

Применение машинного обучения для улучшения модели трат автоматических стратегий назначения ставок в рекламных технологиях

В. И. Королихин¹

Е. Г. Михайлова²

С. К. Мить³

¹Математическое обеспечение и администрирование информационных систем, группа 22.М05-мм, математико-механический факультет, СПбГУ

²к.ф.-м.н., доц, научный руководитель

³руководитель группы ООО «Яндекс Технологии», консультант

10 марта 2024 г.

- Ручная настройка рекламных кампаний сложна и неэффективна
- С появлением автостратегий, бюджет рекламодателя расходуется ориентированно на его продукт, используя больше факторов
- Автостратегии необходимо улучшать и переводить на мл-подход, так как он позволяет анализировать более сложные зависимости

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:

- ❶ Сделать обзор модели трат биддера автобюджета.
- ❷ Реализовать сбор пула для обучения моделей.
- ❸ Подобрать приемочные метрики, функцию потерь и обучить модель.
- ❹ Провести эксперименты, проанализировать результаты и внедрить полученную модель.

Процесс запуска рекламной кампании

- Чтобы запустить рекламу, рекламодатель должен сначала создать набор рекламных материалов
- Дальше он попадает в аукцион. В аукционе рекламные кампании конкурируют за показ на банере, а торгуются между собой при помощи ставок
- В работе рассматривается только способы управления ставками

- Автостратегии, основанные на правилах: используют дополнительный набор правил или эвристик, выставленных рекламодателем, которые определяют, как следует корректировать ставки
- Автостратегии также могут использовать машинное обучение. Такие алгоритмы учатся по историческим данным и корректируют ставки в режиме реального времени

Описание текущего аналитического подхода расчета ставок

- Тогда оптимальной будет такая ставка, которая остается постоянной на протяжении всей работы рекламной кампании
- Интенсивность трат: $EventCost(t) = C \cdot Bid(t)^P \cdot T(t)$
- Формула ставки: $Bid = \left(\frac{WeekLimit - SpentCost(now)}{C(now) \cdot \int_{now}^{week} T(t) dt} \right)^{\frac{1}{P}},$
- Искомый параметр: $C(now) = \frac{\int_{now-week}^{week} EventCost(t) dt}{\int_{now-week}^{week} Bid(t)^P \cdot T(t) dt}$

Недостатки аналитических формул

Колебания из-за:

- Событий связанных, с резким увеличением потока клиентов. Например, черная пятница.
- Изменений рекламодателем настроек кампании.
- Различных внутренних поломок алгоритмов рекламы

Для сбора датасета использовались система MapReduce Ytsaurus.

- 1 Сбор данных из различных источников, отбор нужных признаков и агрегация по часам
- 2 Программа Редьюсера для получения фичей за различные временные промежутки
- 3 Добавление дополнительных фичей-характеристик, не зависящих от рекламной кампании

Некоторые эвристики, которые улучшили качество обучающего датасета.

- 1 Эмулирование оставания данных
- 2 Исключение малоинформативных примеров с отсутствием статистики
- 3 Добавление информации о рабочих часах кампании

Для обучения использовался Catboost.

❶ Кросс-энтропия как функция потерь:
$$\frac{\sum_{i=0}^N -w_i(t_i \log(p_i) + (1-t_i) \log(1-p_i))}{\sum_{i=0}^N w_i}$$

❷ Метрика по приемке llp:

$$best_const = \frac{\sum_{i=0}^N w_i t_i}{\sum_{i=0}^N w_i}$$

$$LL_best_const = \frac{\sum_{i=0}^N -w_i(t_i \log(best_const) + (1-t_i) \log(1-best_const))}{\sum_{i=0}^N w_i}$$

$$LLp = \frac{LL_best_const - LL}{best_const}$$

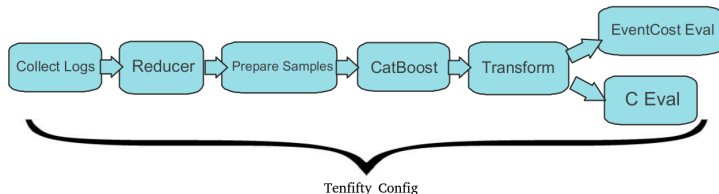
❸ Улучшение по метрике llp на 5%

Общий пайплайн

Результатом работы являются не только обученные модели, но набор отдельных программных модулей, объединенных при помощи конфига. Который позволяет гибко настраивать параметры обучения и обеспечивает воспроизводимость результатов.

Используемые языки

- YQL: это SQL подобный язык, используемый для запуска MapReduce задач
- C++: использовался для реализации программы редьюсера, выполняющего генерацию семплов для обучения
- Tenfifty: внутренняя библиотека для создания конфигов обучения моделей. Набор .yaml файлов



В данной работе рассмотрена предметная область, поставлена задача. Реализован инструмент сбора данных, обучена первая модель, проведены замеры качества.

Планируется:

- 1 Эксперименты и апробация результата