

Рецензия на статью
“The Benchmark Lottery”
От Седашова Данилы, 181 группа

Авторы статьи вводят понятие “лотереи бенчмарков”. Согласно нему, на то, что метод демонстрирует лучшие результаты на каком-либо бенчмарке, не всегда влияет оптимальность алгоритма, а зачастую разное количество неопределённостей. В частности, показатели модели сильно зависят от выбора подмножества заданий. Авторы вводят понятие task selection bias и экспериментально подтверждают свою гипотезу. Также, ссылаясь на прочие исследования, авторы подсвечивают другие проблемы бенчмаркинга: несовершенство разметки, community bias (например, бенчмарки, на которых тестировались сильные модели в оригинальных статьях, становятся популярными). Авторы также предлагают инструкции для создания и использования бенчмарков, а также для процесса ревью статей.

Плюсы статьи:

- Утверждения/гипотезы из статьи подтверждены либо предыдущими, либо собственными исследованиями, всё обоснованно
- Большое количество экспериментов в task selection bias: рассмотрены 4 бенчмарка в разных областях, в каждом большое количество моделей (минимум 11), перебрали различные комбинации подзаданий из бенчмарка
- Вклад: comprehensive survey по проблеме бенчмарков (подсвечивают новую
- проблему для сообщества), плюс вводят понятие task selection bias

Минусы:

- Обзор литературы не в одном месте, а размазан по тексту
- Местами сложно читаемо, есть избыточный текст (даже метафоры)
- Кода нет, эксперименты сложно воспроизвести, гиперпараметры и результаты всех экспериментов не предоставлены — только основные результаты, необходимые для повествования

Оценка: 7, уверенность: 3