

1. Proximal Policy Optimization

- 1.1. Какой недостаток обучения с подкреплением на основе оптимизации политики решают алгоритмы Trust Region Policy Optimization и Proximal Policy Optimization?
- 1.2. Какая суррогатная функция оптимизируется в алгоритме Trust Region Policy Optimization? Какое решение используется для улучшения оптимизации и какие у него недостатки?
- 1.3. Какую модификацию суррогатной функции предлагают авторы Proximal Policy Optimization?

2. Progressive Neural Networks

- 2.1. Напишите формулу скрытого слоя в архитектуре Progressive Network.
- 2.2. Почему Progressive network лучше, чем fine tuning всей модели, справляются с переносом знаний из ортогональной задачи?
- 2.3. Какие зависимости наблюдаются между исходной (source) и целевой (target) задачами в случаях наименее и наиболее успешного переноса?

3. Learning to Navigate in Cities Without a Map

- 3.1. Что такое curriculum learning? Как был применен этот подход для обучения агента навигации в городе?
- 3.2. Как авторы предлагают параметризовать точку в городе, до которой должен дойти агент? Какие преимущества у этого подхода?
- 3.3. Опишите архитектуру модели CityNav.

4. Curiosity-driven Exploration by Self-supervised Prediction

- 4.1. Какую проблему решают авторы статьи Curiosity-driven Exploration by Self-Supervised Prediction? В чем заключается их подход?
- 4.2. Опишите архитектуру модели ICM. Для чего нужны Forward и Inverse части модели?
- 4.3. Выпишите функционал, который оптимизирует модель ICM при обучении, поясните каждую из ее частей.

5. Non-delusional Q-learning and Value Iteration

- 5.1. Как выглядит правило обновления параметров в алгоритме Q-обучения?
- 5.2. В чем заключается проблема delusional bias, которую выделяют авторы статьи? Какие проблемы Q-обучения авторы связывают с delusional bias?
- 5.3. Для каких классов задач обучения с подкреплением применимы алгоритмы Policy-Class Value Iteration и Policy-Class Q-Learning?