

# Поиск частых и нечастых элементов произведения частичных порядков и задача расшифровки двузначной монотонной функции

Н.А. Драгунов<sup>1</sup>, Е.В. Дюкова<sup>2</sup>

<sup>1</sup>Московский Государственный Университет им. М. В. Ломоносова, Ленинские горы 1, Москва, Россия, 119991

<sup>2</sup>Вычислительный Центр ФИЦ ИУ РАН, Вавилова 42, Москва, Россия, 119991

## Аннотация

Рассматриваются две эквивалентные задачи анализа данных: задача поиска по базе транзакций максимальных частых и минимальных нечастых элементов декартового произведения частичных порядков  $P$  и задача поиска верхних нулей и нижних единиц двузначной монотонной функции, определённой на элементах множества  $P$  и принимающей значение 0 и 1 соответственно на частых и нечастых элементах этого множества. Авторами предложен и экспериментально исследован оригинальный подход к решению указанных задач.

## Ключевые слова

декартово произведение частично упорядоченных множеств, максимальный частый элемент, минимальный нечастый элемент, расшифровка двузначной монотонной функции, дуализация над произведением частичных порядков

## 1. Введение

Пусть  $P = P_1 \times P_2 \times \dots \times P_n$  – декартово произведение конечных частично упорядоченных множеств;  $D$  – база данных, содержащая некоторые, не обязательно различные элементы из  $P$ ;  $s \in [0, 1]$ . Элементы множества  $P$ , содержащиеся в  $D$ , называются транзакциями. Элемент  $x \in P$  называется  $s$ -частым, если доля транзакций в  $D$ , следующих за  $x$  не менее  $s$ . Иначе  $x$  –  $s$ -нечастый. Если элемент частый и не следует ни за каким другим частым элементом, то он называется максимальным частым. Если элемент нечастый и при этом он не предшествует никакому другому нечастому элементу, то такой элемент называется минимальным нечастым. Требуется найти множества  $X_{max}$  и  $Y_{min}$ , состоящие соответственно из всех максимальных частых и минимальных нечастых элементов множества  $P$  [1, 2].

Основным приложением методов поиска частых и нечастых элементов в данных является нахождение ассоциативных правил, и вопросы разработки этих методов впервые возникли в связи с задачей анализа потребительской корзины [1]. В [2] предложена идея совместного инкрементального перечисления множеств  $X_{max}$  и  $Y_{min}$ . Инкрементальное перечисление указанных множеств имеет в основном теоретический интерес, уступая по скорости счёта очевидному последовательному способу их построения. В [3] рассмотрен последовательно-совместный метод, являющийся синтезом последовательного и совместного методов.

Поставленная задача эквивалентна задаче расшифровки двузначной монотонной функции  $F_D$ , определённой на  $P$  и принимающей значение 0 и 1 соответственно на частых и нечастых элементах этого множества. В настоящей работе частично представлены результаты, анонсированные авторами в [3], и предложен оригинальный алгоритм расшифровки  $F_D$ .

## 2. Основные результаты

На множестве  $P$  определим функцию  $F_D(x) = 1 - I(v(x) \geq s)$ , где  $I$  – индикаторная функция. Функция  $F_D$  монотонна, то есть, если  $x \preceq y$ ,  $x, y \in P$ , то  $F_D(x) \leq F_D(y)$ . Если

$F_D(x) = 0$ , то  $x$  – ноль функции  $F_D$ . Если  $F_D(x) = 1$ , то  $x$  – единица функции  $F_D$ . Если ноль функции  $F_D$  таков, что любой следующий за ним элемент множества  $P$  является единицей, то такой ноль называется верхним нулем функции  $F_D$ . Если единица функции  $F_D$  такова, что любой предшествующий ей элемент множества  $P$  является нулем, то такая единица называется нижней единицей функции  $F_D$ . Нетрудно видеть, что задача поиска верхних нулей и нижних единиц функции  $F_D$  эквивалентна задаче построения множеств  $X_{max}$  и  $Y_{min}$ . Таким образом, функция  $F_D$  задана при помощи оператора  $B(F_D)$ , который для любого элемента  $x$  из  $P$  возвращает значение  $F_D(x)$  путём вычисления частоты встречаемости  $x$  в  $D$ .

Задача поиска нулей и единиц двузначной монотонной функции  $F(x)$ , заданной при помощи некоторого оператора  $B$ , который для любого  $x \in P$  выдаёт значение  $F(x)$ , называется задачей расшифровки функции  $F(x)$ . Традиционный подход к задаче расшифровки основан на построении оптимального по Шеннону алгоритма (предложен В.К. Коробковым в 1965 г.). Согласно этому подходу сложность алгоритма расшифровки  $F(x)$  следует оценивать числом обращений к оператору  $B$ . Задача построения оптимального алгоритма расшифровки монотонной булевой функции решена Ж. Анселем в 1968 г. В общем случае эта задача решена в [4].

В представляемой работе в качестве основных приводятся результаты сравнительного анализа двух методов расшифровки функции  $F_D$ . Первый использует предложенную в [3] идею последовательно-совместного перечисления  $X_{max}$  и  $Y_{min}$ . Вторым является оптимальным по Шеннону и основан на разбиении множества  $P$  на цепи [4]. На модельных и реальных данных для каждого тестируемого метода оценивается время работы и число обращений к оператору  $B(F_D)$ . Экспериментальные исследования проведены для случая, когда  $P$  – декартово произведение конечных цепей. В последовательно-совместной расшифровке использовался асимптотически оптимальный алгоритм дуализации над произведением цепей.

### 3. Заключение

Исследованы актуальные вопросы снижения временных затрат, возникающие при логическом анализе данных с элементами из декартового произведения конечных частично упорядоченных множеств (произведения частичных порядков). Для задачи поиска максимальных частых и минимальных нечастых элементов произведения частичных порядков и задачи расшифровки двузначной монотонной функции, принимающей значение 0 и 1 соответственно на частых и нечастых элементах рассматриваемого вида данных, показана целесообразность применения асимптотически оптимальных алгоритмов дуализации над произведением частичных порядков.

### 4. Благодарности

Работа выполнена при частичной финансовой поддержке РФФИ (проект № 19-01-00430-а).

### 5. Литература

- [1] Aggarwal, Ch.C. Frequent Pattern Mining / Ch.C. Aggarwal, J. Han. – Springer International Publishing, 2014. – 471 p.
- [2] Elbassioni, K. On Finding Minimal Infrequent Elements in Multi-dimensional Data Defined Over Partially Ordered Sets / K. Elbassioni // ArXiv: 1411.2275 (09.11.2014).
- [3] Драгунов, Н.А. Поиск минимальных нечастых и максимальных частых наборов в частично упорядоченных данных / Н.А. Драгунов, Е.В. Дюкова // Математические методы распознавания образов: Тезисы докладов 19-й Всероссийской конференции с международным участием. – 2019. – С. 10-12.
- [4] Алексеев, В.Б. О расшифровке некоторых классов монотонных многозначных функций / В.Б. Алексеев // Журн. вычисл. матем. и матем. физики. – 1976. – Т. 16, № 1. – С. 189-198.