

Семинарское задание №9

LtR

Мосиенко Константин Викторович

2019

Прочитайте три статьи:

1. SoftRank¹
2. ListNet²
3. ListMLE³

Реализуйте один из предложенных listwise подходов к обучению ранжированию и проверьте свою реализацию в конкурсе <https://www.kaggle.com/c/sda-ir-ltr>. По сути, основной разработки от вас требует реализация listwise функции потерь и построение процесса обучения (неприятностей должен доставить тот факт, что инстанс задачи тут не пара запрос – документ, а пара запрос – список документов), а вот в качестве обучаемой модели разумно взять обыкновенную нейронную сеть.

Описание столбцов:

1. docid - идентификатор документа.
2. quid - идентификатор запроса.
3. rel - оценка релевантности 0, 1, 2. Чем больше - тем лучше.
4. feature_# - числовой признак.

Пример формата посылки можно посмотреть в файле sample.csv. Вам необходимо для каждого quid отсортировать соответствующие docid-ы в порядке убывания вашего ожидания их релевантности. Метрика - NDCG@10⁴. Решения, не проходящие baseline, засчитываться не будут, как и те, в которых реализован не listwise. Топ-10, как всегда, получают дополнительный балл.

После сабмита на kaggle не забывайте присылать мне ваши отчёты, в которых вам следует описать, как вы решали задачу, какие допущения сделали, с какими проблемами столкнулись, как их решили. Ваш код также должен быть в отчёте. Наиболее удобный вариант: ноутбук + сконвертированный в pdf ноутбук для быстрого просмотра.

¹SoftRank: Optimizing Non-Smooth Rank Metrics. Michael Taylor. 2008

²Learning to Rank: From Pairwise Approach to Listwise Approach. Zhe Cao. 2007

³Listwise Approach to Learning to Rank - Theory and Algorithm. Fen Xia. 2008

⁴https://en.wikipedia.org/wiki/Discounted_cumulative_gain