

Секция «Вычислительная математика и кибернетика»

Прогнозирование вероятности кликов на новые рекламные объявления.

Колесников Александр Александрович

Студент

*Московский государственный университет имени М.В. Ломоносова, Факультет
вычислительной математики и кибернетики, Москва, Россия*

E-mail: alekkolesnikov@gmail.com

Основным источником прибыли поисковых систем является контекстная реклама. Главная задача системы показов контекстной рекламы — это отбор баннеров для размещения на странице, которую просматривает пользователь. Целью такого отбора является максимизация прибыли поисковой системы и эффективности рекламы для рекламодателей. Ключевой величиной, используемой в алгоритмах отбора баннеров, является вероятность клика на баннер, при условии что он будет показан конкретному пользователю на конкретной web-странице. Если баннер собрал достаточно большую статистику показов, то вероятность клика по нему можно оценить как отношение количества кликов к количеству показов, и затем скорректировать эту оценку с учетом доступной информации о пользователе и контексте web-страницы, на которой он показывается. Если же баннер новый или имеет слишком короткую историю показов, то предсказание вероятности клика по нему усложняется. В работе рассматривается подход, позволяющий оценивать стартовую вероятность кликов на новые баннеры.

С каждым баннером связан текст, который видит пользователь, и множество фраз, по которым производится отбор баннеров: если пользовательский запрос полностью содержит фразу баннера, то баннер становится кандидатом на показ. Задача состоит в том, чтобы оценить стартовую вероятность клика для каждой пары баннер-фраза. Предлагается регрессионная модель, использующая около 20 признаков следующих типов: средние частоты кликов по баннерам со схожими текстами и фразами, которые набрали значимую статистику; оценки релевантности фразы баннера его тексту; количество слов во фразе; репутация домена, на который ссылается баннер. Кроме того, используется более 100 бинарных признаков присутствия определённых слов во фразе баннера. Признаки для регрессионной модели отбирались с учетом совокупности критериев, оценивающих их полезность для модели. Регрессионная модель для предсказания стартовой вероятности клика строится на основе градиентного бустинга над деревьями регрессии. Целевой функцией при построении модели является частота кликов по баннерам, которые показывались в системе. Предложенная модель показала значительное улучшение качества предсказания вероятности кликов на баннеры по сравнению с базовой моделью, в которой всем парам баннер-фраза приписывалась одна из нескольких констант. Линейная корреляция между прогнозом и вероятностью клика выросла на 75% по сравнению с базовой моделью. Обучение модели производилось на четырех миллионах баннеров, собранных из статистики показов. Данные для исследования предоставлены ООО Яндекс.

Литература

1. Regelson M., Fain D. Predicting click-through rate using keyword clusters. In Proceedings of the Second Workshop on Sponsored Search Auctions, 2006. — Vol. 9623.

2. Richardson M., Dominowska E., Ragno R. Predicting clicks: estimating the click-through rate for new ads. In Proceedings of the 16th international conference on World Wide Web, pages 521–530. ACM, 2007.

Слова благодарности

Выражаю благодарность своему научному руководителю, Воронцову Константину Вячеславовичу, и группе оптимизации рекламы в компании Яндекс за помощь в проведении исследования.