

# Идеи по TrafficPoint

## 1 Основная сложность

Основная сложность задачи — мы не можем произвести надёжные измерения для всех перестановок наших брендов, поэтому надо по имеющимся данным предсказывать, какие будут ctr на других перестановках.

## 2 Возможные подходы

Можно предложить несколько моделей, которые будут определять ctr для неиспробованных перестановок. Во-первых, можно предположить, что ctr зависят только от положения (порядкового номера) бренда и не зависят от того, что это за бренд. Это самое грубое предположение. Однако, из данных я видел, что положение бренда намного больше влияет на ctr, чем его описание.

Предположение второго порядка - ctr для данного бренда зависит только от того, что это за бренд и на каком месте стоит, и не зависят от того, что стоит на других местах и в каком порядке. Например, как в одной из статей, вероятность клика является произведением вероятностей для данного места и данного бренда. Тут можно перебрать (например) все циклические перестановки наших брендов и тем самым получить значения для всех ctr. Этих данных будет достаточно.

Что делать, если у ctr бренда есть зависимость и от того, что за бренды стоят на других местах? Можно ввести предположение, что эта зависимость появляется не просто из-за смотрения на другие места, а из-за того, что пользователь кликал на другие позиции. Однако, как это проверить? Мне не удалось довести эту идею до чего-то полезного.

Я видел в данных по матрасам, что вероятность клика на 3 позицию возрастает в 3 раза, если был клик на 1 позицию. И что? Явной зависимости от того, что это была за позиция, я не увидел. Это говорит о том, что часть пользователей настроена более серьёзно и намеревается посмотреть не только один пункт, а несколько. 2 паттерна поведения — “несерьёзные”, которые никуда не кликают, и те, которые заинтересованы в том, чтобы что-то посмотреть. Вторые имеют тенденцию смотреть больше одного пункта. Если уж пользователь решил посмотреть что-то, то он будет смотреть и первый, и второй пункты.

## 3 Подход через постоянное исследование

Можно постоянно делать exploration, а не только exploitation. Два подхода: всё время усложнять модель (данных становится всё больше, и мы можем

настраивать всё более сложные модели) или выбрать что-то оптимальное и делать транспозиции, набирая статистику, когда мы приходим к улучшению. Например, каждый раз иметь две модели на тесте, и когда наберётся достаточное статистическое доказательство, что одна из моделей лучше, худшую модель выбрасываем. У более хорошей модели делаем случайную транспозицию и смотрим, улучшились ли результаты. Так в каждый момент у нас две модели.

Со “всё более сложными моделями” проблема — я не знаю, как придумать более сложную зависимость для перестановки.

Применить что-то типа “генетического алгоритма”?

Предпочитаемый мной подход — сначала провести все циклические перестановки, а потом, выбрав максимальную по фин. результату перестановку, от неё плясать транспозициями. Чем хуже ожидаемый фин. результат от транспозиции, тем реже (с меньшей вероятностью) её выбирать. Ожидаемые фин. результаты всё время апдейтятся. Продолжаем эксперимент до тех пор, пока с вероятностью 95% не станет ясно, какой фин. результат лучше. После этого переходим к новому эксперименту, выбрав в качестве исходной ту перестановку, которая лучше.

Вопрос: а сколько примерно показов нужно, чтобы отделить две перестановки по фин. результату?

Замечание: если одна из сравниваемых перестановок сильно хуже, то, вероятно, это быстро выяснится. Экспериментов надо будет немного.