, the research process is designed

as five research phases which have been briefly described in Section 1.6. Research activities in each phase were presented in detail in this chapter.

Chapter 5, *Data collection, processing, and the validation of the geographic correlation*, firstly described the data preparation including collecting and filtering VGI instances, cleaning mismatched positioning data, and annotating the credibility and the text content quality labels. Next, through the combination of VGI and associated authoritative geospatial data, VGI instances were enriched in a variety of geographic characteristics. These geographic characteristics are used as predictors for the validation of the geographic correlation and the further D&E of the QA methods. The results of the geographic correlation are presented.

Chapter 6, Development and evaluation of the VGI Quality Assessment methods, firstly describes the used algorithm. Next, two kinds of the QA method were developed and evaluated. The performance of the geographic method is presented and compared. The coefficients of the predictors or the Decision Tree graph are illustrated, and the sensitivity analysis was performed. As for the integrated method, the definition and extraction of the text-related predictors were described. Significant text-related predictors are identified.

Chapter 7, *Conclusion and Recommendations,* concludes the research outcome and the major contribution of this research. This chapter also responds to the research problem and reflects the achievement of addressing the aim and objectives. The potential for the developed QA methods to be used by disaster information managers is discussed. The recommendations for future research directions are proposed.

1.8 Chapter summary

This chapter has comprised the motivation and goal of this research. It presented the background and challenge for using VGI in the context of disaster response, and then it introduced the problem, questions, aim, and objectives of this research. Issues related to VGI quality in the context of disaster management have been highlighted. To solve these issues, development and refinement of VGI QA methods are addressed. Subsequently, the

research approach was briefly described and justified. Finally, this chapter was concluded by delineating the thesis structure.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 2
Volunteered Geographic Information for Disaster Management

2.1 Introduction

This chapter mainly provides a literature review to present the use of VGI in the context of disaster management for addressing the research objective 1-a. Figure 2.1 illustrates the structure of the literature review. Section 2.2 firstly describes the phenomenon of VGI and the underlying components, including the forms and the heterogeneous characteristics of VGI. Then, the reasons why VGI is critical for disaster management and the practical cases are provided in Section 2.3. The VGI collection approaches are extensively reviewed, the characteristics of contributors and their motivations in the literature are presented. The collection and the use of text-based VGI in the Australian emergency management stakeholders are particularly investigated and demonstrated in three cases.

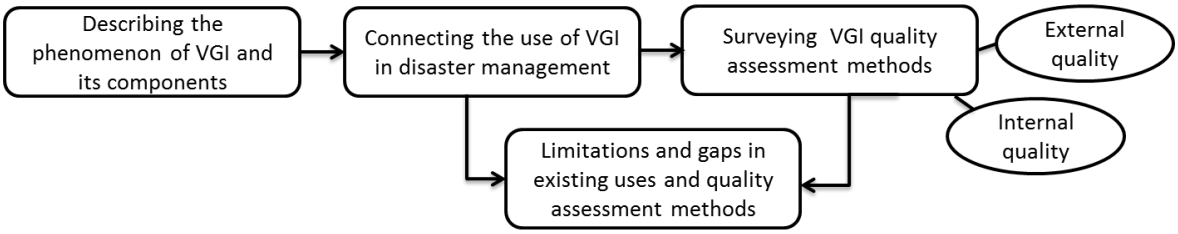


Figure 2.1: Schematic summary of the contents of this chapter

From the literature review, this research found that the VGI quality issue is important and unavoidable. The existing VGI QA methods are further reviewed in Chapter 3.

2.2 Overview of Volunteered Geographic Information

Web 2.0 development provides platforms for flexible communication and sharing timely, relevant, and diverse information. Today, peer production, cloud collaboration, or crowdsourcing are common on the Internet. with and adaptable web-based systems represent an epochal shift in a workflow or a business process of people. Web mapping platforms, GPS devices, and the advance of positioning techniques enable citizens to collect and share GI. Based on these technologies, ten or even hundreds of thousands of people across the globe can collaborate on virtual projects of unimaginable scale and scope

(Graham, 2010). New technological framework with the mass rise of Web 2.0 brings various tools for people to describe or further digitize geographic phenomena.

An important characteristic of these tools is the combination of interactive functions, online data or web services, in order to provide geo-referenced annotations, comments, and observations. Mashup applications (i.e. a single interface with contents from more than one source) such as Google Maps are widely used. These applications facilitate to solve private and business problems (Rajabifard et al., 2010), and create opportunities to collect massive information. Contributors can get rid of the fetters of space and time to provide information at a low cost. A new approach to spatial data collection is formed. GI generated from this approach is more timely and flexible than GI produced by government sectors or commercial mapping companies (Goodchild, 2007). For instance, the satellites have a fixed orbit that the image may not be available at some specific moment; access to authoritative data is often under contradictory licensing restrictions or has a high procurement cost (Antoniou & Skopeliti, 2015). Therefore, this new approach makes human act as sensors to provide complementary information.

Movements in the academia in response to these phenomena have offered a series of varying terminology, with different levels of specificity and connotation, such as User-Generated Geographic Content (UGGC) and Neogeography (Turner, 2006). As such information is usually provided voluntarily by amateur individuals, Goodchild (2007) referred it as Volunteered Geographic Information (VGI) to contrast it with produced and mediated forms of GI and to emphasize the role of information generated through the efforts of volunteers. Over the last decade, VGI has become the most widely used term in the academia particularly in the domain of Geographic Information Science (GIScience). VGI is one research discipline that seeks to identify and leverage the opportunities of regarding citizens as sensors. To provide context for the synthesis of relevant literature particularly in quality issue, this section first reviewed the definition of VGI and a typology of VGI proposed by Craglia, Ostermann, and Spinsanti (2012). Then, the forms of VGI and quality issues are described. Finally, the specific characteristics of VGI are discussed.

2.2.1 Definition and the typology of Volunteered Geographic Information

According to Goodchild (2007), VGI is "*the harnessing of Web tools to create, assemble, and disseminate GI provided voluntarily by individuals."* The voluntary nature of geographic content production should be the specific characterizations of VGI. However, according to a survey of VGI initiatives, it was found that many VGI initiatives where the contributor was involuntary (Elwood, Goodchild, & Sui, 2012). A debate in the terminology was rise. For example, Gorman (2010) criticized that VGI was not an appropriate term as some initiatives paid representatives to organize mapping parties and used authoritative data. Additionally, many studies regarded web-based or social media content as a form of VGI. In portals such as Facebook, Twitter, Flickr, and Youtube, the extraction of VGI was often extended to what the application designed for as worthy geographic content was inferred or added by other methods. Therefore, an ethical line exists and signals important differences in the processes of acquisition and the uses of such content.

Due to this, researchers have tried to make a distinction between various forms of VGI and termed social media content as Contributed GI or Ambient GI (Harvey, 2013; Stefanidis, Crooks, & Radzikowski, 2013). Yet these terms are not widely used in the academic literature. VGI is still a more established term in varying perspectives. Despite these arguments, VGI can be regarded as a subset of User-Generated Content that concerns the characterization of the geographic domain (Elwood et al., 2012).

To provide a comprehensive overview of different modes of VGI initiatives, a typology of VGI was developed (Craglia et al., 2012). This typology has a two-dimensional consideration: the way the information was made available, and the way GI forms part of it (Table 2.1). The typology indicated that if the author contributed a piece of information for a specific purpose in mind voluntarily and publicly available, it is explicitly volunteered. On the other hand, implicitly volunteered information is information publicly available but was not an integral part originally, and is only of secondary concern or derived. For example, social media feeds such as tweets (i.e. Twitter feeds) are implicitly volunteered

information. Likewise, if the information is mainly about a place, it is explicitly geographic. Conversely, if the information is a non-geographic topic but the content has a place to be geocoded, such content is implicitly geographic. Generally, as implicit information requires more processing efforts to evaluate the intention and location of information, it has more uncertainties. Therefore, Craglia et al. (2012) pointed out that most VGI QA methods are done on explicit VGI and a lesser amount of methods are done on implicit VGI, although implicit VGI has more issues regarding its quality. However, this research found that as volunteers can collect implicitly VGI and republish information, the difficulty of distinguishing "explicit" and "implicit" information existed. This typology was not addressed throughout the chapters.

Table 2.1: Typology of VGI (Craglia et al., 2012)

	Explicitly geographic	Implicitly geographic
Explicitly volunteered (active contributions)	A rigorous form of "Volunteered" GI. Examples include OpenStreetMap/Wikimapia.	This can refer to volunteered information about non-geographic topics but contain place names (e.g. Wikipedia articles about non-geographic topics).
Implicitly volunteered (passive contributions)	This can simply refer to UGGC not made public by the author and contributed with a specific purpose. Any public content with the properties of an identifiable place is an example.	This dimension refers to implicitly volunteered UGGC about non-geographic topics but mentioning a place that can be geocoded.

2.2.2 Forms of VGI and quality issues

VGI as a form of spatial data collection, it has limited accuracy by the method of measurement and the device used (Goodchild, 1993). Quality issues exist when a user cannot be sure the information represents the actual nature of the real world (Devillers &

Jeansoulin, 2006; Goodchild, 2009b; Shi, Fisher, & Goodchild, 2002). The factors impacting VGI quality are quite complex. For example, the spatial resolution of the satellite images on OSM is determined by the inherent mapmaking technology and not consistent in every area. This causes uncertain results in the data production. The VGI quality issues require considering all the process of information provision/production.

In different approaches of information provision, VGI can be classified into three heterogeneous forms: (1) map-, (2) image-, and (3) text-based VGI based on the methods that are used to capture the data (Senaratne, Mobasheri, Ali, Capineri, & Haklay, 2016). According to the forms of VGI, various metrics and indicators have been proposed for VGI QA methods (Antoniou & Skopeliti, 2015). Further explanations of these metrics and methods are described in Chapter 3.

Map-based VGI is often used for navigation and mapping specific themes. It concerns two elements, one is the geometries to present on a map such as points, lines, and polygons. Another element is attribute information. The collaboration mapping platform, OpenStreetMap (OSM) is a typical example. As the form of map-based VGI is similar to vector geospatial data, the QA of map-based VGI links to the elements of spatial data quality such as positional accuracy and attribute accuracy (Guptill & Morrison, 1995). Next, image-based VGI is mostly produced implicitly within web portals where contributors take pictures of a geographic object and attach a geo-reference. Quality issues of image-based VGI are highly relevant to positional accuracy, thematic accuracy, and quality of images (e.g. resolutions). Finally, text-based VGI is mostly produced implicitly on social media, such as Twitter or Blogs, where people contribute geographic information by textual descriptions; text-based VGI has usually been assessed in three metrics: relevance, credibility, and text content quality (Craglia et al., 2012; Senaratne et al., 2016).

Note that some VGI instances might consist of multiple forms and can be regarded as hybrid VGI. For example, Text-based VGI such as social media feeds can contain the information of geolocations and images. Moreover, despite the geolocations and attributes of image- and text-based instances are often implicit, such information is potentially

analysed with additional efforts and mapped afterwards. For example, McDougall (2011) collected image-based VGI from web portals and conducted field surveys to identify the peak of the flood from imageries. Image-based instances were transformed to points on a map with flood heights. By comparison with the image of different time periods, the flood heights can be measured without in-situ sensors. Figure 2.2 presents an example of the image-based comparison.

Figure 2.2: A comparison between (a) Image extracted from the Ushahidi platform and (b) Image extracted from Google Street View

2.2.3 Characteristics of VGI

VGI initiatives have a deep link to the cumulative efforts in the GIS/mapping domains. However, compared to similar concepts such as Publication Participation GIS (PPGIS, geodata and knowledge production by local and non-governmental groups), VGI initiatives have specific characteristics. First, VGI initiatives use new tools to collect GI from the crowd, which are increasingly dominated by flexible web-based mapping techniques. These initiatives are predominately self-organized by individuals and Non-Governmental Organizations (NGOs), and governments are seldom involved in (Brown & Kytta, 2014; Dransch, Fohringer, Poser, & Lucas, 2013). Therefore, the data collection is

passive and the data ownership is usually shared under a free/open data license (e.g. OSM). Second, VGI initiatives have heterogeneous regional contexts. VGI might be generated from either urban or rural areas, or either on a broad scale or a local level.

In general, the form and characteristics of VGI caused different research assumptions and concerns. (e.g. legal issue, see: Scassa, 2013). The literature review of this research focused on the use of VGI in the context of disaster management and the quality concern. The correlation between quality and geographic characteristics is one important aspect. For example, what is the relationship between population and the quality of OSM? This question has been asked by Haklay, Basiouka, Antoniou, and Ather (2010), and their case study demonstrated that beyond 15 contributors per square kilometre, the positional accuracy of OSM became very good. Nevertheless, compared to map-based VGI, the relation between text-based VGI quality and geographic characteristics is still not clear and rarely discussed in the literature.

2.3 Using VGI for disaster management

This research focuses on the use of VGI in the context of disaster response and associated quality issues. In this section, the connection between VGI and disaster management is demonstrated. The collection approaches and the associated characteristics of contributors are the essential factors impacting the use and quality of VGI, which were further described.

2.3.1 Role of VGI in disaster management

Disasters take a variety of forms which are routinely divided into natural or human-made. Most of the disasters have a natural origin, including floods, earthquakes, tsunamis, and forest fires. Extreme natural disasters have caused tremendous damage and continue to threaten millions of humans and various infrastructure capabilities each year. On the other hand, man-made or technological disasters such as an explosion, pollution, and so on, are due to the inattention of human or mishandling of dangerous equipment. Note that a specific disaster may spawn extra disaster and cause a complex combination of both natural

and man-made disasters. Examples are food insecurity, epidemics, conflicts and displaced populations. Disaster takes different forms with various duration and intensity scale.

2.3.1.1 Definition of disaster management and disaster response

According to the Queensland Government, disaster management is arrangements about managing the potential adverse effects of an event[1]. Disaster management is " *a process that includes activities before, during and after a hazard event that aim at preventing disasters, reducing their impacts and recovering from their losses* (Poser & Dransch, 2010)". Generally, this process is represented as a cycle consisting of four main phases: Prevention and Mitigation, Preparation, Response, and Recovery. Figure 2.3 shows the breakdown of each phase and concrete activities.

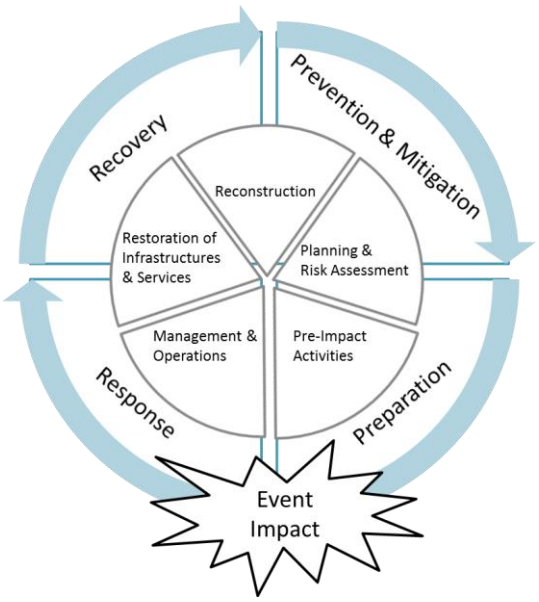


Figure 2.3: Disaster management cycle
(Source: Adapted fromVasilescu, Khan, & Khan, 2008)

[1] https://www.disaster.qld.gov.au/dmg/Glossary/Pages/default.aspx

There are many operational considerations in the disaster management cycle. This research focuses on the information needs of the response phase. The response involves action taken immediately after a disastrous event with the aim of reducing harm, saving lives, or other effects caused by the disaster (Carter, 1991). The response phase starts immediately after the disaster and typically involves a myriad of activities such as search and rescue, evacuations; providing required resources to fulfilling the basic humanitarian needs of the affected population.

2.3.1.2 Why VGI can be critical in disaster response

Access to valid knowledge and information is necessary to accomplish the activities in the response phase in a timely fashion. The required information is expected to be available in disaster agencies through various communication channels such as emergency call service, responders on the ground, or other infrastructures like in-situ sensors or video surveillance. However, in a practical situation, access to real-time data or information is limited. Disaster agencies are hard to confirm the status around stricken regions and face a variety of challenges in collecting GI.

By using information and communication technologies (ICTs) and crowdsourcing, VGI is reciprocal information, and it can be used in emergencies for communication and visualisation of information. Accordingly, VGI has offered important assistance for detecting new events and enhancing situational awareness, particularly during the response phase of extreme events (De Longueville, Annoni, Schade, Ostlaender, & Whitmore, 2010). Note that the use of VGI is not limited in the response phase. In fact, VGI has been used to all the phases of disaster management (Horita, Degrossi, Assis, Zipf, & Albuquerque, 2013).

While the ways to use VGI are increasing with the development of technologies, it is essential to understand the benefits and limitations of using VGI in disaster management. The next section aims to present the benefits and limitations of the different VGI collection approaches.

2.3.2 VGI collection approaches in disaster management

To collect and use VGI in disaster management, generally, there are three prominent approaches: web-based surveys, mapping platforms, and social media (adapted from Dransch et al., 2013). VGI collected in different approaches serving heterogeneous purposes and has various characteristics of benefits and limitations: expected information volume, more or less structured data, low or high data processing effort, and prompt or delayed information provision; and the concern of data reliability and quality focuses on different issues. The relative cases are presented in this section to understand the weakness and strengths of each collection approach.

2.3.2.1 Web-based Surveys

Web-based surveys are online questionnaires. This kind of application is fast and simple to implement for receiving information about a specific topic, which is suited for gathering structured information under conscientious planning. For example, the U.S. Geological Survey's project, "Did You Feel It? (DYFI)" conducted a long-term survey to collect reports about earthquakes from Internet users (Wald, Quitoriano, Worden, Hopper, & Dewey, 2012). Web-based surveys also allow collection at rates and quantities never before considered. Results can be filtered by peer review, and are easy to analyse with statistical methods. For example, in flood management, the survey system can collect numerical information like inundation depth from the crowd to estimate the damage after the flood (Poser & Dransch, 2010).

A disadvantage of web-based surveys is that the systems usually do not allow bi-directional communication. Therefore, information provision might be limited to the predefined questions that information not intended at the moment of questionnaire development will not be collected. Besides, web-based surveys need promotion for attracting contributors. Despite these disadvantages, web-based surveys are very powerful and efficient to collect a large volume of specific data. For example, in the DYFI project mentioned above, questionnaire response rates have ever reached 62,000 per hour (Wald et al., 2012).

2.3.2.2 Mapping platforms

The second approach to collect VGI in disaster management is using the mapping platforms. They have been used for variable purposes. First, high-quality geospatial data can be collected for the need of relief actions. For example, OSM has played a critical role in areas without a digital map of good quality (e.g. the 2010 Haiti earthquake). Second, the mapping platforms are often used for emergency reporting. Volunteers can create web-based applications for communication and enhance situational awareness. With pivotal mapping functions, environmental risk, incident details, and resource requirements can be reported by contributors/affected residents at the exact location without many processing efforts. They can leave text message and media to show the severity and validity of information. Administrators of the platform may collect information from authorities or news media and then share on the platform. These information contributions enhance the understanding of what is happening in existing conditions and how it could change over time. This is essential for saving lives, restoring or maintaining calm, and engendering confidence in the operation of disaster management (Meier, 2012; Reynolds, 2006).

For emergency reporting, Ushahidi is the most representative case. Ushahidi is Swahili for "testimony." It was initially a project developed for mapping reports of violence in Kenya after the post-election in 2008. Within this project, an open-source application "Ushahidi" was developed. As Ushahidi is very easy to deploy and customize, it has been widely used in humanitarian organizations. Figure 2.4 demonstrates the Ushahidi applications for the 2010 Haiti Earthquake.

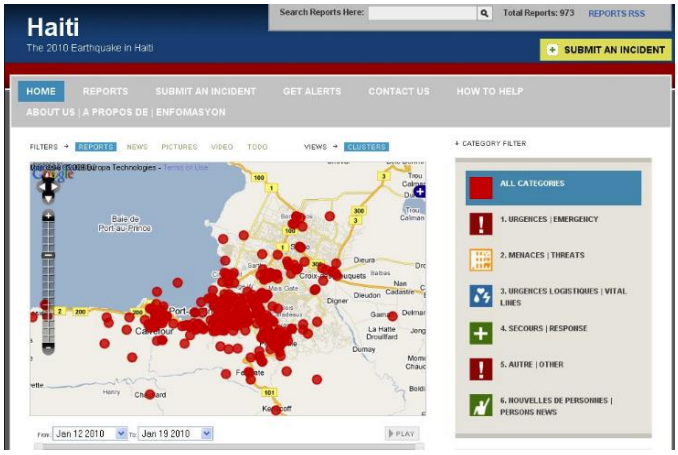


Figure 2.4: Ushahidi-Haiti Map[2]

Although emergency reporting platforms like Ushahidi might be managed by a group of administrators with high skill levels and passions, the quality of collected information is still uncertain. As mentioned in Chapter 1, false reports were identified in the Ushahidi-Chile projects (Iacucci, 2010). And the information provided on mapping platforms was often not detailed enough to correspond with the needs of the stakeholders (Morrow et al., 2011). Besides, according to the author's investigation on the Ushahidi applications for the Queensland floods[3], it was found that the geo-position of a report might mismatch to its location description. Furthermore, similar to web-based surveys, collecting disaster information from the crowd is not easy without the promotion efforts. Therefore, some researchers used social media for collecting VGI as social media have been ubiquitous in people's life.

[2] Source: http://voices.nationalgeographic.com/2012/07/02/crisis-mapping-haiti/screen-shot-2012-06-27-at-8-53-00-am/

[3] In the Australian context of disaster management, the Ushahidi applications have been used in several disaster events. The most important case is the 2011 Queensland Floods. During this event, Australian Broadcasting Corporation (ABC) deployed Ushahidi and collected approximately 1,500 reports

2.3.2.3 Social media

Social media are computer-mediated technologies dedicated to the creation and sharing of information, ideas, forms of expression via virtual communities and networks. According to Kaplan and Haenlein (2010), social media can be categorised into six types: collaborative projects (e.g. Wikipedia), blogs and microblogs, content communities, social networking sites, virtual game worlds, and virtual social worlds. By this definition, mapping platforms can be regarded as a subtype of social media. But in the literature, social media generally refers to the new web portals such as Twitter and Facebook.

Social media sites provide tools and services which allow users to post different forms of content like text, images, or videos and make it available to the public or restricted group. As the growth of the number of users and easy access through mobile devices, social media has played an important role in disaster management, particularly for communication. Textual and media contents are able to be enriched by additional geographic content and contribute to situation awareness. During extreme events such as the 2011 Queensland Floods, social media became connectors for emergency services working with personnel, communities, and volunteers. Emergency service sectors effectively used such technology to improve communicated information and field queries in disaster management (Arklay, 2012; Bird, Ling, & Haynes, 2012).

For assisting disaster management, researchers have tried to use social media for event detection and enhanced situation awareness. However, the implicitly volunteered and geographic contents are the main issues. First, as social media provides a communication channel with multiple functions such as chat and video sharing, the intention of social media feeds is hard to be judged. The use of keywords or "hashtags" is appropriate to detect a trend, but difficult to find specific information. Some keywords can have several meanings (homonyms) depending on the context or be used figuratively. In extreme events, a huge volume of social media feeds with the same hashtags was generated. For example, more than 35,000 tweets containing the #qldfloods hashtag were sent during the period of 2011 Queensland Floods (see: Figure 2.5, Bruns, Burgess, Crawford, & Shaw, 2012). The

retrieval of on-topic, meaningful, or actionable information is more difficult than collecting data from mapping platform or web-based surveys.

Second, most social media data are not geo-referenced. For example, the previous investigation showed that only about 1.5% to 3% of tweets are geotagged; and one percent of all users accounted for 66 percent of georeferenced tweet (Leetaru, Wang, Cao, Padmanabhan, & Shook, 2013; Murdock, 2011). Inferring the location of non-geotagged social media data is a pragmatic challenge to the use of data. A solution is to use toponym recognition, toponym resolution, and geocoding to determine the geographic coordinates of VGI instances. Many researchers have been focused on this issue; Laylavi, Rajabifard, and Kalantari (2016) and Ghahremanlou, Sherchan, and Thom (2014) had contributed to this issue. The assumption behind the geocoding is that the considered VGI instance has a single geographic focus. If the VGI instance has multiple geographic focuses, a hierarchical ontology is suggested (Amitay, Har'El, Sivan, & Soffer, 2004). Note that current geocoding techniques often use a centroid of polyline/polygon or interpolate a position. This approach still has ambiguity. For example, geocoding a long road section in its centroid cannot be regarded as an accurate positioning since the "real described section" might locate either at the beginning or the end of the road, with a distance to the centroid. In some VGI studies, researchers preferred to use geotagged VGI instances (e.g. Albuquerque, Herfort, Brenning, & Zipf, 2015). Moreover, the geographic coordinates of a tweet does not always match the mentioned place as people tweeted something happened in a location might be far away and the GPS coordinate of their devices was recorded.

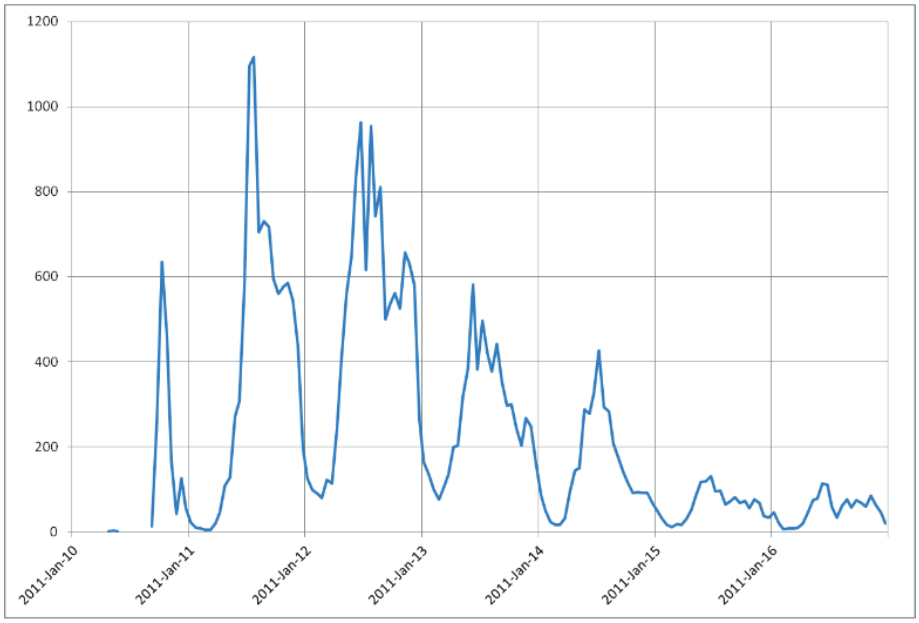


Figure 2.5: #qldflood tweets per hour from 10 Jan. to 16 Jan. 2011 (source: Bruns et al., 2012)

2.3.3 Motivations of contributors and uncertainty of VGI

Since VGI is generally produced by laypeople, it is hard to control the behaviour of contributors. The quality is uncertain. To better understand the linkage between information quality and contributors, researchers have explored motivations of volunteers (Basiouka & Potsiou, 2014; Budhathoki, 2010; Coleman, 2010; Coleman et al., 2009; Gosier, 2011; Haklay & Budhathoki, 2010; Liu, 2014; Starbird, 2011). Through an extensive review of the literature, the motivations can be summarized as (1) intrinsic motivation, such as altruism, personal interest, enhancement of a personal investment, self-expression, and spatial-driven motivation (e.g. pride of place), and (2) extrinsic motivation, such as career opportunity, enhancement of personal reputation/identity, social reward/capital accumulation, monetary return, and improvement of official process. The detailed descriptions of the above motivations are arranged in Table 2.2.

Table 2.2: A summary of the motivations of contributors

Motivation		Description
Intrinsic motivation	Altruism	Contributors seek to benefit others.
	Personal interest	Make a contribution as part of an existing job to showcase professional ability and feel personal satisfaction.
	Enhancement of a personal investment	Contributors seek to learn geospatial technologies/ knowledge to enrich their lives
	Self-expression	Contributors can display knowledge and expertise on mapping and geospatial technologies
	Spatial-driven motivation: pride of place	Adding information about one's own group or community may be good for public relations, tourism, economic development, etc.
Extrinsic motivation	Career opportunity	Crowdsourcing projects provide a means to signal the contributor's knowledge to potential employers.
	Enhancement of personal reputation/identity	Providing the opportunity for contributors to develop identities that are respected, trusted, and valued by a community.
	Social reward/ capital accumulation	Volunteer strengthens his social relation to a community
	Monetary return	Contributors make profit by providing value-added GI products and services and collaboration with required sectors
	Improving official process	Volunteerism can eliminate length of official processing and minimize cost.

While most contributors positively motivate to contribute VGI, it is still noteworthy that vandalism exists and it resulted in biased information. Negative motivations such as mischief, social/political intention, and criminal intent have been identified (Coleman et al., 2009; Ford, 2011). Furthermore, with various levels of contributor's capability or various levels of care taken in processing data, unintentional erroneous information, such as inaccurate measurements, operational mistake, or phenomena described in inconsistent terminology, is inevitable during observation and mapping.

Therefore, in contrast to the enthusiasts of VGI, critics like Keen (2011) concerned that such crowdsourcing has the potential to destroy professionalism and increased difficulties to find high-quality materials. In essence, uncertainty is unavoidable in VGI due to its nature.

2.3.4 The skill level spectrum of contributors

In addition to the aspect of contributors' motivations, Coleman et al. (2009) tried to characterize the behaviours of contributors in a skill level spectrum, within four kinds of VGI initiatives (Table 2.3). They classified contributors as Neophytes, Interested Amateur, Expert Amateur, Expert Professional, and Expert Authority. Within this spectrum, Coleman et al. (2009) addressed that even Neophytes can at least improve the quality of VGI; they can add basic information or identify gaps in map coverage.

Table 2.3: VGI contributors in relevant VGI initiatives (source: Coleman et al., 2009)

	Mapping and Navigation	Social Networks	Civic/ Governmental	Emergency Reporting
Neophyte	Relies on unit to provide directions and follows instructions to add basic point information using the Unit.	Identified gaps in map coverage, familiar with the locale, and has obtained the requisite GPS equipment. Interested in making a first contribution.	Views a GIS map in a town hall meeting around the siting of a power plant in the town	May use cellphone to add basic information detailing location of a potential new wildfire outbreak.
Interested Amateur	Owns a personal system, uses it extensively, has made several contributions. Is aware of both technology strengths & limitations and procedures required to make reliable contributions.	Owns the equipment; familiar with data editing software & processes. Regular contributor of edited map data and may assess other contributions.	Citizen fashions a map to present a counterclaim in a town hall meeting around the siting of a power plant in the town.	May drive from place to place shooting geo-tagged photos showing extent of floodwaters.
Expert Amateur	Familiar with the strengths and weaknesses of multiple systems, has owned more than one. May assess and occasionally amend	Expert with the requisite equipment. Regularly assesses & edits contributions from others. Participates in specification	Individual familiar with conditions in a given neighbourhood and with the operation of the Web-based PPGIS system in use.	Familiar with requirements for data useful to emergency response personnel and may voluntarily travel to sites to provide such information on an 'on-

	Mapping and Navigation	Social Networks	Civic/ Governmental	Emergency Reporting
	the contributions of others.	development & decision-making.		call' basis.
Expert Professional	Mapping or Location-Based Services professional.	Mapping or Location-Based Services professional.	Practising Urban Planner.	Emergency planning and/or response personnel tasked with mapping the position and geographic extent of a given flood or wildfire.
Expert Authority	Specialist consulted by other professionals re: specific problems and/or new developments.		City Planner with extensive knowledge of developments in the area of interest.	Specialist consulted by other professionals re: specific problems and/or new developments.

Note that the skill level requirement for mapping activity is variable depending on the topic of the map (Goodchild, 2009a). Sometimes contributors only need to learn the basic skills. For example, in most Ushahidi applications, contributors are not asked to represent the actual area of a disaster event. They only have to mark a point on the map and describe the information through text contents. In such a case, contributors without skilful capability may have local knowledge to enrich VGI with more attributes. In the context of disaster management, since hazard vulnerability is closely linked to a territory's geographic context, contributors with local knowledge about their physical world and society, or who live in an affected region can make enough sense of situations and provide high-quality information directly.

Although in some specific mapping activity such as cadastral map, mapper has to undertake a lengthy process of training. But the development of useful and well-designed quality assurance methods, tools, and standards will have more significant impacts on the outcome of mapping initiatives (Rahmatizadeh, Rajabifard, & Kalantari, 2016). In summary, the skill level is important but most VGI initiatives are not limited to the skilful contributor. Developing innovative ways of leveraging each other's diverse skills, knowledge, and resources to decrease the uncertainty of VGI is essential in VGI initiatives.

2.3.5 The collection and the use of text-based VGI in the Australian emergency management stakeholders

In addition to the literature review, this research investigated the collection and the use of text-based VGI in three Australian emergency management stakeholders, which were Victoria's State Control Centre, Brisbane City Council, and a volunteer group: Virtual Operations Support Team Victoria (VOST VIC[4]). The investigation found text-based VGI has been used in their information flow as follows:

- Victoria's State Control Centre has begun a process to integrate social media into the structure of command and control for emergency management since Black Saturday (i.e. a series of bushfires burning across Victoria around 7 February 2009). The officers focused on the early identification of threats or emergencies reported on social media and for gaining enhanced situational awareness about known incidents underway. Currently, they have an operational role called "Social Media Intelligence Analyst" and use a manual method to find useful information.
- Based on successful experiences of using VGI in the 2011 Queensland floods, Brisbane City Council built a Ushahidi platform while Cyclone Oswald was struck in 2013. Information from the crowd and the authorities were integrated on the Ushahidi platform. A loose integration of authoritative data and VGI from Emergency AUS can be also visualized on the emergency information mapping platform of Queensland State Government[5].
- The volunteers of VOST monitored disaster events and curated social media data about incidents. For searching for useful data, they have to build a long list of keywords. For example, in a bushfire in Edgecombe/Kyneton of Victoria, January 2016, a complex list including the toponym and important contributors was developed

[4] http://vostvic.net.au/post/virtual-operations-support-team-vost-victoria

[5] http://emergencyqld.info/

(see: Appendix 1). The volunteers could further manually check the searching results for ensuring the relatedness, and stored data as images and shared to official agencies. Retweets were excluded as they were only interested in original messages.

From the stakeholders' experiences, text-based VGI can act as an additional communication channel for citizens and enhance situational awareness. However, all the stakeholders admitted that there is no straightforward methodological process to collect, verify, and corroborate VGI. Currently, using text-based VGI in disaster management requires significant manual efforts and is still not efficient.

2.4 Chapter summary

This chapter firstly presents an overview of VGI, which showcased the heterogeneous nature of VGI through a two-dimensional typology, three forms of VGI, and the variable regional contexts. Next, this chapter highlighted the role of VGI in disaster management. Three VGI collection approaches, web-based surveys, the mapping platforms, and social media were described. The advantages and the limitations among the collection approaches were addressed. And then the characteristics of contributors were described. Finally, the three Australian cases presented the collection and the use of text-based VGI for disaster management. This chapter concluded that VGI has the potential to contribute the information need for disaster management. However, the quality issue is inevitable in the use of VGI.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 3

Quality Assessment Methods of Volunteered Geographic Information

3.1 Introduction

In Chapter 2, the uncertainty of VGI quality was showcased. Therefore, for addressing the research objective1-b and 1-c, reviewing current methods for VGI QA and understanding their advantages and limitations is necessary. This chapter aimed to review the latest developed methods. Researchers have developed methods according to various quality requirements and implemented these methods for specific cases. The advantages and limitations of these developed methods are analysed and described.

First, due to the heterogeneous nature of VGI, there are various VGI quality elements and QA approaches are explained. Next, the author then categorized the VGI QA methods into four systematic types: Comparison, Crowdsourcing, Simple modelling (Multi-criteria assessment), Supervised learning. Relevant case studies are reviewed and described in the subsections of this chapter. Subsequently, advantages and important issues for VGI QA methods are presented. Particularly, the use of geographic characteristics in the existing methods is addressed. Finally, two knowledge gaps in the existing text-based VGI QA methods are summarized.

3.2 VGI quality elements and assessment approaches

VGI in distinct forms is heterogeneous for different uses and consists of various components. Therefore, the quality of VGI is a complex concept. In the literature, VGI can be assessed in its quality either as “internal quality” or “external quality”. Various quality elements have been discussed.

3.2.1 Internal quality and geographic data quality elements

Internal quality, or called “quality-as-accuracy” assesses quality in relation to the absence of errors in the data (Poser & Dransch, 2010); it assumes that information holds a perceivable level of similarity between the data produced and the real-world phenomena described (Devillers, Bedard, Jeansoulin, & Moulin, 2007; Fisher, 1999; Guptill & Morrison, 1995). However, like all measurements, geographic data and VGI have levels of

accuracy that are limited by the ways and the devices used in the process of information provision (adapted from Goodchild, 1993). Various quality measures adhering to the quality principles or specifications have been developed consequently. Most of these specifications have been discussed in the literature about spatial data quality and further defined by the standards organizations. For example, the International Organization for Standardization (ISO) defined geographic information quality as the totality of characteristics of a product that bear on its ability to satisfy stated and implied needs. ISO 19157:2013 (ISO, 2013) provides the principles for describing the quality of geographic data by five data quality elements, which are:

- Completeness — presence and absence of objects. It describes the relationship between objects and the abstract universe of all such objects;
- Logical consistency — degree of adherence to logical rules of data structure (conceptual, logical or physical data structure), attribution and relationships;
- Positional accuracy — the accuracy of the position of objects;
- Temporal accuracy — the accuracy of the temporal attributes and temporal relationships of objects;
- Thematic accuracy – the accuracy of quantitative attributes and the correctness of non-quantitative attributes.

In addition to ISO standard, other data quality elements such as semantical accuracy, topological consistency, and geometric accuracy were also used to measure the quality in the literature (Aalders, 2002; Servigne, Lesage, & Libourel, 2006; van Oort, 2006; Veregin, 1999). These quality elements were also called traditional GI criteria/specification (Bishr & Janowicz, 2010).

While using authoritative data, such specifications are often documented in the metadata. This helps end-users realize data quality. Nevertheless, these quality specifications are absent in VGI datasets. Quality is uncertain and has become a barrier to users. To understand VGI quality, researchers have developed various methods to measure quality specifications such as positional accuracy or completeness. For example, the quality of

OSM has been investigated in many cases. A typical QA method is to compare VGI to a reference dataset (Girres & Touya, 2010; Haklay, 2010; Haklay et al., 2010; Koukoletsos, Haklay, & Ellul, 2012).

However, with respect to the image- and text-based VGI, the two forms often lack geometries and quantitative attributes for assessing accuracy scientifically. Researchers thus asserted that the established measures of traditional GI specifications were not adequate to assess the quality of the image- and text-based VGI (Bishr & Janowicz, 2010; Poser & Dransch, 2010). A complementary view of quality focusing on "fitness of uses" or "external data quality" was addressed.

3.2.2 External quality and information quality criteria

External quality indicates how well the components fulfill user needs for a specific task in a specific context. External QA relies on explicitly stated objectives and requirements of the intended use in a given area. Information quality is defined as the extent to which information can be used by those who apply it (Wang & Strong, 1996). It is measured by the difference between the resource wished for by the user and the actual data produced (Pierkot, Zimányi, Lin, & Libourel, 2011).

The concept of external quality can be further extended to the assessment of individual information pieces. A variety of quality criteria and measurement such as credibility, relevance, and timeliness have been defined, used, and discussed, particularly in the domain of crisis informatics (Bordogna, Carrara, Criscuolo, Pepe, & Rampini, 2014; Eppler, 2006; Friberg, Prödel, & Koch, 2011; Redman & Blanton, 1997; Wang & Strong, 1996). For example, Friberg et al. (2011) discussed information requirements and quality criteria in crisis situations; they conducted a survey of emergency management experts to identify important criteria and their weights for developing a model measuring the information quality score. Craglia et al. (2012) and Albuquerque et al. (2015) addressed the use of text-based VGI and proposed the QA methods. According to the arguments in

the literature, the following information quality criteria were important for emergency management.

(1) "Clarity": the possibility of correct understanding and interpretation of information.
(2) "Conciseness": the terseness of an information object.
(3) "Accuracy": the given information represents the reality and how close something is to the true value.
(4) "Objectivity": the information provides a judgment based on observable phenomena, which is uninfluenced by emotions or personal prejudices
(5) "Completeness": an issue is covered broadly within an information object.
(6) "Credibility" or "Believability": the given information complies with the perceived truth from the perspective of a recipient.
(7) Relevance: the information meets the information needs of a particular user or use case.

Among these, credibility is one of the most discussed criteria. Credibility can be thought that information correctly represents the reality (credibility-as-accuracy) or it complies with the perceived truth from the perspective of information receiver (credibility-as-perception), asserted by Flanagin and Metzger (2008). For example, the majority of people have used Google Maps without complaining even people do not have any quality specifications of Google Maps. They trust it from a long-term experience as they can perceive the Maps were produced by professional experts, and further pass the trust to other users.

For credibility assessment, trustworthiness and expertise are perceived subjectively by the information receiver. However, it is possible to model such a perception. Moreover, credibility can also be assessed from objective characteristics, such as where and when the contributor created VGI (Fogg & Tseng, 1999; Gunther, 1992). Therefore, credibility can be assessed by a quantitative model. The VGI sources (e.g. certified or uncertified contributor), VGI contents, and VGI spatiotemporal characteristics have the potential to formulate the model (Craglia et al., 2012; Goodchild & Li, 2012). The abovementioned

characteristics are used for the assessment of an individual VGI instance. In some case studies, the target of VGI credibility assessment method aims to find credible information about major hazards and incidents in spatiotemporal clusters. Researchers aggregated information from different sources and used cross-validation to ensure the credibility of VGI (De Longueville et al., 2010; Mummidi & Krumm, 2008).

3.2.3 Quality assessment approaches and their relations to the systematic methods

The VGI QA approaches have been widely discussed in the literature. One famous framework is proposed Goodchild and Li (2012), they argued three kinds of approaches: crowdsourcing (the involvement of a group to validate and correct errors of VGI), social (trusted individuals who have made themselves a good reputation with their contributions and act as gatekeepers), and geographic approaches (use of laws and knowledge from geography, such as Tobler's first law to assess the quality). Many works have developed VGI QA methods based on these approaches. Furthermore, Senaratne et al. (2016) reviewed 425 research papers and found a new approach: data mining.

As this research addressrd the validation of the correlation between quality and geographic characteristics of text-based VGI, the geographic approach is a key concept in this research. This approach addressed that geographic fact can constrain information by certain rules and determine VGI quality, it presented a linkage to the tradition of geography thought in the QA methods. For example, if there is a point marked as a building, this point is usually not located on the water. Such constraints rules can be either simple or complex.

Furthermore, although the abovementioned approaches can be used to define most VGI QA methods, the literature review of this research found that many practical cases had mixed approaches. For example, due to the lack of ground-truth data for the geographic approach, crowdsourcing was used to provide the complementary ground-truth data in some studies (e.g. Westrope, Banick, & Levine, 2014, this case is further described in Section 3.3.). Besides, modelling and data ming can consider multiple information factors including geographic facts, social reputations, and linguistic characteristics. Therefore, this

research addressed four types of systematic VGI QA methods in the literature review, which are Comparison, Crowdsourcing, Simple modelling (Multi-criteria assessment), and Supervised learning. Table 3.1 described these methods and some analysis procedures.

Table 3.1: VGI QA approaches and systematic methods

Approach	Systematic method	Description	Analysis procedures
Geographic/ Crowdsourcing	Comparison	Measure quality criteria such as positional accuracy by comparing the VGI with reference from accurate measurement or modelling. In some studies, crowsourccing data quality was compared.	Positional accuracy analysis, semantic consistency check, Geometrical analysis, proximity analysis, integrity constraints
Crowdsourcing /Social	Crowdsourcing	Use crowd or gatekeepers to correct errors and false information.	Manual inspection, manual annotation, peer-reviewing, reputation analysis,
Geographic/ Social/ Crowdsourcing	Simple weighted modelling (Multi-criteria assessment method)	Using variable criteria and associated weights/measures to evaluate quality. Geo-attributes of VGI has been considered in the measures	Quality measures development, weight assignment
Data mining	Supervised learning	Using different types of features to assess credibility and text content quality.	manual annotation, predictor (feature) selection, algorithm selection, model training and evaluation

3.3 Comparison

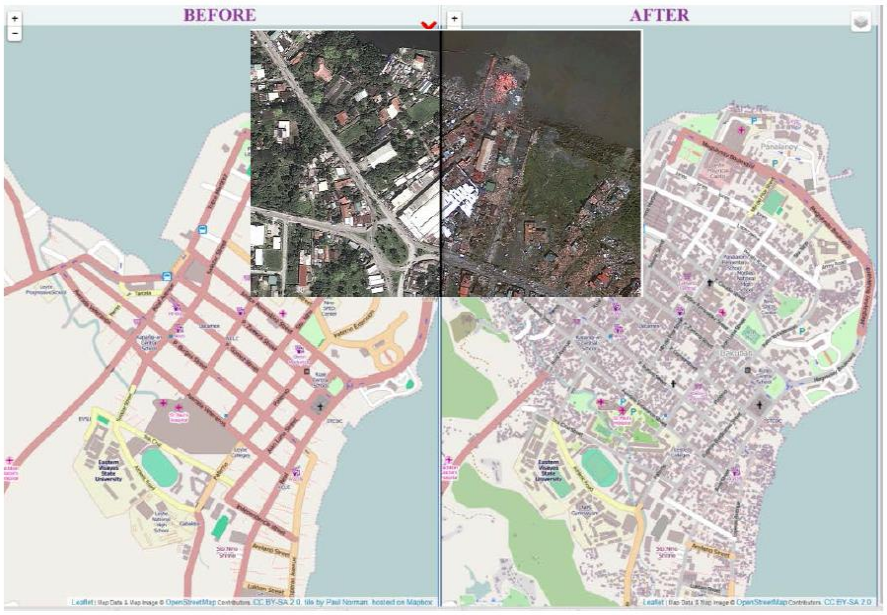
One intuitive way to ensure VGI quality is to compare VGI to high-accuracy references, the "golden standard" (Maué & Schade, 2008). Comparison is often used in map-based VGI as it can quantitatively or qualitatively assess VGI and measure how good it is. For example, many case studies of OSM QA were based on Comparison. The quality of various geographic objects (i.e. point, polyline, polygon) at different scales, times, and regions was compared and measured.

According to the type of geographic objects and the availability of references, various QA methods have been developed. Several studies demonstrate various ways of spatial analysis assessing the quality of OSM road networks, such as using a buffer zone to calculate the percentage of overlap between VGI and references for positional accuracy; calculating the total length of OSM roads and then comparing with the references for the completeness (Haklay, 2010; Haklay et al., 2010; Zielstra & Zipf, 2010). Measures of geometric accuracy, attribute accuracy, and logical consistency were developed in some studies as well (Girres & Touya, 2010). Note that an important concept in the comparison is the matching of geographic objects. As physical objects such as buildings can be represented differently in heterogeneous datasets, a variety of matching methods were developed in the literature (Fan, Zipf, Fu, & Neis, 2014; Hecht, Kunze, & Hahmann, 2013; Kalantari & La, 2015; Koukoletsos et al., 2012).

In the context of disaster management, Comparison can provide a tangible way to measure VGI quality. For example, the Flood Citizen Observatory project in Brazil researchers found the observations by volunteers were statistically equivalent to the sensor data (Brito Moreira, Degrossi, & Albuquerque, 2015; Degrossi, Albuquerque, Fava, & Mendiondo, 2014). However, a limitation of the comparison is that the availability of reference data is uncertain. Reference data are often not available or only present a measurement at a specific moment within a continuous phenomenon (e.g. images of the flood extent). One solution is to use ground surveys. For example, Westrope et al. (2014) investigated the data quality of the OSM building damage assessment project in the Philippines after Typhon Haiyan

(Figure 3.1), and conducted ground surveys in several areas. Westrope et al. (2014), further found that the accuracy of the crowd-tagging was merely 36% compared to their field assessments. Erroneous recognition was mainly due to two limitations: (1) comparative imageries before the event were unavailable, and (2) detailed performance from satellite image was not good enough. Note that ground surveys usually require many efforts and budgets. It is not quite as practical as an alternative.

(a)



(b)

Figure 3.1: The mapping and damage assessment in Tacloban City after Typhoon Haiyan struck in 2014 on OSM: (a) OSM before and after Typhoon Haiyan (b) status of damage in a single building.

The abovementioned cases demonstrate that Comparison can deal with how close the match is between the positions and attribute values of VGI and reference data, or provides more accurate information. And according to the conditions of volunteer engagement, the

QA showed mixed results. Although most related works focused on map-based VGI such as OSM, the study by Poser and Dransch (2010) also indicated that Comparison can be used for other forms of VGI. Moreover, some innovative studies did not focus on the quality of VGI instances but tried to leverage VGI contents to produce information of better quality. Examples can be found in several papers integrating VGI and authoritative data to assess inundation depth and flood surface, which provided a new approach constructing or altering the flooding map in the absence of other inundation data (McDougall, 2011; McDougall & Temple-Watts, 2012; Schnebele & Cervone, 2013; Schnebele, Cervone, & Waters, 2014; Triglav-Čekada & Radovan, 2013)

From the literature review, the author concluded that Comparison can be used to assess the positional accuracy of the VGI instances in this research. But for credibility and text content QA, it cannot provide an efficient process in the response scenario due to the imitated availability of reference data.

3.4 Crowdsourcing

An alternative method for VGI QA, Crowdsourcing, is using crowd intelligence to validate and correct errors. This concept reflects Linus' Law that "given enough eyeballs, all bugs are shallows" (Goodchild & Li, 2012; Hall, Chipeniuk, Feick, Leahy, & Deparday, 2010; Raymond, 2001). Crowdsourcing is often used in a project throughout a wide connection with volunteers. Data assembled by crowd intelligence are assumed to have higher quality as it gathers the individual's ability to validate and correct errors. In VGI initiatives with a social network hierarchy, Crowdsourcing may additionally control quality by trusts individuals acting as gatekeepers (Goodchild & Li, 2012). For example, OSM super-group users track recent changes constantly and flag changes that look wrong to fight with vandalism.

As required information is often absent during a crisis, crowdsourcing has been utilized in different scenarios of disaster management. For example, during an extreme event, it is difficult to classify valuable information or identify specific attributes from available

information. Crowdsourcing, therefore, can be used to deal with content classification problems such as the case of Westrope et al. (2014) mentioned previously. There were several projects providing images for the identification of the required information by crowdsourcing, such as building damage status after an earthquake (Ajmar, Boccardo, & Giulio Tonolo, 2011) or missing aircraft (e.g. Flight MH370[6]). Volunteers were required to tag and curate content so as to organize and filter VGI under predefined structures. Moreover, advanced skills such as snowball sampling and different strategies for using crowdsourcing in crisis situations were introduced by Meier (2011).

The abovementioned case demonstrates that Crowdsourcing has been used in many disaster events for heterogeneous purposes. However, Crowdsourcing has limitations to provide timely results and highly depends on the capacity of human efforts. In a sparsely populated location or underexplored part of the world, the crowdsourcing initiatives are hard to find local contributors for validation (Goodchild & Li, 2012). Therefore, it should be noted that Crowdsourcing does not always pursue a perfect result. Instead, the aims are more relevant to moderate and enhance the information at different levels for further analysis and use. Meier (2011) mentioned that many emergency management experts would rather have some data not immediately verifiable than no data at all. The meaning of crowdsourcing is toward to improve both the volume and the quality of information.

3.5 Simple weighted modelling (Multi-criteria assessment)

In the assessment of information quality, there are some information characteristics relevant to the judgment. These information characteristics can be used in a simple model. For example, researchers hypothesized a collaborative application and proposed a conceptual credibility model; the model not only calculated a trust rating given by contributors but also used spatiotemporal parameters such as the distance between the

[6] See: htttp://www.cbc.ca/news/technology/flight-mh370-search-gets-crowdsourced-help-from-scientists-1.2987450

contributors to the contents' locations (Bishr & Janowicz, 2010; Bishr & Kuhn, 2007; Bishr & Mantelas, 2008). Among several studies in the literature, this research found that the concepts and the techniques of Multi-Criteria Decision Making (MCDM) were often used.

3.5.1 Quality measures and quality score

A general method of MCDM is the Simple Additive Weighting (SAW). Its concept is that the quality of an information instance is determined by the measures of several quality criteria and users can assign distinct weights. The weights of the criteria are multiplied with the values from the given measures to calculate the quality score as the indicator, which can be expressed by the following equation (Friberg et al., 2011):

$$IQS(VGI_j) = \sum_{i=1}^{N} w_i \, m_i(S_{ij})$$

Where N is the number of criteria, w_i is the weight of criterion c_i, m_i is the measure for criterion c_i, and s_{ij} is the score of VGI instance VGI_j for criterion c_i.

Based on this concept, researchers examined the text content, metadata, and source of VGI to develop quality measures. For example, Morris, Counts, Roseway, Hoff, and Schwarz (2012) examined key information elements of tweets for understanding their impact on credibility judgments. In the context of disaster management, Craglia et al. (2012) developed a project called CONAVI (CONtextual Analysis of Volunteered Information), which proposed a semi-automatic workflow to retrieve and assess relevance and credibility of social media data. With the development of CONAVI, Craglia et al. (2012) discussed VGI quality criteria and conceptual measures (Table 3.2). They considered multiple kinds of information factors to determine the quality of individual VGI instances. Moreover, similar to the concepts proposed in CONAVI. Paul (2013) developed a VGI ranking mechanism for crisis situations and addressed varying aspects such as identification of original messages and popular users. Ludwig, Reuter, and Pipek (2015) developed an empirical study for the use of social media during emergencies and then further defined

five criteria and measures to calculate the quality score of social media feeds, which are (1) link, (2) credibility, (3) up-to-datedness, (4) dissemination, and (5) quality of coordinates.

Table 3.2: Criteria and measures of VGI quality assessment (Craglia et al., 2012)

Criteria	Measures
Topicality	(Co-)occurrences of use case-specific keywords, highly mentioned places
Timeliness	A function according to information published time
Spatial proximity	Spatial distance to a location relative to other information such as a known event, high-risk area, or a similar VGI instance
Temporal proximity	Temporal duration to a timestamp relative to other information such as a known event, or a similar VGI instance
Accuracy/granularity	Geographic precision of geo-location, level of detail, the vagueness of boundaries, etc.
Source type/behaviour	Individual account or authoritative account, the reputation of source (e.g. the number of followers), ratings (e.g. number of retweets[7]), certification mechanism, etc.
Geographic context	Characteristics of the location near or around the VGI content (e.g. population density)

3.5.2 Geographic concern of quality measures in text-based VGI assessment

An important aspect of the measures proposed by Craglia et al. (2012) in Table 3.2 is their geographic concern. Due to the limitation of geotagged data availability, the geographic concern was not often absent in text-based VGI QA. Only two cases were found in the author's investigation. One case is the subsequent research projects of CONAVI. The geographic measures were designed and implemented to assessed text-based VGI credibility about forest fires. Forest cover ratio, distance to the nearest hotspot, and population density were used to develop scoring measures (Ostermann & Spinsanti, 2012; Spinsanti & Ostermann, 2013). It was assumed that if a social media feed about forest fire

[7] A Retweet is a re-posting of a Tweet.

was generated from a commune where the number of inhabitants per square kilometre was more than 400, it was not quite credible.

Moreover, Albuquerque et al. (2015) also used a geographic model to identify useful tweets relevant to flooding events; they found that the quality can be assessed by their proximity to flood-affected catchment and height of water level; the results showed tweets up to 10 km to severely flooded areas have a much higher probability of being related to floods.

3.5.3 Issues of the existing multi-criteria assessment methods

The developed multi-criteria assessment methods examine pragmatic factors and sort VGI instances, which are very fast and easy to be implemented. Nevertheless, an issue in the existing works is that the developed measures often lack any evaluation. The influential variables and their threshold values are difficult to be justified. For example, Ludwig et al. (2015) defined a believability measure as "*a tweet with the most retweets is rated 100, a tweet which has not been retweeted is rated 0*". Using this measure, the content from Twitter users without many followers will always get low credibility ratings. Considering more information elements is suggested for the refinement. The geographic measures in CONAVI had the same problem. As mentioned in Section 2.2.3, Spinsanti and Ostermann (2013) argued that population density has a negative impact on the credibility of social media feeds about forest fire event. This assumption cannot be confirmed without evaluations as the effect of demographic factors on fire occurrence can be two-fold, even in the case of forest fires (Oliveira, Pereira, San-Miguel-Ayanz, & Lourenço, 2014). In the U.S.A., the average 10-year percent wildfire starts are 88% human-caused accidentally[8]. Therefore, a further validation of the correlation between quality and the geographic characteristics of text-based VGI is necessary.

[8] The Causal History of Forest Fires, http://forestry.about.com/od/forestfire/a/Causes-Of-Forest-Fires.htm

3.6 Supervised learning method

An alternative approach for VGI QA, particularly for image- and text-based VGI is Supervised learning, a specific Data mining/Machine Learning (ML) technique. The concept of ML is that each information instance has various features (i.e. predictors). A feature-based retrieval model is built to find the relationship between quality indicators and features. Comparing to the other statistical approaches, the emphasis of ML is learning from data.

Based on the ML concept, Supervised learning is often used in the domain of Information Retrieval (IR). IR can be defined as "*finding material (i.e. documents) of an unstructured nature (usually text) that satisfies an information need from within large collections*" (Manning, Raghavan, & Schütze, 2008). As perceived information has certain characteristics, the IR approach uses different mathematical models to recognize characteristics and interpret the specific pattern of information. Algorithms such as Logistic Regression, Decision Tree, Support Vector Machine (SVM) were widely used for building models. Therefore, text-based VGI can be categorised by needs. The advantage of Supervised learning is that it does not require robust evidence of relevant features about a phenomenon. Instead, the developed model can adjust itself from data. Therefore, researchers can train the model based on assumptions, and repeat to test various feature sets. In the existing works, three kinds of VGI features were often used in QA: (1) text-related characteristics (e.g. n-grams[9]), (2) user relationships (e.g. rating of content, number followers/friends), and (3) usage statistics (e.g. number of clicks on information instances) (Agichtein, Castillo, Donato, Gionis, & Mishne, 2008). These features are similar to the

[9] An n-gram is simply a sequence of tokens. In the context of computational linguistics, these tokens are usually words, though they can be characters or subsets of characters. The *n* simply refers to the number of tokens.

factors considered in Multi-criteria assessment. But the ML approach has the advantage to evaluate significant features and deal with the analysis of a huge volume of data.

As the quality indicators are often absent in data, the process is usually based on manual annotations under a Supervised learning framework (Figure 3.2). The quality problem, therefore, can be treated as a classification problem. In the literature, several studies have tried to classify VGI in the context of disaster management. For example, Castillo, Mendoza, and Poblete (2011) tried to identify the credible information cluster of tweets by using a cost-sensitive tree based on the J48 Decision Tree; they found that usage statistics were the best features. Gupta and Kumaraguru (2012) have done a work similar to that of Castillo et al. (2011) based on Logistic Regression but they assessed the credibility level of individual tweets; their case study used a combination of Supervised learning and relevance feedback techniques to rank credible and informative tweets; the results concluded that both text-related characteristics and user relationships were important in credibility assessment.

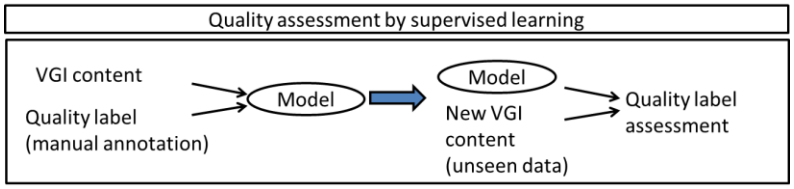


Figure 3.2: Conceptual framework of VGI quality assessment by Supervised learning.

Furthermore, a semi-automated system, AIDR (Artificial Intelligence for Disaster Response) was developed by the Qatar Computing Research Institute. The AIDR research team developed a series of experiments for text-based VGI classification and QA. For example, Imran, Elbassuoni, Castillo, Diaz, and Meier (2013) used crowdsourcing categorized the tweets into (1) “Personal Only”, (2) “Direct Informative and Indirect Informative” (eye-witness or not), (3) other. They defined that the message attracting interest from people beyond the publisher’s social network is informative and developed models. An important part of their experiments is to use syntax-related features, such as Part of Speech (POS) tags and the verbs. The results demonstrated a good performance in

the informative message classification (precision, recall, and AUC were about 0.8). However, the performance of eye-witness tweets identification (i.e. classification of direct informative and indirect informative) was below expectations. Apart from the output of AIDR, several research projects such as Ashktorab, Brown, Nandi, and Culotta (2014) and Truelove, Vasardani, and Winter (2016) also focused on the eye-witness and informative content classification. Yet the performances of their models were not as good as the models of AIDR. The reason is possible that the number and types of input features were not enough for achieving high-performance models. Other notable research projects with similar methodology included Yang, Counts, Morris, and Hoff (2013) and Yin, Lampert, Cameron, Robinson, and Power (2012).

The abovementioned studies have shown that Supervised learning has great potential to solve the information quality problem in crisis situations. However, in comparison with Multi-criteria assessment, the ML systems are quite complex like "black boxes". And most existing efforts only used text-related characteristics; geographic characteristics of VGI were not considered as features. Furthermore, as a reminder, Supervised learning is a data-driven approach so the results highly depend on the data collection. The generalisation of models is uncertain. In other words, if only a small amount of VGI is available, using Supervised learning might be not appropriate.

3.7 Summary of current VGI quality assessment methods

A summary of abovementioned VGI QA methods is listed to present their advantages and limitations. Through the synthesis of the research gaps in the literature, the author argued that there was no enough evidence to supporting the geographic correlation in the text-based VGI QA process.

3.7.1 Advantages and limitations of the established methods

According to the heterogeneous need of information, four major types of VGI QA methods have been reviewed, namely, Comparison, Crowdsourcing, Multi-criteria assessment, and Supervised learning. The advantages and limitations of these methods have been discussed in the previous subsections and summarised in Table 3.3. The selection of an appropriate

method requires considering the availability of reference, the volume of information, and types of quality criteria or quality indicators.

For using VGI in the context of disaster management, as all the developed methods have variable limitations, users accept a certain level of uncertainty in exchange for the ready availability of information (Camponovo, 2013; Posetti, 2012). The VGI approach is suggested to be a complementary approach for the tasks requiring high accuracy of information.

Table 3.3: Advantages and limitations of the VGI QA methods

Method	Advantages	Limitations
Comparison	Appropriate to be used in variable forms and quantitative results are available.	1. Availability of reference data is an issue in the context of disaster management. 2. This method is not appropriate to text-based VGI due to its insufficient structure.
Crowdsourcing	Having showcased this method is practical in many humanitarian activities.	1. Lack of a scientific workflow 2. Time-consuming process 3. Availability and scalability of crowd
Multi-criteria assessment method	Weights and measures can be adjusted to reflect the need of VGI user.	Measures in the existing works lack evaluations and require refinement.
Supervised learning	1. Appropriate to deal with a high volume of VGI. 2. Significant information characteristics impacting on QA can be identified. The performance of models can be evaluated.	1. Data-driven approach; so the generalisation of models is uncertain. 2. Algorithms might be not easy to understand and adjust according to the need of users.

3.7.2 Gaps in the validation of the geographic correlation and problem formulation

From the literature review, Multi-criteria assessment and Supervised learning can generate the QA model in different ways. Researchers have used these methods to develop text-based VGI QA models. A critical finding is that the geographic concern has been addressed in some Multi-criteria measures. However, these measures were rarely validated and require further refinement. The literature review of this research found that only Albuquerque et al. (2015) has presented a statistical correlation from a case study of flood-related tweets in the territory of Germany. A further investigation between geographic characteristics and text-based VGI quality is suggested.

On the other hand, Supervised learning, or other statistical techniques have a rigorous methodology to evaluate the developed methods. The use of these techniques can be complementary to the Multi-criteria method as it can help the justification of measures. However, a problem is that the existing Supervised learning methods were irrespective of the laws and knowledge from geography. The geographic correlation was absent in any work based on Supervised Learning. This research, therefore, proposes to use Supervised learning or auxiliary statistical techniques to validate the geographic correlation and develop the VGI QA methods.

3.8 Chapter summary

This chapter is a literature review of the existing VGI QA methods. First, two approaches of current methods were clarified: the internal quality approach or the external quality approach. Subsequently, four types of QA methods were reviewed in detail. Based on the investigation, the limitations of QA methods in different aspects, such as the lack of scientific workflow, the inflexibility of method adjustment, and the availability of comparable data were listed.

Throughout this chapter, this research concludes the existing work is limited support the assumption of geographic relation and VGI quality in a local area. The refinement of VGI QA methods and validation of the geographic correlation are further investigated.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 4

Research Design and Methodology

4.1 Introduction

This chapter provides the details of the research methodology and design to achieve the research aim and objectives introduced in Chapter 1. This chapter is structured into four parts (Figure 4.1). In the first part, a conceptual design framework is articulated to provide an overview of the research process. Next, this chapter introduces the adopted research methodology, Design Science. The theory and guideline of Design Science are described. Based on the theory and guideline, the four-phase research design is described; key activities are presented and explained. Finally, the associated techniques and tools used in this research are briefly described.

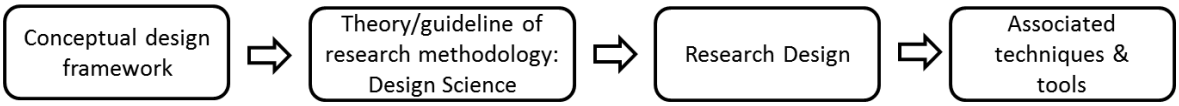


Figure 4.1: The overall chapter structure

4.2 Conceptual design framework

This research aims to provide robust methods for using geographic characteristics of text-based VGI in QA. For achieving this aim, a conceptual design framework is formulated, which is illustrated in Figure 4.2. It shows the relationship between the research context, elements, methods, and outcomes.

The research context addresses the uncertain quality of text-based VGI for disaster response and the lack of validation in the correlation between quality and the geographic characteristics of VGI instances. The summary of the literature review (see: Section 3.6) has outlined these problems. Based on this context, the research elements (i.e. problem, questions, aim, and objectives) set out a blueprint to further find the coherence of knowledge. A case study examining real data in the Australian extreme events is proposed. Research methods based on Supervised learning and modelling are used. The research outcomes are the text-based VGI QA methods based on mathematical models, justification of geographic correlation, and publications.

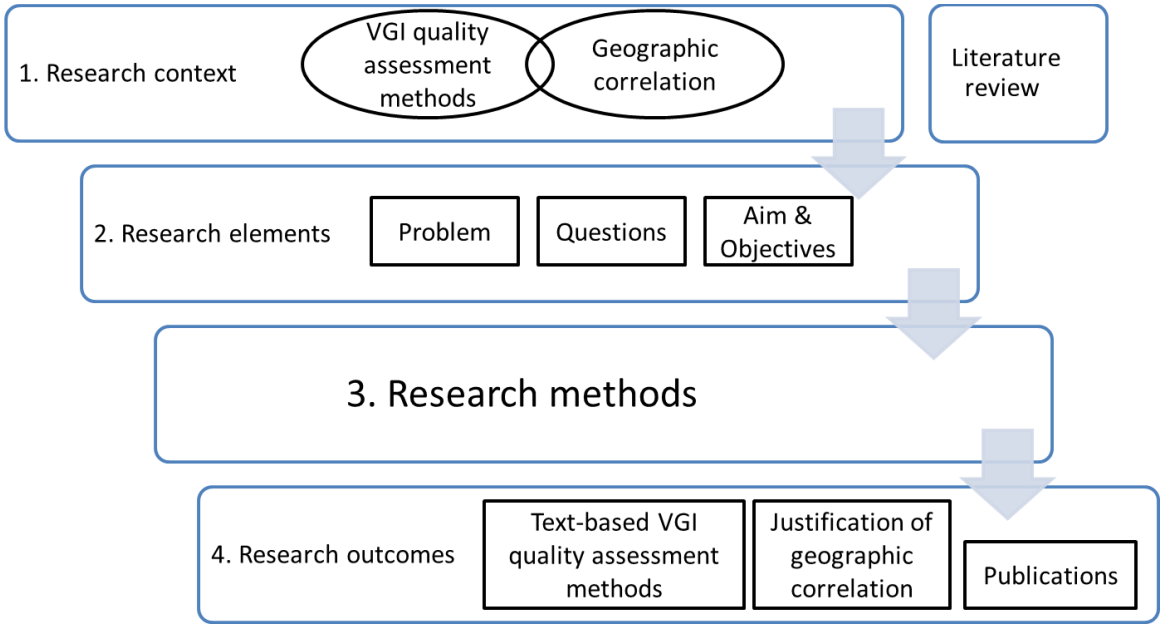


Figure 4.2: The conceptual design framework

4.3 Design Science Research Methodology

The adopted methodology is Design Science Research Methodology (DSRM). DSRM addressed the test and development of hypothesis and design of artificial solutions. This methodology has been used to design information artefacts for disaster management (e.g. Amirebrahimi, 2016), and has been used in some VGI quality studies (e.g. Crowston & Prestopnik, 2013). DSRM is an outcome-based methodology, which offers specific guidelines for evaluation and iteration within research projects; the developed application is most notable in the Engineering and Computer Science disciplines[10]. As this research addressed an iterative process to test the geographic correlation and develop the text-based VGI QA method, DSRM is suitable for this research.

[10] https://en.wikipedia.org/wiki/Design_science_(methodology)

4.3.1 Overview of Design Science Research Methodology

Design Science is a research paradigm in many disciplines, such as engineering and computer science, and it is highly connected to information system research. Information systems are designed and built for supporting the activities of the organization. The main focus of information system research is developing and communicating knowledge concerning both the administration and the use of information technology for managerial and organizational purposes. The implementation and the utility of information system are specific to the organizational or inter-organizational contexts such as the work systems and people (Hevner et al., 2004; Iivari, 2007). Designing information technology artefacts to aid the effectiveness and efficiency of an organization is an instinctive desire of researchers.

Design Science is distinct from natural or social science but a complementary approach (March & Smith, 1995). The typical paradigms of natural or social science seek to develop and justify theories that explain, understand and predict complex phenomena. These paradigms generally have a specific relationship between epistemology, theoretical perspectives, methodology, and methods (Figure 4.3). The research outputs of natural or social science are usually theories for identifying and codifying important principles and laws among natural phenomena or human behaviours.

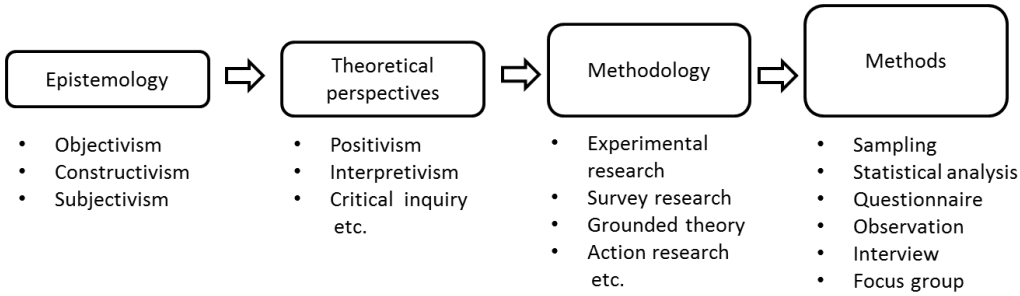


Figure 4.3: Relationship between epistemology, theoretical perspectives, methodology, and research methods (Source: adapted from Crotty, 1998)

From the perspective of information system research, such output often fails to elaborate on the imperative of improving the performance of information technology to design an artefact. On the other hand, justified theories from natural/social science and artefacts are

two sides of the same coin. DSRM addresses to bring together the knowledge base and requirement of the application domain into the design of artefacts in an information system. A combination of design and natural/social sciences theories can provide a problem-solving paradigm from a rigorous base.

4.3.1.1 General outputs of a Design Science project

The core of Design Science is to understand and improve the components in order to construct an artefact for solving a problem based on systematic knowledge (Baskerville, 2008). The artefacts built based on Design Science is broader, including not only material instantiations, but also abstract artefacts such as constructs, models, frameworks, architectures, design principles, methods, and design theories (Gregor & Hevner, 2013; March & Smith, 1995; Vaishnavi & Kuechler, 2015). The variable forms of artefacts are explained in Table 4.1. According to this table, the method developed in this research provides sets of steps to perform text-based VGI QA.

Table 4.1: General outputs of a Design Science project

Artefact	Description
Constructs	The conceptual vocabulary of a problem/ solution domain.
Models	Sets of propositions or statements expressing relationships among constructs.
Frameworks	Real or conceptual guide to serve as support or guide.
Architecture	High-level structures of systems.
Design Principles	Core principles and concepts to guide design.
Methods	Sets of steps used to perform tasks (how-to knowledge)
Instantiations	Situated implementations in certain environments that do or do not operationalize constructs, models, methods, and other abstract artefacts.
Design theories	A prescriptive set of statements on how to achieve a certain objective. A theory usually includes other abstract artefacts such as constructs, models, frameworks, architectures, design principles, and methods

Following the principles and guideline of Design Science, the detailed activities to generate the artefact (i.e. text-based VGI QA methods) and research outcome (i.e. validation of the geographic correlations, and publications) are planned in a rigorous process.

4.3.1.2 Research processes and activities of a Design Science project

DSRM often included two important processes: building and evaluation. Building is the process of artefact construction; evaluation is a validation or proof process in terms of how well a resulting artefact performs to what degree as intended (Gregor, 2006; Hevner et al., 2004). DSRM explores to build innovations that define the ideas, practices, technical capabilities, and products for accomplishing effective and efficient systems with scientific evaluation. The processes focus not only on the productive application of information technology but also on the behaviour of human beings. Therefore, the problems are characterized by unstable requirements and constraints depended on environmental contexts, complex interactions, human cognitive, and social abilities (Brooks, 1987).

The research activities of Design Science within the information system discipline are suggested in a set of guidelines by Hevner et al. (2004), which are summarized as follows:

1. *Design as an artefact: Design Science research must produce a viable artefact in the form of a construct, a model, a method, or an instantiation.*
2. *Problem relevance: The objective of Design Science Research is to develop technology-based solutions to important and relevant business problems.*
3. *Design evaluation: The utility, quality, and efficacy of a design artefact must be rigorously demonstrated via well-executed evaluation methods.*
4. *Research contributions: Effective Design Science Research must provide clear and verifiable contributions in the areas of the design artefact, design foundations, and/or design methodologies.*
5. *Research rigour: Design Science Research relies upon the application of rigorous methods in both the construction and evaluation of the design artefact.*
6. *Design as a search process: The search for an effective artefact requires utilizing available means to reach desired ends while satisfying laws in the problem environment.*

7. *Communication of research: Design Science research must be presented effectively to both technology-oriented and management-oriented audiences.*

4.3.2 Performing Design Science Research in this research

According to the concept and guideline of DSRM, the approach of this research consists of four phases: (1) Conceptual, (2) Design, (3) Development and Evaluation, and (4) Synthesis and Communication. Research activities in these phases can be examined in a three-cycle view proposed by Hevner (2007), which is illustrated in Figure 4.4. The "Relevance Cycle" bridges the contextual environment of the research project with the Design science activities. The "Rigor Cycle" connects the Design science activities with the knowledge base of scientific foundations, theories, experience, and expertise that informs the research project. The two cycles present the core activities in Conceptual phase. The central "Design Cycle" iterates between the core activities of building and evaluating the design artefacts. It includes the major research activities of this research.

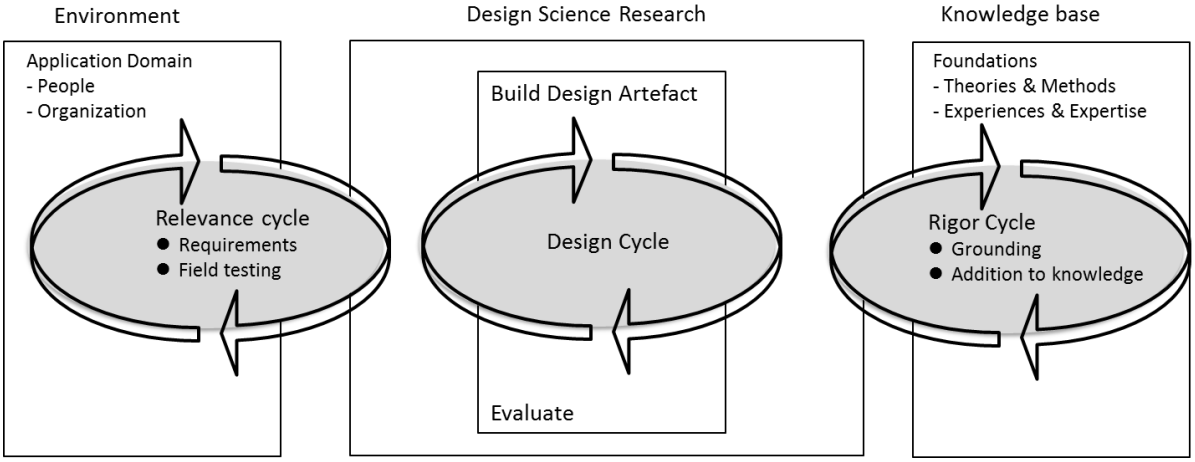


Figure 4.4: Design science research cycles

4.3.2.1 Conceptual phase: identification of problem and requirement and Exploring knowledge base

Conceptual phase initiates Design science research with an application domain that consists of the people, organizational systems, and technical systems (i.e. the Relevance Cycle). It

aims to identify the problem and requirement in the application domain, and defines acceptance criteria for the ultimate evaluation of the research results.

Once the problem is identified, it is important to understand the distinct views in the knowledge base about various dimensions of the problem for a solution (i.e. the Rigor Cycle). Science researches require an ample knowledge base of theories and methods as a foundation. Seeking knowledge to guarantee the innovation of researches is suggested. It is contingent on the researchers to exhaustively research and to reference the existing knowledge base in order to provide research contributions by designing and producing artefacts. Accordingly, decisions and processes in Design Science should be based on grounded behaviour or theories. Appropriate methods of data collection (e.g. interview, survey) and analysis (e.g. statistical model) for constructing the design are important. However, adoptable theories may be undiscovered or incomplete. Therefore the results of design science research may advance the development and generate additions to the knowledge base.

4.3.2.2 Design, development, and evaluation

Design, development, and evaluation (i.e. the Design Cycle) are the keys of DSRM and where the hard work of a design science project is done. Generally, the research activities repeat between the construction of an artefact, its evaluation, and subsequent feedback to refine the design further. Based on the requirements identified and the associated theories and methods from Conceptual phase, the potential alternatives were designed by appropriate research methods such as experimentation, simulation, and case study. The alternatives are further developed and evaluated against requirements until a satisfactory design is achieved (Iivari, 2007; Simon, 1996).

The evaluation requires providing feedback to the construction of artefacts in terms of the quality of the design process and the product under development. Selection of an appropriate evaluation method is essential in the DSRM. Hevner et al. (2004) addressed five major types of evaluation methods: observational, analytical, experimental, testing, and descriptive method, which were summarized in Table 4.2. Accordingly, the artefact

can be evaluated by functionality, completeness, consistency, accuracy, performance, reliability, and usability. The evaluation of the constructed artefact in this research (i.e. the methods for text-based VGI QA) uses the analytical method, including static analysis (e.g. the geographic correlation), dynamic performance (e.g. the model performance), and a simple optimisation analysis (algorithm comparison).

Table 4.2: The evaluation methods in Design Science (Hevner et al., 2004)

Form of evaluation	Methods and techniques
Observational	Case Study: Study artefact in depth in business environment
	Field Study: Monitor use of artefact in multiple projects
Analytical	Static Analysis: Examine structure of artefact for static qualities (e.g., complexity)
	Architecture Analysis: Studying fitness of artefact into technical architecture
	Optimisation: Demonstrate inherent optimal properties of artefacts or provide optimality bounds on artefact behaviour
	Dynamic Analysis: Study artefact in use for dynamic qualities (e.g., performance)
Experimental	Controlled Experiment: Study artefact in a controlled environment for qualities (e.g., usability)
	Simulation: Execute artefact with artificial data
Testing	Functional (Black Box) Testing: Execute artefact interfaces to discover failures and identify defects
	Structural (White Box) Testing: Perform coverage testing of some metric (e.g., execution paths) in the artefact implementation
Descriptive	Informed Argument: Use information from the knowledge base (e.g., relevant research) to build a convincing argument for the artefact's utility
	Scenarios: Construct detailed scenarios around the artefact to demonstrate its utility

4.4 Research design

This research consists of four phases, namely: (1) Conceptual, (2) Design, (3) Development and Evaluation (D&E) phase, and (4) Synthesis and Communication phase. In this section,

the detailed design of the research method is described. The activities of each phase are illustrated in Figure 4.5.

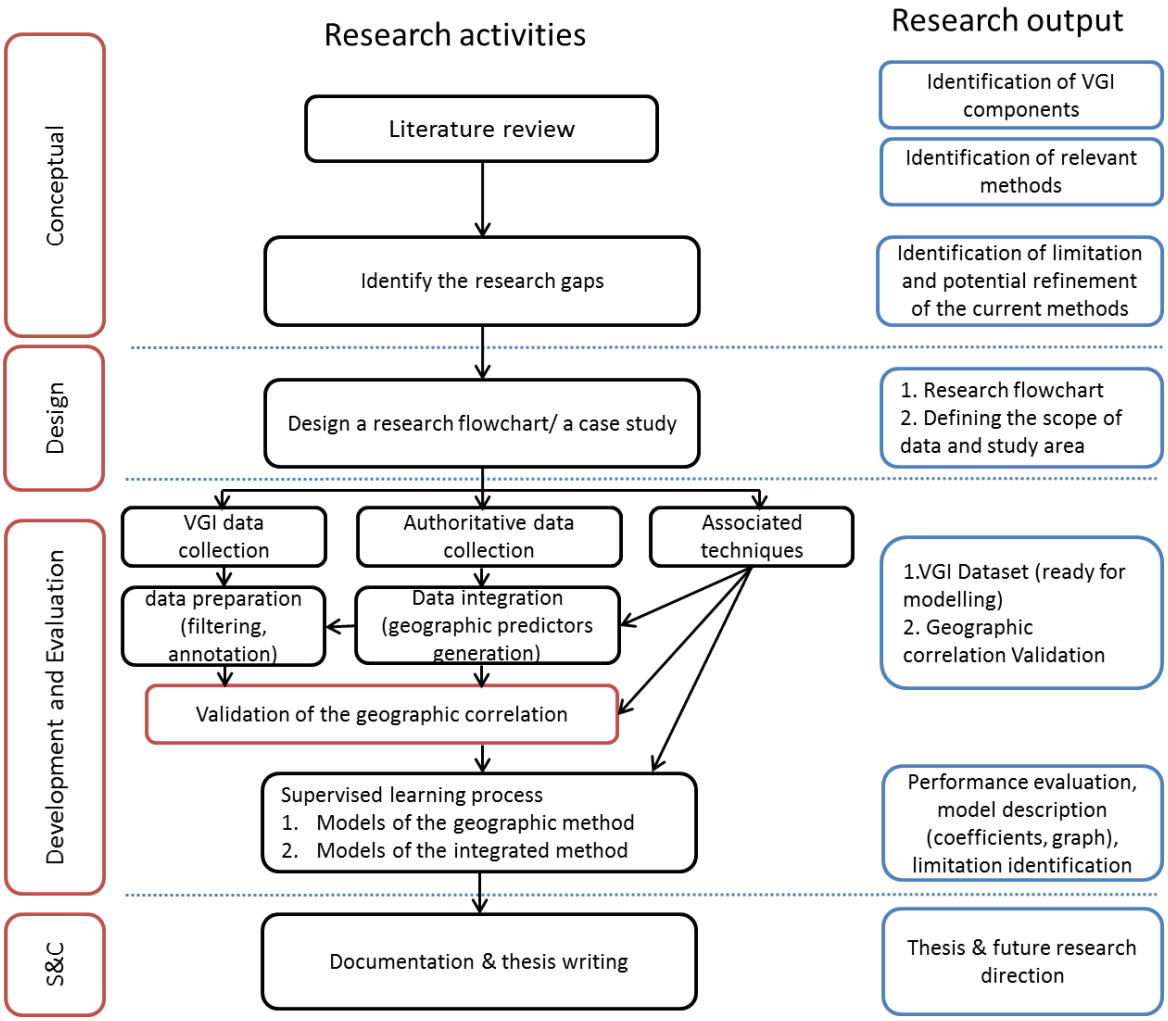


Figure 4.5: Research design and associated activities

4.4.1 Conceptual phase

The conceptual phase formulated the research foundations: problem identification, setting objectives, and exploring knowledge base of this research. The major activity in this phase was conducting a literature review. Meanwhile, the objective 1 of this research, investigation and understanding the use of VGI in the context of disaster management and associated QA methods, was addressed in the literature review.

A literature review is a cornerstone to define a precise problem and understand the knowledge base. Its purpose is to state the current knowledge of published research on a topic objectively through a summary; it conveys to readers a comprehensive overview of issues, knowledge, and ideas established. A decent review might critique each study included, yet this is not a necessary property (Green, Johnson, & Adams, 2006; Helewa & Walker, 2000).

A literature review should include both summary and synthesis writing. A summary is a precise capturing of information in a text; it contains only the main points from the text. On the other hand, a synthesis is where the writer uses two or more summaries to develop an argument on a particular topic, and contains the author's opinion with the summaries as substances.

4.4.1.1 Design of the literature review

The reviewed literature in this research was searched and selected from two academic databases, Google Scholar and the Web of Using search terms "Volunteered Geographic Information" and "User-Generated Geographic Content", about 544 articles were selected until November 30, 2016.

The selected articles were further inspected by the author and categorized as (1) position papers and research agenda (e.g. Flanagin & Metzger, 2008; Goodchild & Glennon, 2010), (2) review papers (e.g. Haworth & Bruce, 2015; Senaratne et al., 2016) (3) case studies about utilizing VGI in the context of disaster management (e.g. Zook et al., 2010), and (4) VGI QA methods (e.g. Poser & Dransch, 2010; Spinsanti & Ostermann, 2013).

4.4.1.2 Findings from the literature review

A narrative literature review had been conducted and synthesized the author's finding in a condensed form in this thesis (i.e. Chapter 2 and Chapter 3). The literature review summarised the concept and the heterogeneous nature of VGI. A specific focus of the literature review was the existing VGI QA methods, which can be categorised into three knowledge levels in Design Science (Iivari, 2007) listed in Table 4.3: (1) conceptual knowledge (e.g. VGI QA approaches, methods, and criteria) (2) descriptive knowledge

(The facts and empirical regularities of the geographic correlation and VGI quality), and (3) prescriptive knowledge (e.g. the associated techniques).

Table 4.3: The knowledge base and the examples of the literature review

Categories of knowledge levels	Illustrations	Knowledge findings in this research
Conceptual knowledge - concepts, constructs - classifications, taxonomies, typologies - conceptual framework	System concepts, ontologies, etc.	Quality issue and QA approaches, methods, and criteria
Descriptive knowledge - observational facts - empirical regularities - theories and hypothesis	- X case A in situation B - X tends to cause A in situation B with probably B	The facts and empirical regularities of the geographic correlation and VGI quality
Prescriptive knowledge - design product knowledge - design process knowledge	The artefact (idea, style functionality, behaviour, architecture) and associated actions.	Techniques for achieving the text-based VGI QA methods

From the literature review, the related cases for using VGI in disaster management were acquired and connected to this research. The final result presented the research objective 1 and the varying perspectives of VGI in the context of disaster management. Particularly, the benefits and issues of utilizing text-based VGI during the response phase were highlighted.

4.4.2 Design phase

From the conceptual phase, two issues were addressed: (1) a lack of efficient methods for text-based VGI QA, and (2) the ambiguous correlation between quality and the geographic characteristics. The two issues formulated the research objective 2 and 3, validation of the geographic correlation and the (D&E) of the text-based VGI QA methods. To achieve the

two objectives, different methods and techniques were adopted. A research flowchart was designed, and then the scenario and the study area were determined.

4.4.2.1 Design a case study and a research flowchart

To achieve the two research objectives, a case study was designed. A case study method enables a researcher to closely examine the data within a specific context, which is often a small geographical area or a very limited number of individuals as the subjects of study. This method is effective to test the validity and comparability of the constructed artefact within the data collection; it also provides the empirical evidence from the investigation of a contemporary phenomenon within the real-life context (Zainal, 2007). Note that the results of a case study might be limited in specific geographical areas.

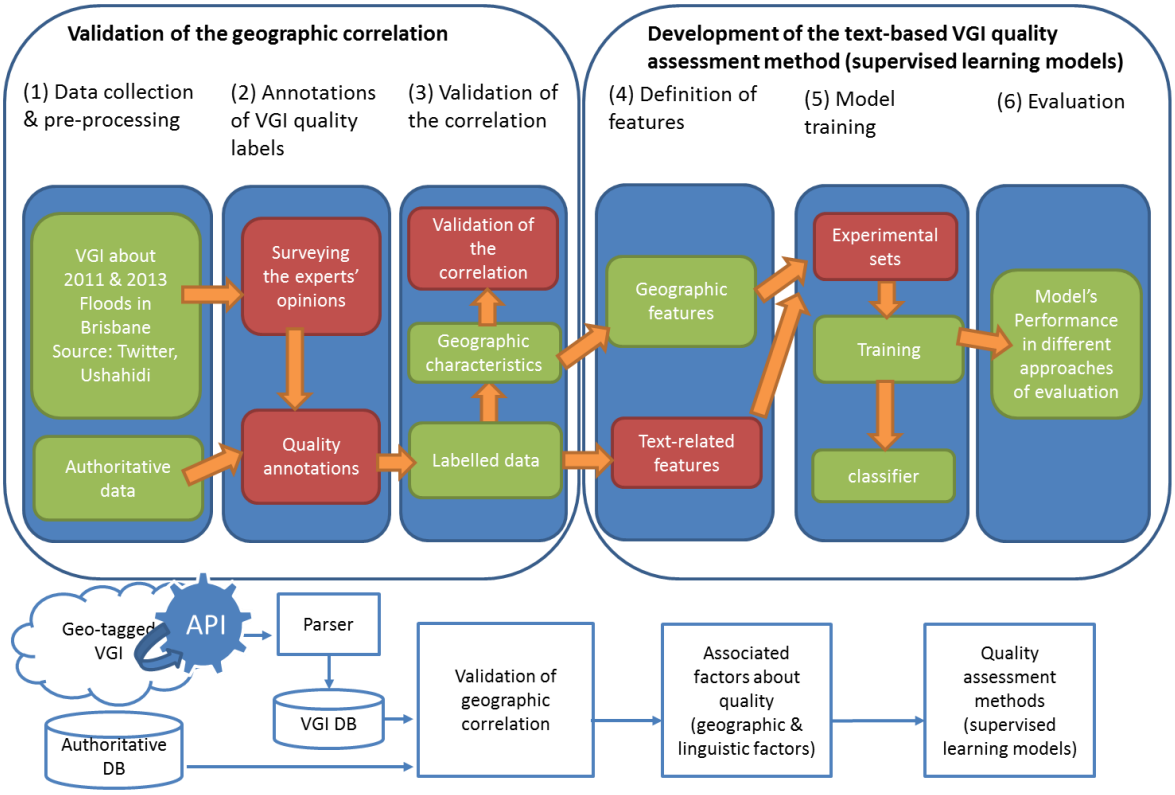


Figure 4.6: Flowchart of the research case study

To complete the case study, a research flowchart was illustrated as the blueprint (Figure 4.6) in this phase. The flowchart categorised the research activities in two parts: (1)

validation of the geographic correlation (i.e. data collection, pre-processing, VGI annotation, and validation of the correlation), and (2) development of the text-based VGI QA methods. The first part defined the scenario and the study area. VGI instances were collected and annotated. Geographic characteristics were analysed for the correlation validation. As it was assumed that the geographic correlation existed, the second part further developed and evaluated the QA methods by using the geographic correlation. Appropriate predictors and algorithms for text-based VGI QA were defined. The evaluation focused on the performance of the QA models by using different settings of training and testing datasets. Research activities are further explained in Section 4.4.3 in detail.

4.4.2.2 Research scenario and study area

This research set the scenario as extreme floods. Flooding was one of the most devastating and common disasters in Australia. Floods caused huge damage in Australia in the past. From 1852 to 2011, at least 951 people were killed by floods and 1326 were injured; the cost of damage reached an estimated $4.76 billion dollars (Carbone & Hanson, 2012). From the literature review in Chapter 2, there had been many cases showcasing how VGI contributed to flood management; VGI can help disaster agencies detect incidents. It also helped some disaster management activities such as the mapping of flooding extent and flood damage estimation (McDougall, 2011; Poser & Dransch, 2010). However, since many research projects used non-geotagged social media feeds (see: Section 0), the geographic correlation was rarely validated and discussed in the literature.

Considering the given scenario and the availability of the geotagged VGI instances, the case study used VGI generated during two previous extreme flooding events in Australia: 2011 and 2013 Queensland Floods. Text-based VGI instances about the two events were collected. Brisbane and its greater metropolitan area were selected as the study area. Figure 4.7 illustrates the administrative zones of Brisbane and all red polygons are the study area.

Brisbane has been built along the Brisbane River, the longest river in South East Queensland, flowing downward from Mount Stanley to Moreton Bay. Brisbane is located in a tropical cyclone risk area. From November to March, thunderstorms are common,

which might be accompanied by heavy rain and destructive winds and cause flooding in severe cases. The Central Business District (CBD) of Brisbane is situated on the northern bank of Brisbane River, during past extreme events, flooding often causes huge damage in the CBD. Floods in the Brisbane area can be categorized into four sources[11]: (1) Creek flooding, (2) River flooding, (3) Overland flow, (4) Coastal flooding.

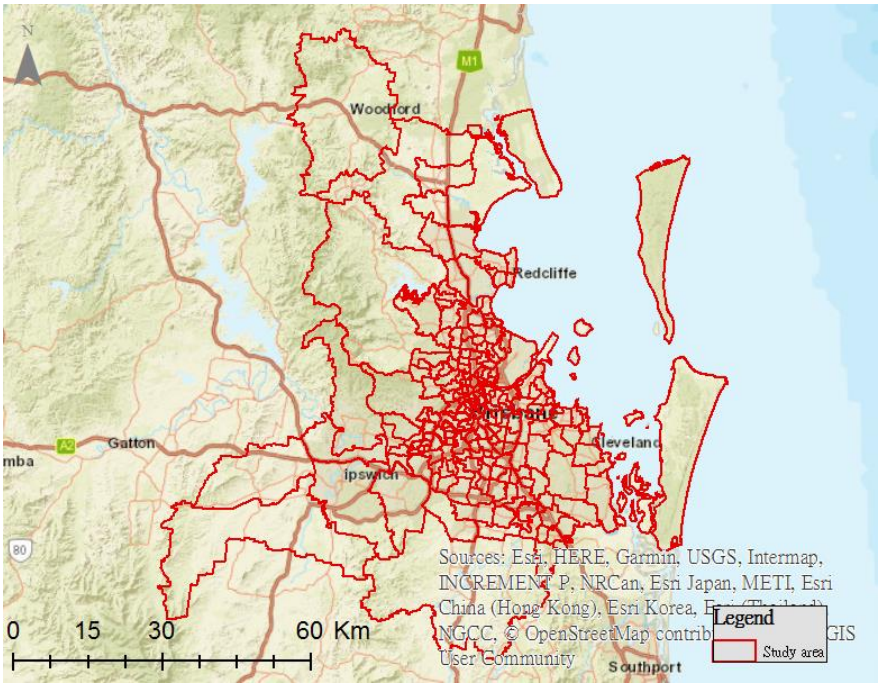


Figure 4.7: The extent of the study area

4.4.3 Development and Evaluation phase

The D&E phase initialized by collecting and preparing data. The geographic correlation was then validated and several experiments based on Supervised learning were further developed and evaluated. The research activities in this phase were:

1. collecting VGI and authoritative data and identifying data issues;

[11] Source: Flooding in Brisbane, a guide for residents, http://www.brisbane.qld.gov.au/sites/default/files/flooding_ in_brisbane_guide_for_residents.pdf, accessed 15 October 2015

2. pre-processing and filtering the collected VGI instance;
3. annotations of the quality labels;
4. defining predictors/features;
5. evaluation of the geographic correlation;
6. development and evaluation of quality assessment methods.

The concepts and the research activities are presented in the following subsections. The associated techniques of the experiments are described in Section 4.5. The implementation of these research activities will be further described in Chapter 5 and Chapter 6.

4.4.3.1 Collecting VGI and authoritative data and identifying data issues

This research proposed to collect two kinds of text-based VGI instances generated during 2011 and 2013 Queensland Floods, which are (1) Ushahidi reports and (2) tweets. The collection method can use the associated API (Application Programming Interface) and tools. An architecture for collecting text-based VGI instances was designed and developed. During the data collection process, some issues were found, such as tweets not related to the floods (off-topic or spam-related contents).Thus several programmed filtering processes and a manual filtering process were designed and used in the next step.

Moreover, mismatched positioning was found. Mismatched positioning was critical for the validation of the geographic correlation. The reason for mismatching was that both Ushahidi reports and tweets may have at least two geo-location identities: “positioning” (i.e. geographic coordinates) and “location description (or toponyms)”. The positioning was generated by different methods. In the Ushahidi platforms, the location description of a report was submitted by contributors by manual typing, and the positioning could be marked in the map manually. However, the positioning might not match the location description (i.e. the distance between the positioning and the extent of the toponym was too long thus the actual position of an event was not clear) Besides, a part of the positioning was not manually marked; it was found some Ushahidi reports compulsorily registered in the centre of the map by the system (i.e. the default extent set by the platform manager). In the case of tweets, the positioning was generally attached by GPS-enabled devices. As VGI

contributors were often moving, and sometimes people in non-affected areas may provide information about affected areas, the positioning of tweets may neither match the location description in the texts.

In the abovementioned situations, the location description presented the correct information that the contributors intended to report, and the positioning not matching to the location should be considered as a "mismatched" error. Mismatching resulted in several issues such as an inconvincible validation. Although location inference and geocoding of non-geotagged data had been discussed widely in the literature, uncertainty still exists in these techniques (see: Section 0). Concerning this issue, non-geotagged VGI instances were excluded in the data collection and a further filtering process was designed.

In addition to the VGI collection, static authoritative data related to the flooding events were collected for several purposes (see: Section 5.3). But Real-time information such as in-situ water level measurement was not used, as the collection and use had different issues from the current approach such as lack of access and interoperability, the availability, and the processing efforts in the response phase (Alamdar, Kalantari, & Rajabifard, 2016)

4.4.3.2 Pre-processing and filtering the collected VGI instances

This step aimed to pre-process and filter the collected data for the subsequent steps. First, VGI instances without any text were excluded. Next, to focus on the issue of credibility and text content quality, only flood-related VGI instances were retained. Relatedness of the collected VGI instances was programmatically and manually determined in this phase. Two programmed methods, keyword- and location-based filtering were developed and used in this step. These methods had been used in the studies about the relatedness judgment (e.g. Craglia et al., 2012; Laylavi, Rajabifard, & Kalantari, 2017).

After the relatedness judgement, concerning the mismatched positioning issue, a data cleaning process was designed. The data cleaning process developed a hierarchical toponym resolution, which located the toponyms into a variety of geometries (i.e. point, polyline, polygon). The hierarchical toponym aimed to solve the ambiguity of place name in a straightforward method (Amitay et al., 2004). The distance between the positioning

and the geo-referenced object was further calculated to exclude mismatched VGI instances. The details of the data cleaning process are explained in Section 5.4. Finally, the pre-processing integrated the two kinds of text-based VGI instances into a point layer for facilitating the experiments.

4.4.3.3 Annotations of the quality labels

As described in Section 1.3, *Research aim*, the VGI quality problem was treated as a classification problem. The quality labels were the values of the outcome variables for the supervised learning experiments. However, the collected VGI instances had no attributes of credibility and text content quality. How to determine the quality labels was critical for the development of QA methods. The annotation process was similar to the "coding" process of social scientists. In this research, it aimed to reflect the need for information managers in disaster response to address the context of this research.

In the existing works, the crowdsourcing method was often used in the annotations (e.g. Imran et al., 2013; Olteanu, Castillo, Diaz, & Vieweg, 2014). However, the author found the crowdsourcing annotations still have uncertainty. Mistaken or inconsistent annotations were found in CrisisLexT6[12] (Olteanu et al., 2014) by the author. And the information requirement during the response of disaster agencies was not addressed in the crowdsourcing. Concerning this, two emergency management experts[13] were invited to rate the quality of the collected VGI examples and provide their opinions for the annotation.

---

[12] For example, (1) the annotations were inconsistent in the tweet published by a bot of Modis flood map (Modis Flood Map Server processed…). (2) A tweet related to the impact of flood with a picture was annoted as off-topic (This pic was our driveway! We are able to get out via another road tho! #qldfloods http://t.co/dAwq48O4)

[13] They two experts are David Williams and Greg Ireton. David Williams is a retired inspector of Victoria Police and a senior research advisor in the Centre for Disaster Management and Public Safety (CDMPS). Greg Ireton is a disaster recovery advisor and worked in Department of Human Services, Victoria for Health and Human Services Emergency Management.

The practical experiences in VGI quality judgment and significant insight were gathered in comparison to the crowdsourcing method. The experts' participation formulated a rigour mechanism and then the collected VGI instances were manually annotated by the author. The number of the annotated VGI instances reached 901 (see 5.6).

In general, the credibility annotation process considered the trustworthiness of collected VGI instances and supporting references for binary classification (low- and high-credibility). On the other hand, the text content quality used several weighted criteria to design the annotation process. The text contents were categorised into three levels. All quality labels are mutually exclusive.

4.4.3.4 Defining predictors/features

From the related work of Spinsanti and Ostermann (2013) and Albuquerque et al. (2015), geographic characteristics had a correlation to the text-based VGI quality. How to define the scope of the geographic characteristics as predictors in the developed method was an important task. This step defined the predictors for the D&E of QA models. "Predictor" is an individual measurable property of a phenomenon being observed in the context. "Feature" is a synonym of "predictor" in Supervised learning. There are two kinds of predictors used in this research, geographic predictors and text-related predictors. Geographic predictors are generated by the integration of VGI and authoritative data such as flood-prone areas. The used of geographic predictors aims to discover knowledge about the "geographic method" in the occurrence of text-based VGI. Note that the collection of authoritative data were based on the availability and the author's assumption.

On the other hand, text-related predictors such as the n-gram model had been widely used in the text classification and shown good capability (Carpenter, 2005; Jurafsky & Martin, 2000; Ye, Zhang, & Law, 2009). The "integrated method" tested both geographic predictors and text-related predictors; its main objective was to test the performance variance compared to the established approach that used textual contents only in the modelling.

4.4.3.5 Evaluation of the geographic correlation

This step examined and validated the "geographic assumption" by evaluating the correlation between VGI quality and the geographic predictors for achieving research objective 2. The concept is similar to Predictor/feature selection (see: Section 4.5.2.2). The Pearson correlation coefficient and the Pearson's chi-squared test were used for the validation. If the correlation existed, the QA method can be formulated.

4.4.3.6 Development and evaluation of the quality assessment method

This step sought to develop and evaluate the text-based VGI QA methods. Supervised learning algorithms were used in the experiments to find appropriate models for the geographic method (i.e. using the available geodata as the extended predictors of text-based VGI QA models) and the "integrated method" (using both geographic predictors and the text-related predictors). With the various experiment design, the performances are recorded. The models' performances are evaluated. The metrics used in the performance evaluation are accuracy, precision, recall, and Area under the ROC Curve (AUC). The definitions of these performance metrics are described in the next section.

Furthermore, during the process of training and evaluation, generalisation is an important concern (see: Appendix 2). ML has an assumption in the generalisation that the distribution of population does not change over time. However, in practice, the distribution may change and the model may not work well in "unseen data (i.e. data not used in the training)". This issue formulated a 2-step evaluation in the experiments. The experiments were designed in two ways: (1) training and evaluation by a single dataset through k-fold cross-validation (i.e. unseen data is from the same event), (2) training and evaluation by using the different event datasets (i.e. unseen data is from the different event).

4.4.4 Synthesis and Communication phase

The final phase was to synthesize the research results and findings in the documents and to present them to interested audiences for communication. The instantiations included this thesis and the journal or conference publications during the development of this research; the journal and conference publications were described in the **List of publications**.

The arrangement of the chapters of this thesis generally followed a schema demonstrated in Table4.4 (adapted from Gregor & Hevner, 2013). It outlined a common structure for communicating a design science-based project. As a publication schema, it was similar to a conventional paper in natural or social science.

Table 4.4: Publication schema for a design science research study

Section	Contents
1. Introduction	Problem definition, motivation, introduction to key concepts, research questions/objectives, scope of study, overview of methods and findings, theoretical and practical significance, structure of remainder of the paper. The problem definition and research objectives should specify the goals that are required of the artefact to be developed.
2. Literature Review	Prior work that is relevant to the study, including theories, empirical research studies and findings/reports from practice. The prior literature surveyed should include any prior design theory/knowledge relating to the class of problems to be addressed, including artefacts.
3. Method	The employed research approach. The specific approach adopted should be explained with reference to existing authorities.
4. Artefact description	A major part of the publication describing the artefact at the appropriate level of abstraction to make a new contribution to the knowledge base. The format should include at least the description of the designed artefact and, perhaps, the design search process.
5. Evaluation	Evidence that the artefact is useful. The artefact is evaluated to demonstrate its worth with evidence addressing criteria such as validity, utility, quality, and efficacy.
6. Discussion	Interpretation of the results including a summary of what was learned, comparison with prior work, limitations, significance, and areas requiring further work. Research contributions are highlighted.
7. Conclusions	Concluding paragraphs that restate the important findings of the work. Restates the contribution and why they are important.

4.5 Associated techniques for the development and evaluation of quality assessment methods

This section describes the concepts of the techniques used for the development of the QA methods, the general methodology of Supervised learning, and the programming libraries and tools.

4.5.1 Statistical and geospatial statistical techniques

In this research, following statistical and geospatial statistical techniques were used to evaluate the geographic correlation and generate geographic predictors. First, the clustering of VGI was assumed to correlate to its quality[14]. To validate this assumption, the Optimized Hot Spot Analysis (OHSA) and the Mean nearest neighbour analysis are used.

1. Optimized hot spot analysis (Getis-Ord Gi*)

The OHSA was used to observe the clustering of VGI. The analysis creates hot and cold spots of the data, which can help to find the areas impacted by floods seriously during the two events.

The OHSA used Getis-Ord Gi* statistics. Getis-Ord Gi* statistics can be described as: Given an area of n suburbs where each suburb is identified with $i(1, 2, \ldots., n)$. Each i is associated with an attribute X_i, which is the count of the VGI instances within the suburb. When we focus on a particular X_i, the remaining attributes in other suburb are denoted as X_j. The Getis-Ord Gi* statistics is established by specifying a null hypothesis that there is no association between the value of X at the suburb i and its neighbours. The null hypothesis states that the sum of values at all the j suburbs within radius d of i is not more or less than one would expect by chance given all the values in the entire study area. The Gi* statistics sum the value of X_i with the j values within d of i. If there is a spatial

[14] Craglia et al. (2012) assumed that VGI quality correlates to the spatial proximity (see: Table 3.1), but this assumption has been not validated in the literature.

correlation, it will be exhibited by spatial clustering of the high or low value of X, and results in positive Gi* or negative Gi*. The form of the Getis-Ord Gi* statistics can be given as:

$$G_i^*(d) = \frac{\sum_{j=1}^{n} w_{ij}x_j - \overline{X}\sum_{j=1}^{n} w_{ij}}{S\sqrt{\frac{n\sum_{j=1}^{n} w_{ij}^2 - (\sum_{j=1}^{n} w_{ij})^2}{n-1}}} \quad (1)$$

Where G_i^*(d) is the z-score to examine the spatial dependency of the region i over all n regions (i.e. suburb of Brisbane area in this research), X_j is the magnitude of variable X at the region j; and w_{ij} is weight value between region i and j that represents their spatial interrelationship.

$$\overline{X} = \frac{\sum_{j=1}^{n} x_j}{n} \quad (2)$$

$$S = \sqrt{\frac{\sum_{j=1}^{n} x_j^{\,2}}{n} - \left(\overline{X}\right)^2} \quad (3)$$

2. Mean nearest neighbour analysis

The nearest neighbour analysis is based on the assumption that the point (i.e. the VGI instances) being measured are free to locate anywhere within a specific area (i.e. a minimum bounding box). It performs a nearest neighbour search and calculates a nearest index based on the mean Euclidean distance from each point to its nearest neighbour. The index is given as:

$$Rn = \frac{\bar{D}_O}{\bar{D}_E} = \frac{\sum_{i=1}^{n} di}{0.5\sqrt{\frac{A}{n}}} \quad (4)$$

Where $\overline{\mathrm{D}}_{\mathrm{O}}$ is the observed mean distance between each point and its nearest neighbour, and $\overline{\mathrm{D}}_{\mathrm{E}}$ is the expected mean distance for the points given in a random pattern; A is the study area, and n is the total number of points. The interpretation of the nearest neighbour analysis is determined by the index Rn where if Rn is less than 1.0, the distribution exhibits

clustering. Conversely, the distribution is toward dispersion. The statistical significance of the analysis result has to be considered. For example, using a 95% confidence level, if the z-score (a measure of standard deviation) is between -1.96 and +1.96 and the p-value is greater than 0.05, the null hypotheses of random distribution cannot be rejected.

3. Pearson correlation coefficient and Pearson's chi-squared test

To evaluate the correlation of predictors/features, the Pearson correlation coefficient (PCC/*r*) and the Pearson's chi-squared test (i.e. chi-squared test) were used. PCC is a measure of the linear correlation between two variables. It has a value between +1 and −1, where 1 is a total positive linear correlation, 0 is no linear correlation, and −1 is total negative linear correlation. It is widely used in the sciences. According to L. H. Cohen (1988) the following table presents the accepted guidelines for interpreting the correlation strength.

Table 4.5: Interpretation of the correlation strength

Correlation	Positive	Negative
Very Weak/NoCorrelation	-0.1 to 0.0	0.0 to 0.1
Weak	-0.1 to -0.3	0.1 to 0.3
Moderate	-0.3 to -0.5	0.3 to 0.5
Strong/Very Strong	-0.50 to 0.1	0.5 to 1.0

On the other hand, the chi-squared test is often used for testing relationships between categorical variables. It assumes a null hypothesis stating that the predictor and quality class are independent; and an alternative hypothesis stating that the predictor and quality class are not independent. By calculating the chi-squared statistics with the selected level of confidence, if the null hypothesis is accepted, the predictor does not have a significant association with quality class. On the contrary, if the null hypothesis is rejected, the predictor has a significant association with quality class.

4.5.2 Supervised learning

Supervised learning is used in the development of the geographic method and the integrated method in this research. Supervised learning focuses on finding how-to knowledge from training data to make classification accurately. A dataset consists of instances with both predictors and output variables is used. The task is to construct an algorithm, which is able to predict the output variable with the set of predictors. The general supervised learning process in this research is illustrated in Figure 4.8. It begins with the data quality problem and then constructing the foundation from data preparation and annotation. Next, it comprises four steps: (1) predictor/feature generation and extraction, (2) predictor/feature selection, (3) algorithm selection, (4) model training and evaluation. Based on these process steps, models are trained and can be further evaluated.

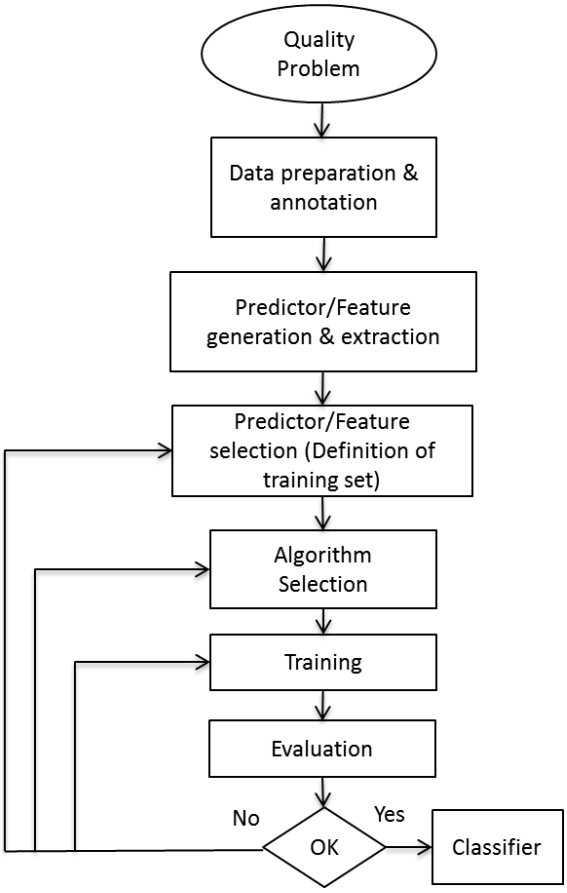


Figure 4.8: Supervised learning process used in this research
(adapted from Kotsiantis, Zaharakis, & Pintelas, 2007)

4.5.2.1 Predictor/feature generation and extraction.

Predictor/feature generation is the process of taking raw, unstructured data and defining predictors for potential use in the Supervised learning model. Predictor/feature extraction is to further test transformations of the generated predictors and select a subset of original and derived predictors for use in the model. The regular techniques of extracting texts into predictors will be used in this research.

4.5.2.2 Predictor/feature selection (evaluation of the significance)

Predictor/feature selection defines the sets/subsets of predictors used in Supervised learning. This step identifies relevant predictors and removes irrelevant and redundant predictors. It is an important process to simplify the model to make an interpretation and enhance the performance (Yu & Liu, 2004).

Predictor selection has a robust methodology in the literature. Generally, predictor selection has three objectives: (1) improving the performance of the model, (2) providing faster and more cost-effective predictors, and (3) providing a better understanding of the underlying process that generated data (Guyon & Elisseeff, 2003). There are a variety of methods for the evaluation. Variable ranking is often used as a baseline method. It applies the statistical measure to assign a predefined metric for ranking predictors, which has advantages of simplicity, scalability, and good empirical success. Examples of the variable ranking methods include the correlation coefficient, the chi-squared test, and information gain.

Note that variable ranking is a filter method. According to Kohavi and John (1997), filter methods select subsets of variables. The metric is independent of the deployed algorithms. The most significant *n* predictors ranked by a filter are not the optimal *n* predictors for an algorithm. The predictor ranking is relative to the defined metric, rather than an optimal set of predictors for the ML algorithm. This is because the predictor interactions are not considered in the filter. Besides, as the metric is independent of the algorithm, the good predictors consider by a filter may be bad predictors for the algorithm on the task.

4.5.2.3 Algorithm selection

Algorithms are techniques for estimating the target function *(f)* to predict the output variable *(y)* given input data *(x)*. Different algorithms make different assumptions about the shape and structure of the function and how best to optimize a representation to approximate it.

The commonly used algorithms in classification can be categorized into three types, which are linear, nonlinear, and ensemble algorithms. Linear algorithms such as Logistic regression assume that the predicted attributes are a linear combination of the input variables. On the other hand, nonlinear algorithms do not make strong assumptions about the relationship between the input variables and the output attribute. They might be more accurate in classification problem but they might learn too well and overfit training data (i.e. overreacting to training data and the prediction on unseen data is not accurate). The popular nonlinear algorithms are Decision Tree and Naïve Bayes. Finally, ensemble algorithms are a powerful and more advanced type, which are techniques that combine the predictions from multiple algorithms in order to provide accurate predictions. This research has reviewed five commonly used algorithms for the classification and documented their advantages and disadvantages in Appendix 2. Among these algorithms, Logistic Regression and Decision Tree (J48) were selected to train the models of the geographic method and the integrated method as the modelling results were both easy to interpret.

4.5.2.4 Model training and evaluation

With the selected predictors and the selected algorithm, a model is trained. The model can be evaluated by several performance metrics. The metrics used in this research and the main evaluation method are described in the next section. If the performance of the model is not good, the experiments can be redesigned. Then the model is re-trained, and the output is evaluated again until a satisfactory performance is reached.

Through the Supervised learning process, the trained models are expected to recognize predictors for text-based VGI QA with satisfactory performance. However, it is noteworthy this process is data-driven, and sometimes it cannot reach a satisfactory result. This might

be due to the use of irrelevant predictors, data size, and imbalanced data. To avoid the issue of imbalanced data, an oversampling technique: SMOTE (Synthetic Minority Over-sampling Technique) is used; the concept of SMOTE will be described in Chapter 6.

4.5.3 Performance metrics of model evaluation

There are five classification performance metrics used in the evaluation of the developed methods: accuracy, precision, recall, F_β-Score, and AUC. These metrics were often used in the evaluation of models (Davis & Goadrich, 2006; Goutte & Gaussier, 2005). The concept are described as follows: when using a binary classification as an example where the real label y = (+1, -1), the model predicts each instance as y' = (+1, -1). For each prediction, there are four possible outcomes as:

- True-Positive (TP), if $y = +1$ and $y' = +1$,
- False-Positive (FP), if $y = -1$ and $y' = +1$,
- False-Negative(FN), if $y = +1$ and $y' = -1$,
- True-Negative (TN), if $y = -1$ and $y' = -1$,

Given a test set of N labelled samples, the performance metrics are explained as follows:

1. Accuracy

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+TF+FN (N)}}$$

Accuracy is the ratio of the predicted results correct matches to the labels, which is plausible for evaluation of the performance of a classifier. However, accuracy is only appropriate to be used in the case of the two classes with equally distributed samples. In imbalanced class case, if most annotated samples belong to the positive class, and the classifier always predicts the positive label independently, then its accuracy will appear high. The accuracy cannot be used for the reference as the model can classify most minority instance as the majority class and still has high accuracy. Due to this issue, precision and recall are required to be further considered.

2. Precision and Recall

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}}, \text{Recall} = \frac{\text{TP}}{\text{TP+FN}}$$

Precision and Recall are two complementary metrics. Precision is the ratio that how accurate are the predictions of the positive samples of the classifier; it is also called Positive Predictive Value. Recall measures how many of the positives samples were discovered; it is also called True Positive Rate (TPR) or Sensitivity[15]. Precision and Recall have an inverse relationship, and they should be used together to indicate the overall performance of the classifier. *In simple terms, high precision means that an algorithm returned substantially more relevant results than irrelevant ones* (highly-sensitive), *while high recall means that an algorithm returned most of the relevant results* (highly-specific) [16]. An ideal classifier is able to find in high precision and recall but one metric is often higher than another in a real case. Therefore, users can choose highly-sensitive models or highly-specific models according to the facts of the particular situation.

3. F_β-Score (F1)

$$F_\beta = (1 + \beta^2)\frac{\text{Precision} \cdot \text{Recall}}{\beta^2 \cdot \text{Precision+Recall}}$$

F_β-Score combines the Precision and Recall in a single measurement. The β parameter is a weight among Precision and Recall depending on the problem under study. F1 score (β=1) is usually applied in the evaluations of classifiers. It treats Precision with the same weight as Recall.

[15] On the other hand, specificity is the true negative rate, which measures the portion of actual negatives that are correctly identified.

[16] https://en.wikipedia.org/wiki/Precision_and_recall

4. AUC (Area under ROC Curve)

AUC is often used to determine the optimized model in the literature. AUC is the area under the Receiver Operating Characteristic (ROC) curve. The ROC curve is a two-dimensional graph used as an evaluation metric in a classification. In a ROC curve, True Positive Rate is plotted on the Y-axis and False Positive Rate is plotted on the X-axis at various threshold settings (Figure 4.9, Fawcett, 2006). The lower-left corner (0, 0) represents a threshold of 1 whereas the upper-right corner (1, 1) represents a threshold of 0. The ideal point (0, 1) represents perfect classification as the true positive rate is 100% and the false positive rate is zero, but most classifiers merely predict close to perfect classification. It has been assumed that in ROC curves the optimal classifier point is the one that maximizes the sum of True Positive Rate and True Negative Rate (i.e. 1 - False Positive Rate or Sensitivity) (Zweig & Campbell, 1993).

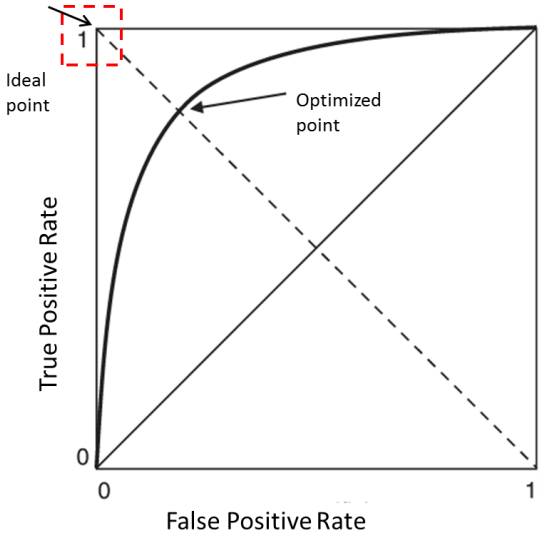


Figure 4.9: An example of ROC curve

AUC is often used for performance assessment and classifier comparison. The AUC of a classifier is equivalent to the probability that the classifier ranks a randomly chosen positive instance higher than a randomly chosen negative instance. Theoretically, the AUC is 0.5 of a random guess is 0.5. The results of AUC were considered excellent for AUC values

between 0.9-1, good for AUC values between 0.8-0.9, fair for AUC values between 0.7-0.8, poor for AUC values between 0.6-0.7 and failed for AUC values between 0.5-0.6 (Lüdemann, Grieger, Wurm, Wust, & Zimmer, 2006).

In this research, the value of the weighted average AUC was used to compare the performances of the models built by Supervised learning algorithms. A model with the highest AUC can be regarded as the optimized one and recommended to be implemented in the disaster information system. Further discussion regarding the benefits and disadvantages of the evaluation metrics is presented by Bekkar, Djemaa, and Alitouche (2013).

4.5.4 Evaluation method: k-fold cross-validation

The evaluation of the built model in classification problem is usually through K-fold cross-validation. K-fold cross-validation is a popular approach and often used as the gold standard in the evaluation of ML models. Figure 4.10 demonstrates the procedure of k-fold cross-validation. It divides the data into k folds and uses k-1 folds as the training data to build a model and then evaluates the model on 1 fold. For example, 10-fold cross-validation means 9 folds (90%) were used for training and 1 fold (10%) for testing. The cross-validation process repeats k-times on different subsets and building up an estimate of the average performance of a model. In comparison with a random splitting of data (e.g. 70% of data as the training dataset and 30% as the testing dataset), k-fold cross-validation is generally the most widely used way to evaluate the model's generalisation performance as it uses the existing data to simulate the process of generalizing to new data[17]. Note that the final model is still built on the full data in Weka.

[17] Referenced from "Evaluatin statistical models", a lecture note of Cosma Shalizi's "Advanced data analysis" class of Carnegie Mellon University. Available from: http://www.stat.cmu.edu/~cshalizi/402/lectures/03-evaluation/lecture-03.pd

Figure 4.10: The procedure of k-fold cross-validation

4.5.5 Programming libraries and Tools

Several programming libraries and tools are used in this research. First, for the collection of Ushahidi data and the flood-related tweets, the Ushahidi public API and an unofficial Java library based on Twitter API, Twitter4J[18] were used. The use of the Ushahidi public API did not require any authentication for accessing the volunteered reports of Ushahidi platforms. On the other hand, Twitter4J requires to set the access key and token of a specific account. By using these tools, raw data on the server can be extracted by sending specific parameters in HTTP requests.

Next, for the processing/data cleaning of the collected VGI, a geoparser, CLAVIN-NERD[19] was used. A gazetteer of 450 Brisbane suburbs was built based on the gazetteer format used in CLAVIN-NERD. ESRI ArcGIS was used to manage the VGI database, analysis (e.g. the Optimized Hot Spot Analysis, OHSA), and to generate specific geographic predictors.

[18] http://twitter4j.org/en/index.html

[19] https://github.com/Berico-Technologies/CLAVIN-NERD

As for the text-related predictor extraction, two NLP tools, LightSIDE[20] and Stanford CoreNLP[21] were used to analyse the textual characteristics of VGI. LightSIDE is developed by the research of Computational Linguistics & Psycholinguistics (CLiPS) centre at the University of Antwerp. It was used to extract the unigram, bigram, and Part of Speech bigrams from VGI instances. The existence of some specific name entity is analysed by Stanford CoreNLP.

Finally, the modelling tool used in this research is Weka[22] (Waikato Environment for Knowledge Analysis). Weka is a modern platform developed at the University of Waikato, New Zealand for applied ML under the GNU GPL. Weka is written in Java and provides a well-documented API. The advantage of Weka is its graphical interface, which is convenient to develop ML project without programming. And all functions of Weka can be used from the command-line interface. Note that this research did not aim to optimize the performance of a specific algorithm so the default model configuration was used.

4.6 Chapter summary

The methodology and the design to conduct this research project were explained in detail in this chapter. At first, a conceptual framework is described to showcase the relationship between the research context, elements, methods, and outcomes. Next, for the development and evaluation of the text-based VGI QA method, DSRM was used. The concept and guideline of DSRM were briefly reviewed for employing DSRM. The research design for achieving these objectives consists of four phases. Research activities in each phase were explained. Having the research methodology and design clarified in this chapter, the next

[20] http://www.cs.cmu.edu/~cprose/LightSIDE.html

[21] Stanford CoreNLP, http://stanfordnlp.github.io/CoreNLP/

[22] http://www.cs.waikato.ac.nz/ml/weka/

chapter presents the research activities for data preparation and the validation of geographic correlation.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 5

Data collection, processing, and the validation of the geographic correlation

5.1 Introduction

This chapter intends to address the research objective 2, validating the correlation between the text-based VGI quality and the geographic characteristics of text-based VGI. The first part describes the data including the VGI including the VGI collection and the authoritative data collection. The second part presents the data preparation including several pre-processing and filtering process. An innovative data cleaning process is specifically introduced. Next, the annotation processes of VGI instances are addressed. To reflect the quality need of disaster information managers in the context of flood response, an investigation was conducted to the emergency management experts. Based on the investigation, the annotation processes were designed. All VGI instances were annotated with a credibility label and a text content quality label for the subsequent modelling method development. Finally, the correlation between the geographic characteristics and quality of VGI is validated by the spatial and statistical analysis.

5.2 VGI collection

This section describes the VGI collection used as the foundation of this research. It first explains the architecture for collecting the VGI instances. The properties and amount of these instances are further described.

5.2.1 Architecture and tools used in the process VGI collection

The architecture of the VGI collection process is illustrated in Figure 5.1. This process collected VGI related to two extreme flooding events in Brisbane, the 2011 floods (January to February 2011) and the 2013 floods (17th January to 29th January 2013) from the Ushahidi platforms and Twitter. The VGI instances were called “2011 event dataset” and “2013 event dataset”. Using the Ushahidi API and Twitter4J, the VGI instances were extracted in JSON (JavaScript Object Notation) format and were further translated to a geodatabase for storage and analysis.

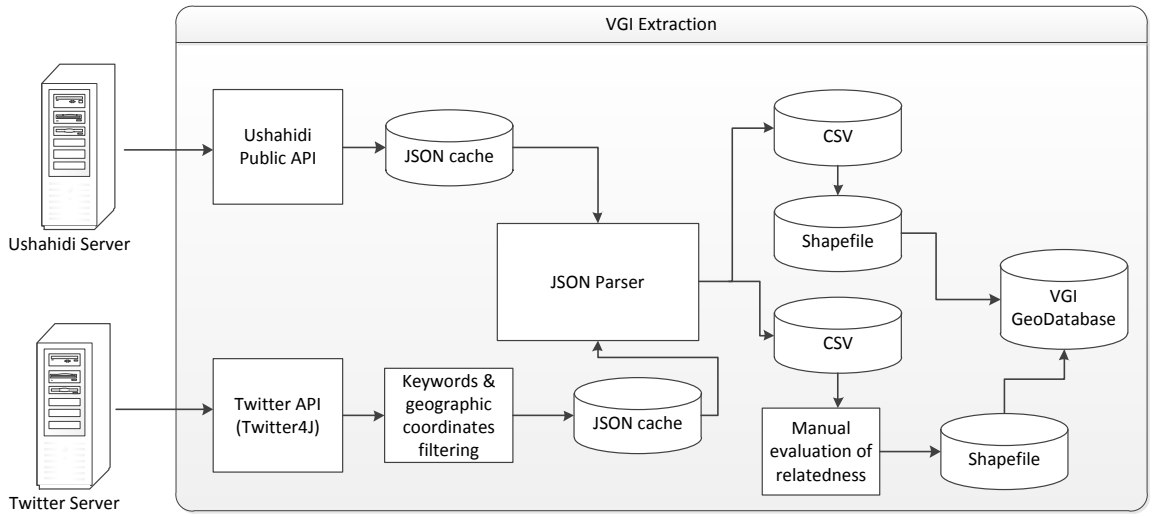


Figure 5.1: The architecture of the VGI collection process

5.2.2 Collecting and filtering of the flood-related tweets

By using Twitter4J, tweets generated during the 2011 and 2013 flooding events respectively were collected and filtered.

5.2.2.1 Relatedness judgement and the amount of the retained tweets

To collect the flood-related tweets, two kinds of programmed filtering were used and only geotagged and original tweets were retained. First, the keyword-based filtering searched tweets by using three keywords: "flood" or "#qldflood" or "#bneflood" regardless of capitalisation. Next, the location-based filtering picked tweets in a given area of a circle by the location of the user's profile and exclude non-geotagged data. The area within a radius of 75 km to the centroid of the Brisbane metropolitan area (i.e. approximately the smallest circle of the study area) was used.

The keyword-based filtering collected 10,920 tweets. After the location-based filtering, 273 geotagged tweets generated during the two events were extracted (10647 were excluded). The number is few so off-topic contents or spam-related contents (e.g. flood insurance advertisements) can be further manually filtered by the author. The manual filtering categorised the tweets as (1) "on-topic" and (2) "off-topic" to ensure the relatedness of information. "On-topic" tweets are tweets related to situational awareness.

In contrast, “off-topic” means the tweets did not relate to the flooding event and situational awareness. Off-topic tweets were excluded in the subsequent experiments. 242 tweets were identified as on-topic and retained. Table 5.1 demonstrates two examples.

Table 5.1: Examples of on-topic and off-topic tweets in the collection process

Classification	Example
On-topic	“Looking towards the botanical gardens #qldfloods #bnefloods #qldfloodsmap”
Off-topic	“if I have to watch our premier one more time talking to us like we are idiots, I am going to combust. #qldfloods #qldpol”

5.2.2.2 Properties of the collected tweets

For all the collected tweets, following properties of tweets were extracted: “Tweet Id”, “User name”, “Textual content”, “Created at (date and time)”, “Geographic coordinates”, and “Number of retweets”. A detail description of the properties can be found in the Javadoc of Twitter4J. The associated methods and return data type were listed in Table 5.2

Table 5.2: Properties of tweets extracted by Twitter4J

Properties	Jave data type	Twitter4J Method	Description
Tweet Id	long	getId()	The unique id of a tweet
User name	java.lang.String	getUser().getScreenName()	User’s name on Twitter (e.g @KuochihHung)
Textual content	java.lang.String	getText()	The text in a tweet

Properties	Jave data type	Twitter4J Method	Description
Created at	Java.util.Date	tweet.getCreatedAt()	The date and time when the tweet was created.
Geographic coordinates	GeoLocation (latitude, longitude, serialVersionUID)	getGeoLocation()	The location that a tweet refers to (if available).
Number of retweets	int	getRetweetCount()	The number of times this tweet has been retweeted

5.2.3 Collecting and filtering of the Ushahidi reports

The Ushahidi reports were collected by using the Ushahidi public API. These reports were published in the platforms managed by Australian Broadcasting Corporation [23] and Brisbane City Council[24] during the 2011 and 2013 flooding events respectively; the reports were submitted from a heterogeneous range of volunteers via various channels (i.e. the web, e-mail, and Twitter).

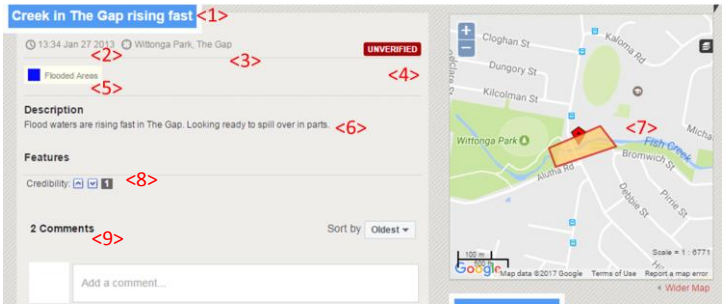
5.2.3.1 Properties of the collected Ushahidi reports and limitations

The properties of the Ushahidi report are illustrated in Figure 5.2. Most properties of a report were accessible and can be extracted by the public API. However, the public API has some limitations. It could not extract the value of participants' credibility ratings and the comments from the Facebook plugin. Besides, in the use of Ushahidi, the contributor could digitize one or multiple features (i.e. point, polyline, and polygon) on the map to

[23] http://queenslandfloods.crowdmap.com/

[24] https://bnestorm.crowdmap.com/

represent the location of an event. Yet the API can only return a centroid of the spatial object(s). And the attached media like photos and video could not be extracted.



<1> Title

<2> Reported time

<3> Location description

<4> Verification status

<5> Category

<6> Textual description

<7> Centroid

<8> Participants' credibility ratings

<9> Comments from the Facebook plugin

Figure 5.2: The properties of a Ushahidi report

5.2.3.2 The amount and the original categorisation of the Ushahidi data

Summarising the collected data, the platform built for the 2011 event contained 1,329 reports. These reports were tagged in 7 categories during the response phase (from 9th January to 15th January 2011): (1) property damage, (2) roads affected, (3) hazards, (4) electricity outages, (5) evacuations, (6) help services (relief centres, hospitals, police stations, others), and (7) trusted reports. By contrast, the Ushahidi platform built for the 2013 flooding event had fewer reports. It owned 480 reports tagged into four categories: (1) flooded area, (2) closed roads, (3) power outage, and (4) sandbag locations. The categorisation was not mutually exclusive so a report can have more than one tag.

To check did spam-related contents exist among the collected Ushahidi reports, the author sampled and inspected five reports from each category. This manual inspection found that all the sampled reports were not spam-related. Since the Ushahidi reports were examined by the volunteers during the events, most contents should relate to the flooding events. However, considering that the reports of "evacuations", "sandbag locations", and "help services" were facilities not impacted by the floods; these reports were assumed to have

weak geographic correlations and thus excluded. Only reports about flood incidents were retained. Additionally, the research design limited the study area in the Brisbane metropolitan area. Thus the reports located outside the study area were excluded. Non-geotagged were also removed. By this filtering process, 811 reports generated during the 2011 event and 389 reports generated during the 2013 flooding event were collected.

5.2.4 Integration of Ushahidi reports and Twitter feeds

This research integrated the collected Ushahidi reports (n = 1200) and the tweets (n =273) as one input dataset for the Supervised learning task. Their similar properties were mapped into data schema and then stored as one table in a geodatabase (ESRI personal geodatabase). The integrated data schema was presented in Table 5.3. Each VGI instance was assigned a short unique ID (OBJECTID). Some property values were Null in this step of integration and further processed in the subsequent steps (e.g. "Toponyms" and "Suburb" of the collected VGI were further identified in the data cleaning process). Moreover, as the tweets did not own a property of thematic group like Ushahidi, they were manually examined by the author and re-categorized into four thematic groups, which are listed in Table 5.4.

This section has described the VGI collection process and the programmed/manual filtering process. After the integration, the Ushahidi data accounted for 83% and tweets accounted for 17% of the collected VGI instances.

Table 5.3: Data schema of the integrated VGI collection

Properties	Data type (GIS)	Description
OBJECTID	Object ID	The unique ID of an individual instance
SHAPE	Geometry (point)	The geometry of the VGI instances
positioning _x	Double	Coordinate of point (X)
positioning _y	Double	Coordinate of point (Y)
Original ID	Short Integer	The unique id of the original VGI source
Source	Short Integer	The source of VGI (Ushahidi = 1; Tweet = 2)

Properties	Data type (GIS)	Description
Text	Text	The textual content of tweets or the title and textual description of Ushahidi reports
Theme	Text	The thematic group of the VGI instances
Location	Text	The name of a place, which is: (1) Location descriptions of the Ushahidi reports, or (2) Toponyms of tweets identified by the hierarchical toponym resolution.
Granularity	Text	The toponym granularity identified by the hierarchical toponym resolution (noted text: "Address", "Road junction", "Park/landmark", "River", "Road section", and "Suburb)
Suburb	Text	The suburb identified by hierarchical toponym resolution
Created at	Date	The date when the VGI instances created
U_Verification	Short Integer	The verification status of the Ushahidi reports
T_ UserName	Text	The user's name on Twitter
T_Retweets	Short Integer	The number of times this tweet has been retweeted
URL	Short Integer	The existence of URL (False = 0; True = 1)

Table 5.4: Thematic groups of VGI collections

Thematic groups	Descriptions	Text examples
Property damage	VGI instances referring to the status of critical infrastructures	Good news no more water on Suncorp Stadium...bad news, pitch is black, caked in sewage
Flooded area	VGI instances referring directly or indirectly to water level measurement or the expansion of flood	Water crossing over bridge belivah rd
Road affected/ Closed Road	VGI instances referring to traffic disruptions	There is a large uprooted tree which has been ripped out of the bitumen footpath in Browning Street which is learning on the tenants properties' building. This is located on the opposite side footpath to the Greek Child Care Centre. I don't know whether this has been reported already.
Other hazards	VGI instances referring to the hazards not direly caused bay inundation	While no contamination of the water supply has been confirmed, Queensland Urban Utilities is advising customers in the above locations to take the precaution.

5.3 Authoritative data collection

Authoritative data were collected from the portals of Australian government agencies[25]. The purposes for using these data were data cleaning, annotating credibility label, and generating geographic predictors to the VGI instances. The collected data sets are described briefly as follows; the details of some data sets can be found in Appendix 3.

1. References for data cleaning

To identify the matched positioning, following vector data were used as the references of the toponyms: (1) Physical road network (polyline of street centerlines), (2) Recreation areas (polygon of parks/landmarks), (3) Statistical local area boundary (polygon of suburbs), and (4) Water resource areas (polygon). These datasets were generated from the Queensland Department of Natural Resources and Mines (DNRM) and the Australian Bureau of Statistics. By using these data sets: (1) a gazetteer based on the Statistical local area boundary was made; (2) the lists of toponyms of Roads, Park, River within the suburbs of Brisbane was made via spatial join.

In addition to the abovementioned data, as a part of the toponyms were address descriptions, and Australian government did not provide address data in open access, Google Maps geocoding service was used for the references of address descriptions.

2. References for VGI credibility annotation (supporting geographic data)

This research found that the credibility of the VGI instances was difficult to determine from the texts (see: Section 5.5.1.2). To support the credibility annotation, flooding data during

[25] Most of the authoritative data used in this research were downloaded from Queensland Spatial Catalogue, http://qldspatial.information.qld.gov.au/catalogue

the two events were collected as references: (1) 2011 and 2013 flood modelling maps[26], and (2) 2011 and 2013 flood-affected streets. A combined display of the flood modelling map, the affected streets, and the VGI instances in 2011 Queensland Floods is illustrated in Figure 5.3.

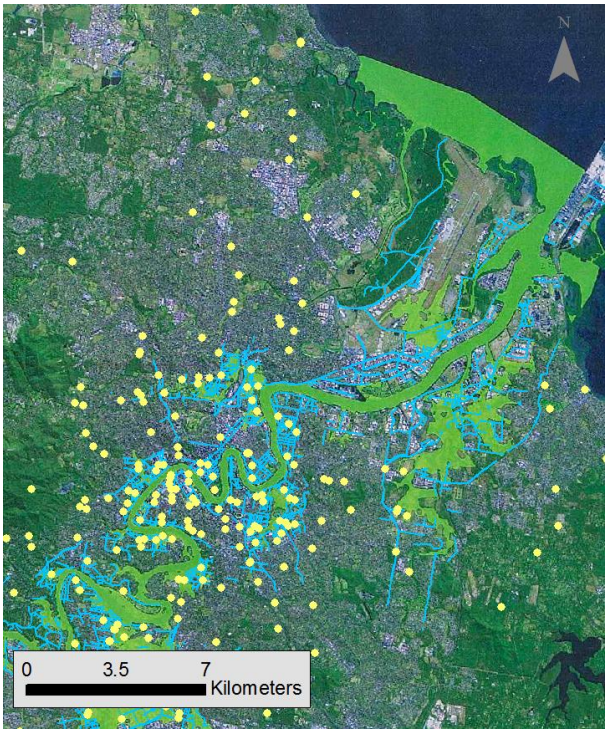


Figure 5.3: A part of the flood modelling map, the affected streets, and the VGI instances in the 2011 Queensland Floods

The flooding data also demonstrated that the impacts of the two flooding events were different. The 2011 event caused a major flood in Brisbane. On the other hand, the 2013 event caused major floods in several areas such as Bundaberg but caused a minor flood in Brisbane. Therefore, the 2011 event caused more damage and loss in Brisbane than the 2013 event. Table 5.5 lists the highest peaks of several flood gauges around the Brisbane

[26] The maps were downloaded from http://www.brisbanefloodinfo.com/mirror.html. The raster modelling map was manually digitized into vector data.

CBD during the two flooding events. Due to the difference in flood severity, the predictive capability of the developed QA methods was investigated.

Table 5.5: Highest peaks of flood gauges along Brisbane River during the two events

Station No.	Name	Stream	2011 peak height (mAHD)	2013 peak height (mAHD)
540198	Brisbane City Alert	Brisbane River	4.455	2.32
540130	Bowen Hills Alert	Breakfast Creek	2.73	2.72
540286	Breakfast Creek Mouth Alert	Breakfast Creek	2.75	2.12
540132	East Brisbane Alert	Norman Creek	3.27	2.71
540071	Corinda High Alert	Oxley creek	9.33	4.05
540274	Oxley Creek Mouth Alert	Oxley creek	9.27	3.36

source: The State of Queensland Department of Natural Resources and Mines (2013); (mAHD: measurement in metres against the Australian Height Datum)

3. Data for the generation of geographic predictors

Although it was assumed that the VGI quality correlated with geographic characteristics, the VGI dataset only had positioning and toponyms. Thus several kinds of authoritative data were collected to generate new geographic attributes in the VGI dataset; these attributes can be further used as predictors.

Based on the author's assumption of the correlation, following data were collected: (1) Digital Elevation Model (DEM), (2) population of each suburb, (3) water resource areas, and (4) flood risk zones.. The spatial distribution of water resource areas, high and medium flood risk zones and the flood extent of 2011 Floods (i.e. the approximate inundation between 13th January and 15th January 2011) are demonstrated in Figure 5.4. It shows that the high and medium flood risk zones cover most areas of the flood extent, and the water resources area is smaller which only covers the water channel.

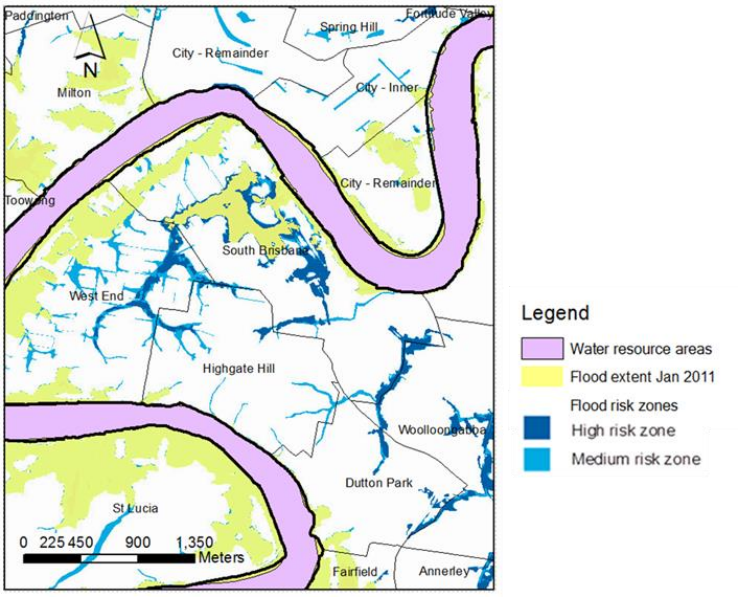


Figure 5.4: The overlay of flood risk zones, water resource areas, and the extent of 2011 floods around South Brisbane

5.4 Data cleaning process

The collected VGI instances had an issue that the positioning might be not precise and accurate (see: Section 4.4.3.1). To evaluate the approximate percentage of mismatched positioning, VGI instances were sampled under 95% confidence level and 10% confidence interval. And then these instances were overlaid with authoritative data and Google Maps in ESRI ArcGIS[27] (as Google Maps was used in the Ushahidi platform). A manual geocoding check found that 39% of the VGI positioning was farther than 100 metres to the toponym and identified as mismatched. Figure 5.5 provides a tweet example that the positioning should be marked in the south of the original positioning. The evaluation also found that six types of toponyms were collected: "Address", "Road intersection",

[27] The integration used a tool developed by GTO software, http://www.gto-software.com/

"Road/street section", "Park/landmark", "River", "Suburb". And the instances described in "Road/street section" accounted for the highest rate.

Figure 5.5: A tweet example of the mismatched positioning

As the validation of geographic correlation was based on the positioning of both Ushahidi reports and tweets, the mismatched positioning could harm the result. Therefore, 454 Ushahidi reports with the compulsory mark by the Ushahidi system were excluded firstly. On the other hand, to exclude mismatched VGI positioning such as Figure 5.5, a semi-automatic data cleaning process was developed. A variety of geo-referenced objects were used to represent the extent of toponyms (point, polyline, and polygon) and identify the ambiguity toponym in tweets. The data cleaning process was designed and illustrated in Figure 5.6.

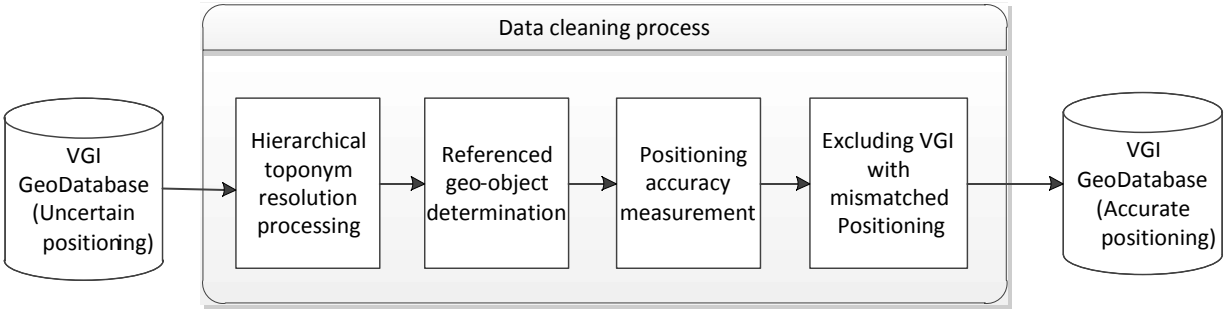


Figure 5.6: The workflow of the data cleaning process

5.4.1 Hierarchical toponym resolution and referenced geo-object determination

In the first step of data cleaning, a hierarchical toponym resolution was developed. It aimed to identify the toponym and the matched geo-referenced objects efficiently. The processing flow of the hierarchical toponym resolution consisted of toponym recognition and toponym resolution, which is illustrated in Figure 5.7. It used a geoparser (i.e. CLAVIN-NERD) and the built gazetteer to recognize and record the suburb in the text, including the name abbreviation and the full name (e.g. Indooroopilly = Indro).

The hierarchical toponym resolution was designed to recognize the correct suburb name in the first step. Since most toponyms of the collected VGI were "Road/street section", the correct suburb has to be identified for finding the right road (i.e. the referential ambiguity of roads[28]). For all instances with one identifiable suburb, they were further classified into the other types of toponym hierarchically. On the other hand, if the geoparser identified more than one suburb or cannot find any suburb from the texts, such a VGI instance was manually examined to assign one identifiable suburb by its positioning.

[28] A road is often across more than one administrative area or might have same name in several areas; there are four roads named "Birdwood Road" in four suburbs of Brisbane and streets of a similar name such as "Birdwood Street" and "Birdwood Terrace". See: street names with suburbs in Brisbane, https://www.data.brisbane.qld.gov.au/data/dataset/85e262dc-9dfe-46e9-87a8-f9f92f6a7780

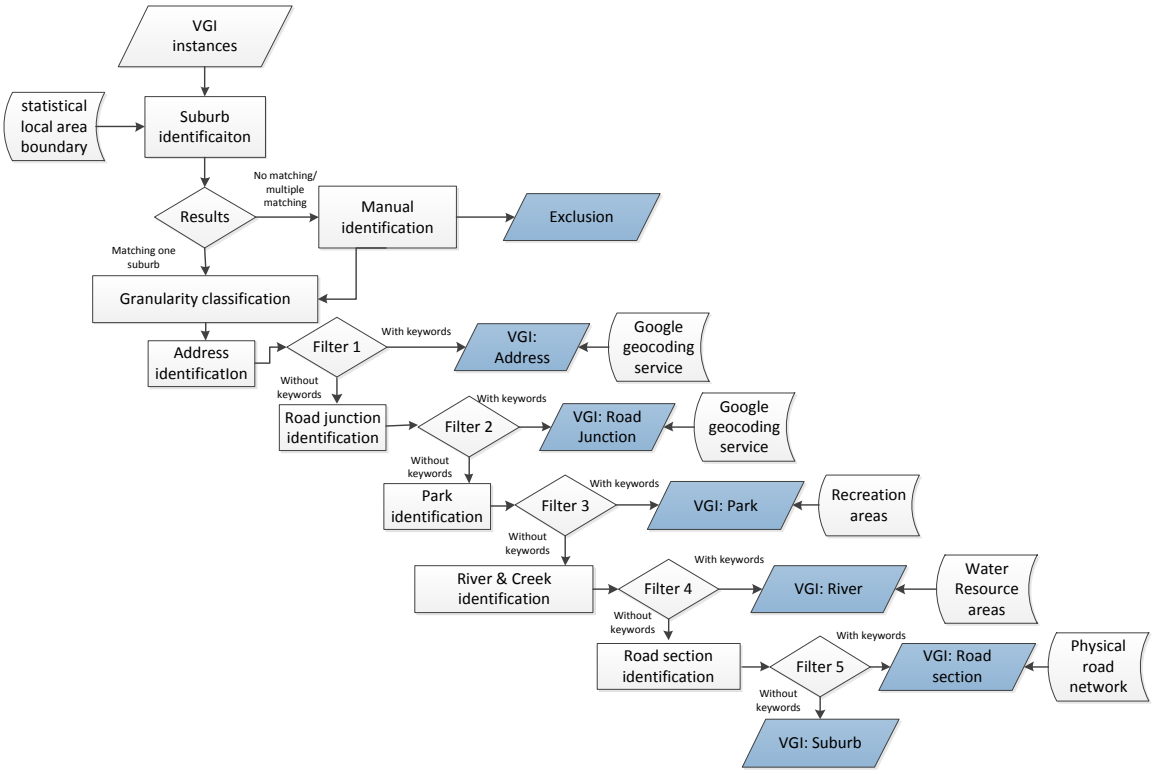


Figure 5.7: The processing flow of hierarchical toponym resolution

After the suburb recognition, a hierarchical string matching method was used. Several keyword-based criteria were used in a program to identify toponyms among the texts and determine five types of granularities. From a finer to a coarser level, the granularities were set as “Address”, “Road junction”, “Park/landmark”, “River”, and “Road section”. If a determinant criterion was not satisfied at a finer level, the determinant criterion at the next level was further used as a filter. In this hierarchical toponym resolution, if the result contained multiple geographic focuses at different granularity, the finer one is set as the toponym; for multiple toponyms at the same granularity, the first toponyms were recorded. Based on this, the toponyms were identified and the references of geographic objects were assigned. Table 5.6 summarizes the toponyms, the geometries, and the sources of the references in this process. The detailed process was described as follows:

Table 5.6: Classification of toponym granularity in the texts of VGI and the references

Granularity	Determinant criteria (keywords)	Geo-referenced objects		
		Geometry	Source	Providers
Address	Existence of number	Point	geocoding service	Google Maps
Road junction	"Junction" or "intersection" or "corner"	Point		
Park/landmark	"Park" or "garden", matching of park name in a given suburb	Polygon	Recreation areas	Government department/ agencies of Australia
River	"River" or "creek", matching of river name in a given suburb	Polygon	Water Resource areas	
Road section	Matching of road/street name in a given suburb	Polyline	Physical road network	
Suburb	Instances did not satisfy abovementioned criteria	Polygon	Statistical Local Areas	

(1) "Address": if the text contained a number, the number was manually identified as house number or postcode or not relevant to a toponym; if the number is a house number, then the toponym was classified as "Address".
(2) "Road junction": if the text contained "junction" or "intersection" or "corner", it was categorised as "Road junction";
(3) "Park/landmark": if the text contained "park" or "garden", it was categorised as "Park";
(4) "River": if the text contained "river" or "creek", it was categorised as "River";
(5) "Road section": for the rest not being classified into any granularity group.

In this process, the VGI instances classified as "Address" or "Road junction" required a further manual identification (e.g. for the identification of postcode and house number), and then the positions from Google Maps geocoding service were used as the references. The accuracy of geocoding was validated by checking if return codes from the geocoding

service were fit to "street_address" or "intersection"[29]. If return codes were not fit, the instance was regarded as mismatching and then excluded. On the other hand, for the instances classified as "Park", "Water Resource Areas", or "Road section", the toponyms were identified and recorded from the texts of VGI instances by the string matching program. A list of toponyms within the given suburb was used. For example, if the VGI located in Rocklea (suburb), the program searched the toponyms in Rocklead such as "Kookaburra" (park), "Egmont" (Rd.), and so forth to find the matched toponym.

If the instances did not match any toponym in abovementioned steps, it was categorized in "Suburb" and further examined by the author to ensure that there was no any toponym in finer granularity. By the hierarchical toponym resolution, the suburb, toponym granularity, and the toponym of the VGI instances were identified and added to the attributes to determine geo-referenced objects. Nearly 80% of VGI instances (n = 714) can be accurately linked to the geo-referenced objects in the first attempt. The other 20% (n=187) requires more processing efforts as the texts were informal or poorly structured or contained typing errors. For example, the toponym "Graceville" in the following text: "Urgent huge supply bottled water #loc inGraceville" is not able to be identified by the default gazetteer. The gazetteer has to add "inGraceville" to match such an informal toponym.

5.4.2 Positioning accuracy measurement and excluding VGI instances with mismatched positioning

After determining the geo-referenced objects, the positioning accuracy could be measured. For each VGI instance, the distance (*d*) between the positioning and the geo-referenced object was calculated by ArcGIS Near tool[30]. The results found that the average *d* was about 30 meters. And the *d* was less than 30 meters among 90% of the collected VGI

[29] There are sever granularity code from the Google Maps geocoding service such as "street_address", "intersection", "country" and so on.

[30] https://desktop.arcgis.com/en/arcmap/10.3/tools/analysis-toolbox/near.htm

instances. However, the Standard Deviation (SD) was approximately 150 meters. This indicates that only a small part of the collected instances had a long-distance, more than 150 meters to the geo-referenced objects. Excluding the instances with a *d* more than 150 meters, the average *d* to the geo-referenced objects would be less than 10 meters, and the SD would decrease to 24 meters. Therefore, a threshold of 150-meter was determined to use for excluding instances with "mismatched" positioning. Two examples being excluded in this process are shown in Figure 5.8. The 150-meter buffer of two VGI instances in the left and middle were not intersected with the geo-referenced objects, so they were excluded. By comparison, the right one demonstrates an example to be included. After this data cleaning process, the final VGI dataset retained 737 Ushahidi reports and 164 tweets (n = 901). The 2011 event contained 606 and the 2013 event contained 295 instances respectively.

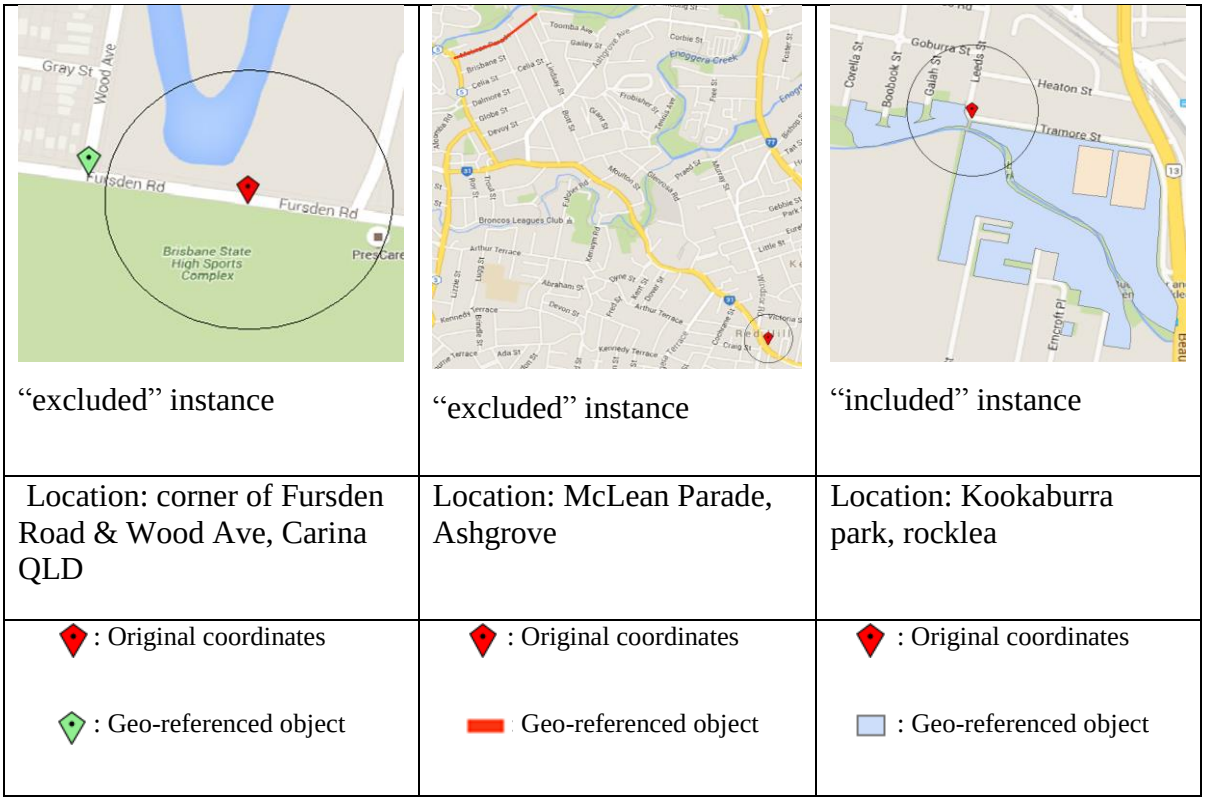


Figure 5.8: VGI examples in data cleaning by the distance to the reo-referenced object (Base Map data ©2015 Google)

5.5 Surveying the experts' opinions for VGI quality annotation

The quality label of each VGI instance was still empty. The quality annotation mechanism was planned to be established from the emergency management experts' opinions. The experts were invited to participate in a questionnaire survey. They rated the sampled VGI instances and provide their opinions on quality annotation.

5.5.1.1 Survey of VGI credibility and text content quality

The questionnaire survey simulated a pragmatic information need in the scenario of flood response. 65 VGI instances were randomly sampled, and the text and the place of these VGI instances were listed. At the beginning of the questionnaire, the definitions of credibility and text content quality were explained (see: Section 1.2). The experts were given a task to identify the credibility and text content quality of VGI samples and assign a rating of credibility and text content quality. If a VGI sample was ambiguous to be rated, it can be rated as "I can't decide". Table 5.7 demonstrates an example generated during the 2011 floods[31] in the questionnaire.

Table 5.7: An example of the sampled instance for the experts' rating

Text	Looking towards the botanical gardens #qldfloods #bnefloods #qldfloodsmap http://twitpic.com/3p8yaw
Place	Botanical gardens
Created at	2011-01-07 19:06
Suburb	Mount Coot-Tha
Question	(1) Is the text credible? ☐Highly credible ☐Seems credible ☐Lowly credible ☐I can't decide. (2) Is the text informative to emergency management stakeholders? ☐ Very informative ☐Seems informative ☐Not informative ☐I can't decide

[31] Source: http:/queenslandfloods.crowdmap.com/reports/view/1001

5.5.1.2 Results of the questionnaire survey

1. Results of the credibility rating

The questionnaire survey found that it was difficult to have consistency in the credibility rating by only reading the limited text. (the kappa statistic = 0.322) (J. Cohen, 1960; Viera & Garrett, 2005). The cases of disagreement had been discussed with the experts, and they expressed that without local knowledge in the described places, they cannot rate credibility precisely. Some detailed information about flooding situation or property damage cannot be judged well[32]. In comparison with the previous work such as Castillo et al. (2011) and Gupta and Kumaraguru (2012), text-related characteristics (e.g. a long text) seemed to not provide consistent perception in the rating process. Due to these facts, reference data of flooding impact (see: Section 5.3) and historical news were collected for the annotation of credibility.

2. Results of the text content quality rating

As for the results of text content quality rating, it showed a moderate agreement (kappa = 0.430). Among VGI instances of the same rating (n=41), 20% was " Very informative", 29% was "Seems informative" and 39% was "Not informative". The results showed that text-related characteristics had a better classification capability in text content quality. Moreover, the experts found that the interpretation between "very informative" and "seems informative" is not clear without more criteria (If regrouping the categories as "informative" and "not informative and other", the ratings had a substantial agreement. kappa = 0.798).

3. Suggestions from the experts

[32] For example, following text was rated as "Highly credible" and "Lowly credible" by the two experts. "why does the BOM Brisbane City River Gauge say its steady/falling at Wednesday 7.30pm when the water is still rising here in walsh street at 7.45pm? it is now nearly 3m deep outside our apartment building with cars completely submerged and oil and petrol seeping up from flooded basements."

In addition to the rating, at the end of the questionnaire, the experts were also asked to answer an open-ended question to provide opinions and insights into their decisions in the rating process. Their suggestions were quoted as follows:

(1) "*Information was aided with specific information including precise locations, time, description of issues. Concern over RT as you can't be sure of time on original tweet from the info provided here. Suburb level info (without street or intersection) close to useless as can't dispatch work crew etc."*

(2) "Informative level is largely affected by quality of description and key location details. Two criteria are used to give a code: whether the contents contain both precise location and time of the event and whether the contents have a reasonable length. Added weight is placed on informative levels by the additions of linked page or content (particularly if they are date and location stamped and the contributor is identified)."

From these comments, there are some specific information needs for text content quality. Precise location, time, and description of text were addressed. And the linked page or media might provide additional weight. Besides, the experts argued that VGI of a finer granularity is more informative to the pragmatic need. Based on the results of ratings and the comments, the annotation process of text content quality was established and further described in the next section.

5.6 VGI quality annotation process

In this section, the process of credibility and text content quality annotation is described. The credibility annotation used the metadata properties of VGI (i.e. U_verificaiton and T_retweets in Table 5.2) and the evidence of flooding impacts. The text content quality was determined by a Multi-criteria method based on a discussion with the experts. The workflow of the annotation process is illustrated in Figure 5.9.

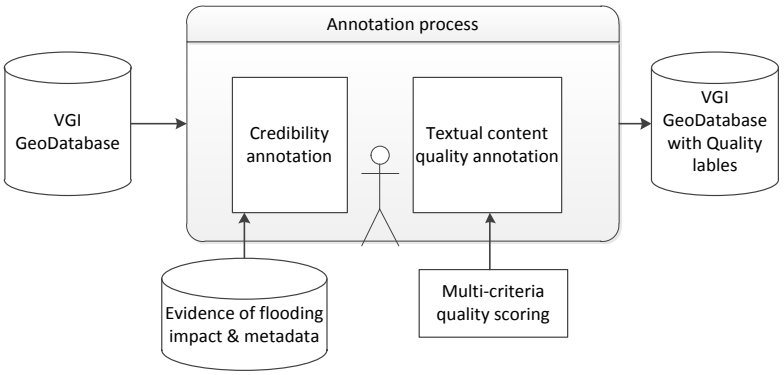


Figure 5.9: Annotation process of the collected VGI instances

5.6.1 Annotation process of credibility

The credibility annotation was a binary classification. VGI instances were classified into two labels: "high-credibility" and "low-credibility". High-credibility class referred the instances had perceivable believability or the described phenomenon of the instances were close to the supporting geographic data. On the other hand, low-credibility class referred to the ambiguous VGI instances, including information "likely to be false" and "certainly false". A two-step annotation process was conducted, which is described as follows.

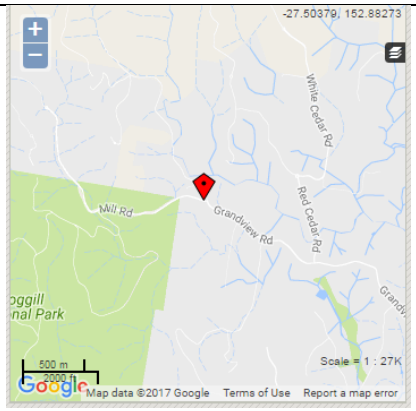
1. Credibility annotation based on the metadata properties

The perceivable believability can be firstly assessed by the available metadata properties of VGI instances. For example, if a Ushahidi report had been verified, or if a tweet had been retweeted by at least twice, such an instance should be more credible and labelled as "high-credibility". Based on this, the properties of VGI instances were examined. This process annotated 55.0% of the Ushahidi reports and 7% of the tweets as "high-credibility".

For the unlabelled Ushahidi reports, as some of them had participant's credibility rating and the comments from the Facebook plugin on the Ushahidi applications, the author manually checked the web page of all the Ushahidi reports (i.e. these properties cannot be extracted by the Ushahidi API). If the participant's credibility rating of a report is positive, it was annotated as "high-credibility". However, if the credibility rating is negative or the comments of the report doubted the credibility of information, the report was annotated as

"low-credibility". For example, Table 5.8 shows a local resident argued that the marked positioning was wrong in a report[33] and thus this report was annotated as "low-credibility".

Table 5.8: A comment pointed out the credibility issue of the Ushahidi report

<table>
<tr><td>Description</td><td colspan="3">Road Closed - Grandview Road, Pullenvale: Grandview Road, Pullenvale - Hazard: Water over road All directions - Closed due to flooding near Pullenvale Primary School. Road closed both directions. No delays expected. Use alternative route.</td></tr>
<tr><td>Location</td><td colspan="3">Grandview Road, Pullenvale QLD 4069, Australia (road section)</td></tr>
<tr><td>Map</td><td> 
</td><td>Comments of this report</td><td>This location should be shown on the map as between Lancing Street and Airlie Road. Also a tree across road on this section of Grandview Road. Access to Airlie Road and further North along Grandview Road from Moggill Road now via Pullenvale Road.</td></tr>
</table>

2. Credibility annotation based on the supporting geographic data

After the previous step, to process unlabelled data, the collected VGI instances were overlaid with the flood modelling maps or the 20-meter buffer areas of the affected streets (i.e. approximate road width). If the instances intersected with these areas, they were labelled as "high-credibility" (Figure 5.10). Then, the author checked the places of

[33] Source: https://bnestorm.crowdmap.com/reports/view/4. Note that although the previous data cleaning process had identified the positioning of this example was correct, yet the comment described the actual flooded area was another place, in a distance of 2 km).

unlabelled data in the historical news. If evidence of flooding was found, the instance was annotated “high-credibility”. Otherwise, the instances were annotated as “low-credibility”.

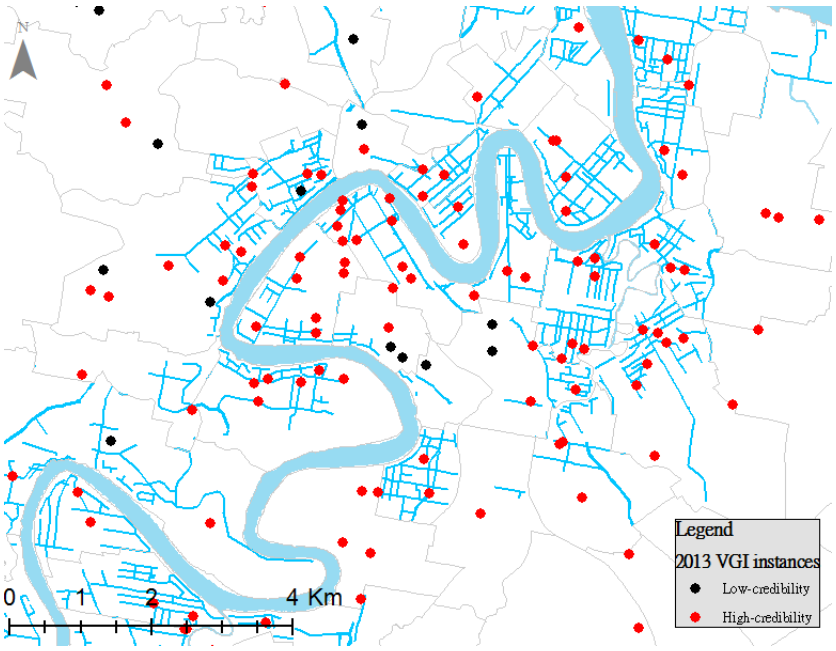


Figure 5.10: The credibility annotation based on the affected streets

The final results showed that among the Ushahidi reports, 615 were annotated as high-credibility and 122 instances were annotated as low-credibility respectively (83.4% and 16.6%). As for tweets, 129 were high-credibility and 35 instances were low-credibility (78.7% and 21.3%). Table 5.9 demonstrates the class distribution in the two events.

Table 5.9: Credibility annotation results

Dataset	Class	Number (percentage)
2011 event	High-credibility	547 (90%)
	Low-credibility	59 (10%)
2013 event	High-credibility	197 (67%)
	Low-credibility	98 (33%)

The annotation process had tried to evaluate credibility with all the ancillary information available. But as noted, the results still can be uncertain because corroborating flood information might be not available. Therefore, false labelling might exist.

5.6.2 Annotation process of text content quality

At the beginning of the text content quality annotation, a binary classification was considered. However, through a discussion with the experts, using more quality labels had the advantages to identify the information of higher priority. Therefore, the annotation of text content quality was a multi-class classification. Three labels (high-, medium-, and low-quality) were used.

To give a clear distinction between the three labels, a discussion was made with the experts and then an annotation process was developed based on the concept of Multi-criteria method. The annotator had to identify the following criteria in text content manually and further calculated the weights of each instance to determine the quality labels:

(1) Quality and specificity of the descriptions: the text descriptions provide the possibility of correct understanding; the content is specific to flooding or emergency management.
(2) Completeness: an issue covered broadly in the texts.
(3) Granularity: the described location has finer granularity such as road section for the requirement of operations (i.e. if the Granularity of instances were noted in a suburb, they were not regarded as the informative instances).
(4) The existence of temporal expressions: the instance has temporal expression recognized from textual description and possibility of real-time use.
(5) The existence of affected entities: the instance describes the people or organization under the influence.
(6) The existence of URLs: tweets often consist of pointless and conventional texts. The existence of URLs (e.g. linked pages or media) may provide useful information indirectly.

The weights of each criterion were defined as follows. The criterion of “Quality and specificity of the description” had the highest weight (+9) in the annotation process. Next,

"Granularity" was in the second place (+6), and "Completeness" and "Temporal expression" were in the third place (+5). Finally, "Affected entity" and "Existence of URL" were the least important criteria (+3). Based on this, the annotator (i.e. the author) read the texts, articulated and recorded the criteria satisfaction, and then summed weights to annotate text content quality. This research defined that the weights between 15 and 19 were annotated medium-quality. High-quality instances satisfying multiple criteria had the weights at least 20. If the weights were below 15, it was regarded as a low-quality. Table 5.10 demonstrates the examples annotated in different quality labels.

Table 5.10: Examples of the text content quality annotation result

Examples of the VGI instances	Criteria satisfaction	weights	Quality Label
a. Jellicoe St was under water this afternoon at 4pm: I arrived home to see our street flooded and our neighbours white Toyota Aurion under flood water.	☒ Quality/Specificity ☒ Granularity ☒ Completeness ☒ Temporal expression ☒ Affected Entity ☐ URLs	28	High-quality (>=20)
b. "Water crossing over bridge belivah rd."	☒ Quality/Specificity ☒ Granularity ☐ Completeness ☐ Temporal expression ☐ Affected Entity ☐ URLs	15	Medium-quality (15~19)
c. The creek in Grantham looks like a battle ground #qldfloods #thebigwet" http://yfrog.com/h3ty1jxj.	☐ Quality/Specificity ☐ Granularity ☒ Completeness ☐ Temporal expression ☐ Affected Entity ☒ URLs	8	Low-quality (<15)

The text content quality labels of all the VGI instances were annotated following this process by the author in 20 hours manually. 23% of the VGI instances were annotated as "High-quality", 58% were "Medium-quality", and 19% were "Low-quality". Table 5.11 shows the class distribution in the two events. An evaluation of the annotation distribution is that the texts came from different sources and had different styles. In the 2011 event, most instances were from volunteers and local people but the texts were usually short and not provided vital information. Medium-quality, therefore, accounted for the highest rate. In the 2013 event, the Brisbane City Council post many reports on the Ushahidi platform so the percentage of high-quality was higher but medium-quality still took up the largest proportion. To further evaluate the annotation process, one expert was asked to label 209 random instances (sample size at the 90% Confidence Level and 5% Margin of Error). The results reached a substantial agreement (kappa =0.756, p=0).

Table 5.11: Text content quality annotation results

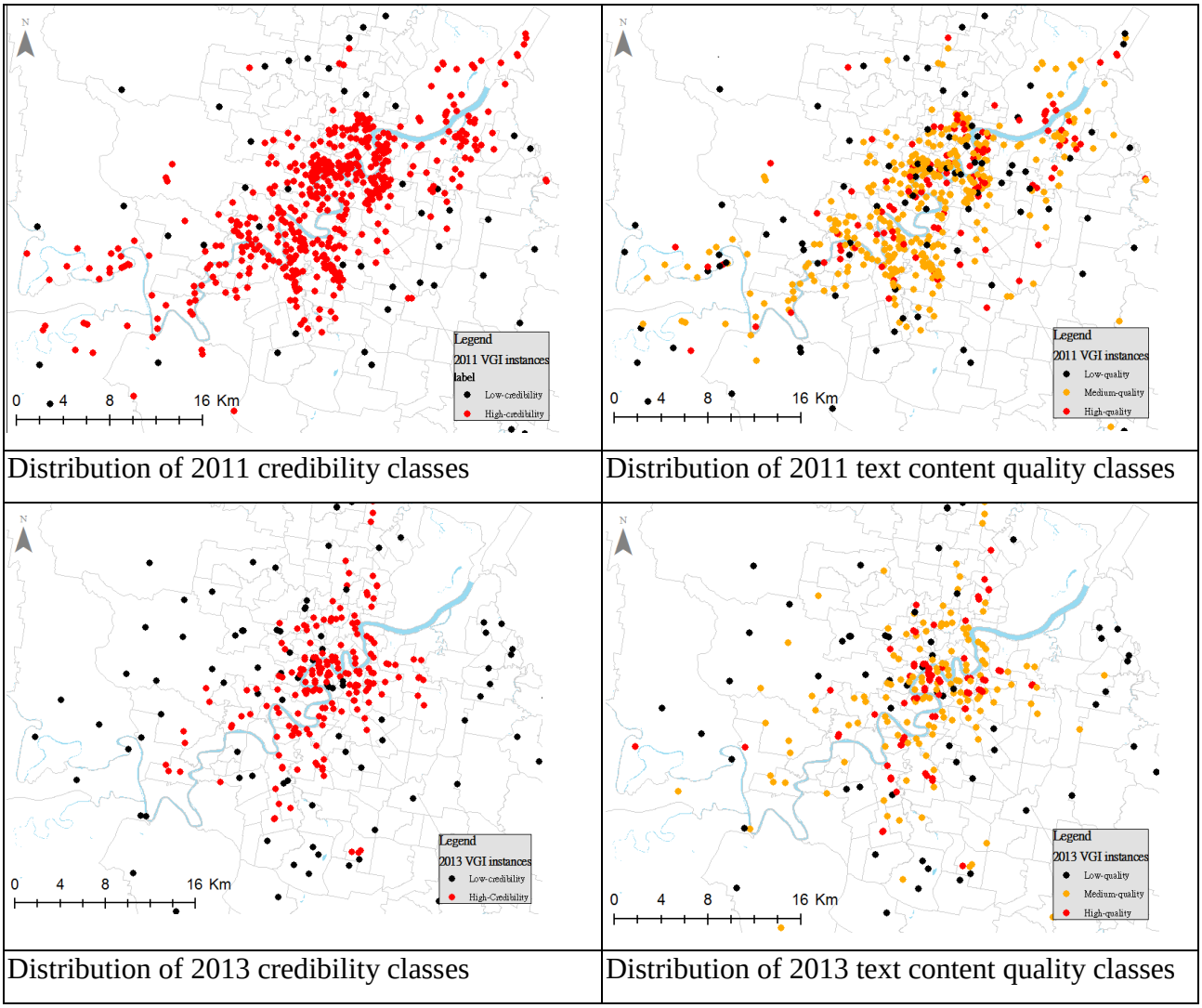
Dataset	Class	Number (percentage)
2011 event	High-quality	123 (20%)
	Medium-quality	381 (63%)
	Low-quality	102 (17%)
2013 event	High-quality	89 (30%)
	Medium-quality	136 (46%)
	Low-quality	70 (24%)

5.7 Correlation evaluation of geographic characteristics, step 1: visual observation and the clustering analysis

After the annotation, the spatial distribution of the VGI instances can be observed, which is illustrated in Figure 5.11. The distribution showed high-quality instances tended to occur around the areas close to the Brisbane River, such as South Brisbane and St. Lucia. Besides, the high-quality instances seemed to cluster. To evaluate the correlation between the quality and the clustering, the Optimized Hot Spot Analysis (OHSA) and the Mean nearest neighbour analysis were performed. Note that here this research did not discuss the use of VGI in emergency activities based on the two analyses. But the analysis can actually help associated activities such as making the evacuation plan or assigning the responders, as hot-spots were supposed to be the most affected areas. And it can also help simple filtering of VGI as VGI from hot-spots was supposed to have a higher priority for taking actions.

5.7.1 Optimized hot spot analysis

The OHSA was performed to find hot and cold spots. The analysis can generate the statistical confidence of hot and cold spots in a nominal label, "Gi_Bin", It could range from -3 (cold spot at 99% confidence) to 3 (hot spot at 99% confidence). To investigate the spatial pattern, a 2-stage analysis was performed. First, the aggregation method was used to create the weighted polygons according to the suburb boundary of Brisbane for a "point-clustering" analysis; the quality label was not used. Subsequently, the quality label was used to see where the high- or low-quality instances clustered.

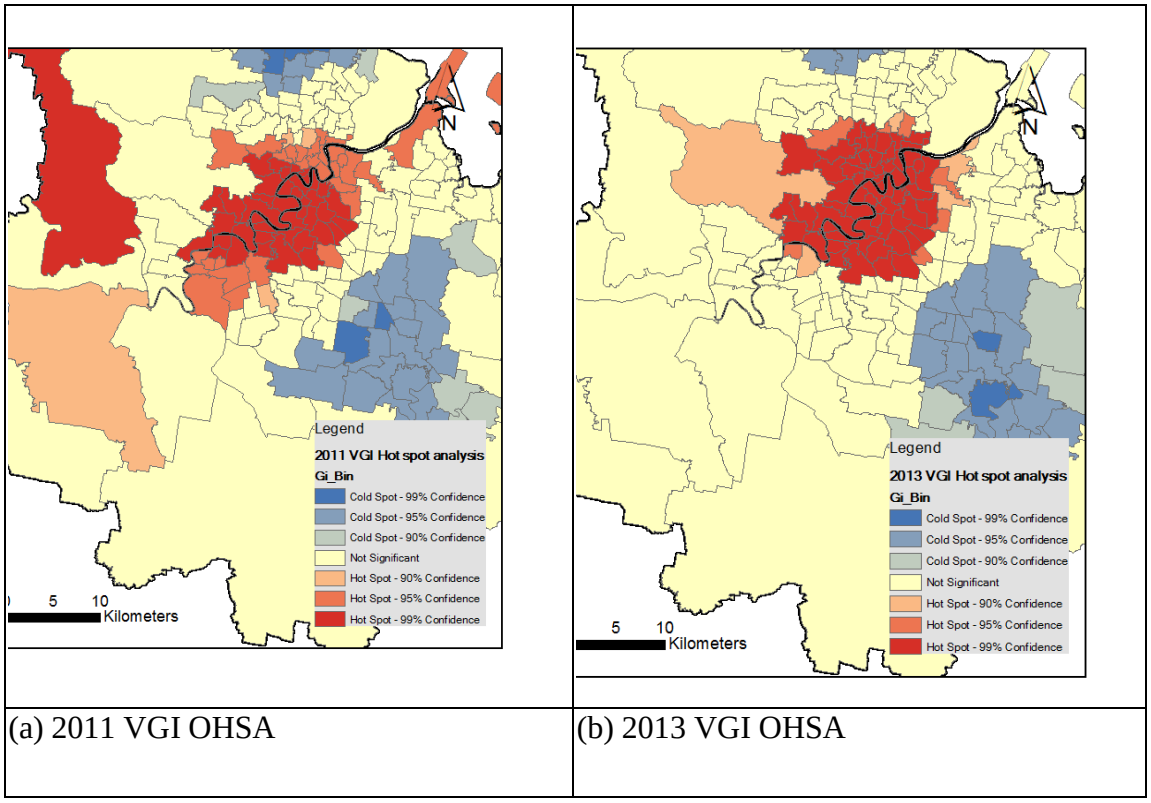


Distribution of 2011 credibility classes	Distribution of 2011 text content quality classes
Distribution of 2013 credibility classes	Distribution of 2013 text content quality classes

Figure 5.11: The credibility class distributions of the VGI instances

Figure 5.12 (a) and (b) illustrates the point-clustering analysis. It shows several Brisbane suburbs along the Brisbane river were hot spots. These hot spots were considered as heavily impacted areas. 75.7% of the 2011 instances and 67.5% of the 2013 VGI instances located within the hot spots. Additionally, the distributions of hot spots in two flooding events are similar. But one visible difference from the distribution is around the suburbs of western Brisbane (i.e. Karana Down and Ipswich Region). The cause of this difference is probably due to a controversial operation of the Wivenhoe Dam during the 2011 floods. The operation caused serious floods out of the predictions in the Ipswich Region.

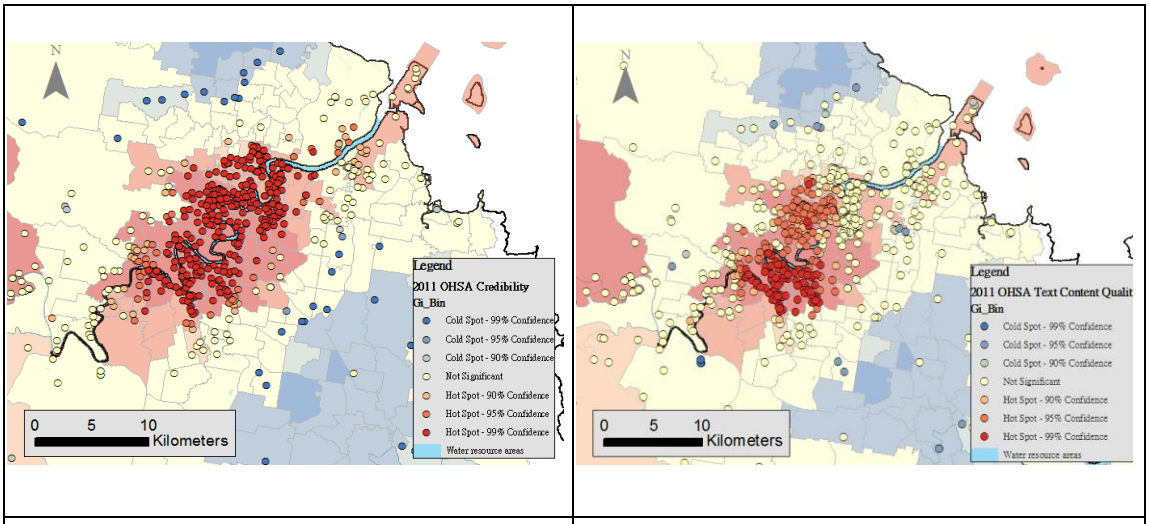
Furthermore, Figure 5.13 (a) and (b) illustrates the analysis based on the quality classes. The results showed that 67% of credibility classes were classified as hot spots; 24% and 9% of them were not significant and cold spots in spatial clustering. The distribution had a similar spatial pattern to the previous point-clustering analysis. As for the text content quality classes, 40% of text content quality classes were classified as hot spots, 55% and 5% of them were not significant and cold spots. In comparison with the distribution of credibility classes, the analysis showed that text content quality classes seemed to have a similar spatial pattern, but fewer instances occurred in the hot spots.



(a) 2011 VGI OHSA

(b) 2013 VGI OHSA

Figure 5.12**:** The results of the OHSA (the quality label was not used)



(a) 2011 VGI OHSA: Distribution of credibility classes

(b) 2011 VGI OHSA: Distribution of text content quality classes

Figure 5.13**:** The results of the OHSA (the quality label was used)

5.7.2 Mean nearest neighbour analysis

To further evaluate the clustering of quality classes, the mean nearest neighbour analysis was performed. The results were listed in Table 5.12, which shows that both the low-credibility instances and low-quality text were not statistically significant in clustering and leaning toward random distribution by using a 95% confidence interval. By comparison, the other classes were not randomly distributed; rather, they were clustered and statistically significant. Besides, the expected and observed mean distances are shorter as well. Since there is a great difference between the measures of mean distances between classes, the "distances to nearest VGI instance" correlates to the quality of VGI, and it should be a significant predictor in the models.

Table 5.12: The results of the Mean nearest neighbour analysis of credibility class

Class	NN index	Expected mean distance	Observed mean distance	z-score	p-value
ALL classes (2011)	0.59	1051(m)	622(m)	-19.21	0
High-credibility (2011)	0.49	944 (m)	464 (m)	-22.75	0
Low-credibility (2011)	0.96	3187 (m)	3047 (m)	-0.64	0.52
High-quality text	0.65	1796(m)	1181(m)	-7.26	0
Medium-quality text	0.53	1087(m)	569(m)	-17.78	0
low-quality text	0.91	2562(m)	2334(m)	-1.71	0.08

5.8 Geographic predictors generation

This research assumed that the quality of a VGI instance correlated to some geographic characteristics. From the results of mean nearest neighbour analysis, the instances of high-credibility and high/mid-quality text were clustered. It was assumed that VGI instances created at a closed spatial range were probably impacted by the same event and thus they were more credible or informative.

On the other hand, the visual observation found that most high-quality classes were located close to the Brisbane River. It was therefore assumed that people who faced the flood severity directly were more likely to provide information about their unsafe situation and their requirements. Therefore, the quality probably correlated to some specific environmental factors such as distance to the water/flood risk zones and local elevation value. Based on this, two types of geographic predictors were defined: (1) clustering predictors, and (2) geographic context (i.e. geo-context) predictors and. All the geographic predictors used in this research are defined in Table 5.13.

Table 5.13: The defined geographic predictors

Predictors		Description (unit, types of data)
Clustering predictors	Distance to nearest VGI	The Shortest Euclidean distance to the nearest VGI neighbour (metre, numeric)
	Hot spot zone	The confidence level of hot/cold spots (number, nominal)
Geo-context predictors	Distance to flood risk zones	The Shortest Euclidean distance to the merged high/medium flood risk zones (metre, numeric)
	Distance to water resource areas	The Shortest Euclidean distance to the water resource areas (metre, numeric)
	DEM value	The local elevation value (metre, numeric predictor)
	Population density (of a suburb)	The population density of the suburb where VGI located (people/square kilometre, numeric predictor)

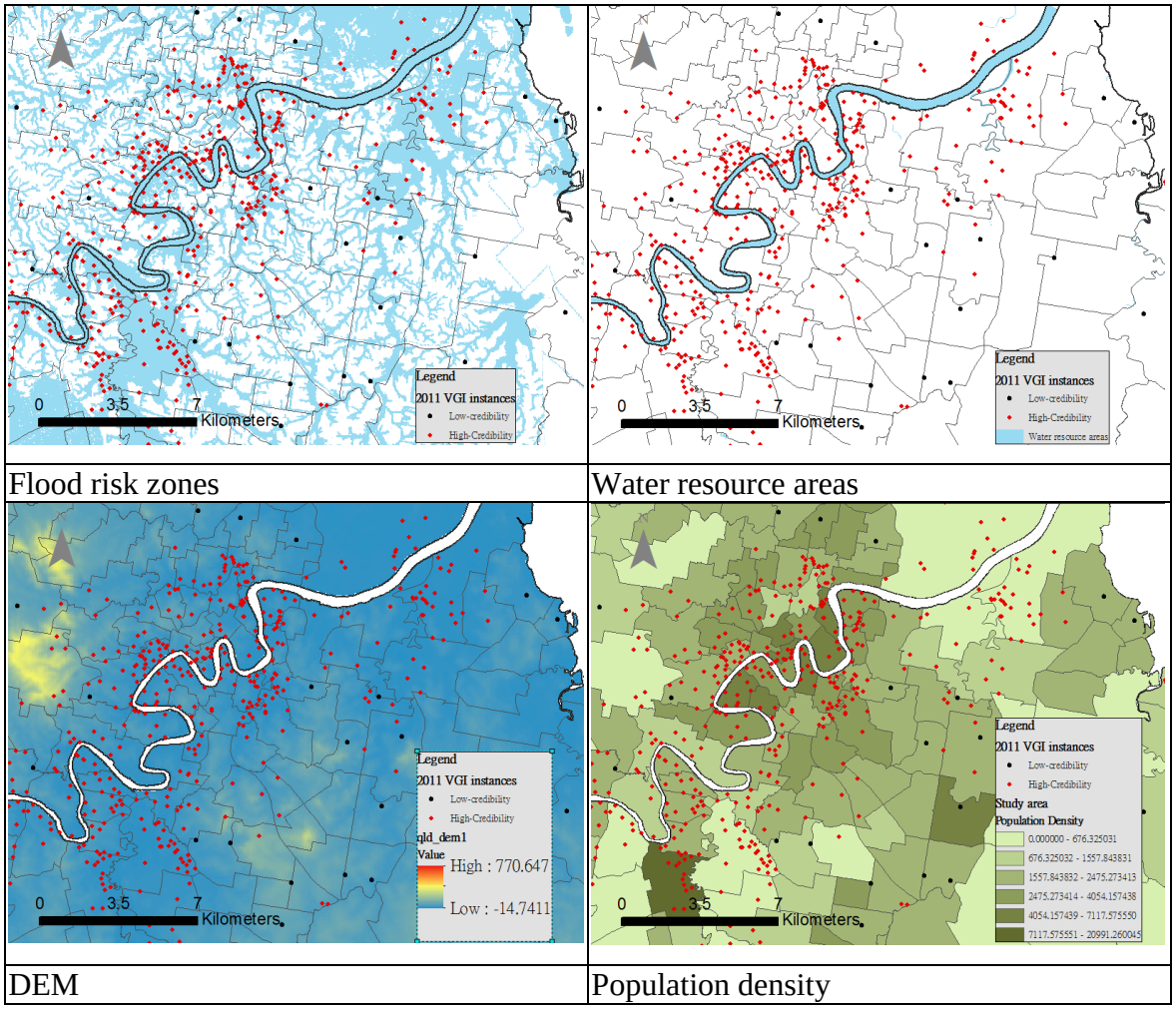
Clustering predictors used in this research include (1) distance to nearest VGI (numeric predictor), (2) suburb hot spot zone b (Gi_bin by OHSA; nominal predictor, Figure 5.12). Geo-context predictors used in this research included (1) Distance to flood risk zones, (2) Distance to water resource areas, (3) DEM value, and (4) Population density of VGI suburb. These predictors were generated from the authoritative data collection. Their definitions are described as follows; Figure 5.14 demonstrates the overlay of the VGI instances from the 2011 event and the geographic context layers.

(1) "Distance to flood risk zone" referred to the distance between VGI and the flood risk zones, which were areas highly potentially being suffered from floods (i.e. at 5%~1% chance occurring in any year);

(2) “Distance to water resource areas” referred to the distance between VGI and the water resource areas such as river, creek, and lake; a VGI instances occurred within or close to the flood risk zones or water resource areas should be more credible or informative (the water resource areas can be a proxy if flood risk data were not available);

(3) “DEM value” referred to the elevation value of the VGI positioning; if the elevation was low, the location was more likely to be affected by floods than other locations;

(4) “Population density of a suburb” is the measurement of the population in a suburb where the VGI instances located within; as more people were affected by the floods, the chance to share high-quality information is higher (high-quality instances were more likely to occur in densely populated areas).



Flood risk zones | Water resource areas

DEM | Population density

Figure 5.14: Overlay map of the 2011 VGI instances and authoritative data

5.9 Correlation evaluation of geographic characteristics, step 2: correlation coefficient and chi-square test

For addressing the research objective 2-b, the correlation between quality and the abovementioned geographic predictors was evaluated; the correlation coefficient evaluator in Weka (CorrelationAttributeEval) and the ranker method was used. For the credibility determination, the ranking order was: (1) Distance to nearest VGI (r = 0.531), (2) Distance to flood risk zones (r = 0.529), (3) Distance to water resource areas (r = 0.488), (4) Hot spot zone (r = 0.388), (5) Population density (r = 0.093), (6) DEM value (r = 0.087).

Next, the chi-square test was performed to further confirm the correlation. The numeric predictors (e.g. distances) were converted to categorical data in 10 classes for the calculation. Since the correlation of DEM value and population density were very weak, the chi-square test found that both of them had a high p-value (p = 0.10 and 0.13) which failed to reach conventional levels of statistical significance (i.e. p > 0.05 when the 95% confidence level is used). Therefore, DEM value and population density were not significantly associated with credibility.

As for text content quality, the ranking order was: (1) distance to nearest VGI (r= 0.354) (2) distance to flood risk zones (r= 0.294), (3) distance to water resource areas (r = 0.265), (4) hot spot zone (r = 0.234), (5) DEM value (r = 0.077), and (6) population density (r=0.056). Subsequently, the chi-square test also found that the p-value of population density and DEM value was beyond 0.05 (p = 0.11 and 0.15).

The results of the correlation evaluation validated the geographic assumption of this research. The four kinds of predictors (clustering predictors and two geographic predictors) correlated to the quality label and can be further used in the development of text-based VGI QA methods. Furthermore, the results demonstrated that the ranking order of the four predictors was similar, but the geographic predictors had a weaker correlation to the text content quality. An evaluation of this result is that some reports out of the hot spots were annotated as high-quality textual contents and worthy to take actions, but the credibility

was uncertain (low-credibility). For example, in the 2013 event, only two VGI instances[34] were collected from Tanah Merah (in the south of Brisbane) and both of them were annotated as high-quality. But as Tanah Merah had no affected streets in the collected data and did not have a serious flood in the historical news so the two instances were annotated as low-credibility.

5.10 Chapter summary

Utilizing text-based VGI is important to engage the broader society in disaster management. To design and develop the VGI QA methods, this chapter presents the data used in this research and the validation of geographic correlation. The major innovative contributions presented in this chapter are: (1) a data cleaning process using a hierarchical toponym resolution, (2) the annotation process of credibility and text content quality, and (3) the validation of correlation between VGI quality and the defined geographic predictors. By the processing and analysis presented in this chapter, the VGI datasets were ready for the development of the models through Supervised learning. The correlated geographic predictors are further used in modelling.

[34] The texts were "Big fallen tree took power line down at Plumbs Rd Tanah Merah. Area is now fenced off for safety. Also power outage in area Silver Court since Sun 27th 11pm" and "Power pole has a tree lying on top of it. Watched a huge explosion last night at the beginning of Drews road."

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 6

Development and evaluation of the VGI Quality Assessment methods

6.1 Introduction

This chapter presents the VGI QA methods based on Supervised learning models for addressing the research objective 3. In the first part of this chapter, the geographic method was demonstrated. The correlated geographic predictors (i.e. Distance to flood risk zones, Distance to water resource areas, Distance to nearest VGI, Hot spot zone) identified in Chapter 5 were used. As the geographic predictors had a weak correlation to text content quality (see: Section 5.9), an integrated method was further proposed in the second part. Text-related predictors and geographic predictors were combined in the integrated method. The scope and the extraction of text-related predictors are described. The modelling results of the two methods are demonstrated throughout this chapter. A part of the detailed models (the coefficients in Logistic Regression and Decision Tree graph) is described. Considering the variances of performances of the geographic method caused by the data size and the used predictors, two sensitivity analyses were performed. Significant text-related predictors in the text content quality classification are identified. The findings of the experiments are briefly discussed and presented.

6.2 Algorithm selection

In this research, a linear algorithm, Logistic Regression, and a non-linear algorithm, Decision Tree (J48) were selected for the test of the geographic method and the integrated method. They were often used for the text classification problem (Aggarwal & Zhai, 2012). Logistic regression is a linear discriminative model. It can obtain a predicted probability score of the VGI instance, which can be used for sorting high-quality data. It is similar to the SVM but easier for the optimization (Aggarwal & Zhai, 2012). On the other hand, Decision Tree has many merits such as easy to interpret, easy to use, computational efficiency. Traditional decision tree researches aimed to maximize the classification accuracy and minimize the classification error. Decision tree can also provide the threshold of a numeric predictor, which can be further used for other QA methods.

1. Logistic Regression

Logistic regression is a fast and simple classification algorithm. It is pragmatic for predicting presence or absence classes based on the values of a set of the predictors. Logistic regression is similar to linear regression, but the prediction value is a probability, matching to a range restriction through the logit transformation, which fits data to a logistic curve. A logistic regression model is built according to the following equation:

$$logit(P) = ln\frac{P}{1-P} = b_0 + b_1x_1 + b_2x_2 + \cdots + b_nx_n \quad (1)$$

Where $x_1, x_2, \ldots, x_n$ are the features, b_0 is the intercept of the model, and b_1, b_2,… , b_n are the estimated coefficients. P is the estimated probability of the event occurrence, and $logit(P)$ is the logit Odds = $\ln(\frac{p}{1-p})$. This equation can be written as:

$$P = \frac{e^{(b_0+b_1x_1+b_2x_2+\cdots+b_nx_n)}}{1+e^{(b_0+b_1x_1+b_2x_2+\cdots+b_nx_n)}} = \frac{1}{1+e^{-(b_0+b_1x_1+b_2x_2+\cdots+b_nx_n)}} \quad (2)$$

Logistic regression proceeds by finding estimates that maximize the resulting conditional distribution, or likelihood function for observations of the predicted classes. In this research, a Logistic Regression adapted by Lecessie and Vanhouwelingen (1992) was used. It used a ridge estimator for dealing with the correlated predictors in regression.

2. Decision tree

Decision tree origins in artificial intelligent research where the aim was to produce a system that could identify patterns and recognize a similar pattern in future (Quinlan, 1986). A decision tree is essentially a hierarchical decomposition of the data space, in which a predicate or a condition on the attribute value is used in order to divide the predictor space into a model (Aggarwal & Zhai, 2012). Such a model can be thought of as a type of automated classification performed by answering a series of questions about the attribute of an observation in a tree (Figure 6.1).

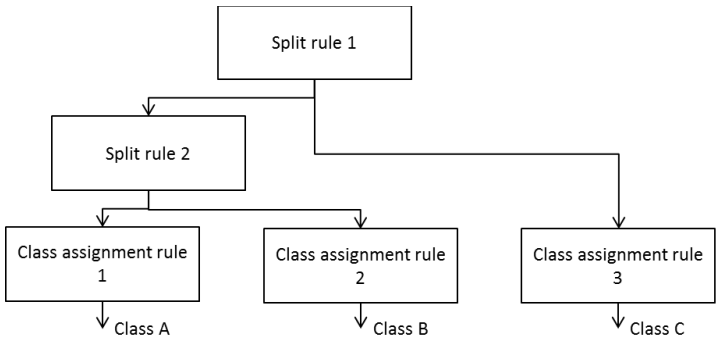


Figure 6.1: A decision tree model

A Top-Down Induction of Decision Tree (TDIDT) is the most common manner. It involves the formulation and comparison of rules for discriminating between instances of different classes. The effectiveness of each rule is evaluated by measures such as Bayesian probability function and Gini-index to see the decrease in impurity achieved by dividing the sample according to that rule. This process is recursively applied to all subsequent nodes and their descendants until the leaf nodes contain a certain minimum number of records, or some conditions on class purity. The construction of a decision tree can be quite complicated if it involves searching for the most efficient tree and reducing overfitting.

6.3 Development and evaluation of the geographic method

Chapter 5 had demonstrated that the geographic predictors correlate to the VGI quality classification. The geographic predictors for VGI QA were used for modelling, which is called "geographic method" in this research. This section demonstrates the experiments of the geographic method and the results.

6.3.1 Training dataset and evaluation dataset of experiments

The geographic method was developed and evaluated in the 2-step experiments by using the different datasets. Table 6.1 summarises the training and testing datasets.

Table 6.1: The experiment design and targets of the geographic method

Exp. type	Training dataset	Testing dataset	Target
1-A	2011	2011	Evaluating the model's performance by using the datasets from an individual event. Only the geographic predictors were used.
1-B	2013	2013	
2	2011	2013	Evaluating the model's predictive capability in the different flooding event. Only the geographic predictors were used.

First, due to the difference between the flood severity and flood areas (see: Section 5.3 and Table 5.5), a positioning of high-credibility instance in one event can be annotated as low-credibility in another event. For example, most of the 2011 VGI instances occurred in Karana Down and Ipswich Region were annotated as high-credibility (positive). But the 2013 instances occurred in the same area were annotated as low-credibility (negative) (see: Section 5.3 and Figure 5.10). This means that if using a model developed by the 2011 instances, 2013 instances from Karana Down and Ipswich Region might become False-Positive. As model developments seek for the generalisation performance, evaluating the model's predictive capability in the different flooding event is an important task.

Consider the generalisation performance, the evaluation can be 2-step. The first experiment tested two training datasets respectively, which were (1) the 2011 event dataset (Exp. 1-A), and (2) the 2013 event dataset (Exp. 1-B). The performances were evaluated through 10-fold cross-validation. Subsequently, the predictive capability in a different event was further evaluated. The second experiment (i.e. Exp. 2) tested a model trained from a dataset of the earlier event in a testing dataset of the later event. So the 2011 event dataset was used for training and by the 2013 event dataset was used for testing.

6.3.2 Using an oversampling technique for imbalanced datasets

The credibility and text content quality annotation results were imbalanced class distribution. This resulted in an issue that the developed models will tend to be overwhelmed by the majority class and ignore the minority class. Different potential

approaches had been proposed for dealing with imbalanced data, such as selecting representative samples (Bai, Lü, Wang, Zhou, & Ding, 2011; King & Zeng, 2001). In this research, an oversampling technique, SMOTE (Synthetic Minority Over-sampling Technique) was used.

SMOTE is the most known oversampling method. It can be used to increase the number of minority class (Chawla, Bowyer, Hall, & Kegelmeyer, 2002). SMOTE takes samples of the predictor space for the target class and identifies its k nearest neighbours, and then it generates "synthetic" instance that combines features of the sample with features of its neighbours. Figure 6.2 illustrates this process, if $k = 4$ and the amount of oversampling needed is 400%, where x_i is the sample from the minority class, x_{i1} to x_{i4} are the four nearest neighbours, the SMOTE generates synthetic instances (r1, r2, r3, r4) by taking the difference between the feature vector of the sample and its nearest neighbour, then multiplying this deference by a random number between 0 and 1 (i.e. $r_i = x_i + (x_{in} - x_i) \times \text{rand}(0\sim1)$). If the amount of oversampling needed is 100%, it will randomly generate a synthetic instance from (r1, r2, r3, r4).

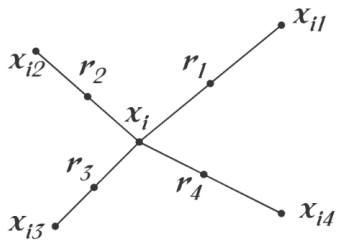


Figure 6.2: The generation of synthetic instances by the SMOTE

SMOTE increases the number of a minority class and makes the distribution of predictor space more general. In this experiment, the number of nearest neighbours k is set at 5; the amount of oversampling depends on the number of the majority class. Table 6.2 and Table 6.3 provide the number of classes before and after SMOTE. The number of VGI instances in each quality class was balanced after SMOTE.

Table 6.2: Number of credibility class after SMOTE

Dataset	Class	Number before SMOTE	Number after SMOTE
2011 event	High-credibility	547	547
	Low-credibility	59	547
2013 event	High-credibility	197	197
	Low-credibility	98	197

Table 6.3: Number of credibility class after SMOTE

Dataset	Class	Number before SMOTE	Number after SMOTE
2011 event	High-quality	123	381
	Medium-quality	381	381
	Low-quality	102	381
2013 event	High-quality	89	136
	Medium-quality	136	136
	Low-quality	70	136

6.3.3 Performance evaluation of Exp. 1

The evaluation of the geographic method by 10-fold cross-validation was showcased in this subsection. Generally, the results indicate that the credibility assessment models trained by geographic predictors reached a good performance. However, as for the text content quality models, the results did not reach a good performance as the correlations of geographic predictors were weaker.

1. Credibility assessment model

The performances of VGI credibility assessment models are summarised in Table 6.4. Generally, the accuracies of the two kinds of algorithms were good (above 82%). The performances of the models trained by the 2011 event dataset (Exp. 1-A) were excellent

(AUC > 0.9, accuracy > 91%). On the other hand, the performances of the models trained by the 2013 event dataset (Exp. 1-B) were a bit lower, (AUC = 0.894/0.884; lower accuracy and F1). Since the collected instances of the 2013 event were fewer than the 2011 event, the training of the models was probably limited. The lower performance was identical with expectation.

Table 6.4: Performances of credibility assessment models (the geographic method)

Algorithm	Exp. Type	Class	Accuracy	P	R	F1	AUC
Logistic	1-A	High	91.1%	0.930	0.901	0.916	0.966
		Low		0.890	0.922	0.905	
	1-B	High	82.2%	0.795	0.868	0.830	0.894
		Low		0.854	0.776	0.813	
J48	1-A	High	94.2%	0.965	0.920	0.942	0.939
		Low		0.912	0.962	0.936	
	1-B	High	85.2%	0.842	0.868	0.855	0.884
		Low		0.863	0.837	0.850	

Note: P = Precision, R= Recall, AUC = weighted Avg. AUC

The coefficients of logistic regression models are listed in Table 6.5. Based on this, the model can be defined according to the equation in Section 6.2(2). For example, the probability equation of Exp. 1-A can be calculated as:

$$P = \frac{1}{1 + e^{-(-2.0561+0.0003x_1+0.0013x_2+0.0001x_3-0.3146x_4-1.3551x_5+\cdots)}}$$

Where the threshold value (cutoff) was at 0.5, P > 0.5 = low-credibility instances; P < 0.5 = high-credibility instances[35]. A calculation example is given here: the P of a high-credibility instance with the distance to ($\mathbf{x_1}$) food risk zones, ($\mathbf{x_2}$) nearest VGI, and ($\mathbf{x_3}$)water

[35] If p = 0.5, the model choose randomly between two labels.

resources areas at 106.35, 478.82, and 208.30; (x_4) Gi_Bin of Hot spot zone at 3 is 0.155. As its P is under 0.5, the model supposes such an instance as high-credibility.

Table 6.5: Coefficients of credibility assessment models (the geographic method)

Predictors	Coefficient (1-A)	Coefficient (1-B)
x_1: Dist. to flood risk zones	0.0003	0.0016
x_2: Dist. to nearest VGI	0.0013	0.0016
X_3: Dist. to water resource areas	0.0001	0.0001
x_4: Hot spot zone (Gi_Bin) = 3	-0.3146	-1.0705
X_5: Hot spot zone (Gi_Bin) = 2	-1.3551	-0.3519
X_6: Hot spot zone (Gi_Bin) =1	-1.3367	0.3814
X_7: Hot spot zone (Gi_Bin) = 0	0.6671	0.1413
X_8: Hot spot zone (Gi_Bin) = -1	4.9829	6.0487
X_9: Hot spot zone (Gi_Bin) = -2	18.0889	5.9398
Intercept	-2.0561	-1.899

Note: If Gi_Bin equals the nominal label, $x_{4\sim9}$ is 1. Otherwise, $x_{4\sim9}$ is zero.

In comparison with Logistic Regression, the accuracies and the AUC of J48 models had small differences. But J48 models had higher Precision, Recall, and F1 generally. Moreover, decision Tree had the advantage to identify specific thresholds. An example of J48 Decision Tree graph trained from the Exp. 1-A is showcased in Figure 6.3. As x_1, x_2, and x_3 were numeric predictors, they were used in the splits multiple times. The interpretation of the tree is intuitive. For example, from the right side, the model was supposed that if a

VGI instance located in a Hot spot zone (Gi_bin = 3) within a distance to a nearest VGI instance and flood risk zones around 538 m and 1054 m respectively can be highly credible.

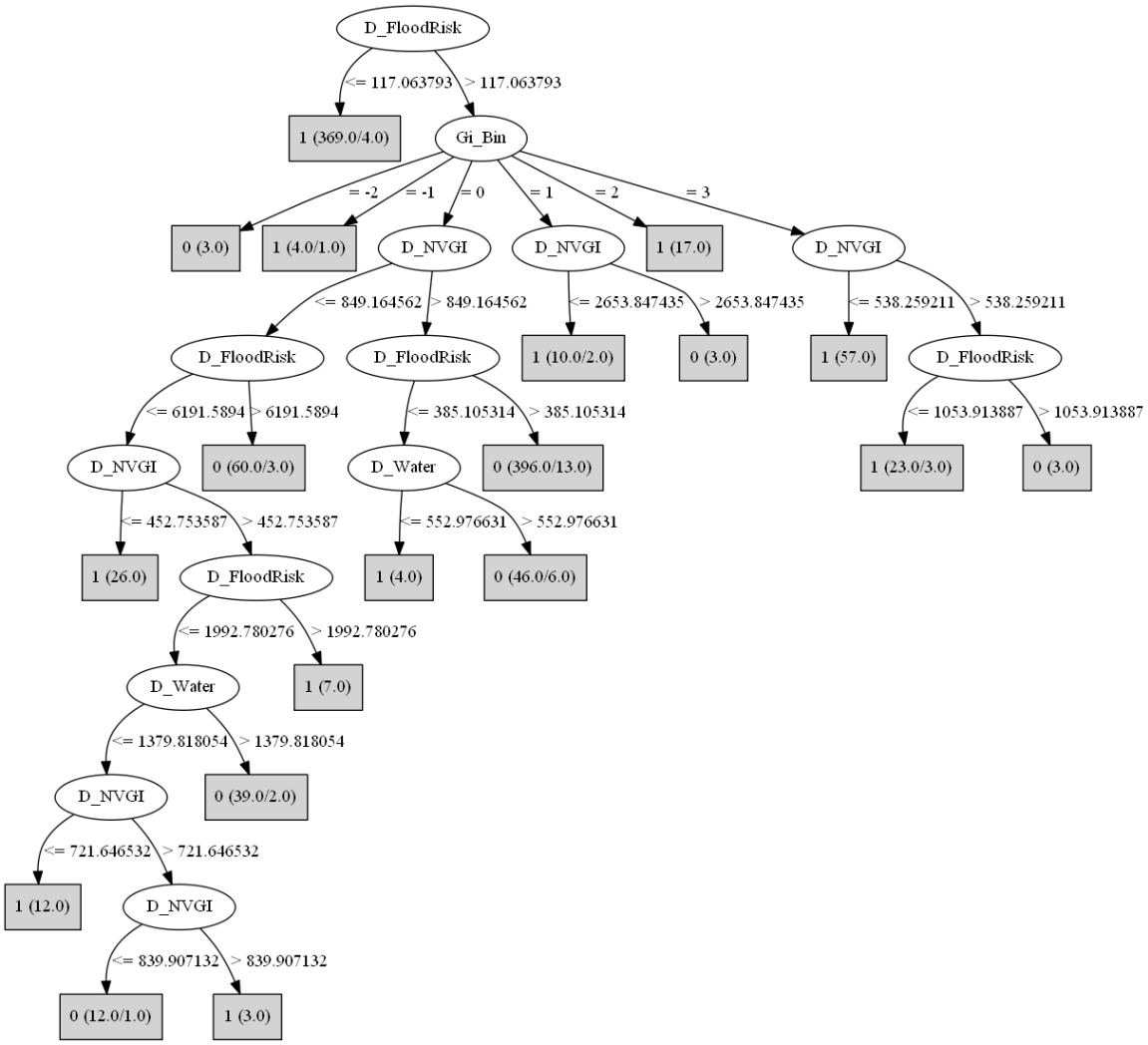


Figure 6.3: J48 tree trained by the geographic method (Exp. 1-A)

2. Text content quality assessment model

The performances of text content quality models developed by the geographic method were subsequently evaluated. Table 6.6 lists the results. Although all the models could reach a fair performance (AUC > 0.7), the accuracies were not high (i.e. < 70%). Precision, recall, and F1 were often low and not consistent. Due to this, this research concluded that the geographic method was not appropriate for text content quality classification.

In comparison to the credibility models, the drop in performance is probably due to the annotation process. Since the text content quality annotation used the specific text-related characteristics, the performance of the geographic method is limited. However, the results still demonstrate a weak geographic correlation to text content quality which requires further investigation.

Table 6.6: Performances of text content quality assessment models (the geographic method)

Algorithm	Exp. Type	Class	Accuracy	P	R	F1	AUC
Logistic	1-A	High	64.5%	0.545	0.530	0.537	0.768
		Medium		0.605	0.564	0.584	
		Low		0.736	0.782	0.758	
	1-B	High	56.4%	0.488	0.458	0.473	0.714
		Medium		0.501	0.606	0.552	
		Low		0.776	0.612	0.684	
J48	1-A	High	64.6%	0.657	0.587	0.620	0.808
		Medium		0.598	0.738	0.661	
		Low		0.746	0.576	0.650	
	1-B	High	56.6%	0.488	0.458	0.473	0.748
		Medium		0.500	0.660	0.569	
		Low		0.842	0.714	0.773	

6.3.4 Performance evaluation of Exp. 2

The Exp. 2 tested the predictive capability of the geographic method. Unseen data from a different event was used as a testing dataset. The performances of VGI models are summarised in Table 6.7. The results showed that the credibility model had a consistent accuracy (about 72% to 75%), and the models reached a fair performance (AUC > 0.7).

Similar to Exp.1, J48 Decision Tree had a better performance in comparison with Logistic Regression.

However, in comparison with the Exp. 1-B (Table 6.4), the performances of the credibility models was lower; Precision of high-credibility instances and Recall of low-credibility were not good. This means that the 2011 models developed by the geographic method were more sensitive to low-credibility instances. The difference was probably caused by two reasons. First, the data size of the 2011 event dataset and 2013 event dataset was different, so the clustering predictors, particularly "distance to nearest VGI instance" had different patterns and thresholds across events. The variances of patterns probably caused a drop in the performances across events, as a part of "high-credibility" instances of the 2013 events with a long distance to the nearest neighbour was classified as "low-credibility" by the 2011 model. Figure 6.4 demonstrates an example of "high-credibility" VGI instances classified as "low-credibility" due to the long distance to its nearest neighbour.

Table 6.7: Predictive capability of the models (the geographic method across events)

Algorithm	Model	Class	Accuracy	P	R	F1	AUC
Logistic	Credibility	High	72.0%	0.673	0.888	0.766	0.741
		Low		0.835	0.566	0.675	
	Text content quality	High	59.4%	0.480	0.304	0.372	0.734
		Medium		0.807	0.545	0.651	
		Low		0.577	0.856	0.689	
J48	Credibility	High	75.1%	0.695	0.858	0.768	0.792
		Low		0.813	0.622	0.705	
	Text content quality	High	60.9%	0.615	0.458	0.525	0.777
		Medium		0.645	0.759	0.697	
		Low		0.701	0.733	0.717	

Figure 6.4: Misclassification of a "high-credibility" instance with long distance to the nearest neighbour

Second, the geographic predictors used did not reflect the dynamic of flooding comprehensively. The 2013 event is a minor flood and the 2011 event is a major flood (see: Table 5.5). The severity of floods probably caused the change of flooded locations and such information may not be seen in the input data totally. Therefore, if using the geographic method for assessing VGI instances collected from future flooding events, an adaptation such as the integration of real-time geographic information and VGI to producing new predictors is suggested.

On the other hand, the results of the text content quality models were similar to the Exp. 1. Due to low accuracies, using text-related predictors in the model is necessary for the improvement of the text content quality models.

6.3.5 Sensitivity analysis of the geographic method

In section 6.34 and 6.3.5, a performance issue of data size and the predictor "distance to nearest VGI instance" in the credibility model was discussed. In this section, a brief sensitivity analysis of this issue is performed. Sensitivity analysis is a way to measure the uncertainty in the output of a mathematical model from the different sources of its inputs. Many methods have been developed to ensure the robustness of models; a review of these methods can be found in Gevrey, Dimopoulos, and Lek (2003). For the interpretability of

models, the sensitivity analysis of the predictors is important. This can be calculated by changing the predictor value or try to ignore it while all the other predictors stay constant which is a transformation method by using a uniform distribution, permutation, and missing values (Naaman, 2019). Besides, a quick sensitivity analysis by changing the input data size (Wickham, 2018) is also helpful to investigate the performance issue.

Therefore, two sensitivity analyses were conducted. First, the data size sensitivity in the credibility models was analysed by resampling and model retraining. Figure 6.5 shows the relative accuracy by changing the input data size (100%, 80%, 50%, 30%), which indicated the models have similar sensitivity on data size as the accuracy dropped with the decrease of data size. It is noteworthy that J48 had better performances with a small data size as the slope between 50% and 30% was less steep. Besides, if the data size is too small (30% in the 2013 event, n =122), the accuracies were under 80%.

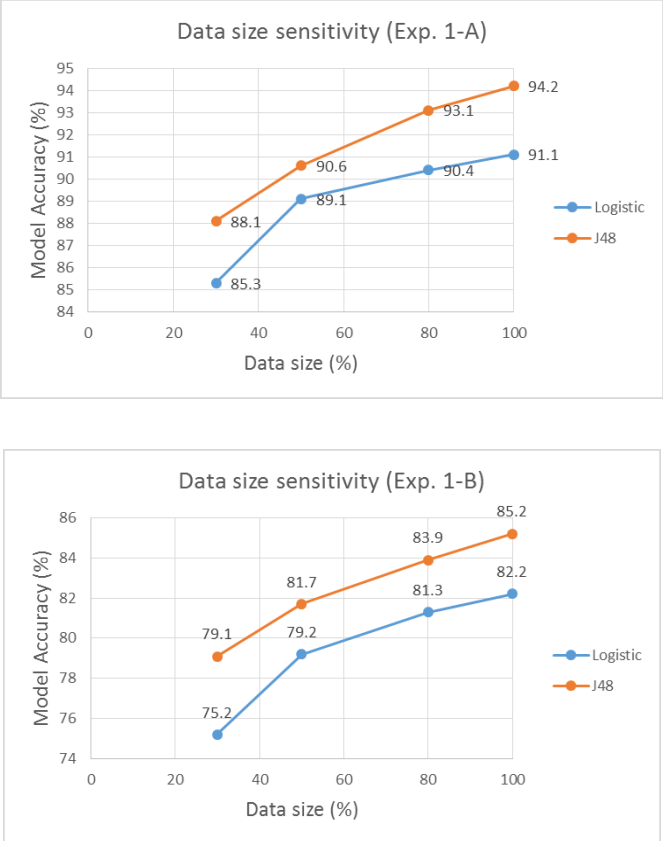


Figure 6.5: Data size sensitivity analysis

The second sensitivity analysis tested the change of model performances by removing one of the geographic predictors and others stay constant (i.e. the missing values). The results of accuracy were listed in Table 6.8. It indicated that the existences of two predictors, "Distance to flood risk zones" and "Distance to nearest VGI" were more sensitive to the model performances as the drops were more significant while removing. Moreover, the Precision of low-credibility classes was provided in the table, as this analysis found that removing "Distance to nearest VGI" generally resulted in a lower Precision. This result indicated that the prediction of low-credibility classes is very sensitive to "Distance to nearest VGI".

Table 6.8: Sensitivity analysis of removing one geographic predictor

		Accuracy (Precision of low-credibility classes)				
Algorithm	Exp.	All predictors	Distance to flood risk zones (x_1)	Distance to nearest VGI (x_2)	Distance to water resource areas (x_3)	Gi_Bin of Hot spot zone ($x_{4\sim} x_9$)
Logistic	1-A	91.1% (0.890)	88.2% (0.873)	88.7% (0.852)	89.7% (0.867)	90.4% (0.882)
	1-B	82.2% (0.854)	80.3% (0.839)	80.9% (0.821)	81.5% (0.832)	81.8% (0.844)
J48	1-A	94.2% (0.912)	91.9% (0.883)	91.3% (0.859)	93.1% (0.901)	91.5% (0.894)
	1-B	85.2% (0.863)	82.8% (0.844)	82.5% (0.835)	84.3% (0.855)	84.7% (0.858)

6.4 Development and evaluation of the integrated method

The integrated method used geographic predictors and the text-related predictors in the modelling. The main assumption of this method is that such integration can help to improve the classification performance of text content quality.

6.4.1 Text-related predictors

Text-related predictors are often used in the standard text classification approach and have demonstrated the proven capability for the text content quality problem. The text-related predictors used in this research were extracted by LightSIDE and Stanford CoreNLP. They were further categorised into two sets. Before the extraction, it requires several processing steps to remove unnecessary and avoidable features for dimensionality reduction of the predictors.

6.4.1.1 Text processing

The text processing in this research included the following steps:

(1) URL detection: the URLs in the texts were manually searched and removed. The existence of URL was annotated in a new field of the data table.
(2) Tokenisation: tokenisation chops the text up into pieces, called tokens. At the same time, punctuation such as commas and stop words (e.g. "and" or "the") were removed, Besides, all upper case strings were converted to lower case.
(3) Stemming: stemming is a process to reduces inflected and derived words to their word stem. Given an example of VGI text "Mundoolun Road, Jimboomba: Flooded in 2 places between Green Ridge Rd and Monmouth Court". After tokenisation and stemming, the text is "mundoolun road jimboomba flood in 2 place between green ridge rd and monmouth court".
(4) Abbreviation conversion: some words in the text were the abbreviation (e.g. "rd/Rd" is the abbreviation for "road"). These abbreviations were manually transformed into a single canonical form.

6.4.1.2 Text-related predictor Set 1

Text-related predictor Set 1 included the common features widely used in established NLP approaches (e.g. Verma et al., 2011)., which were (1) Number of words (numeric) (2) Unigrams (binary feature vector), (3) Bigrams (binary feature vector), and (4) Part of Speech bigrams (binary feature vector). Note that words did not appear in at least three

instances were removed from the unigrams/bigrams. The definitions of these features are described as follows.

1. Number of words

In the use of language, informative contents often have long sentences or many words. Number of words (NoW) is a numeric feature indicating how many words are in a VGI instance, which can be easily extracted. The second row of Table 6.9 demonstrates an example that two lines of VGI instances have 2 and 6 characters separately.

Table 6.9: A simple example of NoW and binary unigram feature vector

Text content	**NoW**	**added**	**near**	**over**	**richmond**	**road**	**water**	**video**
added video	2	T	F	F	F	F	F	T
water over road near Richmond Rd	6	F	T	T	T	T	T	F

Note: NoW-Number of Words, T-True, F-False

2. Unigrams and bigrams

Unigram or bigram is a form of the n-gram model (i.e. "Bag-of-words" model). An n-gram is a contiguous sequence of n words, which is the most common lexical feature (i.e. feature at the word level). To capture the semantic meaning of words, the size of n can be increased. For example, the bigrams (n=2) parses two adjacent words. However, the increase of n adds the dimension of data and requires more computation resources with mixed results. Due to this, only unigram and bigram model were used in this case study. The rationale behind n-gram feature is that texts of the same class tend to share words (Mishra, Mishra, & Sharma, 2013). For example, if the texts contained specific words like "water" and most of them were informative, then the occurrence of "water" is probably significant to the quality judgment.

Examples of the unigram extraction are also demonstrated in Table 6.9. The unigrams (n=1) of the two instances are "added, video" and "water, over, road, near, Richmond, road". The occurrences were presented as "True" or "False" (or the frequency of word occurrences). Moreover, the bigrams are "add video" and "water over, over road, road near, near Richmond, Richmond road". The representation of n-gram model of text-based VGI instances can use a binary feature vector such as the example of Table 6.9, which was also used in this research. Note that the order of words is ignored in a simple unigram model.

3. Part of Speech Bigrams

Part of speech (PoS) tag is a way to classify words according to their functions in context. PoS tags generally have 9 categories, which are noun, pronoun, verb, adjective, adverb, preposition, conjunction, interjection and determiner. PoS is complementary to the n-gram model as PoS examines the syntax of the text.

The PoS tag recognition in this research was extracted by LightSIDE. The tags are quite complex with more than 30 possibilities. An example is demonstrated here. The bigrams of the abovementioned instance "water over road near Richmond Rd" were "BOL_NN (water), NN_IN (water over), IN_NN (over road), road near (NN_IN), near Richmond (IN_NNP), Richmond road (NNP_NN), NN_EOL (road)"; where BOL and EOL represent "Beginning and End of Line"; NN represents a noun; NNP represents a proper noun; and IN represents a subordinating conjunction. More information about the PoS tags in LightSIDE can be referred to the web page of CLiPS[36].

6.4.1.3 Text-related predictor Set 2

In addition to Set 1, there were some specific characters and textual contexts relevant to the text content quality annotation process. These characteristics were called Text-related predictor Set 2, which were only used to train the text content QA classifiers for performance optimisation. Text-related predictor set 2 included: (1) location granularity

[36] http://www.clips.ua.ac.be/pages/mbsp-tags (note that relation tags were not detected by LightSIDE.)

(nominal) (2) the existence of URLs (binary), (3) the existence of inundation depth in texts (binary), and (4) the existence of temporal expression (binary).

These predictors were generated in different ways. First, Location granularity[37] and the existence of URLs had been annotated in the previous processing steps. The two predictors can be directly used. Next, the existence of inundation depth in texts was partially manually annotated on the co-occurrences of numbers and specific terms including "deep", "meter", "cm" in texts (i.e. the occurrences of numbers has been annotated by the PoS Bigrams "CD"). As for the existence of temporal expression, the Name-Entity Recognition (NER) of Stanford CoreNLP was used (Chang & Manning, 2012).

6.4.2 Experiment design of the integrated method

The experiment design of the integrated method was similar to the geographic method. A 2-step experiment was developed and evaluated based on the oversampled VGI instances (Table 6.10). Furthermore, since the standard text classification approach often only used the text contents in the classification, a comparison of the integrated method and a baseline method that only used text contents was addressed here. The models were trained by two kinds of predictor setting for observing the variance in performance. The first one was the baseline that only the text-related predictors (Set 1&2) were used in training. The second one used the two kinds of predictors (i.e. the integrated method).

Table 6.10: The experiment design and targets of the geographic method

Exp. type	Training dataset	Testing dataset	Target
3-A	2011	2011	Evaluating the model's performance by using the datasets from an individual event. Two kinds of the predictor sets were used.
3-B	2013	2013	

[37] Location granularity was categorised into 3 levels, "Address" and "Road junction" =0, "Park/landmark", "River", and "Road section" =1, and "Suburb" = 2

Exp. type	Training dataset	Testing dataset	Target
4	2011	2013	Evaluating the model's predictive capability in the different flooding event. Two kinds of the predictor sets were used.

6.4.3 Performance evaluation of Exp. 3

1. Credibility assessment model

The credibility models trained by the baseline method and the integrated method were listed in Table 6.11. First, compared to the previous credibility models of the geographic method, the performances of the baseline method were lower. For example, the accuracies of the 2011 credibility models in Exp. 3-A were under 79%. But the accuracy was higher than 91% in Exp. 1-A. This shows that the geographic predictors were more significant and useful in the credibility classification.

Next, the performances of the integrated method had mixed results compared to Exp. 1-A. Only the accuracies and AUC in J48 were a little higher than the previous models. For example, the accuracies of Exp. 3-A and 3-B were 94.9% and 85.6% compared to 94.2% and 85.2% of Exp. 1-A and 1-B in Table 6.4 respectively. This is probably because the use of the text-related predictors increased the dimension of predictor space and Logistic regression generally has a limitation to deal with high-dimensional data. Therefore, this research concluded that the integrated method could have a better performance than the geographic method by using an appropriate algorithm such as J48 Decision Tree.

Table 6.11: Credibility assessment models' performances (the integrated method)

Algorithm	Exp. Type	Predictors	Class	Accuracy	P	R	F1	AUC
Logistic	3-A	All	High	86.5%	0.818	0.933	0.872	0.888
			Low		0.925	0.800	0.858	
		T	High	78.3%	0.751	0.811	0.780	0.774
			Low		0.803	0.741	0.771	
	3-B	All	High	84.7%	0.899	0.788	0.840	0.884
			Low		0.805	0.909	0.854	
		T	High	72.4%	0.795	0.673	0.729	0.746
			Low		0.662	0.787	0.719	
J48	3-A	All	High	94.9%	0.958	0.941	0.950	0.951
			Low		0.940	0.957	0.949	
		T	High	78.7%	0.815	0.793	0.804	0.817
			Low		0.754	0.779	0.766	
	3-B	All	High	85.6%	0.891	0.818	0.853	0.889
			Low		0.826	0.896	0.860	
		T	High	73.2%	0.773	0.727	0.749	0.761
			Low		0.687	0.738	0.711	

Moreover, an analysis of the J48 Decision Tree graph trained from the Exp. 3-A showed that the initial nodes of the tree still classified the VGI instances by the geographic predictors. However, some text-related predictors were significant in the splits. For example, the occurrence of the word "flood" at the beginning of a text (BOL_flood, e.g. "flooded road") or the use of past participle verb at the beginning of a text (BOL_VBN e.g. " fallen tree") can help the classification.

2. Text content quality assessment model

Next, the performances of the text content QA models trained in Exp. 3 are listed in Table 6.12. The results demonstrate that the text-related predictors were probably necessary for the text content classification, as the baseline method (T) had good and stable performances in comparison with the previous models of the geographic method (Table 6.4). The baseline method had accuracy around 75% to 85% and good AUC.

On the other hand, as the integrated method had a better weighted AUC and accuracy in Table 6.12, this indicated that the integrated method can enhance the model performances compared to the baseline. This research, therefore, concludes the integrated method can be the optimized way for the text content QA.

Table 6.12: Performances of text content quality (the integrated method)

Algorithm	Exp. Type	Predictors	Class	Accuracy	P	R	F1	AUC
Logistic	3-A	All (integrated method)	High	84.4%	0.811	0.808	0.809	0.899
			Medium		0.802	0.837	0.819	
			Low		0.956	0.900	0.927	
		T	High	83.2%	0.846	0.846	0.819	0.884
			Medium		0.836	0.817	0.826	
			Low		0.876	0.838	0.857	
J48	3-A	All (integrated method)	High	83.1%	0.876	0.751	0.809	0.954
			Medium		0.832	0.829	0.830	
			Low		0.784	0.950	0.859	
		T	High	81.8%	0.810	0.762	0.786	0.951
			Medium		0.888	0.771	0.825	
			Low		0.756	0.954	0.843	

Algorithm	Exp. Type	Predictors	Class	Accuracy	P	R	F1	AUC
Logistic	3-B	All (integrated method)	High	77.0%	0.657	0.596	0.625	0.855
			Medium		0.677	0.759	0.715	
			Low		0.947	0.919	0.933	
		T	High	74.2%	0.636	0.450	0.527	0.843
			Medium		0.622	0.767	0.687	
			Low		0.922	0.956	0.939	
J48	3-B	All (integrated method)	High	76.4%	0.622	0.560	0.589	0.862
			Medium		0.664	0.750	0.704	
			Low		0.970	0.941	0.955	
		T	High	73.7%	0.616	0.560	0.587	0.803
			Medium		0.637	0.681	0.658	
			Low		0.913	0.926	0.920	

To further understand the significances of the used predictors in text content quality model. The correlation coefficient (*r*) ranker method was used here. Table 6.13 listed the value of *r* and the ranking order. The results demonstrated three geographic predictors were ranked in the top 10. Several text-related predictors were ranked in top 10: the granularity label, the occurrence of the hashtag "#qldflood", the existence of temporal expression, and 3 PoS bigrams "DT_NN", "IN_CD", and "NNP_NNP". A short explanation of the significant PoS bigrams can be found in the note of Table 6.13.

Table 6.13: the correlation coefficient and the ranking order of the top 10 significant predictors for text content QA

Text-related predictors	Set 1	Unigrams "#qldflood" (0.333, #3), number of words (0.252, #7) PoS bigrams "DT_NN" (0.251, #8, e.g. Photos of **the flooding** in Maryborough..), "IN_CD" (0.250, #9, e.g. Power out **in 17** mile rocks..), "NNP_NNP" (0.242, #10, e.g. flood house),
	Set 2	Granularity (0.339, #2), existence of temporal expression (0.290, #5),
Geographic predictors:		Distance to nearest VGI instance (0.354, #1), Distance to the flood risk zones (0.294, #4), Distance to water resource areas (0.265, #6)

Note: DT = determiner, NN = noun (singular or mass), IN = conjunction or subordinating or preposition, CD = cardinal number, NNP = noun, proper singular

6.4.4 Performance evaluation of Exp. 4

The Exp. 4 tested the predictive capability of the integrated method. The performances of VGI models are summarised in Table 6.14. In the best case, by using J48, the weighted AUC reached a good performance (>0.8). Yet some issues in other performance metrics were found. The interpretation can be described according to two kinds of quality labels. First, the predictive capability in credibility had an improvement; the accuracy increased about 5% (J48 in Exp. 3-B: 85.6% v.s. Exp. 4 79.4%); other metrics demonstrated a more acceptable result as well. Based on this, this research concluded that the integrated method had a better predictive capability for credibility classification across events.

Second, as for text content quality, the difference between the performance metrics compared to Exp. 3-B is small (e.g. the accuracy of J48 in Exp. 3-B: 76.4% v.s. Exp. 4 75.5%). The predictive capability seemed to be fine in text content QA. Yet a problem here

is the low performance of "High/Medium-quality" classification, which requires a further investigation for the improvement. The assumption here is that the textual pattern of "High/Medium-quality" instances in the 2013 event was not clear. Thus the performances were lower compared to the models trained by the 2011 event dataset.

In brief, the abovementioned facts indicate that the integrated method can perform more consistent in both credibility and text content quality QA. The use of text-related predictors can alleviate the impact of shortcoming in the tested models to some extent, as the predictive capability across events was better than the geographic method.

Table 6.14: Predictive capability of the models (the integrated method across events)

Algorithm	Model	Class	Accuracy	P	R	F1	AUC
Logistic	Credibility	High	76.6%	0.679	0.561	0.615	0.753
		Low		0.799	0.868	0.832	
J48	Credibility	High	79.4%	0.891	0.668	0.764	0.884
		Low		0.736	0.919	0.817	
Logistic	Text content quality	High	73.2%	0.551	0.466	0.505	0.780
		Medium		0.702	0.780	0.739	
		Low		0.923	0.882	0.902	
J48	Text content quality	High	75.5%	0.659	0.500	0.569	0.860
		Medium		0.739	0.774	0.756	
		Low		0.846	0.928	0.909	

6.5 Discussion and chapter summary

In this chapter, the geographic method and the integrated method were developed and evaluated. The results of the geographic method demonstrated that the geographic predictors were powerful in credibility assessment; the credibility models had excellent performance. However, if the geographic method was used in the text content quality models, the performances had low accuracies and not consistent. Therefore, the integrated

method was necessary to deal with text content QA. Compared to the standard text classification approach (models trained by text-related predictors), the integrated method enhanced the performance of text content quality models. Besides, the evaluation of the experiments presents that the integrated method could have better performance than the geographic method by using an appropriate algorithm. The integrated method was therefore recommended as the artefact for text-based VGI QA. Furthermore, an issue of the predictive capability for credibility classification was identified from the experiments. Although two sensitivity analyses were performed, this issue still requires further investigation.

A further discussion of the experiment results will be presented in the next chapter, which is the concluding chapter of this thesis. The overall achievements and findings of this research will present as well.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 7

Conclusions and Recommendations

7.1 Introduction

This chapter concludes the major findings and contribution of this research. In the first part of this chapter, the main aim and the objectives mentioned in Chapter 1 are revisited. The achievements were described. Next, the main outcomes and the contributions of this research are briefly summarised. In the final part, the future research direction for improving the modelling method developed in this research is discussed.

7.2 Achievements of research aim and objectives

The concept of VGI has coined by Goodchild (2007). With the rise of new ICTs such as social media and crowdsourcing platforms, the use of text-based VGI has received increasing attention. For disaster response, there have been many projects showcasing that text-based VGI can provide crucial information. For operations in disaster response, credible and informative content is absolutely essential. As VGI can be overwhelming and contains inherent problems with uncertain properties, various methods have been developed for text-based VGI quality assessment (QA). However, these methods often ignore the correlation between geographic characteristics and quality. Therefore, a new method for finding credible and informative content, and minimizing information overload in crisis situations is critically important. This concept formulated the primary aim of this research. As described in Chapter 1, the aim was:

To develop and evaluate methods based on geographic characteristics of text-based VGI for generating credibility and text content quality labels in the disaster response phase.

The proposed methods, as the main artefact of this research, intend to provide an efficient process for credibility assessment and text content QA. The quality problem is treated as a classification problem and a modelling approach was further designed and developed. A series of research steps were completed in four phases, which were (1) Conceptual, (2) Design, (3) Development and Evaluation (D&E), and (4) Synthesis and Communication phase. In the conceptual phase, the research objectives were formulated. For addressing these objectives, the detailed research activities were designed (see: Section 4.4). The

integration of VGI and authoritative data and Supervised learning techniques have been employed as the primary basis for the development of the modelling method and the validation of geographic correlation.

The VGI QA methods were further designed, developed, and evaluated through a case study of Brisbane Floods. Text-based VGI instances in 2011 and 2013 flooding events were collected. These instances were annotated from a mechanism developed by two emergency management experts' opinions and processed to a balanced dataset. Two kinds of QA methods were further tested. The geographic method used the geographic predictors, and the integrated method tested a combination of the geographic predictors and text-related predictors.

From the D&E of the VGI QA methods, this thesis made two major achievements for using text-based VGI in disaster response: (1) the validation of geographic correlation in VGI quality problem, and (2) efficient methods for VGI QA based on geographic data integration and text-processing. The further details of the achievement of the research aim are described with the achievements of the research objectives discussed in the following subsections.

7.2.1 Objective one

The objective 1 of this research was "investigate and understand the use of VGI in the context of disaster management and the associated QA methods". The objective 1 targeted to provide the foundation of this research: the background and importance of the VGI quality issue. It addressed objectively report the current knowledge through previously published research. The objective 1 comprised three components, which were: (1) the concept of VGI and relevant cases, (2) the existing QA methods in various forms of VGI, and (3) the advantages and limitations of these methods, particularly for text-based VGI QA.

The investigation of these components was undertaken through an exhaustive narrative review of literature about VGI. The results were presented in Chapter 2 and Chapter 3. It first described the definition, forms, and characteristics of VGI. Then, the benefits of using

VGI in disaster management were addressed. Various VGI collection approaches for disaster management and the contributors' characteristics were described. The literature review specifically discussed and summarised current VGI QA methods. According to the review, the author addressed that "VGI quality" is a complex concept as the form and characteristics of VGI is heterogeneous. Therefore, the traditional GI specifications of data quality were not adequate to assess the quality of text-based VGI. The information need of disaster response should be examined from the "external quality" approach and the information quality criteria such as credibility and text content quality.

The literature review further summarised four major types of VGI QA methods: (1) Comparison, (2) Crowdsourcing, (3) Multi-criteria assessment, and (4) Supervised learning. The advantages and limitations of these methods were further identified. First, although researchers had attempted to develop the multi-criteria measures, the developed measures were oversimplified or conceptual; evaluations of the developed measures were often absent. Particularly, the correlation between geographic characteristics of VGI and quality was not tested and evaluated rigorously in the existing efforts. On the other hand, Supervised learning and the associated techniques had the advantage to evaluate the correlation of predictors for model development. However, the use of geographic predictors was not considered in the published research.

Due to the two gaps, refinement and development of rigorous methods for text-based VGI QA were identified. Supervised learning techniques were further employed as the primary basis for developing the modelling method.

7.2.2 Objective two

The objective 2 of this research was to validate the correlation between the text-based VGI quality and the geographic characteristics of text-based VGI in the extreme flooding events. A case study using data of previous extreme flooding events in Brisbane was employed for the validation. Geotagged VGI instances including Ushahidi reports and geotagged tweets, were collected.

A lot of this study's efforts focused on the data preparation and data integration for the development of the QA method. This formulated the first component in the objective 2: design and development of relevant processes for data collection and data preparation in a given context. Several techniques, tools, programming libraries were used (see: Chapter 5). During the data preparation, although only geotagged VGI instances were used, VGI instances with a mismatch positioning were found. Thus a core task, data cleaning process, was developed. This process used a hierarchical toponym resolution with a geoparser and a string matching method to determine the location description/toponym. For excluding mismatched instances, the distance from the positioning to the geo-referenced objects was calculated. Another core task here is quality label annotation. A part of VGI instances was sampled in a questionnaire, and then the experts participated in the questionnaire survey. The experts' opinions were further used for developing the annotation process. Based on this, the credibility was annotated by the metadata properties or the supporting geographic data. On the other hand, the text content quality was annotated by textual description and other criteria.

The second component in the objective 2 is the identification of the significant characteristics impacting credibility and text content quality of VGI for the validation of geographic correlation. The validation used the correlation coefficient to rank the correlation significance and the chi-squared test to test the statistical significance. The result showed that four geographic predictors were correlated, which were (1) distance to nearest VGI, (2) hot spot zone, (3) distance to water resource areas, and (4) distance to flood risk zones.

7.2.3 Objective three

The objective 3 of this research is to design, develop, and evaluate the methods for assessing text-based VGI quality in disaster response. These are two components in the objective 3: (1) D&E of the geographic method and the integrated method based on models of Supervised learning, and (2) testing the predictive capability of the methods in a different flooding event.

The D&E of the proposed method used two kinds of algorithms: Logistic Regression and Decision Tree (J48). In a series of the experiment, several combinations of the training dataset, the testing dataset, and the selected predictors were tested. The results showed that the geographic predictors were critical and highly correlated to credibility; the geographic method had excellent performances in credibility classification. However, the geographic method did not have a high accuracy in text content quality classification (<70%). Due to the difference in flood severity and the size of collected VGI instances, the predictive capability of the credibility model across events was lower.

On the other hand, the integrated methods demonstrated good performances in both credibility and text content quality classification. The integrated method could enhance the performances compared to other experimental settings. The integrated method was therefore recommended for the optimisation of model performances

7.3 Main contributions to the knowledge base

From the achievements addressed in the previous section, a summary of the main contributions of this research are outlined below:

1. The validation of the geographic correlation and the use of geographic predictors

As discussed in Chapter 3, a geographic approach for VGI QA was addressed in the literature (Craglia et al., 2012; Goodchild & Li, 2012). However, most existing efforts are not flexible enough to identify the correlated geographic factors which have an impact on the VGI quality. The methods developed in this research addressed this gap. The correlation between the quality and a variety of geography predictors was validated, and the statistical significances were evaluated. Since most text-based VGI QA methods ignore the geographic characteristics, this finding can contribute a new perspective for a "geographic examination" in text-based VGI quality problem.

The developed QA methods also showcased a better mechanism for understanding the geographic correlation. Coefficients of a specific predictor in a linear function or the thresholds in Decision Tree can be observed. The VGI instances can be filtered and used

more efficiently based on the models. Therefore, the methods can be used as a decision support tool for a number of theoretical and practical applications; they can improve VGI credibility verification and identification of informative contents in time-critical situations. If leveraged strategically, these models can strengthen and augment current crisis management systems.

2. A state-of-art method for text-based VGI quality assessment

Advances in data analytics provide opportunities to receive, process, and make use of text-based VGI more efficiently. In this thesis, the uses of several techniques (i.e. GIS, NLP, and ML) enable an innovative method for adopting high-quality VGI in disaster response. The case study showcased the detailed data input in the models and relevant processing steps such as the data cleaning process and the annotation process. The developed QA methods demonstrated the state-of-art techniques.

A major contribution of the developed method is to reduce the uncertainty in the use of text-based VGI. Unstructured VGI contents were filtered and VGI instances can be assigned the credibility and text content quality labels or indicators by the models. This modelling method contributes to the information need of disaster information managers to assist their actions and decision-making. Future efforts for using web-based information/VGI in disaster management can benefit greatly from this research.

7.4 Recommendations for future research

Throughout conducting this research, a number of issues were identified. Where some of them would contribute to addressing the limitation in this research, others can be extended to future research direction for the improvement of the developed methods.

1. Sensitivity to training data size and the use of clustering predictors

In the experiments, 606 and 295 geotagged VGI instances were collected from 2011 and 2013 Brisbane flooding event. These instances were used to develop the QA models. Though there is no strict requirement of training data size in Supervised learning modelling,

the results demonstrated that the model trained by the 2011 event dataset had better performance than the 2013 event dataset generally. Therefore, this research supposes the models were sensitive to the training data size; a model trained by more VGI instances is expected to have better performance. The data size also impacts the evaluation of predictive capability (Exp. 2 and Exp. 4). Particularly, data size links to the pattern of the clustering. For example, in a smaller training dataset, the clustering could be more implicit. This might limit the use of models in some scenarios. Therefore, the data size sensitivity analysis was performed in 6.3.5, and the number of the collected VGI instances was suggested to be more than 122 for the employment of the developed method. A further investigation of the relationship between data size and performances of models is suggested.

In addition to the data size, the experiments in this thesis ignored the temporal aspect of VGI. To precisely reflect the clustering of VGI instances with the lapse of time, the temporal aspect of information is suggested to be included in the future experiments.

2. Using dynamic flooding information as predictors

In this research, the geo-context predictors were generated from static data. The absence of dynamic flooding information such as in-situ water level sensor measurement should be addressed. Dynamic flooding information was supposed to enhance the model performance as well as the predictive capability across events in different flooding severity. Therefore, the integration of dynamic flooding information in the modelling process is suggested. It is expected that such integration can reflect the flood-affected area more precisely and formulates new geographic facts in the QA process.

3. Availability of "accurate" geotagged data

To validate the geographic assumption, the experiments of this research were established by using geotagged VGI instances. Since the data cleaning process addressed the mismatched positioning issue, the availability of "accurate" geotagged VGI instances is a major limitation to promote the developed methods. First, the data cleaning process in this research spent a lot of time to filter geotagged data (about 3 days for 1473 instance). This

issue limits the use of the developed method in the response. Second, social media were the most convenient platform to harvest text-based VGI but most feeds were not geotagged. Therefore, advanced techniques of geocoding and ensuring positional accuracy are suggested to be developed and tested in the future. For example, geocoding based on the geographic context of location have been discussed in some studies; Huck, Whyatt, and Coulton (2015) tried to use a weighted redistribution by real-world pattern (e.g. population density) to locate non-geotagged VGI in a more precise position; such techniques have the potential to overcome this limitation. Moreover, with the widespread adoption of GPS-enabled devices, the availability of "accurate" geotagged VGI instances is likely to increase in the future.

4. Investigation of the developed methods in other study areas and types of disaster

This research conducted a case study to develop the modelling method for text-based VGI QA. In this case study, the texts of VGI instances were all written in English, and the associated flood-related data were used. As English is the most spoken language and the authoritative data used in this research are usually available in developed countries, the developed methods are expected to be extended and used in other study areas.

However, this research found the scenarios were limited to forest fires and floods. The correlation of geographic characteristics in other types of disasters was suggested to be further investigated. With appropriate supporting geographic data, the methodology used in this thesis has potential to be used broadly.

7.5 Chapter summary and a brief conclusion

On the foundation of the arguments in this chapter, the geographic method and integrated method developed in this research are practical for the information magers to be used in the response. The integrated method is suggested as the best text-based VGI QA solution according to the experiments. By linking VGI with the trend of today's interactive culture in a networked society, this research established a good example to identify useful text-based VGI for disaster response and supporting operations.

THIS PAGE INTENTIONALLY LEFT BLANK

Appendices

Appendix 1. A list for the retrieval of tweets developed by the volunteer of VOST VIC

"edgecombe" OR "woodleigh Heights" OR "bald hill" OR "kyneton" OR from:@stevenzan OR from:@SimonZ72 OR from:@SteveKendall1 OR "pipers creek" OR "pipers ck" OR "baynton" OR "pastoria" OR "sidonia" -"knight of sidonia" -"knights of sidonia" -from:"edgecombe" OR "woodleigh Heights" OR "bald hill" OR "kyneton" OR from:@stevenzan OR from:@SimonZ72 OR from:@SteveKendall1 OR "pipers creek" OR "pipers ck" OR "baynton" OR "pastoria" OR "sidonia" -"knight of sidonia" -from:@baynton_vicki –RT

Appendix 2. Concept of generalisation and a brief comparison of classification algorithms

The ability of a learning algorithm to perform accurately on new or unseen data after having experienced a learning dataset is called generalisation. However, any algorithm has the prediction error, which can be categorized into three parts: (1) bias error, (2) variance error, and (3) irreducible error. The irreducible error cannot be reduced regardless of what algorithm is used. For example, the training process does not have enough predictors to sufficiently characterize the target function; or the training process uses an inappropriate way to sample training set. Therefore, the techniques of any algorithm and modelling focus on bias error and variance error.

Bias error generated from erroneous assumption made by a model to make the target function easier to learn. On the other hand, variance is caused from fluctuations in the training dataset. A simple explanation of the two concepts is to iterate model building process several times (e.g. the input data can be split into different training sets), if the difference between the predicted values and the actual values is continuously large for all repetitions then it suffers from high bias. If the predictions are inconsistent in the repetition then it is said to have high variance (Chakraborty, Elzarka, & Bhatnagar, 2016).

Figure 1 is a graph visualisation of bias and variance by a bulls-eye diagram. A model which always hit close to the centre of the target such as Figure 1 (a) is a perfect prediction model. Ideally, the goal of any algorithm is to achieve low bias and low variance, which means that the algorithm captures the regularities in its training data, but also generalizes well to new or unseen data. However, in reality, the real bias and variance error cannot be calculated because we do not know the actual underlying target function. There is a tradeoff to minimize bias and variance (i.e. bias-variance tradeoff) in any model techniques. Parametric algorithms such as logistic regression have a high bias making them fast to learn and easier to understand. They provide excellent dense model like Figure 1 (b), but generally less flexible as the model might be over-fitting. In turn, they have lower predictive performance on complex problems that fail to meet the simplifying assumptions

of the algorithms bias. By contrast, nonparametric machine learning algorithms such as SVM are more flexible on complex problems and have a high variance. Although such sparse models like Figure 1 (c) might be under-fitting, they are more flexible to deal with complex problems.

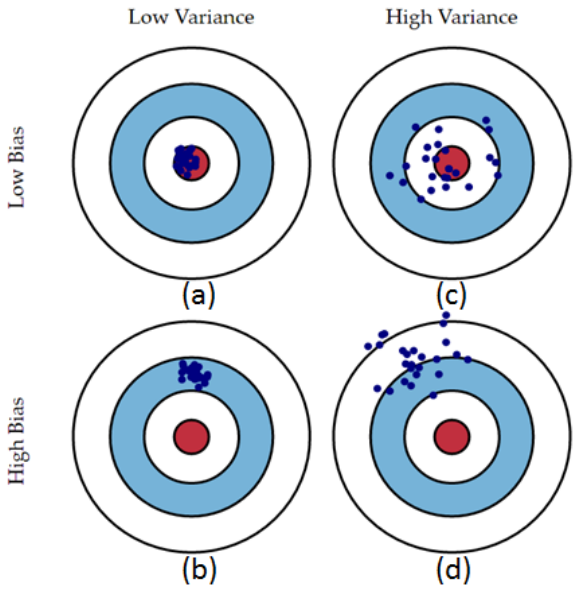


Figure 1: Graphical illustration of bias and variance (source: Fortmann-Roe, 2012)

According to Aggarwal and Zhai (2012) and Brownlee (2016), there are five machine learning algorithms often used in classification and implemented in the selected model tool, Weka of this research. The advantages and disadvantages of the five algorithms are described in the following table.

Table: Advantages and disadvantages of five commonly used classification algorithms

Algorithm	Advantages	Disadvantages
Logistic regression	1. It is easy for implementation (often used as a benchmark), computational efficiency 2. It is easy for the observation of the probability score of samples 3. Issues of multicollinearity can be solved by the regularisation	1. Performance is not good for large feature datasets. 2. It cannot solve non-linear problems 3. Vulnerability to overfitting so a proper presentation of data is critical
Decision tree	1. Easy to interpret, easy to use, computational efficiency 2. It can deal with large feature datasets with good performance and less sensitive to missing data.	1. Vulnerability to overfitting as it makes no assumption on relationships between features (multicollinearity).
Naïve Bayes	1. It was originated from classical mathematical theory and has a solid mathematical foundation and stable classification efficiency; so it explains the results easily. 2. It can deal with large feature datasets with good performance and less sensitive to missing data.	1. Need to calculate the prior probability; 2. There is an error rate in the classification decision; 3. Vulnerability to overfitting so a proper presentation of data is critical
Support Vector Machine	1. It works relatively well when there is clear margin of separation between classes and more effective in high dimensional spaces. 2. It is effective in cases where the number of dimensions is greater than the number of samples.	1. It is not suitable for large data sets. 2. It does not perform very well when the data set has more noise. 3. As it works by putting data points above and below the

Algorithm	Advantages	Disadvantages
		classifying hyper plane, there is no probabilistic explanation for the classification.
Proximity-based algorithm (K-Nearest Neighbors; KNN)	1. It is easy for implementation. 2. Robust with regard to the search space; for instance, classes don't have to be linearly separable. 3. Classifier can be updated online at very little cost as new instances with known classes are presented (no training period). 4. Few parameters to tune: distance metric and k.	1. It is not suitable for large data sets or high dimensional data. 2. Need feature scaling before applying KNN algorithm to any dataset. 3. Sensitive to noisy data, (missing values and outliers).

Appendix 3. The details of authoritative data used in this research

1. References for data cleaning

Physical road network contained attribution of street names, road classification (e.g. freeways, highways, and secondary roads), route numbers, and unique ID. Recreation areas included parks, civic squares, gardens, and so forth in Queensland. "Statistical local area boundary" contained the suburb name of Brisbane. Water resource areas were the natural water channels covering the State of Queensland; the areas were derived from the imageries such as Queensland orthophotography and satellite imagery.

2. References for VGI credibility annotation

2011 and 2013 flood modelling maps were modelling by Bureau of Meteorology based on the forecasted rainfall along the catchment of Brisbane City. The modelling maps indicated the approximate flood inundation during the two events. Brisbane City council referred to the modelling maps and issued the lists of flood-affected streets, which were processed (i.e. marked in a new field) to Physical road network manually.

3. Data for the geographic predictor generation

The format of DEM used in this research was a 25-metre floating-point grid; it was derived from contours and drainage. Population data of each suburb was based on 2015 census, collected from the Australian Bureau of Statistics. Water resources areas have been mentioned in the references for data cleaning. Regarding flood risk zones, the data were merged polygons from several flood studies after the 2011 floods. According to Barton, Wallace, Syme, Wong, and Onta (2015), the flood risk zones are based on hydrologic and hydraulic modelling (2D TUFLOW model) and mapping the run-off. It was calibrated to the 2011 and 2013 events and verified to the historical flooding events (1974, 1996, 1999). The various levels of risk zones were calculated by the estimated flood extents of different Average Recurrence Intervals (ARI). High and medium flood risk zones represented areas with a higher probability of flooding (ARI: 1 in 20 and 100 years respectively). They were merged to measure the proximity from each VGI instance. In contrast, low risk and very

low-risk zones were excluded because flooding is under 0.20% chance of occurring in any year in these areas (ARI: 1 in 500 and 2000 years respectively).

References

Aalders, H. J. G. L. (2002). The Registration of Quality in a GIS. In W. Shi, P. Fisher, & M. F. Goodchild (Eds.), *Spatial data quality* (pp. 186-200). London ; New York: Taylor & Francis.

Aggarwal, C. C., & Zhai, C. (2012). A survey of text classification algorithms. In C. C. Aggarwal & C. Zhai (Eds.), *Mining text data* (pp. 163-222): Springer Science & Business Media.

Agichtein, E., Castillo, C., Donato, D., Gionis, A., & Mishne, G. (2008). *Finding high-quality content in social media.* Paper presented at the Proceedings of the 2008 international conference on web search and data mining.

Ajmar, A., Boccardo, P., & Giulio Tonolo, F. (2011). Earthquake damage assessment based on remote sensing data. The Haiti case study. *European Journal of Remote Sensing, 43*, 123-128.

Alamdar, F., Kalantari, M., & Rajabifard, A. (2016). Towards multi-agency sensor information integration for disaster management. *Computers, Environment and Urban Systems, 56*, 68-85.

Albuquerque, J. P. d., Herfort, B., Brenning, A., & Zipf, A. (2015). A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management. *International Journal of Geographical Information Science, 29*(4), 667-689.

Amirebrahimi, S. (2016). *A framework for micro level assessment and 3D visualisation of flood damage to a building.*

Amitay, E., Har'El, N., Sivan, R., & Soffer, A. (2004). *Web-a-where: geotagging web content.* Paper presented at the Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval.

Antoniou, V., & Skopeliti, A. (2015). *Measures and indicators of VGI quality: an overview*. Paper presented at the ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Science, Volume II-3/W5, La Grande Motte, France.

Arklay, T. M. (2012). Queensland's state disaster management group: An all agency response to an unprecedented natural disaster. *Australian Journal of Emergency Management, The, 27*(3), 9.

Ashktorab, Z., Brown, C., Nandi, M., & Culotta, A. (2014). Tweedr: Mining twitter to inform disaster response. *Proc. of ISCRAM*.

Bai, S., Lü, G., Wang, J., Zhou, P., & Ding, L. (2011). GIS-based rare events logistic regression for landslide-susceptibility mapping of Lianyungang, China. *Environmental Earth Sciences, 62*(1), 139-149.

Barton, C., Wallace, S., Syme, B., Wong, W. T., & Onta, P. (2015). *Brisbane River Catchment Flood Study: Comprehensive Hydraulic Assessment Overview*. Paper presented at the 56th Floodplain Management Australia Conference, Nowra, NSW, Australia. http://www.floodplainconference.com/papers2015/Cathie%20Bartons%20Full%20Paper.pdf

Basiouka, S., & Potsiou, C. (2014). The volunteered geographic information in cadastre: perspectives and citizens' motivations over potential participation in mapping. *GeoJournal, 79*(3), 343-355.

Baskerville, R. (2008). What design science is not. *European Journal of Information Systems, 17*(5), 441-443.

Bekkar, M., Djemaa, H. K., & Alitouche, T. A. (2013). Evaluation measures for models assessment over imbalanced data sets. *Journal Of Information Engineering and Applications, 3*(10).

Bird, D., Ling, M., & Haynes, K. (2012). Flooding Facebook-the use of social media during the Queensland and Victorian floods. *Australian Journal of Emergency Management, 27*(1), 27.

Bishr, M., & Janowicz, K. (2010). *Can we trust information?-the case of volunteered geographic information.* Paper presented at the Towards Digital Earth Search Discover and Share Geospatial Data Workshop at Future Internet Symposium, volume.

Bishr, M., & Kuhn, W. (2007). Geospatial Information Bottom-Up: A Matter of Trust and Semantics. *European Information Society: Leading the Way with Geo-Information*, 365-387. doi:Doi 10.1007/978-3-540-72385-1_22

Bishr, M., & Mantelas, L. (2008). A trust and reputation model for filtering and classifying knowledge about urban growth. *GeoJournal, 72*(3-4), 229-237.

Bordogna, G., Carrara, P., Criscuolo, L., Pepe, M., & Rampini, A. (2014). A linguistic decision making approach to assess the quality of volunteer geographic information for citizen science. *Information Sciences, 258*, 312-327.

Brito Moreira, R. d., Degrossi, L. C., & Albuquerque, J. P. d. (2015). *An experimental evaluation of a crowdsourcing-based approach for flood risk management*. Paper presented at the Conference: 12th Workshop on Experimental Software Engineering (ESELAW), At Lima, Peru.

Brooks, F. (1987). No silver bullet: essence and accidents of software engineering. *Computer, 20*(4), 10-19. doi: doi:10.1109/MC.1987.1663532

Brown, G., & Kytta, M. (2014). Key issues and research priorities for public participation GIS (PPGIS): A synthesis based on empirical research. *Applied Geography, 46*, 122-136. doi:DOI 10.1016/j.apgeog.2013.11.004

Brownlee, J. (2016). How To Use Classification Machine Learning Algorithms in Weka. *Weka Machine Learning.* Retrieved from https://machinelearningmastery.com/use-classification-machine-learning-algorithms-weka/

Bruns, A., Burgess, J. E., Crawford, K., & Shaw, F. (2012). *# qldfloods and@ QPSMedia: Crisis communication on Twitter in the 2011 south east Queensland floods*. Retrieved from Brisbane: http://cci.edu.au/floodsreport.pdf

Budhathoki, N. R. (2010). *Participants' motivations to contribute geographic information in an online community.* University of Illinois at Urbana-Champaign,

Camponovo, M. (2013). *Assessing Uncertainty in Volunteered Geographic Information for Emergency Response.* (Master of Science, Geography), The University of New Mexico, Albuquerque, New Mexico.

Carbone, D., & Hanson, J. (2012). Floods: 10 of the deadliest in Australian history. Retrieved from http://www.australiangeographic.com.au/topics/history-culture/2012/03/floods-10-of-the-deadliest-in-australian-history/

Carpenter, B. (2005). *Scaling high-order character language models to gigabytes.* Paper presented at the Proceedings of the Workshop on Software.

Carter, W. N. (1991). Disaster management: A disaster manager's handbook. In *Disaster management: A disaster manager's handbook*: ADB.

Castillo, C., Mendoza, M., & Poblete, B. (2011). *Information credibility on twitter.* Paper presented at the Proceedings of the 20th international conference on World wide web.

Chakraborty, D., Elzarka, H., & Bhatnagar, R. (2016). Generation of accurate weather files using a hybrid machine learning methodology for design and analysis of sustainable and resilient buildings. *Sustainable Cities and Society, 24*, 33-41. doi:10.1016/j.scs.2016.04.009

Chang, A. X., & Manning, C. D. (2012). SUTIME: A Library for Recognizing and Normalizing Time Expressions. *Lrec 2012 - Eighth International Conference on Language Resources and Evaluation*, 3735-3740.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16*, 321-357.

Coburn, A. W., Spence, R. J. S., & Pomonis, A. (1992). Factors Determining Human Casualty Levels in Earthquakes - Mortality Prediction in Building Collapse. *Proceedings of the Tenth World Conference on Earthquake Engineering, Vols 1-10*, 5989-5994.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement, 20*(1), 37-46. doi:Doi 10.1177/001316446002000104

Cohen, L. H. (1988). Life events and psychological functioning: Theoretical and methodological issues. In (Vol. 90, pp. 11–30): Sage Publications, Inc.

Coleman, D. J. (2010). Volunteered geographic information in spatial data infrastructure: An early look at opportunities and constraints. In R. Abbas, C. Joep, M. Kalantari, & B. Kok (Eds.), *Spatially Enabling Society: Research, Emerging Trends and Critical Assessment*: Leuven University Press, Leuven, Belgium, Leuven.

Coleman, D. J., Georgiadou, Y., & Labonte, J. (2009). Volunteered Geographic Information: the nature and motivation of produsers. *International Journal of Spatial Data Infrastructures Research, 4*(1), 332-358.

Cova, T. J. (1999). GIS in emergency management. In P.A. Longley, M. F. Goodchild, D.J. Maguire, & D. W. Rhind. (Eds.), *Geographical Information Systems: Principles, Techniques, Applications, and Management* (pp. 845-858). New York: John Wiley & Sons.

Craglia, M., Ostermann, F., & Spinsanti, L. (2012). Digital Earth from vision to practice: making sense of citizen-generated content. *International Journal of Digital Earth, 5*(5), 398-416. doi:Doi 10.1080/17538947.2012.712273

Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the research process*: Sage.

Crowston, K., & Prestopnik, N. R. (2013). *Motivation and data quality in a citizen science game: A design science evaluation.* Paper presented at the 2013 46th Hawaii International Conference on System Sciences.

Davis, J., & Goadrich, M. (2006). *The relationship between Precision-Recall and ROC curves.* Paper presented at the Proceedings of the 23rd international conference on Machine learning.

De Longueville, B., Annoni, A., Schade, S., Ostlaender, N., & Whitmore, C. (2010). Digital Earth's Nervous System for crisis events: real-time Sensor Web Enablement of Volunteered Geographic Information. *International Journal of Digital Earth, 3*(3), 242-259. doi:Doi 10.1080/17538947.2010.484869

Degrossi, L. C., Albuquerque, J. P. d., Fava, M. C., & Mendiondo, E. M. (2014). *Flood Citizen Observatory: a crowdsourcing-based approach for flood risk management in Brazil.* Paper presented at the 26th International Conference on Software Engineering and Knowledge Engineering, Vancouver, Canada.

Densham, P. J. (1991). Spatial decision support systems. *Geographical information systems: Principles and applications, 1*, 403-412.

Devillers, R., Bedard, Y., Jeansoulin, R., & Moulin, B. (2007). Towards spatial data quality information analysis tools for experts assessing the fitness for use of spatial data. *International Journal of Geographical Information Science, 21*(3), 261-282. doi:10.1080/13658810600911879

Devillers, R., & Jeansoulin, R. (2006). Spatial data quality: concepts. In R. Devillers & R. Jeansoulin (Eds.), *Fundamentals of Spatial Data Quality* (pp. 31-42). London: ISTE.

Dransch, D., Fohringer, J., Poser, K., & Lucas, C. (2013). Volunteered Geographic Information for Disaster Management. In C. N. Silva (Ed.), *Citizen E-Participation in Urban Governance: Crowdsourcing and Collaborative Creativity* (pp. 98-118).

Elwood, S. (2008). Volunteered geographic information: future research directions motivated by critical, participatory, and feminist GIS. *GeoJournal, 72*(3-4), 173-183.

Elwood, S., Goodchild, M. F., & Sui, D. Z. (2012). Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice. *Annals of the Association of American Geographers, 102*(3), 571-590. doi:Doi 10.1080/00045608.2011.595657

Eppler, M. J. (2006). *Managing information quality: Increasing the value of information in knowledge-intensive products and processes*: Springer Science & Business Media.

Fan, H. C., Zipf, A., Fu, Q., & Neis, P. (2014). Quality assessment for building footprints data on OpenStreetMap. *International Journal of Geographical Information Science, 28*(4), 700-719. doi:10.1080/13658816.2013.867495

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861-874. doi:10.1016/j.patrec.2005.10.010

Fisher, P. F. (1999). Models of uncertainty in spatial data. In P. A. Longley, M. H. Goodchild, & D. J. Maguire (Eds.), *Geographical information systems: Principles and Technical Issues* (2nd edition ed., Vol. 1, pp. 191-205). New York: John Wiley.

Flanagin, A. J., & Metzger, M. J. (2008). The credibility of volunteered geographic information. *GeoJournal, 72*(3-4), 137-148.

Fogg, B., & Tseng, H. (1999). *The elements of computer credibility.* Paper presented at the Proceedings of the SIGCHI conference on Human Factors in Computing Systems.

Ford, H. (2011). Design memo: Verifying information from the crowd. Retrieved from http://www.scribd.com/doc/72271441/Verification-Memo-Latest

Fortmann-Roe, S. (2012). Understanding the bias-variance tradeoff. Retrieved from http://scott.fortmann-roe.com/docs/BiasVariance.html

Friberg, T., Prödel, S., & Koch, R. (2011). *Information quality criteria and their importance for experts in crisis situations.* Paper presented at the Proceedings of the 8th International ISCRAM Conference.

Gevrey, M., Dimopoulos, I., & Lek, S. (2003). Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological modelling, 160*(3), 249-264.

Ghahremanlou, L., Sherchan, W., & Thom, J. A. (2014). Geotagging Twitter Messages in Crisis Management. *The Computer Journal*.

Girres, J. F., & Touya, G. (2010). Quality Assessment of the French OpenStreetMap Dataset. *Transactions in GIS, 14*(4), 435-459. doi:DOI 10.1111/j.1467-9671.2010.01203.x

Goodchild, M. F. (1993). Data models and data quality: problems and prospects. In M. F. Goodchild, Parks B.O., Steyaert L.T. (Ed.), *Environmental Modeling with GIS* (pp. 94-103). Oxford, New York.

Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal, 69*(4), 211-221.

Goodchild, M. F. (2009a). NeoGeography and the nature of geographic expertise. *Journal of Location Based Services, 3*(2), 82-96.

Goodchild, M. F. (2009b). The Quality of Geospatial Context. *Quality of Context, 5786*, 15-24.

Goodchild, M. F., & Glennon, J. A. (2010). Crowdsourcing geographic information for disaster response: a research frontier. *International Journal of Digital Earth, 3*(3), 231-241. doi:Pii 925996051

Goodchild, M. F., & Li, L. (2012). Assuring the quality of volunteered geographic information. *Spatial statistics, 1*, 110-120.

Gorman, S. (2010). Why VGI is the Wrong Acronym. In *Esri DC Development Center blog* (Vol. 2014). Esri DC Development Center blog: Esri.

Gosier, J. (Producer). (2011). Privacy, Security and Motivation of Crowds. [Presentation slides] Retrieved from http://www.slideshare.net/Ushahidi/privacy-security-and-motivation-of-crowds

Goutte, C., & Gaussier, E. (2005). *A probabilistic interpretation of precision, recall and F-score, with implication for evaluation.* Paper presented at the European Conference on Information Retrieval.

Graham, M. (2010). Cloud collaboration: Peer-production and the engineering of the internet. In *Engineering earth* (pp. 67-83): Springer.

Green, B. N., Johnson, C. D., & Adams, A. (2006). Writing narrative literature reviews for peer-reviewed journals: secrets of the trade. *Journal of chiropractic medicine, 5*(3), 101-117.

Gregor, S. (2006). The nature of theory in information systems. *Mis Quarterly*, 611-642.

Gregor, S., & Hevner, A. R. (2013). Positioning and Presenting Design Science Research for Maximum Impact. *Mis Quarterly, 37*(2), 337-+.

Gunther, A. C. (1992). Biased Press or Biased Public - Attitudes toward Media Coverage of Social-Groups. *Public Opinion Quarterly, 56*(2), 147-167. doi:Doi 10.1086/269308

Gupta, A., & Kumaraguru, P. (2012). *Credibility ranking of tweets during high impact events.* Paper presented at the Proceedings of the 1st Workshop on Privacy and Security in Online Social Media.

Guptill, S. C., & Morrison, J. L. (1995). *Elements of spatial data quality* (S. C. Guptill & J. L. Morrison Eds. 1st ed.). Oxford, U.K. ; Tarrytown, N.Y., U.S.A.: Elsevier Science.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research, 3*(Mar), 1157-1182.

Haklay, M. (2010). How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and planning. B, Planning & design, 37*(4), 682.

Haklay, M., Basiouka, S., Antoniou, V., & Ather, A. (2010). How Many Volunteers Does it Take to Map an Area Well? The Validity of Linus' Law to Volunteered Geographic Information. *Cartographic Journal, 47*(4), 315-322. doi:Doi 10.1179/000870410x12911304958827

Haklay, M., & Budhathoki, N. (2010). OpenStreetMap–Overview and Motivational Factors. *Horizon Infrastructure Challenge Theme Day. University of Nottingham, UK.*

Hall, G. B., Chipeniuk, R., Feick, R. D., Leahy, M. G., & Deparday, V. (2010). Community-based production of geographic information using open source software and Web 2.0. *International Journal of Geographical Information Science, 24*(5), 761-781. doi:10.1080/13658810903213288

Harvey, F. (2013). To Volunteer or to Contribute Locational Information? Towards Truth in Labeling for Crowdsourced Geographic Information. In D. Sui, S. Elwood, & M. Goodchild (Eds.), *Crowdsourcing Geographic Knowledge* (pp. 31-42): Springer Netherlands.

Haworth, B., & Bruce, E. (2015). A Review of Volunteered Geographic Information for Disaster Management. *Geography Compass, 9*(5), 237-250.

Hecht, R., Kunze, C., & Hahmann, S. (2013). Measuring completeness of building footprints in OpenStreetMap over space and time. *ISPRS International Journal of Geo-Information, 2*(4), 1066-1091.

Helewa, A., & Walker, J. M. (2000). *Critical evaluation of research in physical rehabilitation: towards evidence-based practice*: WB Saunders Company.

Hevner, A. R. (2007). A three cycle view of design science research. *Scandinavian journal of information systems, 19*(2), 4.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *Mis Quarterly, 28*(1), 75-105.

Horita, F. E. A., Degrossi, L. C., Assis, L. F. G. d., Zipf, A., & Albuquerque, J. P. d. (2013). *The use of Volunteered Geographic Information (VGI) and Crowdsourcing in Disaster Management: a Systematic Literature Review*. Paper presented at the 19th Americas Conference on Information Systems (AMCIS), Chicago, USA.

Huck, J., Whyatt, D., & Coulton, P. (2015). Visualising patterns in spatially ambiguous point data. *Journal of Spatial Information Systems*(10), 47-66.

Iacucci, A. A. (2010). Ushahidi-Chile: an example of crowd sourcing verification of information. In A. A. Iacucci (Ed.), *Diary of a Crisis Mapper* (Vol. 20 April 2014): Ushahidi-Chile: an example of crowd sourcing verification of information.

Iivari, J. (2007). A paradigmatic analysis of information systems as a design science. *Scandinavian journal of information systems, 19*(2), 5.

Imran, M., Elbassuoni, S. M., Castillo, C., Diaz, F., & Meier, P. (2013). *Extracting information nuggets from disaster-related messages in social media*. Paper presented at the Proceedings of the 10th International Conference on Information Systems for Crisis Response and Manageme, Baden-Baden, Germany.

ISO. (2013). ISO19157:2013 Geographic information - Data quality. In. Geneva: ISO.

Jurafsky, D., & Martin, J. H. (2000). N-gram Language Models. In *Speech and language processing*: Prentice Hall.

Kalantari, M., & La, V. (2015). Assessing OpenStreetMap as an Open Property Map. In *OpenStreetMap in GIScience* (pp. 255-272). Switzerland: Springer International Publishing.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons, 53*(1), 59-68. doi:DOI 10.1016/j.bushor.2009.09.003

Kawasaki, A., Berman, M. L., & Guan, W. (2013). The growing role of web-based geospatial technology in disaster response and support. *Disasters, 37*(2), 201-221. doi:DOI 10.1111/j.1467-7717.2012.01302.x

Keen, A. (2011). *The Cult of the Amateur: How Blogs, MySpace, YouTube and the Rest of Today's User Generated Media are Killing Our Culture and*: Nicholas Brealey Publishing.

King, G., & Zeng, L. (2001). Logistic regression in rare events data. *Political analysis, 9*(2), 137-163.

Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence, 97*(1-2), 273-324.

Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. In.

Koukoletsos, T., Haklay, M., & Ellul, C. (2012). Assessing Data Completeness of VGI through an Automated Matching Procedure for Linear Data. *Transactions in Gis, 16*(4), 477-498. doi:DOI 10.1111/j.1467-9671.2012.01304.x

Lüdemann, L., Grieger, W., Wurm, R., Wust, P., & Zimmer, C. (2006). Glioma assessment using quantitative blood volume maps generated by T1-weighted dynamic contrast-enhanced magnetic resonance imaging: a receiver operating characteristic study. *Acta Radiologica, 47*(3), 303-310.

Laylavi, F., Rajabifard, A., & Kalantari, M. (2016). A Multi-Element Approach to Location Inference of Twitter: A Case for Emergency Response. *ISPRS International Journal of Geo-Information, 5*(5), 56.

Laylavi, F., Rajabifard, A., & Kalantari, M. (2017). Event relatedness assessment of Twitter messages for emergency response. *Information Processing & Management, 53*(1), 266-280.

Lecessie, S., & Vanhouwelingen, J. C. (1992). Ridge Estimators in Logistic-Regression. *Applied Statistics-Journal of the Royal Statistical Society Series C, 41*(1), 191-201.

Leetaru, K., Wang, S., Cao, G., Padmanabhan, A., & Shook, E. (2013). Mapping the global Twitter heartbeat: The geography of Twitter. *First Monday, 18*(5).

Liu, S. B. (2014). Crisis Crowdsourcing Framework: Designing Strategic Configurations of Crowdsourcing for the Emergency Management Domain. *Computer Supported Cooperative Work (CSCW), 23*(4-6), 389-443.

Ludwig, T., Reuter, C., & Pipek, V. (2015). Social Haystack: Dynamic Quality Assessment of Citizen-Generated Content during Emergencies. *ACM Transactions on Computer-Human Interaction (TOCHI), 22*(4), 17.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval* (Vol. 1): Cambridge university press Cambridge.

March, S. T., & Smith, G. F. (1995). Design and natural science research on information technology. *Decision Support Systems, 15*(4), 251-266.

Maué, P., & Schade, S. (2008). *Quality of geographic information patchworks.* Paper presented at the Proceedings of AGILE.

McDougall, K. (2011). *Using volunteered information to map the Queensland floods.* Paper presented at the Proceedings of the 2011 Surveying and Spatial Sciences Conference: Innovation in Action: Working Smarter (SSSC 2011).

McDougall, K., & Temple-Watts, P. (2012). The use of LiDAR and volunteered geographic information to map flood extents and inundation. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 1*, 251-256.

Meier, P. (2011). Verifying crowdsourced social media reports for live crisis mapping: An introduction to information forensics. In *iRevolution* (Vol. 2013): iRevolution blog.

Meier, P. (2012). Crisis mapping in action: How open source software and global volunteer networks are changing the world, one map at a time. *Journal of Map & Geography Libraries, 8*(2), 89-100.

Miller, C. C. (2006). A beast in the field: The Google Maps mashup as GIS/2. *Cartographica: The International Journal for Geographic Information and Geovisualization, 41*(3), 187-199.

Mishra, M., Mishra, V. K., & Sharma, H. (2013). Question classification using semantic, syntactic and lexical features. *International Journal of Web & Semantic Technology, 4*(3), 39.

Morris, M. R., Counts, S., Roseway, A., Hoff, A., & Schwarz, J. (2012). *Tweeting is believing?: understanding microblog credibility perceptions.* Paper presented at the Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work.

Morrow, N., Mock, N., Papendieck, A., & Kocmich, N. (2011). Independent evaluation of the Ushahidi Haiti project. *Development Information Systems International, 8*.

Mummidi, L. N., & Krumm, J. (2008). Discovering points of interest from users' map annotations. *GeoJournal, 72*(3-4), 215-227.

Murdock, V. (2011). Your mileage may vary: on the limits of social media. *SIGSPATIAL Special, 3*(2), 62-66.

Murphy, R. R., Tadokoro, S., Nardi, D., Jacoff, A., Fiorini, P., Choset, H., & Erkmen, A. M. (2008). Search and rescue robotics. In *Springer Handbook of Robotics* (pp. 1151-1173): Springer.

Naaman, E. (2019). Understand Your Black Box Model Using Sensitivity Analysis - Practical Guide. Retrieved from https://medium.com/@einat_93627/understand-your-black-box-model-using-sensitivity-analysis-practical-guide-ef6ac4175e55

Oliveira, S., Pereira, J. M., San-Miguel-Ayanz, J., & Lourenço, L. (2014). Exploring the spatial patterns of fire density in southern Europe using geographically weighted regression. *Applied Geography, 51*, 143-157.

Olteanu, A., Castillo, C., Diaz, F., & Vieweg, S. (2014). *CrisisLex: A Lexicon for Collecting and Filtering Microblogged Communications in Crises.* Paper presented at the ICWSM.

Ostermann, F., & Spinsanti, L. (2012). Context Analysis of Volunteered Geographic Information from Social Media Networks to Support Disaster Management: A Case Study on Forest Fires. *International Journal of Information Systems for Crisis Response and Management, 4*(4), 16-37. doi:10.4018/jiscrm.2012100102

Paul, A. (2013). *Extracting important information from a social network stream during crisis*. Paper presented at the Internet Research 14 conference, Denver, USA.

Pereira, J. M. C., & Duckstein, L. (1993). A Multiple Criteria Decision-Making Approach to Gis-Based Land Suitability Evaluation. *International Journal of Geographical Information Systems, 7*(5), 407-424. doi:Doi 10.1080/02693799308901971

Pierkot, C., Zimányi, E., Lin, Y., & Libourel, T. (2011). *Advocacy for external quality in GIS.* Paper presented at the International Conference on GeoSpatial Sematics.

Poser, K., & Dransch, D. (2010). Volunteered geographic information for disaster management with application to rapid flood damage estimation. *Geomatica, 64*(1), 89-98.

Posetti, J. (2012). The twitterisation of ABC's emergency and disaster communication. *Australian Journal of Emergency Management, 27*(1), 34-39.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning, 1*(1), 81-106.

Rahmatizadeh, S., Rajabifard, A., & Kalantari, M. (2016). A conceptual framework for utilising VGI in land administration. *Land Use Policy, 56*, 81-89.

Rajabifard, A., Crompvoets, J., Kalantari, M., & Kok, B. (2010). Spatially enabled societies. In *Spatially enabled society* (pp. 16): Leuven University Press.

Raymond, E. S. (2001). *The Cathedral and the Bazaar*: O' Reilly.

Redman, T. C., & Blanton, A. (1997). *Data quality for the information age*: Artech House, Inc.

Reynolds, B. (2006). Response to best practices. *Journal of Applied Communication Research, 34*(3), 249-252. doi:Doi 10.1080/00909880600771593

Scassa, T. (2013). Legal issues with volunteered geographic information. *Canadian Geographer-Geographe Canadien, 57*(1), 1-10. doi:DOI 10.1111/j.1541-0064.2012.00444.x

Schnebele, E., & Cervone, G. (2013). Improving remote sensing flood assessment using volunteered geographical data. *Natural Hazards and Earth System Sciences, 13*(3), 669-677. doi:10.5194/nhess-13-669-2013

Schnebele, E., Cervone, G., & Waters, N. (2014). Road assessment after flood events using non-authoritative data. *Nat. Hazards Earth Syst. Sci., 14*(4), 1007-1015. doi:10.5194/nhess-14-1007-2014

Senaratne, H., Mobasheri, A., Ali, A. L., Capineri, C., & Haklay, M. (2016). A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science*, 1-29.

Servigne, S., Lesage, N., & Libourel, T. (2006). Quality components, standards, and metadata. *Fundamentals of Spatial Data Quality*, 179-210.

Shi, W., Fisher, P., & Goodchild, M. F. (2002). *Spatial data quality*. London ; New York: Taylor & Francis.

Simon, H. A. (1996). *The sciences of the artificial*: MIT press.

Spinsanti, L., & Ostermann, F. (2013). Automated geographic context analysis for volunteered information. *Applied Geography, 43*, 36-44. doi:http://dx.doi.org/10.1016/j.apgeog.2013.05.005

Starbird, K. (2011). *Digital volunteerism during disaster: Crowdsourcing information processing.* Paper presented at the Paper presented at the CHI '11 Workshop on Crowdsourcing and Human Computation at the 2011 Conference on Human Factors in Computing Systems (CHI 2011), Vancouver.

Stefanidis, A., Crooks, A., & Radzikowski, J. (2013). Harvesting ambient geospatial information from social media feeds. *GeoJournal, 78*(2), 319-338.

Tadokoro, S. (2009). *Rescue Robotics: DDT Project on Robots and Systems for Urban Search and Rescue*: Springer Science & Business Media.

The State of Queensland Department of Natural Resources and Mines. (2013). *Ex-TC Oswald Floods, January and February 2013*.

Triglav-Čekada, M., & Radovan, D. (2013). Using volunteered geographical information to map the November 2012 floods in Slovenia. *Natural Hazards and Earth System Sciences Discussion, 1*, 2859-2881. doi:10.5194/nhessd-1-2859-2013

Truelove, M., Vasardani, M., & Winter, S. (2016). *Introducing a Framework for Automatically Differentiating Witness Accounts of Events from Social Media*. Paper presented at the Locate Conference 16, Melbourne.

Turner, A. (2006). *Introduction to neogeography*: O'Reilly Media.

Vaishnavi, V. K., & Kuechler, W. (2015). Introduction to Design Science Research in Information and Communication Technology. In *Design science research methods and patterns: innovating information and communication technology* (2nd ed., pp. 24): Crc Press.

van Oort, P. (2006). *Spatial data quality: from description to application.* (PhD thesis), Wageningen University,

Vasilescu, L., Khan, A., & Khan, H. (2008). Disaster management cycle–a theoretical approach. *Management & Marketing-Craiova*(1), 43-50.

Veregin, H. (1999). Data quality parameters. *Geographical information systems, 1*, 177-189.

Verma, S., Vieweg, S., Corvey, W. J., Palen, L., Martin, J. H., Palmer, M., . . . Anderson, K. M. (2011). *Natural Language Processing to the Rescue? Extracting" Situational Awareness" Tweets During Mass Emergency.* Paper presented at the ICWSM.

Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. *Fam Med, 37*(5), 360-363.

Wald, D. J., Quitoriano, V., Worden, C. B., Hopper, M., & Dewey, J. W. (2012). USGS "Did you feel it?" internet-based macroseismic intensity maps. *Annals of Geophysics, 54*(6).

Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems, 12*(4), 5-33.

Westrope, C., Banick, R., & Levine, M. (2014). Groundtruthing OpenStreetMap building damage assessment. *Procedia Engineering, 78*, 29-39.

Wickham, M. (2018). Integrating models. In *Practical Java Machine Learning: Projects with Google Cloud Platform and Amazon Web Services*: Apress.

Yang, J., Counts, S., Morris, M. R., & Hoff, A. (2013). *Microblog credibility perceptions: comparing the USA and China.* Paper presented at the Proceedings of the 2013 conference on Computer supported cooperative work.

Ye, Q., Zhang, Z., & Law, R. (2009). Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. *Expert systems with applications, 36*(3), 6527-6535.

Yin, J., Lampert, A., Cameron, M., Robinson, B., & Power, R. (2012). Using social media to enhance emergency situation awareness. *IEEE Intelligent Systems*(6), 52-59.

Yu, L., & Liu, H. (2004). Efficient feature selection via analysis of relevance and redundancy. *Journal of machine learning research, 5*(Oct), 1205-1224.

Zainal, Z. (2007). Case study as a research method. *Jurnal Kemanusiaan, 9*.

Zielstra, D., & Zipf, A. (2010). *A comparative study of proprietary geodata and volunteered geographic information for Germany.* Paper presented at the 13th AGILE international conference on geographic information science.

Zook, M., Graham, M., Shelton, T., & Gorman, S. (2010). Volunteered geographic information and crowdsourcing disaster relief: a case study of the Haitian earthquake. *World Medical & Health Policy, 2*(2), 7-33.

Zweig, M. H., & Campbell, G. (1993). Receiver-Operating Characteristic (Roc) Plots - a Fundamental Evaluation Tool in Clinical Medicine. *Clinical Chemistry, 39*(4), 561-577.