Секция 8

ВЫЧИСЛИТЕЛЬНЫЕ МЕТОДЫ В ДИСКРЕТНОЙ МАТЕМАТИКЕ

УДК 519.7 DOI 10.17223/2226308X/10/60

АЛГОРИТМ ПОСТРОЕНИЯ НЕИЗБЫТОЧНОГО МИНИМАКСНОГО БАЗИСА СТРОГИХ АССОЦИАТИВНЫХ ПРАВИЛ

В. В. Быкова, А. В. Катаева

Ассоциативные правила — тип зависимостей между данными, которые отражают, какие признаки или события встречаются совместно и насколько часто это происходит. Строгие ассоциативные правила представляют интерес для тех приложений, где требуется высокая степень уверенности в установленных зависимостях между данными, например в информационной безопасности, анализе компьютерных сетей и медицине. Чрезмерно большое число выявленных правил существенно усложняет их экспертный анализ и применение. Для решения этой проблемы предложен алгоритм MClose, формирующий для заданного бинарного контекста неизбыточное множество минимаксных строгих ассоциативных правил. Алгоритм основан на свойствах замкнутых множеств.

Ключевые слова: *соответствия Галуа, замкнутые множества, строгие ассоциативные правила, неизбыточность, минимаксный базис.*

При поиске ассоциативных правил анализируемое множество данных обычно описывается бинарным контекстом — матрицей, строки которой отвечают объектам предметной области, а столбцы — признакам этих объектов. Единичное значение элемента матрицы трактуется как наличие у объекта соответствующего признака, а нулевое — как его отсутствие. Бинарное представление данных существенно расширяет математический аппарат для их исследования. Многие современные методы поиска строгих ассоциативных правил базируются на свойствах замкнутых множеств [1, 2]. Строгие ассоциативные правила имеют достоверность 1 и представляют интерес для тех приложений, где требуется высокая степень уверенности в установленных зависимостях между данными, например в информационной безопасности и анализе компьютерных сетей [3, 4].

Главная проблема при поиске ассоциативных правил — это огромное число правил, возникающих при анализе больших контекстов. Для решения этой проблемы используются различные меры значимости. С их помощью правила фильтруются и для анализа предъявляются только те, для которых значения мер значимости превышают заданные пороговые значения. Подобная фильтрация, конечно, уменьшает число правил, но не решает проблему размерности полностью. Часто после фильтрации всё равно остается значительное число правил, при этом многие из них избыточные. Ассоциативное правило считается избыточным, если его удаление из множества правил не приводит к потере информации о связях между данными в рассматриваемой предметной области. Множество ассоциативных правил, не содержащее избыточных (в некотором смысле) правил, принято называть базисом.

В настоящее время известен ряд алгоритмов, позволяющих строить различные базисы для строгих ассоциативных правил. Наиболее значимыми базисами являются канонический и минимаксный. Канонический базис (или базис Дюкена — Гига) состоит из минимального числа строгих ассоциативных правил, рекуррентно описываемых в терминах псевдосодержаний [5]. Канонический базис математически глубоко исследован, однако все предложенные на сегодняшний день алгоритмы его построения в большей степени представляют теоретический, чем практический интерес. Минимаксный базис состоит из строгих ассоциативных правил, имеющих минимальную посылку и максимальное следствие. Для построения минимаксного базиса имеются хорошо апробированные практикой алгоритмы, к которым относится алгоритм Close [6]. В данной работе предлагается алгоритм MClose, расширяющий возможности алгоритма Close. Алгоритм MClose формирует для бинарного контекста неизбыточное множество минимаксных строгих ассоциативных правил.

Для описания сути предлагаемого алгоритма введём необходимые понятия и обозначения. Пусть для предметной области определены два непустых конечных множества G и M объектов и признаков соответственно. Предполагаем, что все объекты в G и признаки в M различны. Пусть задано отношение $I \subseteq G \times M$ инцидентности между G и M. Тройку $K = (G, M, I)$ принято называть контекстом предметной области. Считаем, что существование в I пары (g, m) означает, что объект g имеет признак m, и наоборот, признак m присущ объекту g.

Выберем два произвольных элемента $g \in G$ и $m \in M$. Определим для них два отображения ϕ и ψ: $\phi(g) = \{m \in M : (g, m) \in I\}$ — множество признаков, присущих объекту g; $\psi(m) = \{g \in G : (g, m) \in I\}$ — множество объектов, которые обладают признаком m. Отображения ϕ и ψ обобщаются на $A \subseteq G$ и $B \subseteq M$ следующим образом:

$$\phi(A) = \bigcap_{g \in A} \phi(g) = \{m \in M : \forall g \in A \ ((g, m) \in I)\},$$
$$\psi(B) = \bigcap_{m \in B} \psi(m) = \{g \in G : \forall m \in B \ ((g, m) \in I)\}.$$

Таким образом, $\phi(A)$ — множество признаков, общих для всех объектов из A; $\psi(B)$ — множество объектов, которые обладают всеми признаками из B. Если для отображений ϕ и ψ применить единое обозначение $(\cdot)'$, то формулы для $\phi(A)$, $\psi(B)$ записываются так:

$$A' = \bigcap_{g \in A} g' = \{m \in M : \forall g \in A \ ((g, m) \in I)\},$$
$$B' = \bigcap_{m \in B} m' = \{g \in G : \forall m \in B \ ((g, m) \in I)\}.$$

Из определения этих отображений вытекают свойства, которые формально можно выразить в виде следующих утверждений.

Утверждение 1. Для всякого контекста $K = (G, M, I)$ и любых $B_1, B_2 \subseteq M$ верны следующие свойства:

— *антимонотонность*: если $B_1 \subseteq B_2$, то ${B_2}' \subseteq {B_1}'$;

— *экстенсивность*: $B_1 \subseteq {B_1}''$, где ${B_1}'' = ((B_1)')' \subseteq M$.

Утверждение 2. Для всякого контекста $K = (G, M, I)$ и любых $A_1, A_2 \subseteq G$ верны следующие свойства:

— *антимонотонность*: если $A_1 \subseteq A_2$, то ${A_2}' \subseteq {A_1}'$;

— *экстенсивность*: $A_1 \subseteq {A_1}''$, где ${A_1}'' = ((A_1)')' \subseteq G$.

В силу утверждений 1 и 2, отображения ϕ и ψ составляют пару соответствий Галуа между множествами 2^G и 2^M, частично упорядоченными по включению [1]. Двойное применение отображения «$'$» определяет оператор замыкания на 2^M в алгебраическом смысле [2]. Множество признаков $B \subseteq M$, для которого $B = B''$, называется замкнутым относительно оператора «$''$» в контексте $K = (G, M, I)$.

Ассоциативным правилом контекста $K = (G, M, I)$ называется упорядоченная пара $r = (X, Y)$, где $X, Y \subseteq M$. Ассоциативное правило $r = (X, Y)$ записывается в виде $X \Rightarrow Y$, а множества X и Y называются посылкой (или причиной) и заключением (или следствием) соответственно. Ассоциативное правило $X \Rightarrow Y$ количественно характеризуется с помощью поддержки и достоверности. Поддержка $\delta(X)$ множества $X \subseteq M$ в контексте $K = (G, M, I)$ определяется как отношение числа объектов, которым присущи признаки X, к общему число объектов, представленных в этом контексте: $\delta(X) = |X'|/|G|$. В силу антимонотонности отображения «$'$», функция поддержки также удовлетворяет свойству антимонотонности: для любых $X, Y \subseteq M$ при $X \subseteq Y$ верно неравенство $\delta(Y) \leqslant \delta(X)$. Множество признаков $X \subseteq M$ считается частым в $K = (G, M, I)$, если его поддержка больше или равна пороговому значению $\delta_0 \in [0, 1]$. Если $\delta(X) \geqslant \delta_0$ и $X = X''$, то множество X называется частым замкнутым множеством признаков в $K = (G, M, I)$.

Поддержка ассоциативного правила $X \Rightarrow Y$ относительно $K = (G, M, I)$ — величина $\delta(X \Rightarrow Y) = |(X \cup Y)'|/|G|$, указывающая, какая доля объектов этого контекста имеет признаки $X \cup Y$. Достоверность ассоциативного правила $X \Rightarrow Y$ — отношение числа объектов, обладающих всеми признаками из $X \cup Y$, к числу объектов, которым свойственны только признаки X: $\gamma(X \Rightarrow Y) = |(X \cup Y)'|/|X'|$. Всегда $0 \leqslant \gamma(X \Rightarrow Y) \leqslant 1$. Чем ближе $\gamma(X \Rightarrow Y)$ к 1, тем с большей уверенностью можно сказать, что признаки Y появляются в $K = (G, M, I)$ вместе с признаками X. Строгие ассоциативные правила имеют достоверность 1.

Известно большое число алгоритмов поиска ассоциативных правил. Основополагающими являются алгоритмы Apriori и Close [6]. Алгоритм Apriori работает с частыми множествами признаков и базируется на свойстве антимонотонности функции поддержки. Алгоритм Close использует свойства частых замкнутых множеств. Множество минимаксных строгих ассоциативных правил, полученных в результате работы алгоритма Close, образует минимаксный базис. Этот базис допускает дальнейшее упрощение на основе следующий теоремы.

Теорема 1. Для всякого контекста $K = (G, M, I)$ и любых $X, Y, Z, W \subseteq M$ справедливы следующие свойства строгих ассоциативных правил:

D_1. *Рефлексивность*: $X \Rightarrow X$.

D_2. *Пополнение*: если $X \Rightarrow Y$, то $X \cup Z \Rightarrow Y$.

D_3. *Аддитивность*: если $X \Rightarrow Y$ и $X \Rightarrow Z$, то $X \Rightarrow Y \cup Z$.

D_4. *Проективность*: если $X \Rightarrow Y$ и $Z \subseteq Y$, то $X \Rightarrow Z$.

D_5. *Транзитивность*: если $X \Rightarrow Y$ и $Y \Rightarrow W$, то $X \Rightarrow W$.

D_6. *Псевдотранзитивность*: если $X \Rightarrow Y$ и $Y \cup Z \Rightarrow W$, то $X \cup Z \Rightarrow W$.

Выводимости D_1–D_6 позволяют из некоторого множества строгих ассоциативных правил вывести многие другие строгие ассоциативные правила без дополнительного сканирования контекста. Выводимости, подобные D_1–D_6, справедливы и для функциональных зависимостей, имеющих место в теории реляционных баз данных, где их принято называть аксиомами Амстронга. В [7] выводимости D_1–D_6 доказаны применительно к функциональным зависимостям. В данной работе они доказаны на основе

соответствий Галуа и свойств замкнутых множеств. Доказано также, что выводимости D_1, D_3, D_4, D_5 гарантируют сохранение поддержки: результатом применения их к строгим ассоциативным правилам с поддержкой не менее чем δ_0 всегда являются строгие ассоциативные правила с таким же порогом поддержки. Именно выводимости D_1, D_3, D_4, D_5 применяются в алгоритме MClose для распознавания избыточных строгих ассоциациативных правил и построения неизбыточного минимаксного базиса. Показано, что алгоритм MClose по времени работы сопоставим с алгоритмом Close. Между тем на практике он более чем в 2 раза уменьшает мощность минимаксного базиса, формируемого алгоритмом Close.

Подробное изложение представленных результатов можно найти в [8].

ЛИТЕРАТУРА

1. *Биркгоф Г., Барти Т.* Современная прикладная алгебра. СПб.: Лань, 2005. 400 с.
2. *Гуров С. И.* Булевы алгебры, упорядоченные множества, решетки: определения, свойства, примеры. М.: Книжный дом «ЛИБРОКОМ», 2013. 352 с.
3. *Батура Т. В.* Модели и методы анализа компьютерных социальных сетей // Программные продукты и системы. 2013. № 3. С. 130–137.
4. *Платонов В. В., Семенов П. О.* Методы сокращения размерности в системах обнаружения сетевых атак // Проблемы информационной безопасности. Компьютерные системы. 2012. № 3. С. 40–45.
5. *Кузнецов С. О.* Автоматическое обучение на основе анализа формальных понятий // Автоматика и телемеханика. 2001. № 10. С. 3–27.
6. *Zaki M. J and Hsiao C.-J.* Efficient algorithms for mining closed itemsets and their lattice structure // IEEE Trans. Knowledge Data Eng. 2005. V. 17. No. 4. P. 462–478.
7. *Майер Д.* Теория реляционных баз данных. М.: Мир, 1987. 608 с.
8. *Быкова В. В., Катаева А. В.* О неизбыточном представлении минимаксного базиса строгих ассоциативных правил // Прикладная дискретная математика. 2017. № 36. С. 113–126.

УДК 519.7 DOI 10.17223/2226308X/10/61

ОБРАЩЕНИЕ КРИПТОГРАФИЧЕСКИХ ХЕШ-ФУНКЦИЙ С ИСПОЛЬЗОВАНИЕМ НЕСБАЛАНСИРОВАННЫХ ПРИБЛИЖЕНИЙ РАУНДОВЫХ ФУНКЦИЙ[1]

И. А. Грибанова

Представлены результаты решения задач обращения неполнораундового варианта криптографической хеш-функции MD4 с использованием новой техники, которая включает в себя следующие этапы: замену некоторых раундовых подфункций MD4 несбалансированными булевыми функциями; решение полученной изменённой задачи; использование части информации из решения изменённой задачи для перехода к решению исходной задачи. Предлагаемая техника комбинируется с дополнительными условиями на переменные сцепления, введёнными ранее Г. Доббертином. Проведённые вычислительные эксперименты демонстрируют работоспособность предлагаемого подхода в применении к задаче обращения 39-шаговой версии MD4 (MD4-39).

Ключевые слова: *криптоанализ, обращение хеш-функций, MD4, SAT.*

[1]Работа поддержана грантом РНФ № 16-11-10046.