

Одной из классических проблем в рекомендательных системах является позиционное смещение (position bias) - эффект, заключающийся в увеличении средней фактической конверсии с ростом позиции. Этот эффект негативно влияет на качество рекомендательного алгоритма машинного обучения, поскольку при наличии данного эффекта у отдельных объектов (item-ов) завышается (или занижается) историческая конверсия и среднее значение целевой переменной, что приводит к завышению (или занижению) прогноза при применении данной модели. Помимо того, что позиционное смещение приводит к ухудшению качества прогнозирования, данный эффект вызывает ещё одну проблему: если у одного объекта и средняя позиция показа, и историческая конверсия выше, то как правило не существует возможности объективно определить, за счёт чего именно обусловлена разница в конверсиях - за счёт более высокой позиции или за счёт большей релевантности.

Исследование включало в себя следующие этапы: Сначала была предложена вероятностная модель построения обучающей выборки, содержащей эффект позиционного смещения. Эта модель основана на том, что распределение позиции для конкретного объекта можно смоделировать распределением Дирихле. Далее был предложен метод повышения качества произвольного алгоритма машинного обучения, основанный на симметричной модификации признаков и целевых переменных в зависимости от позиции в выдаче. Эффективность данного подхода оценивалась при помощи метрики RAUC [1, 2], используемой для оценки качества работы алгоритма с учётом наличия позиционного смещения в обучающих и тестовых данных. На последнем этапе исследования была предложена метрика, оценивающая наличие позиционного смещения в выборке только

на основании совместного распределения объектов и позиций. Было показано, что данная метрика имеет высокую корреляцию с увеличением качества прогноза за счёт метода, предложенного на предыдущем этапе. Это особенно важно в прикладных исследованиях, поскольку позволяет оценить эффективность применения метода до фактической его реализации.

Deep Position-wise Interaction Network for CTR Prediction:
<https://arxiv.org/abs/2106.05482>

Position Bias Estimation for Unbiased Learning to Rank in Personal Search: <https://research.google/pubs/position-bias-estimation-for-unbiased-learning-to-rank-in-personal-search/>