

Санкт-Петербургский государственный университет

Группа 22.М05-мм

Королихин Владимир Игоревич

Предсказание параметров в формулах
управляющих воздействий автоматических
стратегий назначения ставок в рекламных
кампаниях

Отчёт по учебной практике

Научный руководитель:
к.ф.-м.н., доц Е.Г.Михайлова

Консультант:
руководитель группы ООО «Яндекс Технологии» к.т.н. Е.А. Самосват

Санкт-Петербург
2023

Оглавление

Введение	3
1. Постановка задачи	5
2. Обзор	6
2.1. Процесс запуска рекламной кампании	6
2.2. Обзор видов автостратегий	7
2.3. Описание текущего аналитического подхода расчета ставок	8
2.4. Недостатки аналитических формул	10
3. Результаты	12
Список литературы	13

Введение

В прошлом создание контекстной рекламы требовало от рекламодателя сложной ручной настройки. Для максимизации своих бизнес-показателей он должен был тщательно выбирать поисковые запросы, релевантные объявлению, а затем самостоятельно выставить ставки для участия в аукционе (алгоритм, решающий какую рекламу показать). Однако в последние годы произошел сдвиг в сторону автоматизации и упрощения работы с рекламой с использованием методов машинного обучения. Это привело к разработке автотаргетинга, который автоматически подбирает поисковые запросы для таргетирования рекламы, и автоматических стратегий, позволяющих рекламодателям доверить процесс эффективного расходования своих рекламных бюджетов рекламной площадке. В данной работе рассматриваются способы улучшения автоматических стратегий в Яндекс Директе. Это сервис для размещения объявлений контекстной рекламы на Яндексе и на сайтах-партнерах его рекламной сети.

Рекламодатель может выбирать различные виды автостратегий и их параметры исходя из целей бизнеса и знания стоимости собственных продуктов или определенных событий. С технической точки зрения он накладывает ограничения на некоторую функцию, которую необходимо оптимизировать. Ее результатом является набор управляющих воздействий, использующихся непосредственно в аукционе. Кроме выдерживания ограничений, такая функция должна быть адаптивна к внезапным изменениям в ее параметрах. Например, рекламодатель может изменить среднюю цену за клик. При такой постановке задачи точные аналитические формулы начинают работать не очень хорошо. В частности, не учитывается ситуация холодного старта - когда кампания только запустилась и данных, чтобы предугадать ее поведение в аукционе недостаточно. Для решения такой проблемы лучше всего подходит машинное обучение, которое позволяет эффективно находить сложные и не поддающиеся формализации структурные закономерности входных данных. Таким образом, исследование способов перехода от ана-

литического подхода решения задачи к моделям машинного обучения является актуальной задачей.

Можно доказать, что оптимальной моделью трат, максимизирующей конверсии при ограничениях на недельный бюджет или среднюю цену конверсии является такая, в которой ставка остается постоянной на протяжении всей недели. Вследствие этого возникает задача минимизации дисперсии ставок. Кроме того, ставки рекламных кампаний подчиняются степенному закону, что означает с одной стороны большое количество маленьких кампаний и малое количество больших, а с другой разброс целевой переменной по абсолютной величине. Такие ограничения сужают набор функций потерь и приемочных метрик. Само значение ставки вычисляется на основе оцененных коэффициентов, которые посчитаны по накопленной статистике из прошлого или же характеристик рекламной кампании в случае отсутствия такой статистики. Таким коэффициентом может выступать, например, временной таргетинг.

В данной работе рассматриваются различные подходы по переходу от аналитических формул в алгоритмах выставления ставок к алгоритмам машинного обучения, которые бы могли оценивать отдельные коэффициенты этих формул. Они позволяют адаптировать рекламную кампанию к изменениям в аукционе, более точно проводить аналитику кампании и улучшить модель трат, приводя ее к оптимальному состоянию.

1. Постановка задачи

Целью данной работы является адаптация текущих аналитических формул вычисления коэффициентов, участвующих в формулах управляющих воздействий автобюджета, на модели машинного обучения. Для достижения цели были поставлены следующие задачи:

1. Сделать обзор модели трат биддера автобюджета.
2. Реализовать сбор пула для обучения моделей.
3. Подобрать приемочные метрики, функцию потерь и обучить модель.
4. Провести эксперименты, проанализировать результаты и внедрить полученную модель.

2. Обзор

В этом разделе рассматривается работа автостратегий, возможные способы прогнозирования ставок, а также принцип работы рекламы в Яндекс Директе.

2.1. Процесс запуска рекламной кампании

Чтобы запустить рекламу, рекламодатель должен сначала создать набор рекламных материалов, которые будут рассказывать потенциальным клиентам о преимуществах продуктов или услуг. Материалы включают в себя рекламные тексты, изображения или видео, а также набор ссылок, которые будут вести по рекламным банерам. Следующим шагом рекламодатель должен задать таргетинг - набор критериев, которые позволят определить аудиторию, на которую будет рассчитана реклама. Существует также автотаргетинг, который автоматически подбирает запросы пользователя под заданные рекламные банеры. По данному набору ограничений решается попадет ли реклама в рекламный хит. Хит это процесс загрузки страницы с рекламными банерами. В случае положительного решения реклама начинает участие в аукционе, который в этом хите проводится. В аукционе рекламные кампании конкурируют за показ на банере, а торгуются между собой при помощи ставок [2]. Алгоритм, который выбирает ставку в каждом хите называется биддинг. В данной работе рассматривается только этап биддинга. Будем считать, что рекламные материалы нам заданы, а на таргетинг повлиять никак не можем. Единственное чем мы можем управлять это ставками, меняя которые можем добиваться эффективности рекламы.

Исторически самым первым способом биддинга являлось ручное управление ставками, которое перекладывало на рекламодателя необходимость задавать ставку для каждой ключевой фразы. В зависимости от ключевой фразы, по которой произошел таргетинг выставлялась ставка с которой реклама заходит в аукцион.

В последние годы произошел переход в сторону автоматизации выставления ставок, которые называются автостратегиями. Рекламодатель

тель с помощью ограничений ставит оптимизационную задачу на алгоритм автостратегии, которые в свою очередь считают оптимальную ставку. Например рекламодатель может поставить задачу максимизировать конверсии при заданной средней цене конверсии и ограничения на недельный бюджет.

2.2. Обзор видов автостратегий

2.2.1. Автостратегии, основанные на правилах

Автостратегии, основанные на правилах похожи на стратегии с ручным управлением ставок, но используют дополнительный набор правил или эвристик, выставленных рекламодателем, которые определяют, как следует корректировать ставки.

Такие автостратегии часто используются в ситуациях, когда существуют конкретные ограничения или требования, которые должны быть выполнены. Например, рекламодатель может хотеть, чтобы его реклама показывалась только в определенное время суток или в определенные дни недели. Более сложные правила могут включать калибровку ставки на определенную величину, если показатель кликабельности по его объявлению падает ниже определенного порога.

Одним из преимуществ автостратегий, основанных на правилах, является то, что они часто относительно просты в реализации и могут быть настроены в соответствии с конкретными потребностями рекламодателя. Однако они могут не учитывать такие факторы как поведение пользователей, конкуренция и меняющиеся рыночные условия.

Автостратегии, основанные на правилах наиболее эффективны при использовании в сочетании с другими алгоритмами [3], в частности с машинным обучением.

2.2.2. Автостратегии, использующие машинное обучение

Автостратегии также могут использовать машинное обучение. Такие алгоритмы учатся по историческим данным и корректируют ставки

в режиме реального времени на основе различных факторов, таких как размещение рекламы или таргетинг на аудиторию. Например в статье [1] изучен метод прогнозирования поведения ставок с использованием градиентного бустинга на решающих деревьях. Для последующих корректировок авторы используют линейную регрессию.

В работе [2] авторы использовали обучение с подкреплением для обучения модели автоматического назначения ставок, которая показала хорошие результаты на A/B тестировании.

К сожалению, переиспользовать какие-либо вышеперечисленные подходы не представляется возможным из-за сложившихся внутри компании процессов по обучению моделей, в частности используется Catboost. С другой стороны, в данной работе не ставится цель предсказывать ставки, а лишь коэффициенты в формулах. Конечно, в будущем планируется перейти на полностью машинно-ориентированный подход, но это более сложная задача.

2.3. Описание текущего аналитического подхода расчета ставок

Рассмотрим, какой будет являться оптимальная ставка при следующих условиях:

Теорема: Если выполнено хотя бы одно из следующих ограничений:

1. Задана средняя цена конверсии
2. Задано ограничение на период работы рекламной кампании

Тогда оптимальной будет такая ставка, которая остается постоянной на протяжении всей работы рекламной кампании

Доказательство:

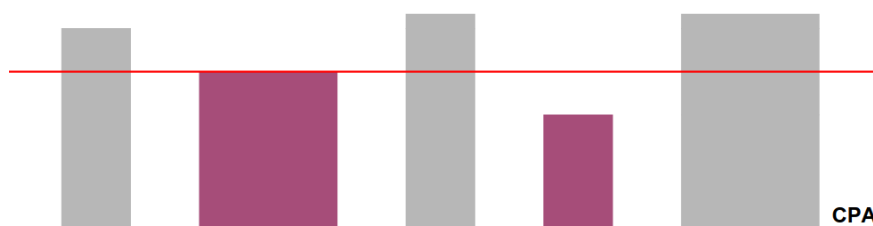


Рис. 1: количество конверсий и цена конверсии

На рисунке изображены хиты, которые произойдут в течение недели. Ширина столбика это количество конверсий, а высота это цена конверсии. Таким образом график в терминах CPA, т.е. Cost per action - средняя цена конверсии.

Для начала покажем какой будет искомая ставка. Постепенно поднимая линию выше, мы будем покупать все более дорогие конверсии. Вследствие этого вырастет бюджет, который мы за неделю потратили и средняя цена конверсии. Таким образом, в какой-то момент мы достигнем одного из ограничений. Ставка, которая при этом получится и будет оптимальной.

Предположим существует другой алгоритм, который смог купить больше конверсий, чем наш. Значит он должен был купить какое-то количество конверсий над чертой, так как все что ниже купил наш алгоритм. Возможно он отказался от части конверсий, которые купил наш алгоритм, однако ему нужно их будет докупить из тех, что находятся над чертой и соответственно дороже. Следовательно вырастет либо потраченный недельный бюджет, либо средняя цена конверсии, что противоречит условию.

Примечание: в доказательстве алгоритма мы пользуемся предположением, что нам заранее известно ожидаемое количество конверсий в каждом хите.

В Яндекс Директе существует большое количество автостратегий. Можно максимизировать конверсии/установки/клики/показы с различными видами ограничений на среднюю цену события (например конверсий) и/или на недельный/периодный бюджет с оплатой за показы/клики/конверсии. Все перечисленные автостратегии объединяет общая схема подсчета ставок. Для определенности в работе рассматривается оптимизация конверсий с ограничением на недельный бюджет.

Рассмотрим текущий аналитический подход расчета ставки для данного случая:

- Пусть $T(t)$ - средняя интенсивность трафика, не зависящая от заказа, а зависящая от количества пользователей в интернете в

данный момент времени. Например, ночью этот параметр будет меньше, чем днем.

При этом: $\int_{week} T(t)dt = 1$

- Тогда интенсивность трат в текущий момент времени ($EventCost$) можно связать со ставкой на заказ (Bid), где C - некоторая константа, зависящая от заказа, а $P \geq 1$ - степень к ставке, которая позволяет захватывать более дорогие места показа, чтобы быстрее прийти к стационарному состоянию:

$$EventCost(t) = C \cdot Bid(t)^P \cdot T(t) \quad (1)$$

- Проинтегрировав формулу (1), можно выразить ставку на заказ на оставшийся период работы кампании:

$$Bid = \left(\frac{WeekLimit - SpentCost(now)}{C(now) \cdot \int_{now}^{week} T(t)dt} \right)^{\frac{1}{P}}, \quad (2)$$

где $WeekLimit$ - ограничение на недельный бюджет, а $SpentCost$ - траты к текущему моменту времени

- Аналогично можно выразить и константу C :

$$C(now) = \frac{\int_{now-week}^{week} EventCost(t)dt}{\int_{now-week}^{week} Bid(t)^P \cdot T(t)dt} \quad (3)$$

2.4. Недостатки аналитических формул

Ранее было показано, что в идеальном случае ставка на заказ должна быть постоянной на протяжении всей работы кампании. К сожалению, в реальном мире такого добиться очень сложно из-за множества проблем. Например:

- События связанные, с резким увеличением потока клиентов. Например, черная пятница.
- Изменения рекламодателем настроек кампании.
- Различные внутренние поломки алгоритмов рекламы

Для вычисления ставки, исходя из формулы (2) нужно знать $C(now)$, которая оценивается по историческим данным согласно формуле (3). Из этого вытекает несколько недостатков текущего подхода:

1. В случае сезонных скачков (например, черной пятницы) изменится и объем трафика, приходящий по данному заказу, что повлияет на ставки. Вернуться после таких изменений к прежним значениям может быть проблематично
2. В случае старта кампании в формуле (3) возникает неопределенность $\frac{0}{0}$, решить которую помогают прайорные значения в числителе и знаменателе. По мере получения статистики прайоры уменьшаются, однако точность C на старте кампании остается под сомнением.

Решить данную проблему позволит машинное обучение, которое сможет эффективно использовать имеющиеся характеристики кампании для предсказания начальных ставок. А в будущем и накопленную статистику для адаптивной реакции на сезонные изменения.

3. Результаты

В данной работе рассмотрена предметная область, поставлена задача и начата реализация сбора обучающего датасета. Составлен план для будущего решения:

1. Собрать и предобработать датасет
2. Придумать метрики приемки и обучить на полученных данных Catboost
3. Эксперименты и апробация результата

Список литературы

- [1] Li Wei, Cui Ying Grace, Zhang Ruofei, and Mao Jianchang. Bid landscape forecasting in online advertising. — 2012. — Apr. 5. — US Patent App. 12/895,773.
- [2] Amin Kareem, Kearns Michael, Key Peter, and Schwaighofer Anton. Budget optimization for sponsored search: Censored learning in mdps // arXiv preprint arXiv:1210.4847. — 2012.
- [3] Berg Jordan, Greenwald Amy, Naroditskiy Victor, and Sodomka Eric. A first approach to autonomous bidding in ad auctions // Workshop on Trading Agent Design and Analysis at the 11th ACM Conference on Electronics Commerce. — 2010.