

ДИСЦИПЛИНА	Проектирование интеллектуальных систем (часть 1/2)
ИНСТИТУТ	информационных технологий
КАФЕДРА	вычислительной техники
ВИД УЧЕБНОГО МАТЕРИАЛА	Материалы для практических/семинарских занятий
ПРЕПОДАВАТЕЛЬ	Холмогоров Владислав Владиславович
СЕМЕСТР	6, 2023-2024

Практическая работа № 1

«Анализ ассоциативных правил»

по дисциплине «Проектирование интеллектуальных систем (часть 1/2)»

Цели: приобрести навыки поиска ассоциативных правил при анализе наборов информационных исторических данных.

Задачи:

1) Провести анализ ассоциативных правил исторических данных, выполнив следующие подзадачи:

- определить предметную область решаемой задачи, ею могут выступать потребительские корзины любых розничных и интернет-магазинов, истории просмотров определённых ресурсов, посещения мест (экскурсии, туры и т.п.), симптомы заболеваний, социальные сети и группы и т.д.;

- выбрать или сгенерировать соответствующий набор данных, включающий списки с унифицированным названиями (чтобы название одного и того же элемента в разных строчках было одинаково) или группами (если в наборе данных один и тот же предмет всё-таки имеет разные названия, то он должен хотя бы относиться к одной и той же группе для реализации возможности поиска ассоциативных правил);

- найти ассоциативные правила с количеством объектов не меньше 2-ух и рассчитать для них метрики «поддержки» (*support*), «доверия, уверенности» (*confidence*) и «убеждённости» (*conviction*), пояснить суть каждой метрики в целом и смысл в данной задаче предметной области;

- реализовать алгоритм поиска ассоциативных правил (см. Примечание 1), описать и проанализировать его суть и результаты.

2) В качестве дополнительного задания реализовать хотя бы один из следующих пунктов (один на оценку 4, все – на оценку 5):

- рассчитать метрики лифта (*lift*) и рычага (*leverage*), описать их суть и смысл в данной задаче предметной области;

– реализовать несколько алгоритмов поиска ассоциативных правил, сравнить их основные показатели производительности: качество результатов, скорость работы, требуемое количество памяти.

– провести экспертную кластеризацию на основе результатов анализа ассоциативных правил (см. Примечание 2).

ПРИМЕЧАНИЕ:

1) Алгоритмы поиска ассоциативных правил генерируют из начального набора данных решения(наборы)-кандидаты (наборы совместно встречающихся элементов) по заданным параметрам, как правило, на основе минимальных значений показателей поддержки и (иногда) уверенности, так алгоритм Apriori на основе поиска в ширину позволяет находить наборы длины $i+1$ каждую следующую итерацию; алгоритм Eclat генерирует наборы-кандидаты на основе поиска в глубину, причём для проверки каждого из правил должны быть предварительно проверены и составляющие правила на соответствие минимальному значению поддержки λ каждого из них (для правила-множества «огурцы \rightarrow помидоры \rightarrow сметана» должны выполняться правила-подмножества «огурцы», «помидоры», «сметана», «огурцы \rightarrow помидоры», «помидоры \rightarrow сметана», «огурцы \rightarrow сметана»), алгоритм FP-growth позволяет предварительным упорядочиванием элементов внутри каждой транзакции на основе показателей их поддержки построить ассоциативное дерево, которое в одной структуре отображает разом все наборы часто встречаемых элементов, имеет модифицированный вариант FPMax, акцентированный на большие наборы. Среди специфических правил ассоциативного анализа можно выделить GUNA, позволяющий находить ассоциативные правила на основе расширенного списка логических операций, и алгоритм поиска неупорядоченных данных OPUS, представляющий (как и FP алгоритмы) собой граф, основанный на переупорядочивании пространства поиска с последующий обрезкой неудовлетворяющих ветвей, а также

алгоритмы CARTE, объединяющий анализ ассоциативных правил и кластеризацию, Tertius, объединяющий анализ ассоциативных правил и дерево решений, и стохастический COBWEB, выполняющий инкрементную кластеризацию и иерархическую категоризацию.

2) При достаточно высоких значениях поддержки и доверия ассоциативные правила могут указать на скрытые коллаборативные зависимости, представляющие собой невыявленные группы (кластеры), которые в результате их осмысления могут стать явными группами (классами) при дальнейшем анализе данных. Например, продукты: огурцы, помидоры, зелень и сметана совместно образуют группу «ингредиенты летнего салата», программное обеспечение: редактор текста, редактор таблиц и корпоративная информационная система могут входить в группу «офисное программное обеспечение» и т.п. Целью анализа ассоциативных правил является выявление таких взаимосвязей, а задачей аналитика данных является их обработка с последующим выделением в отдельный осмысленный класс товаров, что позволяет ориентироваться на определённую целевую аудиторию, увеличивать продажи и строить рекомендации на основе знаний о предметной области конкретных товаров.

ОСОБЫЙ БОНУС (доступен только в том случае, если выполнены пункты 2-го задания):

Выполнить поиск ассоциативных правил с помощью специфических методов поиска вроде GUHA и OPUS, либо реализовать методы, позволяющие разом выполнить и задачу анализа ассоциативных правил, и кластеризацию (вроде CARTE).