

ОБРАБОТКА БОЛЬШИХ ОБЪЕМОВ ДАННЫХ В АВТОМАТИЗИРОВАННЫХ СИСТЕМАХ УПРАВЛЕНИЯ АВТОМОБИЛЬНЫМ ПОТОКОМ

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

Швец В.И., Хадасевич А.И.

Теслюк В.Н. – к.ф.-м.н., доцент

Ввиду постоянного роста автомобилизации на дорогах применяются новейшие технологии, призванные повысить уровень безопасности. Одно из приоритетных направлений – развитие автоматизированных систем управления дорожным движением, способных в режиме реального или близком к реальному времени считывать постоянно изменяющуюся обстановку на дорогах, оценивать её по заданным параметрам (интенсивность, средняя скорость, заполненность полосы или дороги и т.д.), распознавать возникающие нештатные ситуации (заторы, сложные метеорологические условия, дорожно-транспортные происшествия), а затем принимать меры по регулированию потока, недопущению аварийных ситуаций или минимизации их негативных последствий.

Автоматизированная система управления автомобильным дорожным движением представляет собой комплекс программно-аппаратных средств, способных в режиме реального или близком к реальному времени считывать постоянно изменяющуюся обстановку на дорогах, оценивать её по заданным параметрам (интенсивность, средняя скорость, заполненность полосы или дороги и т.д.), распознавать возникающие нештатные ситуации (заторы, сложные метеорологические условия, дорожно-транспортные происшествия), а затем принимать меры по регулированию потока, недопущению аварийных ситуаций или минимизации их негативных последствий.

Сбор данных о дорожной обстановке осуществляется с помощью специально разработанной аппаратуры: камер скорости, видеокамер, дорожных метеостанций, детекторов транспортного потока, датчиков и радаров. Регулирование потока выполняется с использованием электронных информационных панелей, указателей и табло, светофорных объектов, шлагбаумов. Вся вышеперечисленная аппаратура располагается непосредственно на участках дорог, обмен информацией между ней и центром управления дорожным движением осуществляется по проводным и беспроводным каналам связи.

Наиболее сложным процессом является автоматизация анализа данных и принятия решений. Ещё в недалёком прошлом данные, получаемые в центре управления дорожным движением, обрабатывались операторами - специально обученными людьми. Современные же системы, применяемые в крупных городах, могут насчитывать десятки и сотни тысяч объектов взаимодействия с водителями автомобилей. Данные от них поступают регулярно через короткие промежутки времени, и обработать их человеку или даже группе людей не под силу. В результате чего необходимо применять алгоритмы обработки больших объёмов данных, способные за приемлемое время выдавать наиболее оптимальный результат. В качестве примера можно привести ситуацию, когда в случае распознавания затора система с помощью светофорных объектов и информационных табло перенаправляет водителей на соседние улицы, если те свободны. Другой пример – временное ограничение максимально допустимой скорости движения в сложных погодных условиях (снегопад, гололёд).

При использовании автоматизированных систем управления нельзя не учитывать специфику конкретного города и его районов, зависимость транспортных потоков от времени суток и периода года. Ввод в систему ограничений и параметров требует длительного наблюдения за дорожной обстановкой со стороны человека. Более предпочтительным в данном вопросе видится применение алгоритмов машинного обучения. Система с течением времени сама накапливает данные, производит статистический анализ и предлагает оператору системы наиболее оптимальные варианты поведения для конкретного участка дороги в распространённых ситуациях. Например, если в зимний период года система фиксирует снижение средней скорости движения, в ручном или автоматическом режиме могут быть изменены интервалы светофорного регулирования. Необходимость принятия решения оператором в таких ситуациях сводится к минимуму.

Следует отметить, что алгоритмы обработки больших объёмов данных могут быть применены не только для анализа дорожной обстановки в центрах обработки данных. К вышеуказанной категории также относятся алгоритмы распознавания изображений. В результате распознавания картинки с дорожных камер можно определить, например, тип транспортного средства, автомобильный номер. Алгоритмы анализа изображений применяются также для распознавания правонарушений: движение с непристёгнутым ремнём безопасности, выезд за стоп-линию и т.д. Эта информация при необходимости может быть передана взаимодействующим компьютерным системам.

Для поддержания высокой скорости обработки данных, стабильности и бесперебойности работы подобные системы создаются распределёнными, т.е. работают на кластерах из сотен и тысяч узлов. Обрабатываемые данные дублируются и в случае отказа одного узла/кластера не теряются, процесс обработки не замедляется. Горизонтальная масштабируемость систем позволяет использовать необходимое количество недорогих узлов для увеличения производительности, не прибегая к покупке мощных серверов и дорогих сетей хранения данных. Классическим примером фреймворка для распределённого хранения и обработки данных является проект ApacheHadoop.

Хранение огромного массива данных в таких системах требует комплексного использования как классических SQL баз данных, так и СУБД с подходом NoSQL. Для решения определённых задач с высокой скоростью доступа может быть полезным использование графовых баз данных.

Таким образом, применение методов хранения и алгоритмов обработки данных больших объёмов в современных автоматизированных системах управления дорожным движением является не просто необходимым, но и закономерным шагом для их эволюции. Благодаря этому поддерживается высокая скорость доступа и обработки данных, получение результатов в режиме реального времени, автоматическая адаптация под изменяющиеся условия на базе алгоритмов машинного обучения, преобразование форматов данных с помощью алгоритмов распознавания изображений. Ко всему этому добавляется высокая отказоустойчивость системы в результате распределённой обработки данных.

Список использованных источников:

1. Развитие автоматизированной системы управления дорожным движением Минска как части интеллектуальной транспортной системы города. [Электронный ресурс]. – Электронные данные. – Режим доступа: https://rep.bntu.by/bitstream/handle/data/27593/Razvitie_avtomatizirovannoj_sistemy_upravleniya_dorozhnym_dvizheniem_Minska.pdf
2. Автоматизированная система управления дорожным движением «АГАТ» [Электронный ресурс]. – Электронные данные. – Режим доступа: <https://agat.by/wp-content/uploads/2015/05/АСУДД-АГАТ-презентация.pdf>
3. АСУДД: Что висит над дорогой? [Электронный ресурс]. – Электронные данные. – Режим доступа: <https://habr.com/post/124249/>

ПОСТРОЕНИЕ РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЫ НА ОСНОВЕ АЛГОРИТМОВ ОБРАБОТКИ ЕСТЕСТВЕННОГО ЯЗЫКА

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

Хопова А.А., Волорова Н.А.

Волорова Н.А. – к.т.н., доцент

В данном докладе рассматриваются методы тематического моделирования текстов, а также оценки качества получаемых результатов. При этом в качестве исходных данных используются тексты микроблогов, которые существенно отличаются от традиционных текстов книг, статей и пр. Так же рассматриваются система рекомендаций на основе схожести контента и с учётом реакций других пользователей (лайки, репосты и т.д.). Такая система является более эффективной, чем традиционный подход (фильтрация), благодаря использованию дополнительных метрик при формировании рекомендации. Применение такой системы позволит пользователям находить релевантные материалы, хранящиеся в социальных сетях.

Введение

В наше время количество информации постоянно растёт, большая её часть представляет собой неупорядоченные текстовые данные, например, различные WEB-ресурсы, социальные сети, блоги, форумы, новостные сайты и тому подобное. Человеку сложно самостоятельно их обрабатывать. Более того, ручной анализ неэффективен для больших объёмов текста, так как он ограничен скоростью, погрешностями и ошибками, обусловленными человеческим фактором. Следовательно, требуются методы, способные автоматически обрабатывать такие данные. Методы могут основываться на извлечении определенной информации, например, списках тем, содержащихся в текстах.

Для того чтобы выделить из текста основные темы, человеку достаточно его прочитать. В условиях постоянно увеличивающегося количества информации, в частности, текстовой (так называемый, информационный бум), приходится анализировать данные такого объема, которые человек не в силах обработать. Поэтому необходимы методы, позволяющие автоматически извлекать темы из большого набора данных. Теоретически обоснованным и активно развивающимся направлением в анализе текстов на естественном языке, является тематическое моделирование коллекций текстовых документов.

Построение тематической модели может рассматриваться как задача одновременной кластеризации документов и слов по одному и тому же множеству кластеров, называемых темами. В терминах кластерного анализа тема – это результат би-кластеризации, то есть одновременной кластеризации и слов, и документов по их семантической близости. Как правило, выполняется нечёткая кластеризация, то есть документ может принадлежать нескольким темам в различной степени. Таким образом, сжатое семантическое описание слова или документа можно представить в виде распределения на множестве тем. Процесс нахождения этих распределений и называется тематическим моделированием.

В 2003 году Д.Блей предложил модель скрытого размещения Дирихле (Latent Dirichlet Allocation, LDA). Это одна из первых и широко используемых вероятностных тематических моделей. Основной идеей таких моделей является наличие генеративного процесса – процесса, порождающего документы с использованием предопределённых тем. Задача заключается в том, чтобы подобрать темы таким образом, чтобы вероятность сгенерировать данный набор документов была максимальной.

В вероятностных тематических моделях темы представляются в виде распределений над словами. Оценить качество полученных тем можно вручную: можно выбрать слова с наибольшей вероятностью и понять, что они вместе означают. При большом количестве тем требуется много времени, чтобы оценить, насколько понятными для человека они получились.

Алгоритмы поиска наиболее правдоподобных скрытых параметров делятся на две категории: на основе сэмплирования и вариационные методы. Алгоритмы первой группы пытаются собрать конечную выборку переменных, на которой ищется максимум.