

- априорной неопределенности относительно законов распределения шумов;
- ограниченных выборок. Решение аналогичной задачи методом нелинейной фильтрации приводит к увеличению времени наблюдения в несколько раз при одинаковой точности оценивания.

#### ЛИТЕРАТУРА

1. **Прокис Дж.** Цифровая связь / Пер. с англ. под ред. Д.Д. Кловского — М.: Радио и связь, 2000. — 800 с.
2. **Корни Г. и Корн Т.** Справочник по математике для научных работников и инженеров. — М.: Наука, 1973. — 832 с.

3. **Шлома А.М.** О решении операторных уравнений при неполной информации // Журнал вычислительной математики и математической физики. — 1996. — Т. 36. №3. — С. 15—27.
4. **Шлома А.М., Смердова Н.Е.** Применение теории операторов функционального анализа для задач нелинейной фильтрации // Радиотехника и электроника. — 1999. — Т. 44. № 2. — С. 190—198.

Получено 21.03.08

УДК 621.396.62.019.4

## ПОИСК АССОЦИАТИВНЫХ ПРАВИЛ ПРИ АНАЛИЗЕ ЗАГРУЗКИ СЕТЕЙ СОТОВОЙ СВЯЗИ

М.В. Терешонок, аспирант МТУСИ

**Анализ загрузки сетей сотовой связи в целом является достаточно сложной задачей, состоящей из нескольких частных подзадач. Одна из таких подзадач — определение устойчивых групп абонентов. Под группой в данном случае подразумеваются два или более абонента, совершающие совместные перемещения по городу. Данная задача, в известном смысле, аналогична задаче поиска ассоциативных правил. Поэтому были исследованы вопросы применения алгоритмов поиска ассоциативных правил для определения устойчивых групп абонентов сетей сотовой связи.**

**Постановка задач.** Целью поиска ассоциативных правил является нахождение закономерностей между связанными событиями в базах данных. Множество произошедших одновременно событий принято называть транзакцией. Транзакционная или операционная база данных представляет собой двумерную таблицу, которая содержит номер транзакции и список событий, произошедших во время этой транзакции. Номер транзакции — уникальный идентификатор, определяющий транзакцию.

Ассоциативное правило имеет вид: «Из события А следует событие В». Оно устанавливает закономерность следующего вида: «Если в транзакции встретился набор элементов А, то можно сделать вывод, что в этой же транзакции должен появиться набор элементов В».

Основными характеристиками ассоциативного правила являются поддержка и достоверность правила. Правило «Из события А следует событие В» имеет поддержку  $P$ , если  $P\%$  транзакций из всего набора содержат одновременно наборы элементов А и В. Достоверность этого правила показывает, какова вероятность того, что из события А следует событие В. Правило «Из А следует В» справедливо с достоверностью  $C$ , если  $C\%$  транзакций из всего множества, содержащих набор элементов А, также содержат набор элементов В.

С помощью алгоритмов поиска ассоциативных правил можно получить все возможные правила вида: «Из А следует В» с различными значениями поддержки и достоверности. Однако в большинстве случаев количество правил необходимо ограничивать заранее установленными минимальными и максимальными значениями поддержки и достоверности.

Правила с очень большими значениями поддержки и достоверности могут оказаться очевидными и хорошо известными. С другой стороны, правила с низкими их значениями могут оказаться в большей части необоснованными. Таким образом, необходимо определить такой интервал («золотую середину»), который, с одной стороны, обеспечит нахождение неочевидных правил, а с другой — их обоснованность.

**Алгоритм поиска ассоциативных правил.** Проведенный анализ различных алгоритмов поиска ассоциативных правил выявил наиболее эффективный на сегодняшний день алгоритм Apriori [1—3]. Приведем данный алгоритм в приложении к задаче поиска устойчивых групп абонентов мобильной связи.

Будем рассматривать так называемые события  $x_j$ , представляемые, например, в виде чисел. Пусть есть некоторое количество транзакций  $T = \{T_1, T_2, \dots, T_n\}$ , где каждая транзакция  $T_i$  — некоторый набор событий  $X^i = \{x_1, x_2, \dots, x_m\}$ , произошедших одновременно ( $i = 1..n, m$ , — количество событий в  $i$ -й транзакции,  $n$  — количество транзакций). Общее число различных событий во всех рассматриваемых транзакциях равно  $M$ .

Требуется найти правила типа: «Если в транзакции встретился набор элементов А, то можно сделать вывод, что в этой же транзакции должен появиться набор элементов В».

Алгоритм Apriori состоит из следующих шагов:

1. Формирование одноэлементных наборов и расчет поддержки этих наборов. Одноэлементным набором называется набор, содержащий только одно событие. Число таких наборов равно  $M$ . Под поддержкой одноэлементного набора понимается следующая величина:

$$P(x_j) = \frac{N(x_j)}{n}, \quad (1)$$

где  $N(x_j)$  — количество транзакций, в которых встречается одноэлементный набор  $x_j$ .

2. Отбраковка одноэлементных наборов по заданному порогу поддержки  $P_{\min}$ . Если  $P(x_j) < P_{\min}$ , то набор событий  $x_j$  в дальнейшей работе алгоритма не используется. Данное пра-

вило позволяет уменьшить пространство поиска. Отсечение элементов происходит на основе предположения о том, что у часто встречающегося набора событий все подмножества должны быть часто встречающимися. Если в наборе находится подмножество, которое на предыдущем этапе было определено как нечасто встречающееся, то оно уже не включается в формирование и подсчет кандидатов.

3. Формирование двухэлементных наборов путем перебора всех возможных комбинаций, оставшихся после отбраковки одноэлементных событий, подсчет их поддержки и достоверности и отсечение наборов с уровнем поддержки ниже  $P_{\min}$  или уровнем достоверности ниже  $C_{\min}$ . Достоверностью набора  $\vec{x}_j$  называется величина:

$$C(\vec{x}_j) = \frac{N(\vec{x}_j)}{N(x_j)}, \quad (2)$$

где  $N(x_j)$  — количество транзакций, в которых встречается хотя бы одно событие из набора  $\vec{x}_j$ ;  $N(\vec{x}_j)$  — количество транзакций, в которых одновременно встречаются все события из набора  $\vec{x}_j$ .

4. Исключение из списка часто встречающихся одноэлементных наборов тех наборов, которые не входят в найденные часто встречающиеся двухэлементные наборы.

5. Формирование из оставшихся одноэлементных наборов  $n$ -элементных ( $n \geq 3$ ) наборов с уровнем поддержки ниже  $P_{\min}$  и исключение из списка часто встречающихся одноэлементных наборов тех наборов, которые не входят в найденные часто встречающиеся  $n$ -элементные наборы. Данный шаг алгоритма повторяется до выполнения любого из двух критериев остановки, указанных ниже.

Критерии остановки:

- на очередном шаге алгоритма поддержка всех наборов событий становится меньше порога  $P_{\min}$ ;
- алгоритм доходит до формирования  $M$ -элементного набора событий.

Результатом работы алгоритма является совокупность всех наборов событий, содержащих более одного события, имеющих поддержку, превышающую порог  $P_{\min}$ , и достоверность, превышающую порог  $C_{\min}$ .

Для составления таблицы транзакций используется информация о регистрации абонентов в различных (географически разнесенных) сотах сети. В качестве анализируемых используются события регистрации абонентов сети с уникальными номерами. Соответственно транзакция в данном случае представляет собой набор событий регистрации в одной соте абонентов с разными номерами, произошедших од-

новременно (в пределах заданного временного интервала), и может быть записана в виде (Абонент- $N$ , Абонент- $K$ ,... Абонент- $M$ ). Выбор указанного временного интервала определяется особенностями анализируемой системы связи.

Сформированный таким образом набор транзакций подвергается анализу при помощи алгоритма, приведенного выше. Полученные в результате работы алгоритма ассоциативные правила вида (Абонент- $X$ , Абонент- $Y$ ,... Абонент- $Z$ ) представляют собой составы выявленных групп абонентов сетей сотовой связи. О степени устойчивости данных групп можно судить по достоверности соответствующих ассоциативных правил. Об относительной частоте перемещения соответствующих групп абонентов позволяют судить значения поддержки данных ассоциативных правил.

Таким образом, использование алгоритмов поиска ассоциативных правил позволяет определить состав групп абонентов сотовой связи, степень устойчивости этих групп и относительную частоту их перемещений.

На рисунке показаны события регистрации абонентов в одной соте в течение половины рабочего дня.

Видно, что выделенные абоненты проходили регистрацию в рассматриваемой соте одновременно. Анализ показанных на рисунке событий регистрации с помощью алгоритма Argiogi, реализованного в разработанном авторами программном пакете DView, позволил автоматически выявить выделенные на рисунке красным цветом и пронумерованные группы абонентов. Сама процедура анализа заняла менее 1 с машинного времени ПЭВМ на базе процессора Pentium 4 3,2 ГГц. Поддержка ассоциативных правил, соответствующих данным группам, составляет единицы процентов в связи с тем, что большая часть абонентов регистрируется по одиночке (относительная частота групповых регистраций мала), а достоверность — приведена в таблице.

Таблица

№ группы	1	2	3	4	5
Достоверность правила, %	83,3	100	66,7	66,7	100

Из таблицы видно, что группы 2 и 5 были наиболее устойчивы в течение анализируемого временного интервала, в то время как абоненты из менее устойчивых групп 3 и 4 регистрировались и в индивидуальном порядке.

**Анализ полученных результатов показал, что ресурсоемкость алгоритма представляется авторам приемлемой, а результаты его работы достаточно просто интерпретировать. Эти обстоятельства позволяют сделать вывод о целесообразности применения алгоритма поиска ассоциативных правил Argiogi для определения устойчивых групп абонентов сетей сотовой связи.**

#### ЛИТЕРАТУРА

1. Agrawal R., Imielinski T., Swami A. Mining Associations between Sets of Items in Massive Databases / Proc. of the ACM-SIGMOD 1993 Int'l Conference on Management of Data, Washington D.C., may 1993.
2. Agrawal R., Srikant R. Fast Algorithms for Mining Association Rules / Proc. of the 20th Int'l Conference on Very Large Databases, Santiago, Chile, sept. 1994
3. Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И. Методы и модели анализа данных: OLAP и Data Mining. — СПб.: БХВ-Петербург, 2004. — 336 с.

