Contribution to discussion of ISR paper

F. Godtliebsen, S.O. Skrøvseth and S. Wei

We would like to congratulate the authors with an excellent and comprehensive review of statistical scale-space methods. Here we discuss the potential contribution of scale-space perspectives for reinforcement learning, a type of machine learning algorithm, with a focus on some important applications in medicine.

Within medicine, information about patients is often gathered over time, and the optimal state or the relevant choice of scale are not a priori clear. A relevant example is controlling the blood glucose levels of patients with type 1 diabetes, which is challenging for the patients, particularly due to the need for taking multiple lines of evidence into consideration before prescribing an appropriate action in terms of insulin dose. While scale-space methods have traditionally proved to be useful as visualization devices, we conjecture here that they may also be used as a feature generating process, e.g., in reinforcement learning algorithms. Such an approach, whereby scale-space methods are used in combination with reinforcement learning, could provide, for multiple chronic diseases where the input space is large and complex, improved care both for the patient self-management and health care delivery.

Specifically, reinforcement learning is a type of machine learning concerned with how a learner chooses actions to control a system. It is well suited to situations where decisions are made sequentially along a timeline, actions depend on the observed state, effects manifest at later points in time than the actions that induced them, and there is some notion of preferred state(s). The principle of reinforcement learning is based on the interaction between a decision-making agent and its environment, where the goal is to train an agent to take actions that result in desirable states [1]. The traditional reinforcement learning framework consists of the following elements. At each decision time step t , we have

- State $s_t \in \mathcal{S}$, the set of possible states of the environment.
- Action $a_t \in \mathcal{A}$, the set of actions available.
- Reward r_{t+1} , the reward received one time step later as a consequence of the action.

A typical sequence of observations would be $s_t, a_t, r_{t+1}, s_{t+1}, a_{t+1}, \dots$. Note that in this setup the reward is collected when the next state is revealed. Thus, at each decision time point, the agent receives information about the state of the environment and chooses an action. The environment reacts to this action and transitions to a new state, which determines whether the agent receives a positive or negative reinforcement.

The goal in reinforcement learning is for the learner to determine a **policy** π , a mapping $\mathcal{S} \rightarrow \mathcal{A}$, that maximizes some notion of cumulative rewards, for which a commonly used concept is that of a discounted **return**:

$$R_{t-1} = \sum_{i=0}^{T-t} \gamma^i r_{t+i} \quad (1)$$

with a terminal state T (possibly $T \rightarrow \infty$) and discount rate $\gamma \in [0, 1]$. The discount rate indicates how much future rewards matter.

In the determination of the optimal policy, a useful quantity is the action-value function for a policy π defined as

$$Q_t^\pi(s, a) = \mathbb{E}_\pi \{ R_{t+1} + \gamma V_{t+1}^\pi(s) \mid s_t = s, a_t = a \} \quad (2)$$

This is the expected return of taking the action a from state s , and thereafter following policy π . The *optimal* action-value function is defined to be

$$Q_t^*(s, a) = \max_{\pi} Q_t^\pi(s, a). \quad (3)$$

The optimal Q function can be used to determine the optimal policy in an obvious way. At decision time t , given a state s , the optimal policy takes the action $\max_{a \in \mathcal{A}} Q_t^*(s, a)$. Q-learning, one of the most important breakthroughs in reinforcement learning algorithms, offers a way to estimate Q^* in (3). One of the main features of Q-learning is that it directly approximates Q^* while following any sufficiently explorative policy.

As with any machine learning task, data representation is key to the success of an algorithm. Also known as feature learning, this is the task for finding a representation of the input that will result in good performance of the machine learning algorithm [2]. Preliminary results indicate that scale-space methods can be an important tool in feature learning of reinforcement learning. It is also clear that the discount rate γ in (1) plays an important role in reinforcement learning algorithms. Since this parameter controls how future rewards are taken into account in the learning process, it has some resemblance to a scale-space parameter. We encourage the authors to discuss if and potentially how γ can be utilized through a scale-space point of view. Finally, we mention that reinforcement learning also can be put into a Bayesian framework, see e.g. Chapter 11 of [3] for details. It would be great if the authors could shed light on how scale-space methods can be utilized in the Bayesian reinforcement learning approach.

References

- [1] Richard S. Sutton and Andrew G. Barto. *Introduction to Reinforcement Learning*. MIT Press, Cambridge, MA, USA, 1st edition, 1998.
- [2] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(8):1798–1828, 2013.
- [3] Marco Wiering and Martijn van Otterlo (Eds.) *Reinforcement learning : State-of-the-art*. Springer-Verlag Berlin Heidelberg, 2012.

Author Manuscript

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Godtliebsen, F;Skrovseth, SO;Wei, S

Title:

Statistical Scale Space Methods

Date:

2017-04-01

Citation:

Godtliebsen, F., Skrovseth, S. O. & Wei, S. (2017). Statistical Scale Space Methods. INTERNATIONAL STATISTICAL REVIEW, 85 (1), pp.36-37. <https://doi.org/10.1111/insr.12193>.

Persistent Link:

<http://hdl.handle.net/11343/291529>