



В.В. Матвеев, А.А. Столбова

СРАВНЕНИЕ АЛГОРИТМОВ ПОЛУЧЕНИЯ АССОЦИАТИВНЫХ ПРАВИЛ ДЛЯ РЕШЕНИЯ ЗАДАЧИ АНАЛИЗА ПОКУПАТЕЛЬСКИХ КОРЗИН

(Самарский университет)

Анализ содержимого покупательских корзин является важной частью ведения успешного бизнеса в области розничной торговли. За годы торговли в базах данных торговых сетей накоплены обширные массивы проведенных транзакций [1]. Каждая такая база может использоваться для увеличения прибыли при помощи анализа существующих транзакций, выделения правил зависимости и применения их для предсказания новых покупок или распределения товаров по торговым точкам. Одним из способов анализа покупательской корзины является получение ассоциативных правил. Алгоритмы поиска ассоциативных правил активно разрабатываются и улучшаются, так что для реализации рекомендательной системы необходимо провести сравнение алгоритмов между собой по времени работы и количеству находимых правил [2].

В качестве исходных данных в исследовании используется файл Market-Basket-Optimisation.csv, содержащий в себе 7501 транзакцию и 119 различных товаров. На рисунке 1 изображена структура файла, открытого в табличном процессоре MS Excel. Данный файл является одним из эталонных образцов и пригоден для оценки эффективности работы алгоритмов [3].

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	shrimp,almonds,avocado,vegetables mix,green grapes,whole weat flour,yams,cottage cheese,energy drink,tomato juice,low fat yogurt,green tea,													
2	burgers,meatballs,eggs													
3	chutney													
4	turkey,avocado													
5	mineral water,milk,energy bar,whole wheat rice,green tea													
6	low fat yogurt													
7	whole wheat pasta,french fries													
8	soup,light cream,shallot													
9	frozen vegetables,spaghetti,green tea													
10	french fries													
11	eggs,pet food													
12	cookies													
13	turkey,burgers,mineral water,eggs,cooking oil													
14	spaghetti,champagne,cookies													
15	mineral water,salmon													
16	mineral water													
17	shrimp,chocolate,chicken,honey,oil,cooking oil,low fat yogurt													
18	turkey,eggs													
19	turkey,fresh tuna,tomatoes,spaghetti,mineral water,black tea,salmon,eggs,chicken,extra dark chocolate													
20	meatballs,milk,honey,french fries,protein bar													
21	red wine,shrimp,pasta,pepper,eggs,chocolate,shampoo													
22	rice,sparkling water													
23	spaghetti,mineral water,ham,body spray,pancakes,green tea													
24	burgers,grated cheese,shrimp,pasta,avocado,honey,white wine,toothpaste													
25	eggs													
26	parmesan cheese,spaghetti,soup,avocado,milk,fresh bread													
27	ground beef,spaghetti,mineral water,milk,energy bar,black tea,salmon,frozen smoothie,escalope													
28	sparkling water													
29	mineral water,eggs,chicken,chocolate,french fries													
30	frozen vegetables,spaghetti,yams,mineral water													
31	herb & pepper,tomato sauce,light cream,magazines													
32	mineral water.chocolate.avocado.eggs													

Рисунок 1 – Структура файла исходных данных



Сравнение проводится для алгоритмов Apriori и Eclat. Являясь одними из старейших алгоритмов такого рода, они, тем не менее, по сей день активно применяются в отраслях получения данных [4].

Оценка производится путём замера времени обработки набора данных штатными средствами среды исполнения. Всего проведено 5 опытов для каждого алгоритма с предварительной эмпирической оценкой величины поддержки, при которой будет заметна задержка. Данные по результатам экспериментов занесены в таблицу 1, графическая интерпретация эксперимента приводится на рисунке 2. По оси абсцисс располагается значение поддержки, по оси ординат – значение времени в секундах.

Таблица 1 – Сравнение алгоритмов по значению поддержки

Поддержка	Время работы алгоритма Eclat, с	Время работы алгоритма Apriori, с
0,1	0,00	0,00
0,01	0,01	0,00
0,0001	3,15	5,06
0,00001	5,6	180,45
0,000001	7,2	460,6

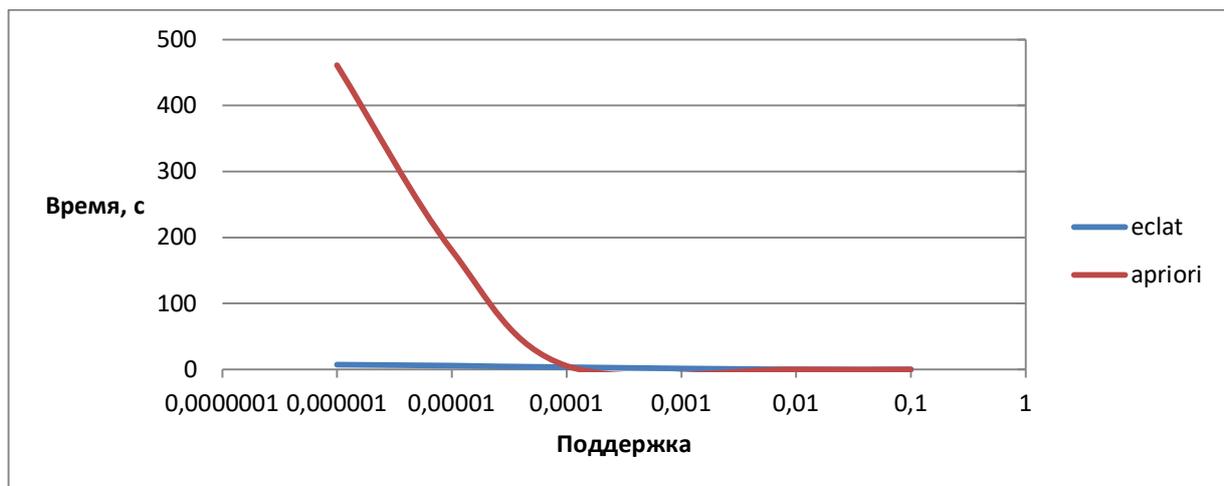


Рисунок 2 – График зависимости эффективности алгоритмов от значения поддержки

Кроме того, проводится оценка зависимости количества сгенерированных правил от размеров исходной выборки. Для этого значение поддержки остаётся фиксированным, а значения количества обрабатываемых транзакций меняется от 500 до 7500 с шагом в 1000. Значения результатов эксперимента приведены в таблице 2. На рисунке 3 приводится графический результат обработки эксперимента.



Таблица 2 – Сравнение алгоритмов по объёму данных

Объём данных	Алгоритм Eclat	Алгоритм Apriori
500	682674	5593919
1500	43654	256001
2500	10383	21057
3500	8786	15092
4500	8289	12974
5500	7883	11639
6500	7306	10265
7500	6778	8944



Рисунок 3 – График зависимости количества выводимых правил от общих размеров выборки

Исходя из полученных данных, следует вывод, что алгоритм Apriori эффективен при сравнительно больших значениях поддержки, после чего следует резкое снижение эффективности. Причиной эффективности алгоритма Eclat является обход в глубину, за счёт чего снижается количество обращений к исходной базе. Кроме того, заметна обратная зависимость между объёмами выборки и количеством выводимых правил, что свидетельствует о том, что большие объёмы позволяют предсказывать и обобщать точнее.

Общим выводом по результатам данных экспериментов может служить утверждение, что алгоритм Eclat в большей степени применим для решения задачи анализа покупательской корзины для больших объёмов исходных данных и при низких значениях поддержки, что позволяет ему находить более редкие правила. На высоких значениях поддержки алгоритмы Eclat и Apriori сравнимы.

Литература

1 Ivaschenko A., Stolbova A., Golovnin O. Data Market Implementation to Match Retail Customer Buying Versus Social Media Activity //Science and Information Conference. – Springer, Cham, 2020. – С. 363-372.



2 Ассоциативные правила, или пиво с подгузниками [Электронный ресурс] // Хабрахабр. – URL: <https://habr.com/ru/company/ods/blog/353502/> (дата обращения: 18.04.2021).

3 Market Basket Optimization [Электронный ресурс] // Kaggle. – URL: <https://www.kaggle.com/gameilsaad/market-basket-optimisation> (дата обращения: 18.04.2021).

4 Алгоритм Apriori (Apriori algorithm) [Электронный ресурс] // Loginom. – URL: <https://wiki.loginom.ru/articles/apriori.html> (дата обращения: 18.04.2021).

А.В. Мельников, Л.А. Великанова, Е.С. Легостаева

СИСТЕМА МОНИТОРИНГА И АНАЛИЗА ПЕРСПЕКТИВНЫХ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ В ИНТЕРЕСАХ ВООРУЖЕННЫХ СИЛ РОССИЙСКОЙ ФЕДЕРАЦИИ

(Военный учебно-научный центр Военно-Воздушных сил
«Военно-Воздушная академия имени профессора Н.Е. Жуковского
и Ю.А. Гагарина», г. Воронеж)

В целях совершенствования информационного обеспечения организаций и учреждений Министерства обороны Российской Федерации в части развития и применения современных информационных технологий (ИТ) в интересах Вооруженных сил Российской Федерации (ВС РФ), в настоящее время ведется разработка системы мониторинга и анализа информационных технологий (СМАИТ). Мониторинг ИТ является необходимым условием для устойчивого развития ВС РФ, так как на его основе возможно получить оперативную информацию о текущем состоянии ИТ, необходимых для развития вооружения, военной и специальной техники (ВВСТ).

Согласно организационно-методическим указаниям по проведению работ на заключительном этапе научно-исследовательской работы (НИР) «Исследование вопросов развития и применения современных информационных технологий военного и двойного назначения в интересах ВС РФ», шифр «Дальновидец», при выполнении исследований по мониторингу и анализу ИТ должны учитываться [1, 2]:

- государственные приоритеты, определяющие направления и тенденции развития ИТ, их внедрения в автоматизированные системы военного назначения (АС ВН) (образцы ВВСТ);

- отечественный и зарубежный опыт, со сравнительным анализом существующих АС ВН (образцов ВВСТ) и ИТ, применяемых в них;

- анализ применения ИТ в различных АС ВН (образцах ВВСТ) видов и родов войск ВС РФ с учетом возможности их заменяемости;

- опыт применения ИТ при эксплуатации АС ВН (образцах ВВСТ) в различных условиях, в том числе в сложных (критических) условиях (высокие и