

«Цифровая экономика». - №5. - 2021. Электронный ресурс.
<https://www.digitaleconomy.ru/journal/modern-approaches-to-the-development-of-educational-applications-on-react>

4. Улучшение пользовательского опыта в сфере образования: стратегии успеха стартапа// Электронный ресурс.<http://surl.li/muddgd>

5. 20 лучших сайтов для изучения программирования//
<https://ultahost.com/blog/ru/20-lutsih-saitov-viuciti-progromirovanie/>

УДК 004.522

Ш.Ю. Тедженов

Институт телекоммуникаций и информатики Туркменистана
Ашхабад, Туркменистан

АНАЛИЗ ПОТОКА ДАННЫХ

***Аннотация.** В связи с распространенным использованием датчиков и сетей инструменты мониторинга, «большие данные» сегодня перемещаются по конвейерам обработки данных предприятия в потоковом режиме мода. Хотя некоторые компании предпочитают размещать свои данные инфраструктура обработки и услуги в виде частных облаков, другие полностью передают эти услуги публичным облакам.*

***Ключевые слова:** потоки данных, обработка сложных событий, ассоциация майнинг правил.*

Sh.Y. Tedjenov

Institute of Telecommunications and Informatics of Turkmenistan,
Ashgabat

DATA STREAM ANALYSIS

***Abstract.** Due to prevalent use of sensors and network monitoring tools, big volumes of data or “big data” today traverse the enterprise data processing pipelines in a streaming fashion. While some companies prefer to deploy their data processing infrastructures and services as private clouds, others completely outsource these services to public clouds.*

***Keywords:** Data streams, Complex Event Processing, Association Rule Mining.*

Мы живем в эпоху, когда тенденции не держатся долго. Поэтому временные аспекты рекомендаций чрезвычайно важны. К сожалению, когда объемы потоковых данных и правила вывода увеличиваются, аналитики данных реагируют на увеличение значений поддержки и

достоверности, чтобы получить меньше правил с более сильным подъемом. Тем не менее, правила, которые демонстрируют высокие значения подъема в течение длительного периода времени (например, месяца), возможно, уже устареют к концу этого периода. Например, продажа мороженого, прохладительных напитков и пластиковых стаканчиков будет чрезвычайно популярна в самый жаркий месяц в году, так как мешки с песком и лопаты востребованы во время урагана. После того, как тренд исчезнет, возможностей для продаж не будет. Эти временные закономерности возникают еще быстрее в случае онлайн-продаж или на фондовых рынках, где каждую секунду происходят миллионы транзакций. Поточковый анализ правил может выявить такую тенденцию, как «когда акции HPQ и MSFT падают более чем на 1%, DELL следует за ними» в течение одного часа или дня. Предположим, что недавняя тенденция (т. е. правило с минимальной поддержкой и доверием) возникает только через определенный период времени. Его доверительное значение может не соответствовать самый высокий в мире, но его ценность для местного бизнеса может быть довольно высокой. Предлагаемые нами системы разработаны с учетом этих правил. Другие приложения потокового анализа включают кластеризацию и классификацию потоков [7].

Подумайте о переворачивающемся окне, которое настолько велико, что может охватить все данные, используемые для автономного анализа. В этом случае оффлайн-анализ и онлайн-анализ с одним большим окном дадут одинаковые результаты набора правил. В системах CEP и меняющихся окнах достаточно одного параметра (т.е. размера окна), чтобы переключиться с автономного анализа на анализ в режиме, близком к реальному. Если от анализа скользящего временного окна ожидаются дополнительные выгоды, переключение типа окна опять же будет незначительным усилением.

С. Реализация запроса ARM

Мы получили Java-реализации алгоритмов Apriori и FP-Growth из библиотеки правил ассоциации известного инструмента машинного обучения Weka [12] и интегрировали эти алгоритмы в движок Esper, который также основан на Java. Для добавления новых операторов необходимо реализовать специальную функцию агрегирования в Esper (класс AggregationSupport). Мы реализовали этот интерфейс, чтобы добавить алгоритмы непосредственно в Esper для анализа правил потока:

ЗАПРОС 1:

```
SELECT Apriori(parameters, table.feature1, table.feature2)
```

FROM event.win:length(5) AS table

Параметры, которые мы использовали для инициализации алгоритмов Weka Apriori внутри Esper, были следующими:

```
'-N 10 -T 0 -C 0,9 -D 0,05 -U 1,0 -M 0,1 -S -1,0 -c -1'
```

где N — количество выводимых правил, T — тип метрики, по которой ранжируются правила (0 = уверенность | 1 = рост | 2 = рычаг | 3 = убежденность), C — минимальная оценка метрики (например, минимальная достоверность = 0,9). правила, U/M — верхняя/нижняя границы минимальной поддержки (по умолчанию = 1,0 и = 0,1), D — дельта, на которую уменьшается минимальная поддержка на каждой итерации (по умолчанию = 0,05), S — уровень значимости. , а c — индекс класса (по умолчанию = последний).

Можно присвоить равные значения границам U/M (например, 0,3), чтобы избежать итераций, что было бы правильным выбором для сред потоковой обработки. Однако в этом случае пользователь должен хорошо знать домен и правильно установить значения, чтобы найти нужное количество правил для каждого временного окна. В средах динамической потоковой передачи фиксированные ручные настройки могут привести к тому, что будет извлечено слишком много или слишком мало правил. Поэтому мы предпочли, чтобы эту динамическую настройку выполняла система Weka. Выбранные функции описаны в части D.

ЗАПРОС2:

```
SELECT FPgrowth(parameter, table.feature1, table.featt2)
```

```
FROM event.win:length(5) AS table
```

Параметры, которые мы использовали для инициализации алгоритмов Weka FP-Growth внутри Esper, были следующими:

```
'-P 2 -I -5 -N 10 -T 0 -C 0,9 -D 0,05 -U 1,0 -M 0,7'
```

где P — индекс атрибута для двоичных атрибутов в обычных плотных экземплярах (используется индекс по умолчанию 2 для разреженных экземпляров), I — максимальное количество элементов, включаемых в большие наборы элементов (и правила) (по умолчанию = -1, т. е. без ограничений). .), N — необходимое количество выходных правил, T — тип метрики, по которому ранжируются правила (0 = достоверность), C — минимальная оценка метрики (например, минимальная достоверность = 0,9) правила, D — дельта, по которой минимальная поддержка уменьшается на каждой итерации (по умолчанию = 0,05), U/M — верхняя/нижняя границы минимальной поддержки (по умолчанию = 1,0 и = 0,1).

D. Набор данных LastFM и предварительная обработка

Данные LastFM содержат информацию примерно о 1000 людях

(набор данных Lastfm-1K) [10], которые слушают песни в базах данных LastFM. В этом наборе данных размером около 3 ГБ около 75 000 уникальных исполнителей, несколько сотен тысяч уникальных песен и миллионы транзакций. Кратко, поля включают в себя <идентификатор пользователя, временную метку, mbid исполнителя, имя исполнителя, mbid песни, название песни>. На этапе предварительной обработки мы сначала очистили записи с отсутствующей информацией об исполнителе и удалили поля временных песен, которые не способствовали извлечению правил. Этот процесс выполнялся в автономном режиме, и наша будущая работа включает в себя предварительную онлайн-обработку. Мы использовали скользящие окна на основе счетчика. Наконец, у нас были две функции набора данных (<user-id, Artist-mbid>). Поскольку алгоритм Apriori использует большой объем памяти, мы дополнительно обрезали данные, включив в них пользователей, которые прослушивали более 100 песен, и песни, которые в целом были прослушаны более 3000 раз. В результате 967 уникальных пользователей прослушали 1105 уникальных исполнителей.

Е. Результаты деятельности

Результаты офлайн-анализа, приведенные в таблице 2, показывают, что рост FP генерирует результаты «Топ-10 правил» в 75–613 раз быстрее, чем алгоритм Apriori в потоке данных. Это соответствует большинству предыдущих работ [9], поскольку FP-рост позволяет избежать итеративных поколений кандидатов, рассчитанных с помощью Apriori.

	[i:967 a:1105]	[i:1105 a:967]
Apriori	61.403s	226.502s
FPGrowth	0.811s	0.369s

(i: экземпляры, a: атрибуты)

Таблица 2 - Результаты офлайн-анализа

Онлайн-анализ проводился в небольших окнах с скользящим подсчетом пользователей и исполнителей размером 10x10. На рисунке 5 показаны динамические изменения поддержки Weka (minSupport, minConfidence) и соответствующее количество правил, генерируемых в каждом интервале для набора данных LastFM. Ось X для графиков скользящего окна подсчета увеличивается на 1 при каждом 1 отсчете события, тогда как переворачивающиеся окна увеличиваются на 1 при каждом слайде, который перемещает окно на 10 событий. Следовательно, для получения одного и того же временного региона между двумя графиками требуется сопоставление 10 к 1 (например,

88□8 или 9). Мы видим, что операция обновления динамической поддержки Weka (U/M) работает правильно и генерирует наборы правил TOP10 для каждого периода. Мы также обнаружили, что переворачивающиеся окна практически генерируют кумулятивную функцию распределения (CDF) или «агрегированный» набор правил для правил, найденных с помощью скользящих окон. Тем не менее, наиболее важным выводом из этих результатов является то, что в анализе скользящего окна существуют временные наборы правил, которые упускаются из виду онлайн-анализом скользящего окна из-за агрегирования. Наборы правил были рассчитаны примерно за 300–500 мс для каждого интервала, как показано на рисунке 7 для обоих типов окон. Для краткости мы пропускаем результаты с окнами большего размера.

Список использованных источников

1. D. Abadi, D. Carney, U. Cetintemel, M. Cherniack, C. Convey, S. Lee, M. Stonebraker, N. Tatbul, and S. Zdonik. Aurora: A new model and architecture for data stream management. VLDB Journal, 12(2):120–139, August 2003.

C. Giannella, J. Han, J. Pei, X. Yan, P. S. Yu; Mining Frequent Patterns in Data Streams at Multiple Time Granularities; Data Mining: Next Generation Challenges and Future Directions, AAAI/MIT; 2003.

УДК 338.2

А.В. Шеметовец

Белорусский национальный технический университет
Минск, Беларусь

РАЗВИТИЕ ИМПОРТОЗАМЕЩАЮЩИХ ПРОИЗВОДСТВ В РАМКАХ ЭКОНОМИЧЕСКОЙ БЕЗОПАСНОСТИ РЕСПУБЛИКИ БЕЛАРУСЬ

Аннотация. В статье было рассмотрено понятие импортозамещения и указана его важность в современных реалиях, был рассмотрен перечень импортозамещающих товаров в Республике Беларусь. Импортозамещение позволяет укрепить экономическую независимость страны и минимизировать зависимость от внешних поставок.