

Секция «Педагогическое образование и образовательные технологии»

Изучение интеллектуальных алгоритмов в углубленном курсе информатики на примере реализации алгоритма Apriori

Салахова Алёна Антоновна

Студент (бакалавр)

Московский педагогический государственный университет, Москва, Россия

E-mail: aa.salakhova@yandex.ru

В курсе информатики углублённого уровня изучается несколько самообучающихся алгоритмов, позволяющих составлять правила из большого набора данных, относящихся к такому направлению информационных технологий, как искусственный интеллект. Одним из них является Apriori.

Алгоритм Apriori наиболее понятен и востребован для обучающихся, т.к. они сталкиваются с его применением ежедневно, что позволяет связать достаточно сложный теоретический материал с жизнью.

Данный алгоритм позволяет пополнять и формировать базы знаний. Он был разработан в 1993 году Ракешем Агравалом для исследования потребительских корзин, поэтому Apriori иногда называют *market basket analysis*, т.е. анализ потребительских корзин. Задача поиска шаблонов покупок, совершаемых клиентами супермаркетов, актуальна сегодня и, вероятно, не потеряет актуальность в будущем. Apriori находит ассоциативные правила (association rule mining) − правила отношения к чему-либо, используемые для группировки данных, которые проще всего записать конструкцией "if-then". Простым и понятным примером, встречающимся ежедневно, является расположение полок в гипермаркете: сухие сливки лежат рядом с кофе и т.д. Здесь работает следующее правило: если (if) человек купил кофе, тогда (then) он захочет добавить сливки.

Нахождение определённых правил, которые встречаются достаточно редко, позволяют продать потребителю тот товар, на который он бы не обратил внимания, если бы не расположение на витрине. Именно алгоритм Apriori применяется при составлении акций и распродаж и позволяет получить больший доход. Аналогично его можно использовать для анализа продуктивности взаимодействия людей и составления эффективных команд (в менеджменте), выявления динамики пациентов медицинскими экспертными системами, анализа групп риска в отношении суицидов, преступлений и т.п.

У алгоритма Apriori огромное множество реализаций, используемых в различных проектах, например: Orange Data Mining, ARTOOL, Weka (Waikato Environment for Knowledge Analysis) и другие. Использование алгоритма в подобных системах открыло новые возможности для различных сфер, что, несомненно, говорит об его актуальности.

Важным следствием является необходимость умения реализовать Apriori на практике хотя бы для частных случаев, что доступно для понимания и усвоения обучающимися старшей школы в рамках курса информатики углублённого уровня.

Предлагается изучать алгоритм на практических занятиях в три этапа: письменная реализация в тетради, реализация алгоритма в табличном процессоре и, самый сложный уровень, реализация Apriori на языке программирования C++ (без использования специфичных для него структур и библиотек для лёгкого переноса на любой другой язык, например, Pascal или Python). Для каждого этапа приводится подробная инструкция для педагогов и обучающихся.

Данная последовательность была использована при объяснении материала в Лицее г.Реутов при преподавании темы в 11 классе и показала положительный результат усвое-

ния темы школьниками.

Источники и литература

- 1) Калинин И. А., Самылкина Н. Н., Информатика. Углублённый уровень: учебник для 11 класса. М.: БИНОМ, 2013. - 4 Гл.
- 2) Салахова А.А., "Изучение интеллектуальных алгоритмов на примере реализации Apriori в углублённом курсе информатики журнал "Информатика май-июнь/2016, М.: Первое сентября, 2016. - с.3-17
- 3) Divya Bansal, Lekha Bhambhu, "Execution of APRIORI Algorithm of Data Mining Directed Towards Tumultuous Crimes Concerning Women- International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 9, September 2013, p. 54-62
- 4) Jian Pei, Jiawei Han, Micheline Kamber. "Data Mining, Second Edition : Concepts and Techniques", 2nd Edition — U.S.A., The University of Illinois at Urbana-Champaign: "Morgan Kaufmann", January 2006, 722 p.
- 5) Wenjing Zhang, Donglai Ma, Wei Yao. "Medical Diagnosis Data Mining Based on Improved Apriori Algorithm", Journal of Networks, Vol 9, No 5 (2014), 1339–1345, May 2014
- 6) Краковецкий Александр (CEO DevRain Solutions), "Анализ рыночной корзины и ассоциативные правила". [Электронный ресурс] Режим доступа: <http://habrahabr.ru/post/66016/>
- 7) Свободная энциклопедия Wikipedia, статья об Apriori. [Электронный ресурс] Режим доступа: https://en.wikipedia.org/wiki/Apriori_algorithm
- 8) Шахиди Аюбир, "Apriori — масштабируемый алгоритм поиска ассоциативных правил". [Электронный ресурс] Режим доступа: <https://basegroup.ru/community/articles/apriori>
- 9) Anita Wasilewska, "APRIORI Algorithm Stony Brook University [Электронный ресурс] Режим доступа: http://www3.cs.stonybrook.edu/634/lecture_notes/07apriori.pdf

Слова благодарности

Научному руководителю и прекраснейшему специалисту, Самылкиной Н.Н.

Иллюстрации

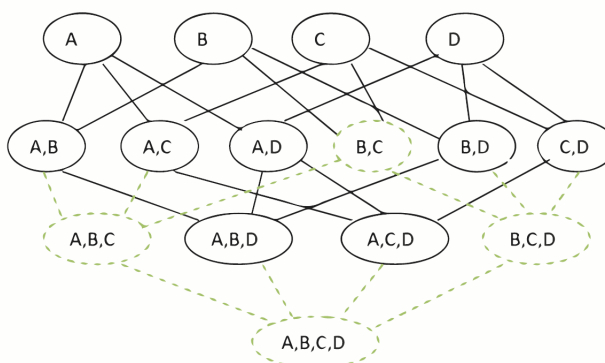


Рис. 1. Схема выбора k-элементных часто встречающихся множеств

=ЕСЛИОШИБКА(ЕСЛИ(ПОИСК(\$B\$6;\$E3:1)-0;1;0);0)

Transaction		ID								
TID	Items	minsup=	3							
101	мясо, хлеб			101	1	2	3	4	5	6
102	хлеб, соль, сахар			102	1	1	0	0	0	0
103	мука			103	0	0	0	0	1	0
104	спички, хлеб, мясо			104						
105	хлеб, спички			105	1	1	0	0	0	1
106	мясо, мука			106	0	1	0	0	0	1
107	соль, сахар			107	1	0	0	0	1	0
108	мясо, хлеб, соль, спички			108	0	0	1	1	0	0
					1	1	1	0	0	1

Рис. 2. Нормализованная таблица покупок в табличном процессоре Excel

№	Правило	Достоверность, %
1	Если покупатель взял мясо, то он купит хлеб.	75
2	Если покупатель взял хлеб, то он купит мясо.	60
3	Если покупатель взял хлеб, то он купит спички.	60
4	Если покупатель взял спички, то он купит хлеб.	100

Рис. 3. Найденные ассоциативные правила при реализации с помощью Excel

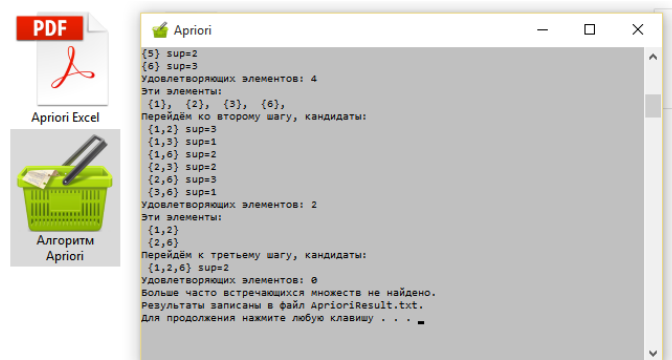


Рис. 4. Программа на C++, реализующая алгоритм Apriori