

изображения отпечатка и выделения особых точек.

Так фильтр Габора показал себя очень хорошо, но его «собрат», основанный на коэффициентах отражения, при тех же данных, дает почти такой же коэффициент качества при своей скорости и легкости реализации.

Что касается бинаризации, то пороговая бинаризация уступает адаптивной по качеству, а скорость и сложность этих алгоритмов отличается незначительно.

Исходя из полученных в ходе исследовательской работы результатов, следует следующие: наиболее оптимальным по качеству, скорости и сложности является алгоритм обработки и анализа отпечатка пальца, основанный на фильтрации с использованием коэффициента отражения и на адаптивной бинаризации.

Список использованных источников:

1. В. В. Жуковский, С. В. Сай, Способ улучшения качества изображения отпечатка пальца, 2009
2. Zin Mar Win, Myint Myint Sein, An Efficient Fingerprint Matching System for Low Quality Images, 2011
3. Rishabh Mishra, Prashant Trivedi, Student Attendance System Based On Fingerprint Recognition and One-to-Many Matching, 2011

ИСПОЛЬЗОВАНИЕ ОБУЧАЮЩИХСЯ АЛГОРИТМОВ ДЛЯ АНАЛИЗА СПРОСА НА ПРОДУКЦИЮ И ОРГАНИЗАЦИИ ЦЕПОЧЕК ПОСТАВЩИКОВ

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Лось Л. А.

Волорова Н. А. – канд. техн. наук, доцент

В последнее время в условиях постоянных изменений в экономической сфере особое значение придается изучению механизмов рынка, мгновенному реагированию на изменение спроса, регулированию цен, экономии временных и финансовых ресурсов различными способами. Нас будут интересовать способы с применением достижений сферы информационных технологий.

Машинное обучение – направление, формально относящееся к искусственному интеллекту, но также включающее в себя использование разработок математической статистики, теории вероятности, методов оптимизации. Это направление предполагает построение алгоритмов, способных обучаться на предоставленных данных. Т. е. при получении данных алгоритм исследовании закономерностей (обучение) и после этого позволяет сделать предсказания о данных, которые могут быть получены в будущем. Алгоритмов машинного обучения в настоящее время разработано большое количество, для разных сфер и разных задач они различаются, имеют свои плюсы и минусы.

Исходя из определения алгоритмов машинного обучения, к ним можно отнести корреляционный и регрессионный анализ. Эти подходы мы будем использовать для прогнозирования спроса на основе информации о совершенных ранее покупках.

Предположим, что у нас есть интернет-магазин с сохраненными в базе данных операциями покупок, просмотров товаров, добавлений товаров в список желаний, информацией о пользователях и т. д. С помощью *корреляционного анализа* мы можем отобрать те факторы, которые влияют на спрос в большей степени (более взаимосвязаны со спросом). Для этого достаточно рассчитать коэффициент корреляции, он может принимать значения от -1 до 1. При коэффициенте корреляции равном 0 мы можем говорить об отсутствии взаимосвязи между выбранными параметрами. Чем ближе абсолютное значение коэффициента к 1, тем о большей степени зависимости мы можем говорить. После определения значимых факторов мы можем приступить к прогнозированию, используя для этого *регрессионный анализ*. Он позволит нам предсказать значения зависимой переменной (величины спроса) с помощью независимых переменных (факторов), определить, какой вклад принесет изменение независимых переменных.

С развитием информационных технологий стало возможным не только наблюдать и прогнозировать колебания спроса, но и оказывать на него влияние, провоцировать увеличение спроса на конкретную продукцию. В этой сфере нам могут пригодиться многие алгоритмы машинного обучения.

Сбор предпочтений пользователей на данный момент осуществляется во многих интернет магазинах, новостных сайтах, сайтах с медиа контентом и д. р. для рекомендации вам товаров, фильмов, музыки, новостей, ссылок на другие сайты и многого другого. Рассмотрим некоторые подходы.

Коллаборативная фильтрация работает следующим образом: рассматриваются большие объемы данных, например, о покупателях. Находятся более мелкие группы со сходными покупками, которые объединяются в один список и сортируются по какому-либо рейтингу (по количеству выбравших этот товар людей). Затем эти товары, за исключением уже приобретенных, рекомендуются покупателю. Для определения схожести каждый покупатель сравнивается со всеми другими и находится коэффициент подобия. Этот

коэффициент может быть рассчитан разными путями, выбирать которые следует в каждом конкретном случае, например, определение *евклидова расстояния*, нахождение *коэффициента корреляции Пирсона*, *коэффициента Жаккарда*, установление *манхэттенского расстояния* и т. д.

В приложениях, обрабатывающих большие объемы данных, часто используется *кластеризация* – метод обнаружения и визуализации групп, связанных между собой предметом, людей. Алгоритм *иерархической кластеризации* на каждом шаге объединяет две самые похожие группы на основе самого близкого расстояния между группами (чем ближе они друг к другу, тем более похожи). Алгоритм продолжает работу до тех пор, пока не останется только одна группа. Результаты такого прохода по данным можно представить в виде графа, на котором видно из чего состоит кластер и расстояния между группами. Минусами алгоритма является то, что древовидная структура не разбивает данные на группы, и на больших наборах этот алгоритм будет работать медленно, так как приходится попарно вычислять соотношения между группами. Эти недостатки не являются при использовании *кластеризации методом К-средних*. В отличие от предыдущего решения, этот алгоритм начинается с выбора *к* случайно расположенных точек, которые станут центрами кластеров. Каждому элементу назначается ближайший к нему центр, и центр переносится в новую точку, вычисленную как среднее по всем его элементам. Цикл продолжается до тех пор, пока центры не перестанут передвигаться. Алгоритмы кластеризации удобно использовать для определения групп сходных товаров предлагаемых покупателю для сравнения.

Цепочка поставок – это совокупность организаций, людей, технологий, информационных и финансовых потоков, в процессе прохождения продукта от поставщика к потребителю. Управление цепочкой поставок включает несколько стадий. В проекции на интернет-магазин они могут выглядеть следующим образом:

- планирование – разработка планов соответствия товаров ожиданиям и потребностям потребителей (будем считать, что методы и алгоритмы, изложенные в предыдущей части, помогли определить выбор продукции и ее объемы);
- выбор поставщика – выбор поставщиков сырья/готового продукта, планирование способов доставки, оплаты, размещения и т.д.;
- производство – допустим, что этот этап для интернет-магазина, реализующего готовую продукцию не актуален;
- поставки – создание сетей складов, организация приема заявок от клиентов, транспортировка и т.д.;
- возврат – организация возврата дефектных товаров организация технического обслуживания.

Организовать цепь поставок с такими этапами можно множеством способов, перебор такого количества возможных вариантов вручную займет не один день. Поэтому в данной ситуации весьма логично будет использование *методов оптимизации*. Такие методы используются в случаях с множеством возможных решений, зависящих от множества переменных. Оптимизация сводится к поиску наилучшего решения путем подбора различных сочетаний переменных и их сравнения между собой. Для решения любой оптимизационной задачи необходимо правильно подобрать целевую функцию. После ее выбора задача сводится к подбору таких входных параметров, которые будут минимизировать значение этой функции.

В нашем случае входными переменными могут быть цены поставщиков на товары, стоимости доставок товаров на склады магазина, время доставки, стоимость аренды складов, состав комплектаций поставок, стоимость транспортировки товара покупателю, время ожидания товара покупателем. Таких факторов можно подобрать множество, и не все их сразу выражены в денежном эквиваленте. Так придется выбрать учитывать ли время ожидания товара или оценивать только стоимость его доставки. Для таких факторов можно придумать их выражение в денежном эквиваленте. Определившись с переменными, влияющими на стоимость нужно составить функцию, в результате которой мы получим одно число (например, общую стоимость всех этапов), - это и будет наша *целевая функция*.

Для решения этой задачи могут использоваться разные алгоритмы оптимизации. *Алгоритм случайного поиска* (генерируется конечное число гипотез и для них значение целевой функции, возвращается наилучшая гипотеза) не гарантирует нахождение оптимального решения.

Алгоритм спуска с горы начинается со случайного решения и ищет лучшие решения возле него. Возможна ситуация нахождения локального минимума, который может не совпадать с глобальным минимумом (оптимальным решением).

Алгоритмы имитации отжига пытаются бороться с проблемой скатывания в локальный минимум. Они включают возможность перехода к худшему решению, чтобы найти в итоге лучшее.

Генетические алгоритмы состоят в том, чтобы создать набор случайных решений (популяцию). Значение целевой функции вычисляется для всех выбранных гипотез на каждом шаге, затем формируется новая популяция. В нее входят лучшие решения из текущей и новые (полученные из лучших с помощью модификации). Итерации повторяются заданное число раз или до прекращения улучшения целевой функции.

От условий задачи зависит, будет ли работать конкретный алгоритм. Так большинство из них основаны на том, что оптимальное решение находится близко к другим хорошим решениям, и если это условие не выполняется, любой из методов оптимизации будет работать не лучше, чем случайный поиск.

Использование алгоритмов машинного обучения во многих сферах дает значительный выигрыш во времени, потраченном на нахождение закономерностей в больших объемах данных, позволяет делать на них прогнозы. Что в свою очередь сокращает расходы, улучшает функционирование компаний и предоставляет преимущество перед конкурентами.

Список использованных источников:

1. Сегаран. Т. Программируем коллективный разум / Сегаран. Т. // Пер. с англ. – СПб: Символ-Плюс, 2008. – 368 с., ил.
2. Введение в Управление цепочками поставок (SCM) [Электронный ресурс] – Режим доступа. – URL

<http://www.galaktika.ru/amm/introduction-to-scm.html> (дата обращения 04.04.2016).

3. Построение алгоритма отбора факторов, влияющих на конъюнктуру товарного рынка [Электронный ресурс] – Режим доступа. – URL <http://dis.ru/library/531/26996/> (дата обращения 04.04.2016).

ВЕРОЯТНОСТНАЯ МОДЕЛЬ РЕЙТИНГОВ ИГРОКОВ В КОМАНДНЫХ ВИДАХ СПОРТА

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Михолап А. А., Стержанов М. В.

Рассмотрена баесова модель оценки уровня игры спортивных команд и отдельных участников. Приведен способ уменьшения вычислительной сложности модели с помощью фактор-графа. Предложен способ повышения предсказательной точности путем агрегации результатов различных систем.

Ключевые слова: системы рейтингов, баесовы методы, вероятностные модели.

Задача ранжирования и построения рейтингов возникает в различных ситуациях и особенно часто в контексте спортивных соревнований. Объективная оценка навыков игроков и команд помогает принимать решения при покупке новых игроков, построении списка лучших команд и прогнозировании исходов будущих матчей.

В 2006 году корпорация Microsoft опубликовала статью с описанием новой рейтинговой системы TrueSkill. Она была разработана коллективом авторов для подбора соперников в многопользовательской онлайн игре Halo 2 и, согласно мнению авторов, может рассматриваться как обобщение Elo.

Недостаток Elo заключается в дискретных уровнях коэффициента скорости изменения рейтингов K . Они были подобраны вручную, на основании истории игр и здравого смысла. Однако, желательно, чтобы система могла подстраивать K автоматически. Например, если новичок играет со старым игроком, уровень игры которого стабилен и достаточно точно известен, то результат этого матча намного полезнее, чем матча против другого новичка, о котором ничего не известно. Чтобы сделать это очевидным, можно упростить и принять уровень навыка опытных игроков за точку. Тогда, сравнивая новичка с ними, мы получим для него довольно точные координаты. Если же случайная величина рейтинга противника также имеет широкий интервал, мы не можем сделать таких выводов, так как оценка очень неопределенная, а сами интервалы даже могут перекрываться.

TrueSkill помимо среднего значения μ варьирует также и дисперсию σ для каждого игрока, благодаря чему может автоматически сжимать и увеличивать интервал неопределенности без ручного вмешательства, основываясь лишь на истории игр.

Второй предпосылкой создания нового алгоритма была необходимость моделирования ничьих в явном виде. Вероятность ничьей это один из критериев качества разбиения игроков на команды в Halo 2 и других играх и видах спорта. Матч наиболее интересен, когда силы обеих сторон равны и исход непредсказуем.

Также в систему встроена детализация на уровне отдельных игроков. Чтобы автоматизировать управление случайными величинами приходится неизбежно усложнять алгоритм. Если в Elo использовалась базовая теория вероятности, логистическая функция и простое правило обновления, в TrueSkill за счет введения переменной дисперсии применяются значительно более сложные конструкции.

С начала двухтысячных годов в статистике начали набирать значительную популярность Баесовы методы. Они представляют собой набор правил работы со случайными величинами, позволяя упростить идеальную вероятностную модель до практически применимого уровня. Это возможно за счет ручного определения значимых факторов (которые описываются случайными величинами) и связей между ними.

TrueSkill использует именно Баесов подход, выражая модель встречи двух команд фактор графом (рисунок 1).

Система инициализируется определенными значениями, называемыми априорной информацией. В нее можно включить всевозможные сведения от экспертов, но в данном случае это просто фиксированный стартовый рейтинг. Важно различать априорную информацию в контексте всей системы и одного матча.

До игры осуществляется прямой проход по фактор графу сверху (от узлов, содержащих рейтинг игроков) вниз (к узлам, обозначающим исход матча). С его помощью оцениваются вероятности различных исходов. После завершения игры случайные величины в нижних узлах заменяются на реальные числовые