

МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ В ПРОЦЕДУРАХ АВТОМАТИЧЕСКОЙ ОБРАБОТКИ ТЕКСТА

Теория машинного обучения зародилась практически одновременно с появлением первых персональных компьютеров и с того момента является активно развивающейся дисциплиной. Ее развитие вызвано ростом мощностей современных компьютеров, увеличением доступных для анализа объемов данных, а также постоянным расширением области ее применения. Машинное обучение работает с объектами – элементарными единицами данных, которые характеризуются наблюдаемыми и скрытыми переменными. Главной задачей машинного обучения является автоматическое определение взаимозависимостей между наблюдаемыми и скрытыми переменными объекта, чтобы для произвольного объекта по его наблюдаемым компонентам можно было предсказать возможные значения скрытых компонентов.

Среди методов машинного обучения выделяют контролируемое и неконтролируемое обучение (т.е. обучение с учителем и обучение без учителя). Контролируемое обучение необходимо в тех случаях, когда некоторое свойство доступно для определенного массива данных (обучающей выборки), но на данный момент отсутствует и должно быть предсказано для других случаев. Каждый случай представляет собой пару «объект – ответ». К наиболее распространенным алгоритмам обучения с учителем относятся следующие способы формальной обработки данных. Во-первых, дерево принятия решений, т.е. средство поддержки принятия решений, которое использует древовидный граф или модель принятия решений с минимальным количеством вопросов «да/нет», ответив на которые, можно сделать верный выбор. Оно позволяет подойти к решению проблемы со структурированной и систематической стороны, чтобы в итоге прийти к логическому выводу. Во-вторых, нужно отметить наивный байесовский классификатор, основанный на применении теоремы Байеса со строгими (наивными) предположениями о независимости функций. Можно выделить следующие сферы применения теоремы Байеса: определение спама в электронной почте; сегментация новостных статей по их тематике; определение эмоциональных (оценочных) составляющих массива текстов.

Неконтролируемое обучение (обучение без учителя) используется для обнаружения неявных отношений в немаркированном наборе данных. В этом случае ответы не задаются, и необходимо искать зависимости между объектами. Основная задача неконтролируемого обучения сводится к кластеризации. Она заключается в том, чтобы сгруппировать объекты в кластеры, используя данные об их попарном сходстве, чтобы поместить максимально похожие между собой элементы в одну группу (кластер). Алгоритмов

кластеризации довольно много. Самые популярные представлены в виде нейронных сетей, которые активно используются для глубокого обучения на основе сотен тысяч и миллионов объектов. Обученные таким образом сети не просто позволяют моделировать сложные объекты (например, тексты или изображения), но и генерируют в процессе обучения информативные признаки описания, которые могут быть использованы другими, более простыми алгоритмами машинного обучения в качестве наблюдаемых переменных объекта.

Д. Киселева

ОБ ОСОБЕННОСТЯХ НАТИВНОЙ РЕКЛАМЫ

Интернет – это идеальный рынок для продвижения товаров и услуг. Он не имеет территориальных и временных ограничений, позволяет собирать и использовать информацию о пользователях, предлагает большое число разнообразных площадок, предоставляет гибкость в ведении рекламной кампании благодаря сервисам и технологиям, которые позволяют реализовать коммуникацию с целевой аудиторией, влиять на результат проведения рекламной кампании и менять ее ход в любой момент времени.

Разновидностью интернет-рекламы является контекстная реклама – вид сетевой рекламы, которая демонстрирует текстовые рекламные объявления или баннеры в поисковых системах и на сайтах-партнерах рекламной сети поисковой системы. Относительно новым типом контекстной рекламы можно считать нативную (естественную) рекламу, по форме и содержанию соответствующую веб-странице, на которой она размещена. К основным форматам нативной рекламы относятся, во-первых, реклама в поисковых системах – рекламное сообщение, содержание которого соответствует поисковому запросу пользователя. Подобное сообщение выдается со специальной пометкой, что это реклама. Во-вторых, в качестве нативной рекламы можно рассматривать рекламную публикацию в новостной ленте – спонсорский пост в социальной сети, который отображается в новостной ленте пользователя и отвечает его интересам и нуждам (т.е. является таргетированным). Такие публикации отмечаются цветом или специальной пометкой. В-третьих, нативной рекламой считается рекомендательный блок – блок со ссылками на другие статьи, среди которых размещена ссылка на рекламную статью. Блок формируется алгоритмами на основе интересов пользователей и помещается внизу страницы. В-четвертых, к нативной рекламе относится спонсорская статья, написанная в партнерстве с рекламодателем. Данный тип контента должен соответствовать материалам сайта и интересам целевой аудитории. Разновидностью спонсорских статей является премиальная нативная реклама, когда интернет-площадка публикует у себя рекламные материалы, которые привлекают внимание пользователей. Размещение