

Информатика и кибернетика

УДК 004.021

**Дараган Евгений Николаевич**

*магистрант*

*Белорусского государственного университета*

*информатики и радиоэлектроники*

## **ДИСКРЕТНО-СОБЫТИЙНОЕ МОДЕЛИРОВАНИЕ РАНЖИРОВАНИЯ ТОВАРОВ И УСЛУГ В СИСТЕМАХ ЭЛЕКТРОННОЙ КОММЕРЦИИ**

***Аннотация.** Исследованы практические аспекты моделирования алгоритмов и методов ранжирования в системах электронной коммерции. Показан алгоритмически оптимальный способ генерации событий для имитации событий купли/продажи. Показаны способы использования структур данных для получения оптимальной реализации алгоритмов ранжирования.*

***Ключевые слова:** событийное моделирование, ранжирование, алгоритмы и структуры данных.*

Системы электронной коммерции занимают существенное место в мировом товарообороте. Так, по статистике 2015 года в развитых странах электронная коммерция занимает более 10 % всего товарооборота [1] и продолжает неуклонно расти. С ее ростом также растут потребности бизнеса в качественных инструментах и открываются новые рубежи в сферах логистики, скорости и качества обслуживания покупателей, легкости вступления в ряды продавцов, разнообразия представленных товаров.

Ранжирование товаров и услуг является неотъемлемой частью современных систем электронной коммерции. Существует множество вариаций методов ранжирования, которые основываются на различных предположениях, однако не все из них одинаково хороши. Приведем ряд критериев, соответствие которым является обязательным для качественной реализации алгоритма ранжирования:

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

Далее будет приведена более подробная информация по каждому из предлагаемых критериев с обоснованием его необходимости.

### **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.



**Рис. 1. График Amazon Sales rank от времени rank на примере товара из раздела eBooks. Составлено автором на основе [2]**

### **Моделирование алгоритмов ранжирования.**

Задача моделирования алгоритма ранжирования сводится к моделированию двух процессов: процесса продаж и процесса ранжирования.

Для моделирования процесса продаж необходима информация о типовом распределении популярности товаров и услуг. Согласно исследованию [3], реальное распределение продаж различных категорий товаров на сайте Amazon в первом приближении описывается распределением Ципфа с экспонентой, близкой к единице. Процесс продаж для отдельно взятого товара хорошо описывается пуассоновским процессом с математическим ожиданием, равным величине популярности товара.

Моделирование массива независимых пуассоновских процессов можно эффективно произвести следующим образом. В структуре данных «куча» (heap) будут храниться времена стартовых событий продаж для каждого товара, помещаемые туда на старте алгоритма, при этом ключом для вычисления минимума в куче будет являться время планируемой продажи. На каждом шаге моделирования из кучи извлекается факт продажи с минимальным временем, и туда же помещается следующий факт продажи для извлеченного товара. Время следующего факта продажи вычисляется случайным образом исходя из распределения межсобытийного интервала для пуассоновского процесса, которое является экспоненциальным распределением с экспонентой, обратно пропорциональной величине ожидаемой популярности продажи. Далее процесс повторяется с шага извлечения. Алгоритмическая сложность такого моделирования равна  $O(M \log N)$ , где  $M$  – количество смоделированных событий, а  $N$  – общее количество товаров.

Моделирование самого алгоритма ранжирования зависит от того, к какому классу он принадлежит: к классу динамических (т. е. с

возможностью получать обновленный ранк товара сразу после обработки алгоритмом факта продажи), или к классу статических (когда возможно только получение сразу всех ранков товаров в фиксированный момент времени).

Эффективное моделирование динамических алгоритмов возможно при помощи структуры данных «список с пропусками» (skip list), реализация которого должна поддерживать операцию RANK, возвращающую ранк объекта (позицию в отсортированном по критерию ранжирования списке). Алгоритмическая сложность такой реализации будет составлять  $O((K + M) \log N)$ , где  $K$  – количество запросов на ранк конкретного товара ( $K = kN$  при получении ранков сразу всех товаров  $k$  раз),  $M$  – количество обновлений показателя ранжирования, а  $N$  – количество всех товаров.

Статические методы ранжирования, которые могут получать ранжированный список только сразу для всех товаров выгодно моделировать напрямую при помощи сортировки всех товаров по показателю ранжирования. Алгоритмическая сложность такого моделирования будет составлять  $O(kN \log N)$ , где  $k$  – количество выборок результата ранжирования,  $N$  – количество всех товаров.

Для расчета показателей ранжирования, которые опираются на количество продаж в фиксированном временном окне можно использовать следующий алгоритм, обеспечивающий сложность  $O(1)$  в пересчете на одно событие продажи. Пусть структура данных «дека» (deque) хранит последовательность событий продажи в порядке их появления, а также ассоциированный с ним товар, а отдельный ассоциативный массив хранит количество событий в deque для каждого товара. При появлении нового события мы помещаем его в очередь, увеличивая при этом счетчик событий для соответствующего товара, а после извлекаем из другого конца очереди события до тех пор, пока временная разница между новым и

извлеченным событием перестанет превышать длину фиксированного временного окна алгоритма, попутно уменьшая счетчики событий для товаров, ассоциированных с извлеченными событиями. Таким образом, очередь будет состоять только из событий, произошедших в интересующем нас временном окне, а массив счетчиков событий будет содержать актуальные данные после каждого нового события.

### **Литература**

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). Режим доступа: <https://cyberleninka.ru/article/n/analiz-mirovogo-i-rossiyskogo-rynka-elektronnoy-torgovli-tendentsii-i-problemy-razvitiya>
2. [Электронный ресурс]. – Режим доступа: <https://www.novelrank.com/>
3. 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.