

Министерство образования Республики Беларусь
Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК 004.021

Дараган Евгений Николаевич

**МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ СИСТЕМ
ЭЛЕКТРОННОЙ КОММЕРЦИИ**

АВТОРЕФЕРАТ

на соискание степени магистра информатики и вычислительной техники
по специальности 1-40 81 04 «Обработка больших объемов
информации»

Научный руководитель
Волорова Н. А.
к.т.н., доцент

Минск 2019

Введение

Одной из метрик, позволяющей справедливо и беспристрастно оценить спрос на тот или иной вид товаров или услуг, является объем продаж (фактов оказания услуги), как полных (за все время существования товара или услуги) или недавних (за последние N дней). Но зачастую такая информация оказывается коммерческой тайной, так как позволяет с легкостью оценить основные характеристики бизнеса, такие как товароборот и прибыль, и поэтому практически не раскрывается в системах электронной коммерции, и тем более в реальном бизнесе. Альтернативой такой метрике может служить *ранг* товара или услуги в общем списке таковых на той или иной платформе электронной коммерции, упорядоченном по какому-либо критерию, объективно отражающему необходимую величину. Такая ранговая метрика будет однозначно отражать популярность товара (но действовать исключительно в сравнении с другими товарами), и обладать ценным свойством *не-раскрытия* реальных данных о продажах, сохраняя их в виде коммерческой тайны.

Данная работа посвящена исследованию практических реализаций различных методов ранжирования и имитационного моделирования их поведения.

Краткое содержание работы

Глава 1 «Алгоритмы и методы ранжирования» посвящена описанию некоторых алгоритмов ранжирования и критериев их корректности.

В разделе 1 «Общие сведения» дан обзорный вид на ранжирование в применении к различным системам — в социологии, спорте, электронной коммерции. Также даны определения понятиям ранжирования и ранга, взятые из контекста социологии, приведены примеры ранжирования и рангов.

В разделе 2 «Алгоритмы ранжирования популярности в системах электронной коммерции» дана более подробная перспектива на применение алгоритмов ранжирования в системах электронной коммерции, приведены практические примеры из крупных торговых площадок вроде Amazon, даны отсылки на исследования и патенты, касающиеся Amazon sales rank, показана практическая важность такого рода исследований.

В подразделе 1 «Базовые принципы» приведены теоретические основы ранжирования в соответствии с критерием ранжирования.

В подразделе 2 «Пример простой реализации» приведен пример простой практической реализации алгоритма ранжирования на языке SQL, на примере которой показаны определенные недостатки алгоритмов ранжирования такого

рода. Суть данной реализации состоит в ранжировании товаров по количеству продаж в некотором окне ранжирования, в примере равного одним суткам.

В подразделе 3 «Критерии корректности» приводятся критерии корректности, которыми должна обладать качественная реализация алгоритма ранжирования для того, чтобы иметь хорошие практические характеристики. Такими критериями являются:

1. Нестарение
2. Гладкость
3. Отсутствие априорной информации о распределении популярности товаров
4. Справедливость

Далее приведены более подробные описания каждого из критериев.

Нестарение

Нестарение – это важное свойство алгоритма ранжирования, заключающееся в том, что при полной остановке продаж (по технической или по реальной причине) величины ранка популярности товаров должны оставаться неизменными. Тем не менее, полная остановка продаж может повлиять на то, как будут меняться ранки товаров после возобновления продаж – например, усиливая прыжки ранка после факта продажи.

Гладкость

Критерий гладкости отвечает за монотонное и гладкое увеличение ранка популярности товара в случае отсутствия продаж у этого конкретного товара и наличия типовой продажной активности в целом. Демонстрацию критерия гладкости можно увидеть на рисунке 1. На этом графике видно, как товар постепенно, гладко и непрерывно теряет популярность при отсутствии продаж. Факты продаж при этом выглядят как резкие скачки популярности в лучшую сторону.

Отсутствие априорной информации о распределении популярности товаров

Отсутствие априорной информации о распределении популярности товаров ограничивает возможность знания о распределении реальных данных, запечатленного в алгоритме ранжирования. Отсутствие такой информации с большей вероятностью позволяет алгоритму адаптироваться к отклонениям, неизбежно случающимся в реальных данных. Другими словами данный критерий можно сформулировать как предпочтение непараметрических методов и формулировок параметрическим.

Справедливость

Справедливостью можно назвать свойство алгоритма ранжировать товары с более высокими продажами в общем случае выше. Данное свойство алгоритма нельзя назвать объективными, т. к. не существует однозначного метода сравнения продаж товаров, и оно будет приобретать те или иные особенности в зависимости от примененного метода оценки продаж.



Рисунок 1 - График Amazon Sales rank от времени rank на примере товара из раздела eBooks.

Далее в подразделе 4 «Улучшенная реализация ранжирования и ее свойства» показано, каким образом можно улучшить простую реализацию из предыдущего раздела, чтобы она могла лучше соответствовать критерию гладкости. Однако также показано, что несмотря на улучшения, критерий гладкости в улучшенном методе тем не менее не соблюдается. Суть улучшений при этом состоит в том, чтобы использовать дополнительный критерий для ранжирования в виде даты последней продажи товара и упорядочивать товары по ней при совпадении основного критерия в виде количества продаж в окне агрегации.

В подразделе 5 «Метод N последних продаж» дано описание другого алгоритма ранжирования, в котором в качестве критерия ранжирования используется средний темп продаж для последних N товаров. Однако далее в тексте работы обосновано показано, что данный метод также нарушает критерий гладкости.

В подразделе 6 «Модифицированный метод N последних продаж» показано, как метод N последних продаж может быть изменен для того, чтобы соблюдать критерий гладкости. Суть модификации состоит в том, чтобы добавить временную дистанцию до последнего факта продажи в формулу расчета критерия ранжирования, что улучшает некоторые свойства алгоритма.

В подразделе 7 «Комбинированный метод N последних продаж» показано, как можно скомбинировать модифицированный метод N последних продаж с простым методом для получения комбинированного метода, обладающего преимуществами обоих методов — длинным окном агрегации для товаров с низкими продажами и фиксированным окном агрегации для товаров с высокими продажами.

В подразделе 8 «Метод линейной свертки» описан класс методов, использующих операцию линейной свертки над фактами продаж для получения критерия ранжирования. Показано несколько способов, которыми можно получать критерий ранжирования, а также продемонстрированы недостатки в виде низкой численной точности для удаленных по времени фактов продажи.

В главе 2 «Имитационное моделирование» описаны некоторые положения, касающиеся априорных знаний, применяемых при построении моделей и процесса моделирования, а также практические соображения по реализации моделей. Также в данной главе даны подробные описания применяемых технологий и сделаны выводы о применимости алгоритмов в реальном окружении.

Раздел 1 «Распределение Ципфа как распределение популярностей товаров» описывает важность распределения Ципфа в мире торговли и электронной коммерции. В данном разделе приведены приблизительные параметры распределения, выявленные в исследованиях платформы Amazon, а также зафиксированы параметры для использования в модели.

Раздел 2 «Практические аспекты реализации алгоритмов ранжирования» описывает структуры данных и технологии, применяемые при реализации моделей и примеров. Среди структур данных, заслуживающих внимания при реализации методов ранжирования стоит отметить кучу (heap), список с пропусками (skip list), а также деку (deque). Технологии, применяемые при реализации методов включают в себя язык запросов SQL и язык программирования C++.

Раздел 4 «Результаты моделирования» демонстрирует результаты имитационного событийного моделирования работы алгоритма ранжирования, описанного в разделе 1.2.2.

Заключение

В результате исследования алгоритмов ранжирования в системах электронной коммерции были получены следующие результаты:

- Показана практическая ценность и необходимость алгоритмов ранжирования в различных контекстах.
- Было предложено несколько вариаций алгоритмов ранжирования, основывающихся на различных метриках, а также их улучшения.
- Были выведены критерии корректности алгоритмов ранжирования, соблюдение которых гарантирует хорошие практические качества алгоритмов.
- Было смоделировано поведение одного из алгоритмов ранжирования в условиях, близких к реальным.

Дальнейшее направление работы в данном направлении может включать в себя разработку и моделирования большего числа алгоритмов, а также отказоустойчивую практическую реализацию, пригодную для быстрого внедрения в реальные проекты систем электронной коммерции. Также одним из возможных направлений исследований может оказаться реверс-инжиниринг систем ранжирования крупных интернет-компаний (например, Amazon или Barnes & Noble) для проведения рыночных исследований по примеру работ [3], [4].

Список использованных источников

1. Хаванова Н.В., Бокарева Е.В. Анализ мирового и российского рынка электронной торговли: тенденции и проблемы развития // Сервис в России и за рубежом. 2017. №3 (73). URL: <https://cyberleninka.ru/article/n/analiz-mirovogo-i-rossiyskogo-rynka-elektronnoy-torgovli-tendentsii-i-problemy-razvitiya> (дата обращения: 10.12.2018).
2. Бабич Н.С., Батыков И.В. Ординальное шкалирование. Краснодар, 2004.
3. Touzani S., Van Buskirk R. Estimating sales and sales market share from sales rank data for consumer appliances //Physica A: Statistical Mechanics and its Applications. – 2016. – Т. 451. – С. 266-276.
4. Iba T. et al. Power-law distribution in Japanese book sales market //Fourth Joint Japan-North America Mathematical Sociology Conference. – 2008. – Т. 137.
5. Brynjolfsson E., Hu Y. J., Smith M. D. The longer tail: The changing shape of Amazon's sales distribution curve. – 2010.

6. Hanks S., Spils D. Increases in sales rank as a measure of interest : пат. 7058599 США. – 2006.
7. Hunter J. S. The exponentially weighted moving average //Journal of quality technology. – 1986. – Т. 18. – №. 4. – С. 203-210.
8. Deriche R. Recursively implementating the Gaussian and its derivatives : дис. – INRIA, 1993.
9. Gastal E. S. L., Oliveira M. M. High-Order Recursive Filtering of Non-Uniformly Sampled Signals for Image and Video Processing //Computer Graphics Forum. – 2015. – Т. 34. – №. 2. – С. 81-93.
10. Gabaix X. Zipf's law for cities: an explanation //The Quarterly journal of economics. – 1999. – Т. 114. – №. 3. – С. 739-767.
11. Li W. Random texts exhibit Zipf's-law-like word frequency distribution //IEEE Transactions on information theory. – 1992. – Т. 38. – №. 6. – С. 1842-1845.
12. Antipov E. A., Pokryshevskaya E. B. Rank-sales relationship in electronic commerce: Evidence from publicly available data on 11 product categories //Electronic Commerce Research and Applications. – 2016. – Т. 16. – С. 1-6.

Список публикаций соискателя

1-А. Евгений Дараган. Дискретно-событийное моделирование ранжирования товаров и услуг в системах электронной коммерции//Наука онлайн: Международный научный электронный журнал. - 2019. - №1. - <https://nauka-online.com/publications/informatsionnye-tehnologii/2019/1/diskretno-sobytijnoe-modelirovanie-ranzhirovaniya-tovarov-i-uslug-v-sistemah-elektronnoj-kommertsii/>