

DOI:

УЛУЧШЕНИЕ УПРАВЛЕНИЯ ВЗАИМООТНОШЕНИЯМИ С КЛИЕНТАМИ НА ОСНОВЕ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА МОДЕЛЕЙ ПОВЕДЕНИЯ ПОЛЬЗОВАТЕЛЕЙ

Монастырская М.М., Соловьев В.И.

Финансовый университет при Правительстве РФ,

Россия, г. Москва, Щербаковская ул., 38

182271@edu.fa.ru, vsoloviev@fa.ru

Аннотация: Предложен комплекс моделей регрессионного и кластерного анализа, а также поиск ассоциативных правил, для поиска закономерностей в поведении пользователей относительно маркетинговых кампаний с учетом характеристик пользователей и финансово значимых метрик. На примере двух организаций удалось с достаточной точностью прогнозировать стоимость привлечения клиентов (CPA, cost-per-action). Использование предложенных моделей позволяет организациям совершенствовать настройки рекламных объявлений для повышения эффективности интернет-рекламы.

Ключевые слова: регрессионный анализ, кластерный анализ, поиск ассоциативных правил, маркетинговые метрики, цена за действие, интернет-реклама.

Введение

Конкуренция в XX в. росла прежде всего за счет того, что рынок производителя сменился рынком потребителя. Чтобы становиться более ориентированными на клиентов, организации стали внедрять различные информационные технологии, обеспечивающие персонализацию и удобство взаимодействия с клиентами.

В 1980-х гг. организации начали активно внедрять CRM-системы (customer relationship management systems), предназначенные для автоматизации взаимоотношений с клиентами, прежде всего, в части маркетинга и продаж. CRM-системы позволили не только вести учет клиентов, контрагентов, реализуемых товаров и услуг, но и автоматизировать многие процессы. На сегодняшний день CRM-системы широко распространены как среди крупного, так и среди среднего и малого бизнеса.

Сегодня на привлечение клиентов компании тратят значительные бюджеты. Крупнейшие рекламодатели на российском рынке, среди которых МТС, Lamoda, Мегафон, Mail.RU Group и другие, тратят на интернет-рекламу более чем по 2 млрд руб. в год. Доля интернет-рекламы за последние 15 лет существенно выросла: с 2% в 2005 г. до 40% в 2019 г. Интернет-реклама фактически стала основным каналом продвижения, а ее использование привело к росту конкуренции.

Интернет-реклама позволяет отслеживать все клики пользователей, в том числе переходы по рекламным объявлениям и по страницам веб-сайтов. Также она позволяет собирать информацию о каждом клиенте, включая город, возрастную категорию, тип используемого устройства и др. Эта информация образует весьма существенные наборы данных о маркетинговых кампаниях, пользователях и их поведении.

Несмотря на широкую функциональность CRM-систем, значительная часть бизнес-задач этими функциональными возможностями не решается. Среди таких задач – актуализация и обогащение данных о клиентах, скоринг клиентов, управление маркетинговыми кампаниями и др.

При этом технологии машинного обучения широко используются в различных областях маркетинга и продаж, однако в сфере управления взаимоотношениями с клиентами алгоритмы машинного обучения используются достаточно редко.

Сегодня большинство маркетологов управляют рекламой на основе отчетности, построенной вручную, что позволяет проанализировать только нескольких основных разрезов, но такой подход не всегда является эффективным, так ручной перебор и анализ всех комбинаций требует слишком больших трудозатрат. Более того, такую активность необходимо проводить на регулярной основе, поэтому технологии машинного обучения представляются эффективным инструментом оптимизации и совершенствования процессов управления интернет-рекламой.

Данная работа посвящена совершенствованию управления взаимоотношениями с клиентами на основе анализа паттернов поведения пользователей с помощью алгоритмов машинного обучения. Предлагается подход, который заключается в построении комбинации моделей регрессионного и кластерного анализа для поиска закономерностей в поведении пользователей относительно маркетинговых кампаний с учетом характеристик пользователей и финансово значимых метрик.

Объектом исследования в данной работе является процесса привлечения клиентов, а предметом выступает поиск паттернов поведения пользователей в интернет-маркетинге.

Под паттерном понимается систематически повторяющийся, устойчивый элемент (фрагмент поведения) либо последовательность таких элементов. Выявление закономерностей в поведении пользователей может позволить корректировать маркетинговые кампании в реальном времени путем установления ограничений и затрат на клик (в том числе по ограничению аудитории, которой показываются объявления) и повысить эффективность кампаний.

Для построения модели был проведен сравнительный анализ работы алгоритмов, таких как линейная регрессия, деревья решений, случайные леса, AdaBoost и XGBoost, для предсказания значения метрики эффективности маркетинговых кампаний, и K-средних, EM-алгоритм и иерархическая кластеризация для поиска сегментов в поведении пользователей, а также выполнен поиск ассоциативных правил для анализа выявленных кластеров. С точки зрения практического применения подхода, были проведены сбор и обработка данных по интернет-рекламе двух организаций, была реализована модель по прогнозированию стоимости привлечения CPA (cost-per-action). Для анализа закономерностей в системе был применен кластерный анализ и поиск ассоциативных правил по маркетинговым объявлениям. Такая модель позволит организациям совершенствовать настройки рекламных объявлений для повышения эффективности интернет-рекламы в целом.

1 Материалы и методы

В данной работе рассматривается применение методов машинного обучения для анализа кампаний контекстной рекламы. Контекстная реклама – вид размещения рекламных объявлений по запросам пользователей или на ресурсах, близких по специфике к объявлению, т.е. рекламодатель изначально размечает рекламные объявления таким образом, чтобы она была показана только потенциально заинтересованным пользователям.

Самым распространённым примером такой рекламы выступают объявления на страницах поисковых систем, таких как Яндекс и Google. На рынке российской рекламы в интернете более 50% занимает Яндекс.Директ и более 35% – Google AdWords. При вводе пользователем поисковой фразы сервисы показывают платные объявления помимо основной поисковой выдачи. Стоит отметить, что для данного канала привлечения критически важна правильная настройка целевой аудитории, что вручную возможно сделать только на верхнем уровне в связи с большим количеством параметров настройки.

На сегодняшний день используются следующие основные подходы оценки эффективности маркетинговой активности:

- *CTR* (click-through rate, или показатель кликабельности);
- *CPA* (cost per action, или цена за действие);
- *CPO* (cost per order, или стоимость заказа);
- *CRR* (cost revenue ratio, или доля рекламных расходов).

Каждая из метрик является показательной и применяется в зависимости от задач и специфики деятельности предприятия.

CTR рассчитывается как отношение количества кликов на баннер или рекламное объявление к общему количеству показов, т.е. в рамках контекстной рекламы эта метрика показывает, насколько показываемое объявление соответствует запросу пользователя.

CPA считается как отношение расходов на рекламное объявление к количеству совершенных целевых действий, данная метрика особо важна с точки зрения оптимизации процессов, так как отражает финансовую составляющую эффективности маркетинговых кампаний.

Если целевым действием выступает приобретение товара или услуги, то метрика называется *CPO* и отражает производительность рекламы. Если значение *CPO* больше прибыли с заказа, то такой канал привлечения является нерентабельным, однако не всегда возможно однозначно оценить значения переменных данного показателя.

Наконец, *CRR* рассчитывается как отношение расходов на рекламу к доходам от рекламы, чем ниже значение показателя, тем эффективнее реклама. Для данной метрики корректность расчета переменных является такой же затруднительной, как для *CPO*.

CTR выступает основной метрикой, которая показывает отношение показов объекта маркетинговой активности и кликов на него. Данный показатель помогает оценить эффективность каналов взаимодействия с клиентами.

Ключевой особенностью метрики *CPA* является то, что компания сама решает, в отношении какого целевого действия считается эффективность активности. Часто помимо особенностей бизнеса, которые влияют на определение целевого действия, на его выбор также могут повлиять системные ограничения, если в одном интернет-магазине с доставкой можно отследить действия клиента от отклика до покупки товара, то в другом интернет-магазине возможна аналитика только до промежуточного действия до покупки, например, записи на примерку в случае отрасли продажи одежды и обуви.

Значения данных метрик необходимо не только измерять, но и уметь предсказывать, так как это влияет на эффективность маркетинговых активностей и стратегии в целом. Далее для поиска закономерностей в различных кластерах по данным из сервисов веб-аналитики (Google Analytics), а также из CRM-системы и внешних источников, реализован анализ паттернов путем объединения алгоритмов прогнозирования маркетинговых метрик и кластеризации клиентов для получения достоверной картины об эффективности маркетинговых кампаний.

Разработка модели осуществляется на основе данных конкретной организации по рекламным кампаниям. Данные содержат информацию об активности пользователей после перехода по рекламному объявлению в разрезе различных описательных характеристик пользователей. Источником данного набора данных является личный кабинет Google Adwords данной организации.

Набор данных содержит 1143 записи с 11 признаками:

- *ad_id* – уникальный идентификатор рекламного объявления;
- *xyzcampaignid* – уникальный идентификатор маркетинговой кампании, связанный с рекламным объявлением;
- *fbcampaignid* – уникальный идентификатор маркетинговой кампании в Facebook, связанный с рекламным объявлением;
- *gender* – пол пользователя, которому было показано объявление;
- *age* – возраст пользователя, которому было показано объявление;
- *interest* – код, определяющий категорию интересов, к которой относится пользователь;
- *Impressions* – количество показов рекламного объявления;
- *Clicks* – количество кликов на рекламное объявление.
- *Spent* – расходы компании на показы рекламного объявления;
- *Total conversion* – количество пользователей, которые были заинтересованы в товаре/услуге после просмотра рекламного объявления;
- *Approved conversion* – количество пользователей, которые купили товар/приобрели услугу после просмотра рекламного объявления.

В этот набор данных был добавлен синтетический признак, отражающий значение *CPA* и рассчитывающийся как отношение стоимости рекламы к количеству целевых действий:

$$(1) \quad CPA = \frac{Spent \cdot Impressions}{Approved conversion},$$

где стоимость рекламы рассчитывается как произведение расходов на показ объявления и количество показов, а количество целевых действий как количество пользователей, которые совершили покупку. Так как целевые действия могли быть не совершены, то считаем, что минимальная цена за действие будет при совершении хотя бы одного целевого действия, т. е. при значении параметра *Approved conversion*, равном 1.

Предсказание значения цены за действие (*CPA*) представляет собой задачу регрессии.

Для решения данной задачи необходимо построить модель регрессионного анализа. Для получения наиболее точного прогноза сравнивались следующие алгоритмы: линейная регрессия [1], деревья решений [2], AdaBoost [3], случайный лес [4], FTRL-Proximal [5], а также XGBoost [6].

В задаче предсказания значения цены за действие лучшим показал себя алгоритм XGBoost. При глубине деревьев, равной 3, и 90 деревьям, данный алгоритм дал среднюю абсолютную ошибку предсказания $MAE = 1\,737\,275$, что меньше, чем все остальные алгоритмы. Для сравнения, линейная регрессия показала $MAE = 3\,598\,575$, случайный лес дал $MAE = 1\,814\,407$, для бустинга на основе алгоритма AdaBoost после подбора гиперпараметров средняя абсолютная ошибка предсказания составила $MAE = 1\,775\,887$.

Регрессионный анализ показателя *CPA* позволяет получить четкое представление о соотношении маркетинговых кампаний и пользовательских сегментов, а также его влиянии на показатели эффективности рекламы.

Поиск паттернов поведения основан на поиске закономерностей в откликах пользователя на рекламное объявление с учетом финансового анализа показателей для повышения эффективности изменений в маркетинговых кампаниях.

Для сегментации клиентов использован кластерный анализ. Для оптимального количества кластеров, равного 5, была проведена сегментация алгоритмом К-средних [7] и EM-алгоритмом [8]. Показатель силуэта для метода К-средних составил 0,8, что лучше чем для EM-алгоритма (у которого он оказался равен $-0,38$).

Важно отметить, что помимо того, что разные группы пользователей могут по-разному проявлять активность с точки зрения заинтересованности в предлагаемых товарах и услугах, они по-разному могут вести себя с точки зрения совершения целевого действия, то есть высокий интерес не всегда ведет к совершению покупки, например, компания может привести на сайт большое количество специалистов, работающих в той же сфере, или пользователей, которые не готовы к расходам на товары из предлагаемой ценовой категории, и в итоге они не совершат покупку, что негативно отражается на эффективности маркетинговой кампании.

Главная задача разрабатываемой модели заключается в том, чтобы найти финансово эффективные и неэффективные сегменты и настроить маркетинговые кампании в соответствии с полученными результатами, таким образом, на какие-то категории товаров стоит привлекать меньшее количество клиентов, но которые потенциально являются покупателями, или, например, каким-то пользователям необходимо предлагать другую категорию товаров, чтобы они в итоге совершили целевое действие.

В результате был получен ряд конкретных практических выводов. Например, наименьшая цена за действие оказалась в кластере под номером 3. Этот сегмент свойствен для маркетинговой кампании «*huз_campaign_id_1178*» с аудиторией, состоящей из мужчин в возрасте от 30 до 39 лет. Соответственно, рекомендуется увеличить ставки для данной рекламной кампании в данном сегменте.

Помимо классических методов кластерного анализа в данной работе предлагается решить задачу поиска закономерностей на основе ассоциативных правил, позволяющих находить связи между переменными в исходном наборе данных на основе меры интенсивности.

Для реализации поиска закономерностей был применен алгоритм Apriori [9]. Гиперпараметры функции позволяют настраивать, насколько сильными должны быть связи между объектами для определения ассоциативных правил, а именно минимальный уровень поддержки, минимальное значение достоверности для правил, минимальный уровень зависимости, а также максимальное количество элементов в правиле. Были подобраны следующие оптимальные значения гиперпараметров: 0,003, 0,2, 4 и 2 соответственно.

Было проведено сравнение паттернов поведения для наборов данных с добавлением и без добавления признака *CPA*.

Оказалось, например, что если не принимать во внимание эффективности маркетинговой кампании, то товары и услуги, относящиеся к категории «916», наиболее востребована среди мужчин с интересом «15». При добавлении синтетического признака *CPA* в набор данных мы получаем, что достаточно низкое значение *CPA* характерно для рекламной кампании «1178» среди пользователей с интересом «16», а также среди пользователей возрастной категории «45-49».

Для внесения изменений в настройки рекламных кампаний критически важно понимать не только общие тенденции среди пользователей, но и то, как найденные закономерности связаны с их финансовой эффективностью, так как отдать предпочтение необходимо не наиболее активному, а самому прибыльному сегменту.

Таким образом, на основе кластерного анализа и поиска ассоциативных правил модель определяет паттерны поведения пользователей в разрезе их активности и совершения целевых действий с точки зрения предприятия. На основе полученных результатов можно определить новые правила распределения маркетингового бюджета и перенастроить маркетинговые кампании. После внесения изменений можно построить поиск закономерностей для построения процесса непрерывного улучшения онлайн-рекламы организации.

Таким образом, на основе кластерного анализа и поиска ассоциативных правил модель определяет паттерны поведения пользователей в разрезе их активности и совершения целевых действий с точки зрения предприятия.

На основе полученных результатов можно определить новые правила распределения маркетингового бюджета и перенастроить маркетинговые кампании. После внесения изменений можно построить поиск закономерностей для построения процесса непрерывного улучшения онлайн-рекламы организации.

2 Результаты и обсуждение

На основе полученных результатов можно определить новые правила распределения маркетингового бюджета и перенастроить маркетинговые кампании. После внесения изменений можно повторить процедуру с учетом новых данных и изменения поведения пользователей при новых настройках рекламы и построить процесс непрерывного улучшения онлайн-рекламы организации.

Если модель используется для оптимизации и модернизации интернет-рекламы, а именно повышения эффективности используемого бюджета или сокращение бюджета при сохранении текущих показателей рекламы, то при использовании метрики *CPA* важно четко определить, какое действие является целевым.

Например, *Google Analytics* позволяет настраивать аналитику по достижению целей пользователями. При настройке целей в системе процесс сбора и обработки данных может быть значительно упрощен.

Так, целевым действием для интернет-магазина часто является оформление заказа, компаниям также рекомендуется передавать в систему статус финального заказа, чтобы считать эффективность рекламной кампании относительно выкупленных заказов.

Для организаций, предлагающих услуги, например, медицинских учреждений, салонов красоты и т. п., целевым действием часто выступает запись на прием.

Для компаний, предоставляющих товары и услуги в B2B секторе, целевым действием может выступать скачивание материалов для клиентов с сайта или оформление заявки на обратный звонок. Так как на оформление финальной сделки влияют много факторов и выявить влияние рекламы на совершение сделки сложно, то более объективно в данном случае оценивать промежуточное действие.

Что касается интерпретации полученных результатов, то необходимо привлечения специалиста по маркетингу для определения гипотез для совершенствования рекламных кампаний. Рекомендуется ограничить число экспериментов, проводимых в один момент времени, т. е. предлагается выбрать наиболее эффективные или наиболее убыточные сегменты рекламных объявлений.

Далее необходимо выбрать стратегию по внесению изменений (например, отключение рекламного объявления или изменение целевой аудитории для данного объявления) и определить период времени, в течение которого будет проводиться эксперимент.

Чтобы получить наиболее объективные результаты по эффективности проверяемой гипотезы, рекомендуется придерживаться стратегии A/B-тестирования.

A/B тестирование – метод маркетингового исследования, в рамках которого контрольная группа элементов сравнивается с тестовым набором данных, в котором изменены какие-то параметры относительно контрольного набора, по определенным ключевым показателям, чтобы выявить наилучшую стратегию в рамках исследования.

Такой подход позволяет сравнить разные активности в один и тот же момент времени, что позволяет наиболее объективно проверить гипотезы с учетом различных факторов, влияющих на значение показателей, например, сезонность, доступность ассортимента и так далее.

Для проведения A/B тестирования в маркетинговых кампаниях можно прибегнуть к ручной настройке дублей рекламных объявлений с изменением параметров относительно исходных объявлений, по итогу проведения эксперимента необходимо оставить наиболее эффективный вариант.

После внесения финальных изменений в рекламу компании рекомендуется повторить работу алгоритма с учетом новых настроек на актуализированном наборе данных. Далее весь процесс необходимо повторить.

Для дальнейшего совершенствования работы алгоритма рекламные объявления можно также кластеризовать на основе машинного обучения, например, по изображению или тексту рекламного объявления, однако на начальном этапе рекомендуется объединять рекламные объявления вручную.

При использовании предложенной модели, помимо настройки сбора данных, необходимо уделить особое внимание синхронизации внесения изменений в настройки маркетинговых кампаний с результатами анализа модели, т. е. при экспорте данных необходимо указывать корректный период времени, на основе которого проводится анализ.

Другое решение данного вопроса заключается в добавлении дополнительного атрибута в набор данных, чтобы модель могла различать различные версии рекламных объявлений.

Предложенная модель может быть применена в разных организациях для решения задачи совершенствования управления взаимоотношений с клиентами, например, привлечения целевых пользователей на сайт. Более того, разработанные рекомендации позволяют на основе стандартных

инструментов веб-анализа строить отчеты, которые могут быть использованы в качестве данных для разработанной модели.

Литература

1. *Соловьев В.И.* Анализ данных в экономике: Теория вероятностей, прикладная статистика, обработка и визуализация данных в Microsoft Excel. М.: КНОРУС. 2018.
2. *Воронцов К.В.* Лекции по логическим алгоритмам классификации. URL : <http://www.machinelearning.ru/wiki/images/3/3e/Voron-ML-Logic.pdf> (дата обращения 05.05.2020).
3. *Freund Y., Schapire R.E.* A decision-theoretic generalization of on-line learning and an application to boosting // *Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence)*. Vol 904. 1995. P. 23-37.
4. *Breiman L.* Random Forests // *Machine Learning*. Vol. 45. 2001. P. 5-32.
5. McMahan H.B., Holt G., Sculley D., Young M., Ebner D., Grady J., Nie L., Phillips T., Davydov E., Golovin D., Chikkerur Sh., Liu D., Wattenberg M., Hrafinkelsson A.M., Boulos T., Kubica J. Ad click prediction: a view from the trenches // *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '13)*. New York, NY, USA: Association for Computing Machinery, 2013. P. 1222–1230.
6. *Chen T., Guestrin C.* XGBoost: A scalable tree boosting system // *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*. New York, NY, USA: Association for Computing Machinery, 2016. P. 785–794.
7. *Steinhaus H.* Sur la division des corps materiels en parties // *Bulletin de l'Académie. Polonaise des Sciences, C1. III*. Vol IV. 1956, P. 801—804.
8. *Dempster A.P., Laird N.M., Rubin, D.B.* Maximum likelihood from incomplete data via the EM algorithm // *Journal of the Royal Statistical Society, Series B*. Vol. 39. 1977, P. 1–38.
9. *Agrawal R., Srikant R.* Fast algorithms for mining association rules // *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB '94)*, Santiago, Chile, 1994. P. 487-499.