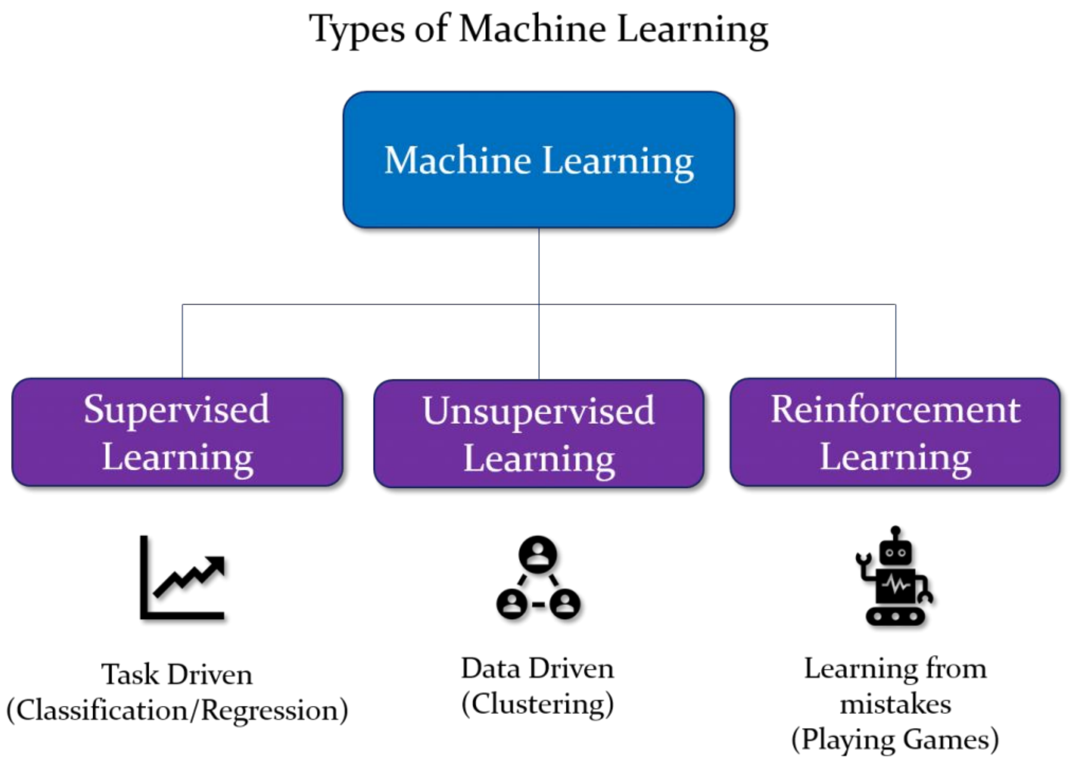
Reinforcement Learning

In Python

MasoudKaviani.ir

Interdisciplinary Schools

Types of Machine Learning



Types of Machine Learning

Machine Learning

Supervised Learning

Unsupervised Learning

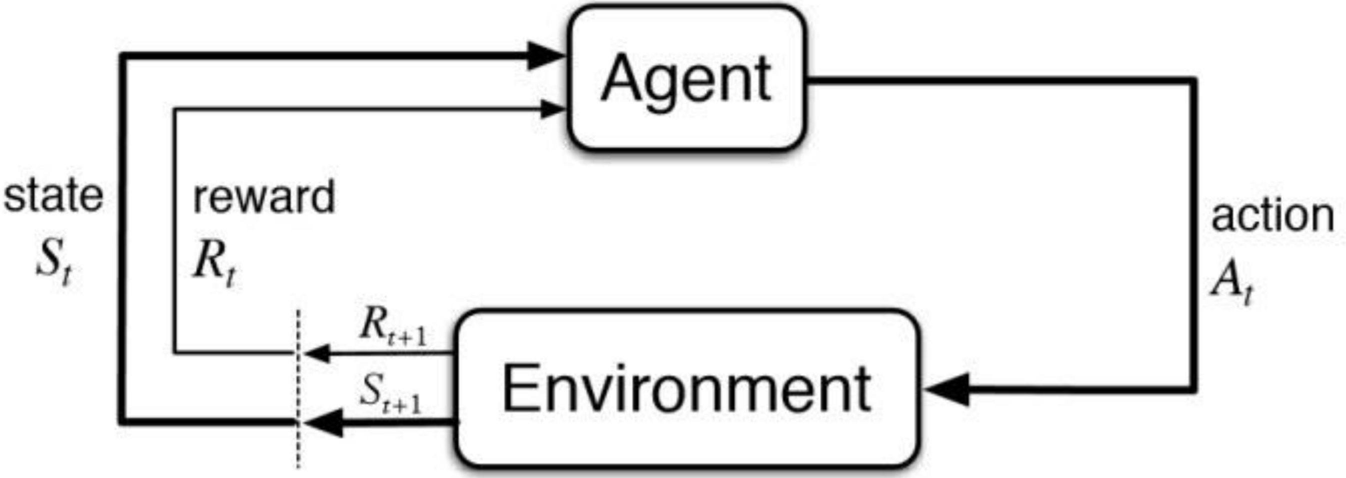
Reinforcement Learning

Task Driven
(Classification/Regression)

Data Driven
(Clustering)

Learning from mistakes
(Playing Games)

Reinforcement Learning



Key Terms

1. Environment
2. State
3. Reward
4. Policy
5. Value (Future Reward)

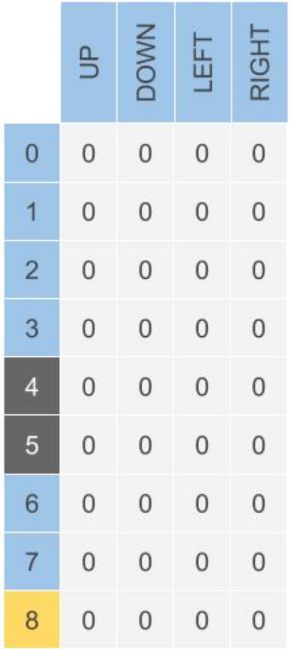
RL Algorithms

---

1. Monte Carlo Tree Search (MCTS)
2. Temporal Difference (TD)
3. Asynchronous Actor Critic Agent (A3C)
4. SARSA
5. **Q-Learning**
6. Deep Q-Learning

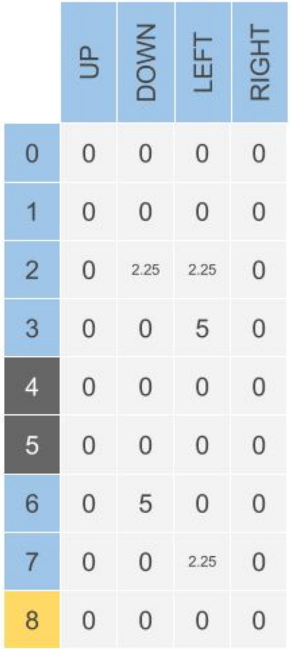
Q-Learning

Q-table initialised at zero



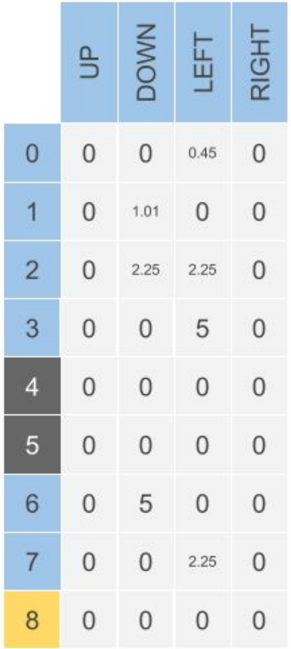
	UP	DOWN	LEFT	RIGHT
0	0	0	0	0
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0
5	0	0	0	0
6	0	0	0	0
7	0	0	0	0
8	0	0	0	0

After few episodes



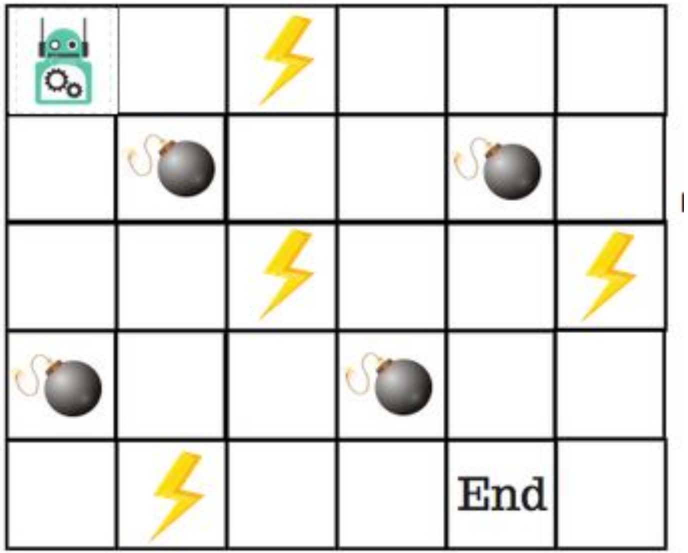
	UP	DOWN	LEFT	RIGHT
0	0	0	0	0
1	0	0	0	0
2	0	2.25	2.25	0
3	0	0	5	0
4	0	0	0	0
5	0	0	0	0
6	0	5	0	0
7	0	0	2.25	0
8	0	0	0	0

Eventually

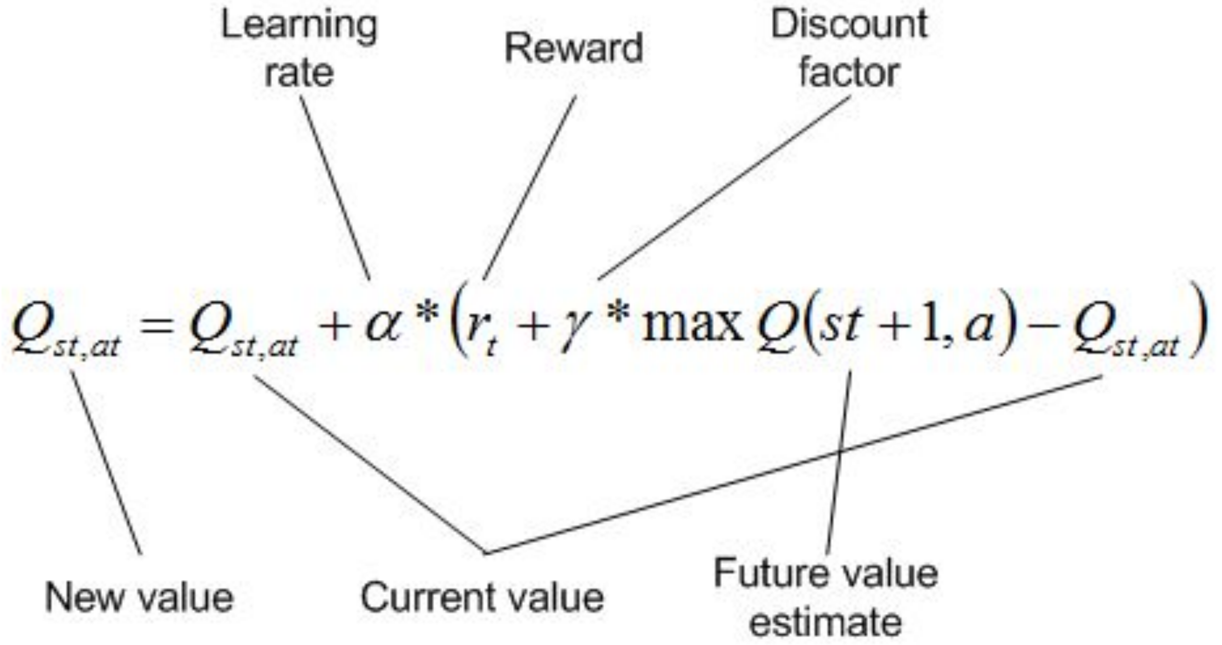


	UP	DOWN	LEFT	RIGHT
0	0	0	0.45	0
1	0	1.01	0	0
2	0	2.25	2.25	0
3	0	0	5	0
4	0	0	0	0
5	0	0	0	0
6	0	5	0	0
7	0	0	2.25	0
8	0	0	0	0

Q-Learning



Q-Learning



$$Q_{st,at} = Q_{st,at} + \alpha * \left(r_t + \gamma * \max Q(st+1, a) - Q_{st,at}\right)$$

Learning rate

Reward

Discount factor

New value

Current value

Future value estimate

Applications of RL - Self Driver Car

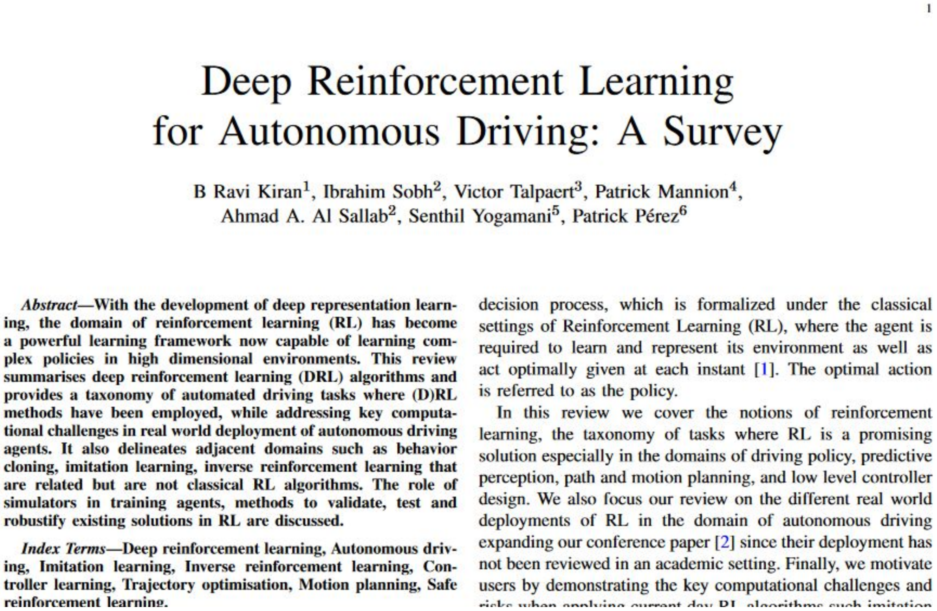
Applications of RL - Self Driver Car

Startups:

- **wayve.ai**
- **zoox.com**
- **getcruise.com**
- **pony.ai**

Applications of RL - Self Driver Car

arxiv.org/abs/2002.00444



Deep Reinforcement Learning for Autonomous Driving: A Survey

B Ravi Kiran[1], Ibrahim Sobh[2], Victor Talpaert[3], Patrick Mannion[4],
Ahmad A. Al Sallab[2], Senthil Yogamani[5], Patrick Pérez[6]

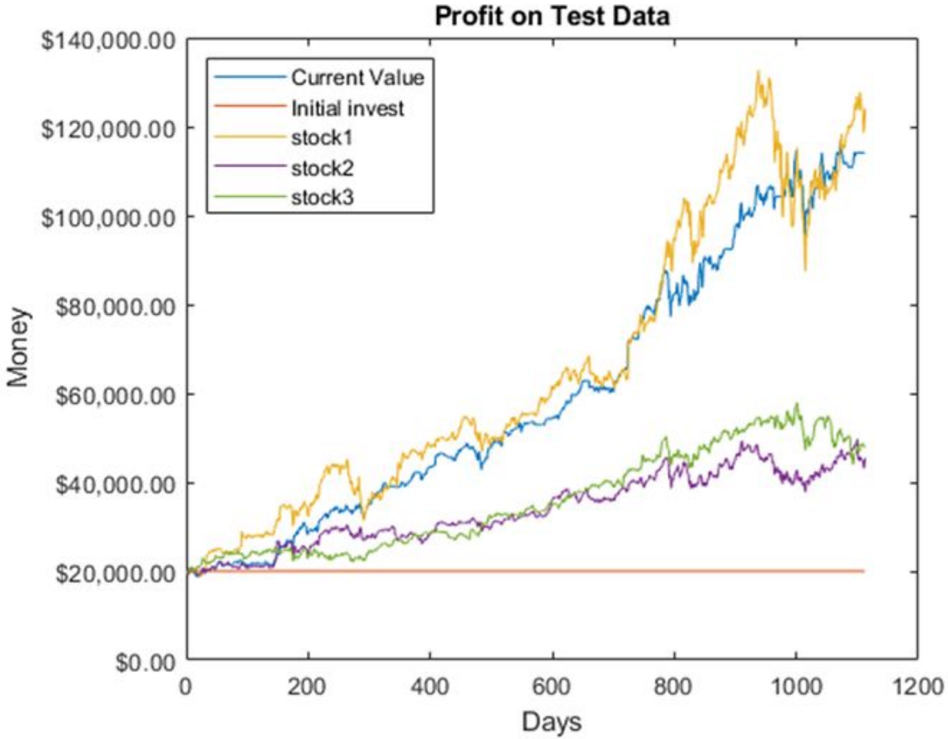
***Abstract*—With the development of deep representation learning, the domain of reinforcement learning (RL) has become a powerful learning framework now capable of learning complex policies in high dimensional environments. This review summarises deep reinforcement learning (DRL) algorithms and provides a taxonomy of automated driving tasks where (D)RL methods have been employed, while addressing key computational challenges in real world deployment of autonomous driving agents. It also delineates adjacent domains such as behavior cloning, imitation learning, inverse reinforcement learning that are related but are not classical RL algorithms. The role of simulators in training agents, methods to validate, test and robustify existing solutions in RL are discussed.**

***Index Terms*—Deep reinforcement learning, Autonomous driving, Imitation learning, Inverse reinforcement learning, Controller learning, Trajectory optimisation, Motion planning, Safe reinforcement learning.**

decision process, which is formalized under the classical settings of Reinforcement Learning (RL), where the agent is required to learn and represent its environment as well as act optimally given at each instant [1]. The optimal action is referred to as the policy.

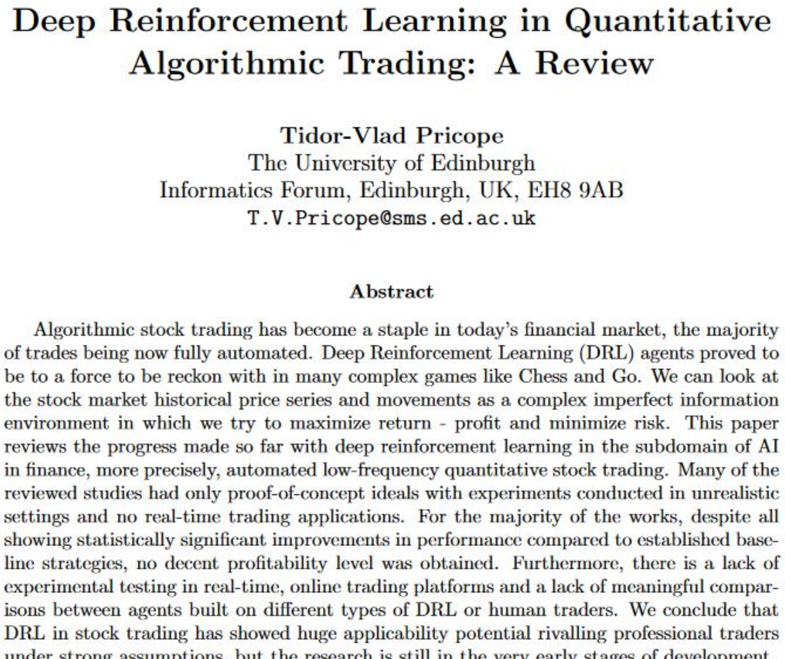
In this review we cover the notions of reinforcement learning, the taxonomy of tasks where RL is a promising solution especially in the domains of driving policy, predictive perception, path and motion planning, and low level controller design. We also focus our review on the different real world deployments of RL in the domain of autonomous driving expanding our conference paper [2] since their deployment has not been reviewed in an academic setting. Finally, we motivate users by demonstrating the key computational challenges and risks when applying current day RL algorithms such imitation

Applications of RL - Trading and Finance



Applications of RL - Trading and Finance

arxiv.org/abs/2106.00123



Deep Reinforcement Learning in Quantitative Algorithmic Trading: A Review

Tidor-Vlad Pricope
The University of Edinburgh
Informatics Forum, Edinburgh, UK, EH8 9AB
T.V.Pricope@sms.ed.ac.uk

Abstract

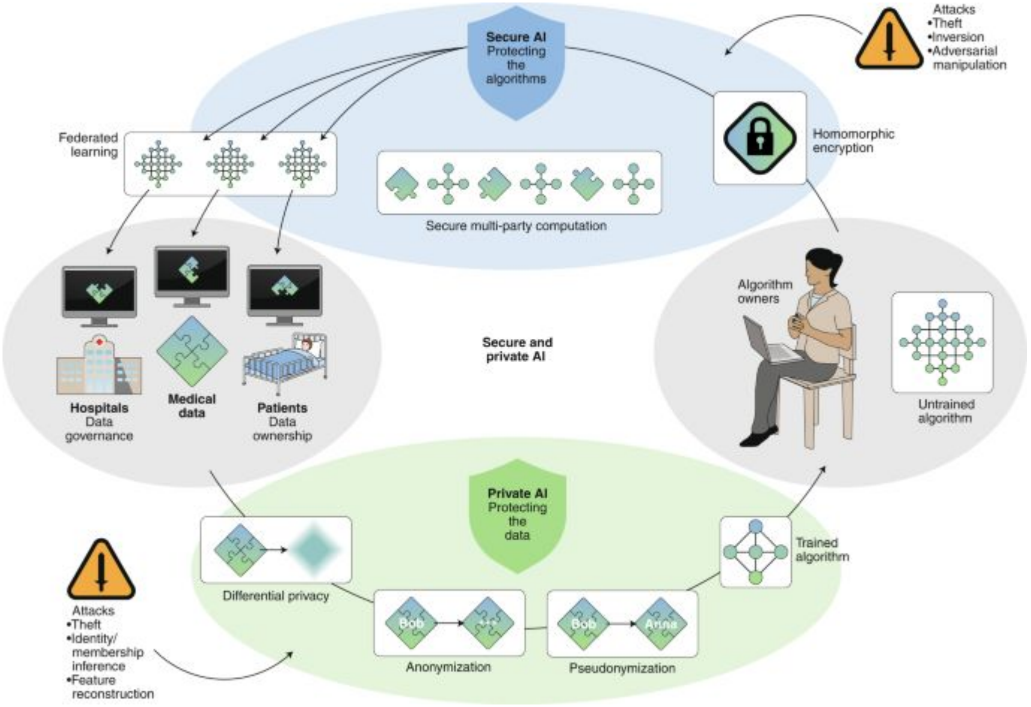
Algorithmic stock trading has become a staple in today's financial market, the majority of trades being now fully automated. Deep Reinforcement Learning (DRL) agents proved to be to a force to be reckon with in many complex games like Chess and Go. We can look at the stock market historical price series and movements as a complex imperfect information environment in which we try to maximize return - profit and minimize risk. This paper reviews the progress made so far with deep reinforcement learning in the subdomain of AI in finance, more precisely, automated low-frequency quantitative stock trading. Many of the reviewed studies had only proof-of-concept ideals with experiments conducted in unrealistic settings and no real-time trading applications. For the majority of the works, despite all showing statistically significant improvements in performance compared to established baseline strategies, no decent profitability level was obtained. Furthermore, there is a lack of experimental testing in real-time, online trading platforms and a lack of meaningful comparisons between agents built on different types of DRL or human traders. We conclude that DRL in stock trading has showed huge applicability potential rivalling professional traders under strong assumptions, but the research is still in the very early stages of development.

Applications of RL - Trading and Finance

Course:

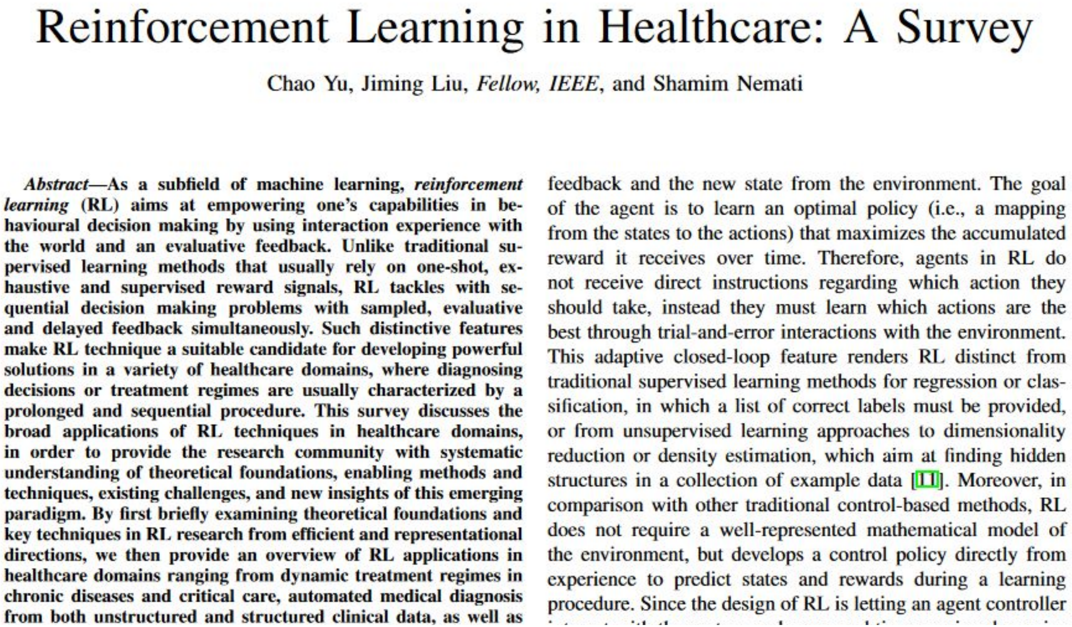
coursera.org/learn/trading-strategies-reinforcement-learning

Applications of RL - Healthcare



Applications of RL - Healthcare

arxiv.org/abs/1908.08796



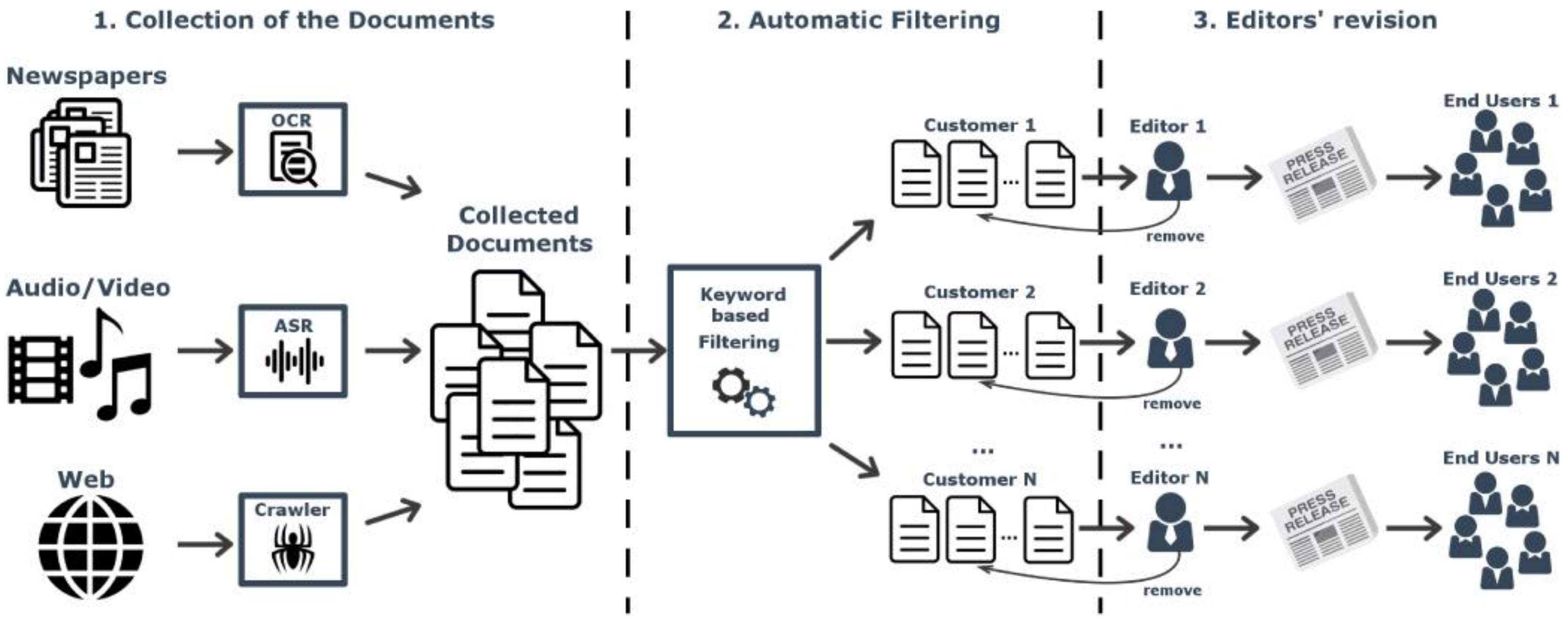
Reinforcement Learning in Healthcare: A Survey

Chao Yu, Jiming Liu, *Fellow, IEEE*, and Shamim Nemati

***Abstract*—As a subfield of machine learning, *reinforcement learning* (RL) aims at empowering one's capabilities in behavioural decision making by using interaction experience with the world and an evaluative feedback. Unlike traditional supervised learning methods that usually rely on one-shot, exhaustive and supervised reward signals, RL tackles with sequential decision making problems with sampled, evaluative and delayed feedback simultaneously. Such distinctive features make RL technique a suitable candidate for developing powerful solutions in a variety of healthcare domains, where diagnosing decisions or treatment regimes are usually characterized by a prolonged and sequential procedure. This survey discusses the broad applications of RL techniques in healthcare domains, in order to provide the research community with systematic understanding of theoretical foundations, enabling methods and techniques, existing challenges, and new insights of this emerging paradigm. By first briefly examining theoretical foundations and key techniques in RL research from efficient and representational directions, we then provide an overview of RL applications in healthcare domains ranging from dynamic treatment regimes in chronic diseases and critical care, automated medical diagnosis from both unstructured and structured clinical data, as well as**

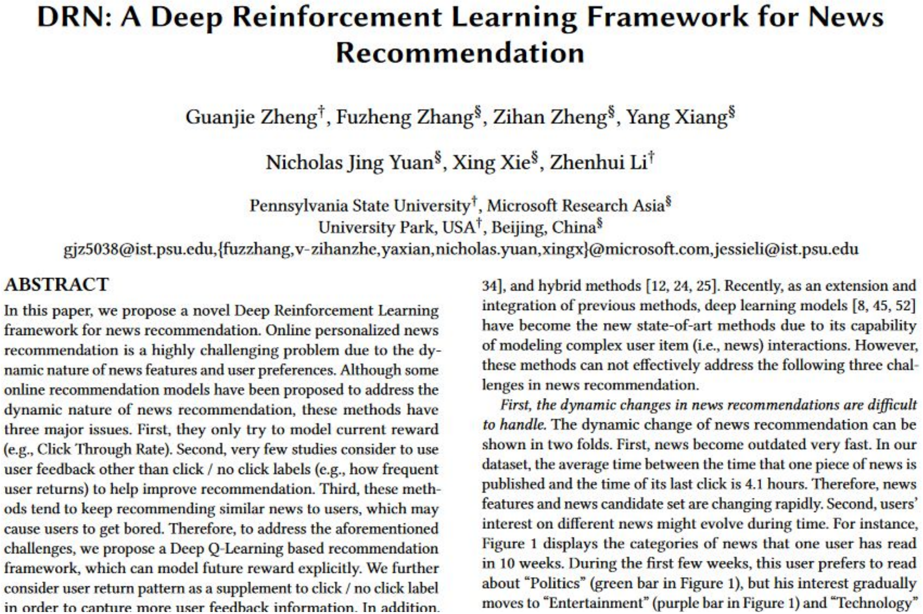
feedback and the new state from the environment. The goal of the agent is to learn an optimal policy (i.e., a mapping from the states to the actions) that maximizes the accumulated reward it receives over time. Therefore, agents in RL do not receive direct instructions regarding which action they should take, instead they must learn which actions are the best through trial-and-error interactions with the environment. This adaptive closed-loop feature renders RL distinct from traditional supervised learning methods for regression or classification, in which a list of correct labels must be provided, or from unsupervised learning approaches to dimensionality reduction or density estimation, which aim at finding hidden structures in a collection of example data [1]. Moreover, in comparison with other traditional control-based methods, RL does not require a well-represented mathematical model of the environment, but develops a control policy directly from experience to predict states and rewards during a learning procedure. Since the design of RL is letting an agent controller

Applications of RL - News Recommendation



Applications of RL - News Recommendation

dl.acm.org/doi/fullHtml/10.1145/3178876.3185994



DRN: A Deep Reinforcement Learning Framework for News Recommendation

Guanjie Zheng[†], Fuzheng Zhang[§], Zihan Zheng[§], Yang Xiang[§]

Nicholas Jing Yuan[§], Xing Xie[§], Zhenhui Li[†]

Pennsylvania State University[†], Microsoft Research Asia[§]
University Park, USA[†], Beijing, China[§]
gjz5038@ist.psu.edu,{fuzzhang,v-zihanzhe,yaxian,nicholas.yuan,xingx}@microsoft.com,jessieli@ist.psu.edu

ABSTRACT

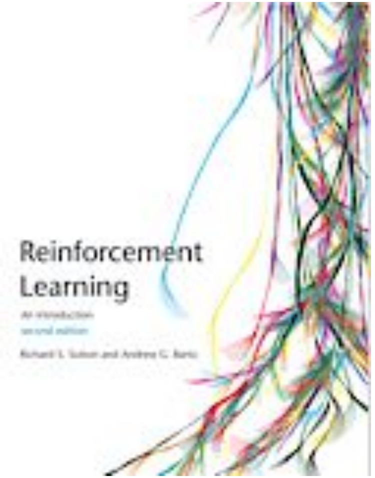
In this paper, we propose a novel Deep Reinforcement Learning framework for news recommendation. Online personalized news recommendation is a highly challenging problem due to the dynamic nature of news features and user preferences. Although some online recommendation models have been proposed to address the dynamic nature of news recommendation, these methods have three major issues. First, they only try to model current reward (e.g., Click Through Rate). Second, very few studies consider to use user feedback other than click / no click labels (e.g., how frequent user returns) to help improve recommendation. Third, these methods tend to keep recommending similar news to users, which may cause users to get bored. Therefore, to address the aforementioned challenges, we propose a Deep Q-Learning based recommendation framework, which can model future reward explicitly. We further consider user return pattern as a supplement to click / no click label in order to capture more user feedback information. In addition,

34], and hybrid methods [12, 24, 25]. Recently, as an extension and integration of previous methods, deep learning models [8, 45, 52] have become the new state-of-art methods due to its capability of modeling complex user item (i.e., news) interactions. However, these methods can not effectively address the following three challenges in news recommendation.

First, the dynamic changes in news recommendations are difficult to handle. The dynamic change of news recommendation can be shown in two folds. First, news become outdated very fast. In our dataset, the average time between the time that one piece of news is published and the time of its last click is 4.1 hours. Therefore, news features and news candidate set are changing rapidly. Second, users' interest on different news might evolve during time. For instance, Figure 1 displays the categories of news that one user has read in 10 weeks. During the first few weeks, this user prefers to read about "Politics" (green bar in Figure 1), but his interest gradually moves to "Entertainment" (purple bar in Figure 1) and "Technology"

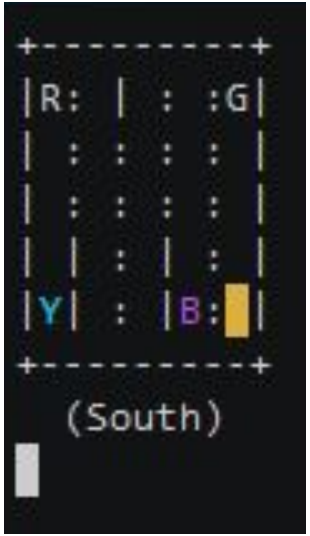
RL Book

- incompleteideas.net/book/the-book.html
- Hands On Reinforcement Learning with Python



RL in Python

- gym.openai.com
- github.com/facebookresearch/ReAgent



```
+---------+
|R: | : :G|
| : : : : |
| : : : : |
| | : | : |
|Y| : |B: |
+---------+
  (South)
```

Sources and Image ref.

- newtechdojo.com/3-types-of-machine-learning
- kdnuggets.com/2018/03/5-things-reinforcement-learning.html
- towardsdatascience.com/practical-reinforcement-learning-02-getting-started-with-q-learning-582f63e4acd9
- freecodecamp.org/news/an-introduction-to-q-learning-reinforcement-learning-14ac0b4493cc
- developer.ibm.com/articles/cc-reinforcement-learning-train-software-agent
- theverge.com/2020/4/22/21231168/cruise-gm-av-ev-renewable-energy-solar-taxis
- ww2.mathworks.cn/matlabcentral/fileexchange/74176-reinforcement-learning-for-financial-trading?s_tid=FX_rc3_behav
- dl.acm.org/doi/fullHtml/10.1145/3350546.3352510