Author/s:
Hur, B;Hardefeldt, LY;Verspoor, K;Baldwin, T;Gilkerson, JR

Title:
Overcoming challenges in extracting prescribing habits from veterinary clinics using big data and deep learning

Date:
2022-05-01

Citation:
Hur, B., Hardefeldt, L. Y., Verspoor, K., Baldwin, T. & Gilkerson, J. R. (2022). Overcoming challenges in extracting prescribing habits from veterinary clinics using big data and deep learning. Australian Veterinary Journal, 100 (5), pp.220-222. https://doi.org/10.1111/avj.13145.

Persistent Link:
https://hdl.handle.net/11343/336354

Hardefeldt Laura Yvonne (Orcid ID: 0000-0001-5780-7567)
Gilkerson James (Orcid ID: 0000-0003-1801-4817)

Overcoming challenges in extracting prescribing habits from veterinary clinics using big data and deep learning.

B Hur[a,b*], LY Hardefeldt[a], K Verspoor[b,c], T Baldwin[b], JR Gilkerson[a]

[a]Asia-Pacific Centre for Animal Health, Melbourne Veterinary School, University of Melbourne

[b]School of Computing and Information Systems, University of Melbourne

[c]Health and Biomedical Informatics Centre, University of Melbourne

[*]Corresponding author

Abstract

Understanding antimicrobial usage patterns and encouraging appropriate antimicrobial usage is a critical component of antimicrobial stewardship. Studies using VetCompass Australia and Natural Language Processing (NLP) have demonstrated antimicrobial usage patterns in companion animal practices across Australia. Doing so has highlighted the many obstacles and barriers to the task of converting raw clinical notes into a format that can be readily queried and analysed.

We developed NLP systems using rules-based algorithms and machine learning to automate the extraction of data describing the key elements to assess appropriate antimicrobial use. These included the clinical indication, antimicrobial agent selection, dose, and duration of therapy.

Our methods were applied to over 4.4 million companion animal clinical records across Australia on all consultations with antimicrobial use to help us understand what antibiotics are being given and why on a population level. Of these approximately only 40% recorded the reason why antimicrobials were prescribed, along with the dose and duration of treatment.

NLP and deep learning might be able to overcome the difficulties of harvesting free text data from clinical records, but when the essential data is not recorded in the clinical records then this becomes an insurmountable obstacle.

Introduction

Antimicrobial resistance (AMR) is a growing threat that results in morbidity and mortality through decreasing the effectiveness of antimicrobial agents in human and veterinary patients.[1–3] Antimicrobial stewardship programs have been shown to be useful in combating resistance but require data to inform them for both identifying targets for stewardship interventions and for assessing effectiveness and implementability.[4–7]

VetCompass Australia is a large database that collects data from over 180 clinics across Australia.[8] However, much of its data is in free text form, making it difficult to readily query. Many challenges are due to the record keeping of veterinary practitioners and the current software systems used to record and report such information. Use of free text, abbreviations and clinical shorthand are all designed to make work easier for the practitioner, but all of these practices mean that extracting meaningful data from the dataset is problematic. In addition, clinical records with missing data (patient weight, antimicrobial dosage and duration of treatment) make utilising big data sets even more difficult.

Natural Language Processing is an area of computer science that can help structure this free text.[9,10] Recent breakthroughs in neural networks and deep learning have resulted in performance that meets or exceeds human level performance on many tasks related to extracting data out of texts.[11,12]

Our work has explored the use of NLP to extract data out of VetCompass and use this data to inform antimicrobial stewardship efforts in Australia.[13–15] In this work, over 4.4 million

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/avj.13145

records were extracted between 2012-2017 inclusive to describe the usage patterns in Australia.

Many of the techniques leveraged deep learning to structure the data for analysis. While the performance of the models is good overall, there were several obstacles that were impossible to overcome. This paper explores those obstacles and potential paths forward for them to be overcome. Specifically, extracting the specific elements of AMU events:

1. Ingredient of the antimicrobial agent
2. Indication for the administration
3. Dose and duration of therapy

Extracting antimicrobial ingredients

The first step in this research was to develop methods to extract the ingredient (antimicrobial agent) from the clinical records stored within VetCompass Australia.[13] To do this, we manually created a comprehensive dictionary that contained the antimicrobial drugs available to veterinarians in Australia and the various agents that these products contain. A rules-based method was then developed which matched text segments within the prescriptions to the dictionary. We overcame the challenge of misspelled text segments by using an algorithm that would evaluate the length of the text divided by the difference in the string measured by Levenshtein (edit) distance.[16] This method was validated on the dataset being used for the study, however the products matched to dictionaries is an ever-evolving list. Therefore, any future research using these methods need to have the accuracy evaluated and likely have the dictionary updated with additional items as they become available on the market.

Determining the correct indications

To develop a system for extracting the correct indication for an antimicrobial agent, we needed to develop a novel neural network model to examine the clinical notes and extract data describing the antimicrobial agent prescribed and the clinical indication for administration.[14] Training machine learning models to understand the reason for an antimicrobial administration relies on specific information being incorporated in the clinical records in a consistent manner. To train these models requires annotations to be made to clinical records so that the model can recognise the information recorded as free text. Such labels were unavailable and having an expert annotate the records by hand to train such models is cost prohibitive and extremely time consuming.

To overcome this, we leveraged the amount of data available within VetCompass Australia. This data was used to first pretrain the model to learn the syntax and vocabulary used within the clinical notes before focusing the model on the specific task of identifying the indication for the antimicrobial agent being used. The model was then optimised to maximise the annotation effort of the clinicians annotating the data through an instance selecting pre-processing step to select the most appropriate records that should be annotated. However, even with the optimisation of the records selected, the indication for antimicrobial administration was found to be present in fewer than 85% of records. This information is critical to any antimicrobial being dispensed and should be clearly indicated in the record.

Calculating the dose and duration of antimicrobials given

To extract the dose and duration out of the clinical records we used rules-based algorithms to extract the amount of each medication being given, the frequency of administration, and weight of the patient.[17] However, when evaluating the records themselves, it was found that the weight was missing from over 40% of clinical records evaluated. Furthermore, when

assessing the overall ability to extract the necessary information, including the indication, it was found that fewer than 40% of the clinical records had all the necessary information to perform the task. This severely limits what can be reasonably extracted out of a clinical document.

Evaluation

This research has enabled the automation of extracting of ingredients, indication for use, dose, and duration of antimicrobials being dispensed which enables the understanding of companion animal AMU patterns at scale. This enhances the understanding of what is currently being done in companion animal practice at the consultation level and critical component to implementing AMS programs for which it has been used to inform.

While the NLP methods were developed to extract AMU patterns, they are also capable of extracting other information about other happenings within the clinical record. The deep learning models used in the study can be repurposed with additional training to classify any information that is relevant and present within the text. While it would require additional labels to train the models for such tasks, we have also created methods to minimise the labelling required to train the models for a specific task.[14]

The primary limits of what could reasonably be extracted out of the clinical records was due to other factors. This was either the lack of information being available in the clinical records or the lack of resources, such as a database of drug names that is maintained.

Improving Clinical Records

Mapping drug names to the ingredient currently requires a comprehensive list of existing medications with their trade names mapped to the ingredients. There are standardised terminologies that contain medications, such as RxNorm and many veterinary medications included within this database.[18] However, there is no requirement to have veterinary medications registered within RxNorm. Having the requirement for veterinary medications to be entered into standardised vocabularies such as RxNorm would greatly improve the ability to extract the medications out of clinical records.

Another area that could be improved is through the practice management software used by clinicians. For example, when a prescription label is created, requiring an indication at the time of creating a prescription label would ensure the data is there. Additionally, the weight of an animal could be copied to the current clinical record automatically if the patient was not weighed at the current consultation, or there could be a forced requirement for an estimated weight. Finally, the dosage estimate could also automate based on the amount of medication being prescribed.

Conclusions

We developed various systems using NLP techniques, neural networks, and VetCompass Australia to examine the clinical notes, and extract data describing the antimicrobial agent prescribed, the dose, duration, and the clinical indication for administration. The effectiveness is limited based on the data recorded within the clinical record.

Acknowledgements

We would like to thank VetCompass Australia, which is supported by the Australian Research Council (LE160100026). We would also like to thank all participating clinics, including Greencross, for donation of data to VetCompass Australia.

Conflicts of interest

None to declare.

References

1. Roca I, Akova M, Baquero F et al. The global threat of antimicrobial resistance: science for intervention. *New Microbes and New Infections* 2015;6:22–29.

2. Duff A, Keane S, Hardefeldt LY. Descriptive study of antimicrobial susceptibilty patterns from equine septic synovial structures. *Proceedings of the 39th Bain Fallon Memorial Lectures*. Equine Veterinarians Australia,.

3. Johnston G, Lumsden J. Antimicrobial susceptiblity of bacterial isolates from 27 thoroughbreds with arytenoid chondropathy. *Proceedings of the 39th Bain Fallon Memorial Lectures*. Equine Veterinarians Australia,.

4. Arda B, Sipahi OR, Yamazhan T et al. Short-term effect of antibiotic control policy on the usage patterns and cost of antimicrobials, mortality, nosocomial infection rates and antibacterial resistance. *Journal of Infection* 2007;55:41–48.

5. Cisneros JM, Neth O, Gil-Navarro MV et al. Global impact of an educational antimicrobial stewardship programme on prescribing practice in a tertiary hospital centre. *Clinical Microbiology and Infection* 2014;20:82–88.

6. Pulcini C, Botelho-Nevers E, Dyar OJ et al. The impact of infectious disease specialists on antibiotic prescribing in hospitals. *Clinical Microbiology and Infection* 2014;20:963–972.

7. Hardefeldt LY. Implementing antimicrobial stewardship programmes in veterinary practices. *Veterinary Record* 2018;182:688–690.

8. McGreevy P, Thomson P, Dhand NK et al. VetCompass Australia: A National Big Data Collection System for Veterinary Science. 2017;15.

9. Kao A, Poteet SR. *Natural Language Processing and Text Mining*. Springer Science & Business Media, 2007.

10. Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. *J Am Med Inform Assoc* 2011;18:544–551.

11. Devlin J, Chang M-W, Lee K et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv:181004805 [cs]* 2018; *http://arxiv.org/abs/1810.04805*.

12. Howard J, Ruder S. Universal Language Model Fine-tuning for Text Classification. 2018; *https://arxiv.org/abs/1801.06146v5*.

13. Hur B, Hardefeldt LY, Verspoor K et al. Using natural language processing and VetCompass to understand antimicrobial usage patterns in Australia. *Australian Veterinary Journal* 2019;97:298–300.

14. Hur B, Baldwin T, Verspoor K et al. Domain Adaptation and Instance Selection for Disease Syndrome Classification over Veterinary Clinical Notes. *Proceedings of the 19th*

SIGBioMed Workshop on Biomedical Language Processing. Association for Computational Linguistics, Online, 2020:156–166. *https://www.aclweb.org/anthology/2020.bionlp-1.17*.

15. Hur BA, Hardefeldt LY, Verspoor KM et al. Describing the antimicrobial usage patterns of companion animal veterinary practices; free text analysis of more than 4.4 million consultation records. *PLOS ONE* Public Library of Science, 2020;15:1–15.

16. Levenshtein VI. Binary Codes Capable of Correcting Deletions, Insertions and Reversals. *Soviet Physics Doklady* 1966;10:707.

17. Karystianis G, Sheppard T, Dixon WG et al. Modelling and extraction of variability in free-text medication prescriptions from an anonymised primary care electronic medical record research database. *BMC Medical Informatics and Decision Making* 2016;16:18.

18. Liu S, Ma W, Moore R et al. RxNorm: prescription for electronic drug information exchange. *IT Professional* 2005;7:17–23.