Владимир Биллиг
vladimir-billig@yandex.ru

Алгоритм Apriori_Goal построения ассоциативных правил В базах данных с классификацией

Суперкомпьютерные дни в РФ 2025

Ассоциативные правила. Алгоритм Apriori

- Ассоциативные правила: X => Y
- Классический алгоритм Apriori
- Apriori_Goal vs Apriori:
  - База данных
  - Монотонность vs Антимонотонность
  - Критерии
  - Эффективность

База данных. Препроцессор

- База данных – реляционная БД – описание объектов
- Входные параметры + Целевой параметр
- Препроцессор – важная часть алгоритма:
  - Вход: db, def_db
  - Преобразование, кодирование, сортировка
  - Выход: dbf = {dbf1, dbf2, …dbfk}
  - Запись – целое число, сохраняющее информацию

Монотонность vs Антимонотонность

- Антимонотонность Apriori:
  - Not(X – частый набор) => Not (E(X) – частый набор)
- Монотонность Apriori_Goal
- Теорема 2 (Монотонность критериев корреляции и достоверности)
  - Когда положительно коррелированное правило с посылкой X расширяется , используя положительно коррелированного кандидата Y , достоверность расширенного правила может только возрастать относительно достоверности исходных правил.

Ассоциативные правила. Критерии

- Ассоциативные правила : $X \Rightarrow Goalk$
- Два критерия частоты правила:
  - Fg = Support(X, dbk) / nk; Fall = Support(X, dbk) / N
- Два критерия надежности правила:
  - Confidence = Support(X, dbk) / Support(X, db)
  - Lift = P(Goalk / X) / P(Goalk)
  - Correlation = f(lift) -1 <= correlation <= 1
- Обобщенный критерий

Эффективность

- Факторы, обеспечивающие эффективность:
  - Распараллеливание
  - При построении правил не требуется обходить всю базу данных – достаточно обхода подмножества
  - Кодирование позволяет операции над множествами заменить операциями над целыми числами: X & Y == Y

Экспериментальная оценка времени построения правил на базе данных diabetes Python

Число записей	Число построенных правил	Время построения правил (сек)
44200	13	0.015
44200	22	0.025
442000	13	0.17
442000	22	0.20
4420000	13	1.85
4420000	22	2.04

Заключение

- Для баз данных с целевым параметром алгоритм Apriori_Goal позволяет строить ассоциативные правила с возрастающей степенью достоверности.
- Алгоритм строит правила как с высокой частотой появления, так и редкие правила, которые могут представлять особый интерес.
- Алгоритм позволяет строить правила, доказывающие высокую вероятность появления цели при заданной посылке, так и опровергающие появление цели.
- Алгоритм обладает достаточно высокой эффективностью по времени построения правил.
- Алгоритм успешно использовался при анализе различных медицинских баз данных.

Спасибо!

Вопросы