

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение
высшего образования
«КРЫМСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ имени В. И. ВЕРНАДСКОГО»
(ФГАОУ ВО «КФУ им. В. И. Вернадского»)

Таврическая академия (структурное подразделение)

Факультет математики и информатики
Кафедра информатики

Сечин Михаил Александрович

НЕИЗБЫТОЧНОМ ПРЕДСТАВЛЕНИИ МИНИМАКСНОГО
БАЗИСА СТРОГИХ АССОЦИАТИВНЫХ ПРАВИЛ

Выпускная квалификационная работа

Обучающегося 4 курса

Направления подготовки 01.03.02. Прикладная математика и информатика

Форма обучения очная

Научный руководитель
доцент кафедры информатики,

кандидат технических наук

Ильченко.А.В

К ЗАЩИТЕ ДОПУСКАЮ:

Заведующий кафедрой,
доктор физико-математических наук,
профессор

В. И. Донской

Симферополь, 2019

Оглавление

Введение.....	3
1. Минимальные генераторы формального контекста	3
1.1. Формальный контекст	3
1.2. Контекстное следование признаковых подмножеств	5
1.3. Оператор контекстного замыкания	5
1.4. Отношение контекстной равносильности и минимальные генераторы .	5
1.5. Минимаксные ассоциативные правила.....	7
1.6. Поддержка предикатного подмножества	8
1.7. Свойства строгих ассоциативных правил	10
1.8. Алгоритм <i>CLOSE</i> и <i>M-CLOSE</i>	13
1.9. Программная реализация алгоритма MClose и результаты.....	16
Заключение	19
Список литературы	19
Приложение	20

Введение

1. Минимальные генераторы формального контекста

1.1. Формальный контекст

Пусть для предметной области определены два непустых конечных множества G (объектов) и M (свойств). Предполагаем, что все объекты в множествах G и свойства в M различны. Пусть задано бинарное отношение $I \subseteq G \times M$. Один из способов задания предметной области с помощью **таблицы 1**, где $\{g1, g2, g3, g4, g5\} \in G$ объекты, $\{m1, m2, m3, m4, m5\} \in M$ свойства. Считаем, что пара существует (g, m) в I означает, что объект g имеет свойство m , и наоборот — свойству m соответствует объект g (в таблице отмечено 1).

Таблица 1.

	$m1$	$m2$	$m3$	$m4$	$m5$
$g1$	1	1	0	0	1
$g2$	0	1	0	1	1
$g3$	1	0	0	0	1
$g4$	1	1	1	1	1

Тройку $K = (G, M, I)$ принято называть **формальным контекстом** предметной области.

Выберем в $K = (G, M, I)$ два произвольных элемента $g \in G$ и $m \in M$. Определим для них два отображения φ и ψ : $\varphi(g) = \{m \in M : (g, m) \in I\}$, $\psi(m) = \{g \in G : (g, m) \in I\}$, где $\varphi(g)$ — множество свойств, которые есть у объекта g , а $\psi(m)$ — множество объектов, обладающих признаком m .

Тогда отображения φ и ψ обобщаются для произвольных множеств $A \subseteq G$ и $B \subseteq M$.

$$\varphi(A) = \bigcap_{g \in A} \varphi(g) = \{m \in M \mid \forall g \in A (g, m) \in I\}$$

$$\psi(B) = \bigcap_{m \in B} \psi(m) = \{g \in G \mid \forall m \in B (g, m) \in I\}$$

Для отображений φ и ψ определим единое обозначение $()'$ и назовем его оператором Галуа.

Оператор Галуа A' , где $A \subseteq G$ ставит множеству объектов в соответствие максимальное множество признаков так, что каждый объект находится в отношении с каждым признаком. Например, для примера из **Таблицы 1** $\{g1, g3\}' = \{m1, m2, m5\}$. Аналогично, оператор Галуа ставит множеству признаков в соответствие максимальное множество объектов, так что каждый признак находится в отношении с каждым объектом. Например, для примера из **Таблицы 1**, $\{m2\}' = \{g1, g2, g4\}$.

Тогда формулы отображения φ и ψ записываются так:

$$g' = \{m \in M: (g, m) \in I\}, m' = \{g \in G: (g, m) \in I\}$$

А для произвольных множеств:

$$\begin{aligned} A' &= \bigcap_{g \in A} g' = \{m \in M \mid \forall g \in A (g, m) \in I\} \\ B' &= \bigcap_{m \in B} m' = \{g \in G \mid \forall m \in B (g, m) \in I\} \\ (A1 \cup A2)' &= A1' \cap A2'; \\ (B1 \cup B2)' &= B1' \cap B2'. \end{aligned}$$

Из определения отображения «'» следуют эти свойства:

Утверждение 1. Для любого 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 любого 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 2^G и 2^M — системами всех подмножеств множеств G и M соответственно, частично упорядоченными по теоретико-множественному включению [4, 5]. Известно, что для соответствий Галуа φ и ψ справедливы равенства [5]

$$\varphi(\psi(\varphi(A))) = \varphi(A), \psi(\varphi(\psi(B))) = \psi(B)$$

или, то же самое, в единых обозначениях

$$((A')')' = (A'')' = A',$$

$$((B')')' = (B'')' = B'.$$

1.2. Контекстное следование признаков подмножеств

?

1.3. Оператор контекстного замыкания

Определение 2. Определим оператор замыкания $A'' = (A')'$. При применении оператора Галуа (\prime) дважды мы получим замыкание исходного множества.

Для примера применения оператора замыкания рассмотрим в Таблицы 1, $(\{g1, g3\})' = \{m1, m2, m5\}' = (\{g1, g2, g3, g4, g5\})$. Оператор замыкания обладает следующими свойствами:

- *Рефлексивность*: для любого $B \subseteq M$ всегда $B \subseteq B''$;
- *Монотонность*: если $B_1 \subseteq B_2 \subseteq M$, то $B_1'' \subseteq B_2'' \subseteq M$;
- *Идемпотентность*: для любого $B \subseteq M$ всегда $(B'')'' = B'$.

Справедливость этих свойств вытекает из утверждений 1 и 2.

Если $B = B''$, то множество признаков $B \subseteq M$ называется замкнутым относительно оператора Галуа в контексте. Множество $B'' = \varphi(\psi(B))$ можно трактовать как набор признаков, которые всегда появляются в объектах контекста K вместе с признаками из B , причём это множество является наибольшим по включению в пределах K . Очевидно, что $(\emptyset)'' = \varphi(\psi(\emptyset)) = G'$, где G' — множество признаков, свойственных всем объектам контекста K . Если $B' = \emptyset$, то всегда $B'' = \varphi(\psi(B)) = \varphi(\emptyset) = M$. Если $B' \neq \emptyset$, то, исходя из (1) – (4), замыкание для $B \subseteq M$ можно вычислить за один просмотр контекста K по форму

$$B' = \cap_{g \in G} \{g' | B \subseteq g'\}.$$

В анализе формальных понятий пара множеств $(A, B), A \subseteq G, B \subseteq M$, таких, что $A' = B$ и $B' = A$, называется формальным понятием контекста $K = (G, M, I)$ с объёмом A и содержанием B [6]. В формальном понятии (A, B) множества A и B всегда замкнуты относительно «''» в этом контексте: $A = A''$ и $B = B''$. Замкнутые множества также нашли широкое применение в поиске ассоциативных правил [7, 14, 15].

1.4. Отношение контекстной равносильности и минимальные генераторы

Определение 3. Пусть $A, B \subseteq M$. Подмножества признаков A и B равносильны ($A \equiv B$) в контексте $K = (G, M, I)$, если их отображение Галуа совпадают (подмножества объектов $A', B' \subseteq G$ равны):

$$A \equiv B \Leftrightarrow A' = B' \quad (1.3)$$

Это отношение будем называть отношением равносильности K (по имени контекста).

Пусть A, B и C произвольные подмножества признаков. Очевидно, отношение контекстной равносильности обладает следующими свойствами:

$$- \text{Рефлексивность: } A \equiv A \quad (1.4)$$

$$- \text{Симметричность: если } A \equiv B, \text{ то } B \equiv A \quad (1.5)$$

$$- \text{Транзитивность: если } A \equiv B \text{ и } B \equiv C, \text{ то } A \equiv C \quad (1.6)$$

Таким образом, в силу наличия свойств (1.4) – (1.6), отношение контекстной равносильности есть отношение эквивалентности и, следовательно, разбивает семейство признаков подмножеств на классы K - равносильных подмножеств.

$[A]_K = \{B \subseteq M : B \equiv A\}$ – обозначение всех признаков подмножеств, K - равносильных признаковому подмножеству A .

$\langle [\cdot]_K, \subseteq \rangle$ – обозначение произвольного класса K - равносильных признаков подмножеств вместе с унаследованным отношением порядка " \subseteq ".

Каждый класс K - равносильных признаков подмножеств, вместе с отношением порядка " \subseteq ", есть множество, упорядоченное.

Каждый класс K - равносильных признаков подмножеств обладает следующими свойствами.

1. Если $A, B \in [\cdot]_K$ элементы из одного класса, то и их объединение $A \cup B \in [\cdot]_K$ также является элементом того же самого класса.

2. $\bigcup_{A \in [\cdot]_K} A \in [\cdot]_K$ – объединение всех элементов одного класса K - равносильных элементов, есть элемент этого же класса.

3. В каждом классе K - равносильных элементов существует единственный максимальный (по включению) элемент. Этот элемент совпадает с объединением всех элементов этого класса.

4. В каждом классе K - равносильных элементов существует не меньше одного минимального (по включению) элемента.

Пусть $X \subseteq M$ – замкнутое подмножество, то есть $X'' = X$. Подмножество признаков $Y \subseteq M$ называется генератором замкнутого подмножества X , если замыкание Y совпадает с множеством X :

$$Y'' = X.$$

Утверждение 3. Каждый элемент класса равносильности является генератором максимального элемента этого класса.

Следствие 3. Минимальный элемент класса равносильности является генератором максимального элемента этого класса.

Определение 4. Такой, минимальный, элемент класса принято называть *минимальным генератором*.

Множество $\rho \subseteq M$ называется минимальным генератором замкнутого набора признаков $Y \subseteq M$, $Y = Y''$, тогда и только тогда, когда $\rho'' = Y$ и не существует другого множества $\tau \subseteq M$, такого, что $\tau \subset \rho$ и $\tau'' = Y$.

Иначе, генератор замкнутого набора признаков Y наименьшее по мощности множество признаков, имеющее замыкание Y . Количество признаков, из которых состоит генератор ρ , называется мощностью этого генератора. Если $|\rho| = k$, то ρ является k -генератором.

1.5. Минимаксные ассоциативные правила

Ассоциативным правилом на множестве признаков M контекста $K = (G, M, I)$ называется упорядоченная пара множеств $r = (X, Y)$ где $X, Y \subseteq M$. Записывают ассоциативное правило $r = (X, Y)$ в виде $X \Rightarrow Y$, где множество X посылка и Y называют заключением (или следствием) [16].

Ассоциативное правило $X \Rightarrow Y$ называется *минимаксным* в контексте $K = (G, M, I)$, если не существует другого ассоциативного правила $X^* \Rightarrow Y^*$, такого, что $X^* \subseteq X$ и $Y \subseteq Y^*$ и

$$\delta(X^* \Rightarrow Y^*) = \delta(X \Rightarrow Y), \gamma(X^* \Rightarrow Y^*) = \gamma(X \Rightarrow Y).$$

Минимаксный базис – это базис состоящий из минимаксных строгих ассоциативных правил, у которых минимальная посылка и максимальное следствие. Для построения минимаксного базиса существует ряд алгоритмов, например, алгоритм Close.

Любое ассоциативное правило $X \Rightarrow Y$ количественно можно охарактеризовать с помощью поддержки $\delta(X \Rightarrow Y)$ и достоверности $\gamma(X \Rightarrow Y)$ [17]. Эти числовые функции определяются через понятие поддержки множества признаков.

1.6. Поддержка предикатного подмножества

Определение 1.6.1 Пусть дан формальный контекст $K = (G, M, I)$. *Поддержкой* подмножества признаков $X \subseteq M$ называется часть объектов множества G объектов контекста, содержащих признаки X к общему числу объектов:

$$\text{supp}(X) = |X'|/|G|. \quad (7)$$

Из формулы (7) следует, что для любого $X \subseteq M$ значение $\text{supp}(X)$ находится в границах $0 \leq \text{supp}(X) \leq 1$.

Чаще всего задается пороговое значение $\text{minsupp} > 0$ для того чтобы отсеять те подмножества признаков, которые не являются решающими.

Определение 1.6.2 Пусть дан формальный контекст $K = (G, M, I)$. *Частое множество* — это множество признаков $X \subseteq M$ поддержка которого больше порогового значения: $\text{supp}(X) \geq \text{minsupp}$.

Утверждение 1. Для всякого контекста $K = (G, M, I)$ и любого множества $X \subseteq M$ поддержка множества X совпадает с его замыканием X'' :

$$\text{supp}(X'') = \text{supp}(X).$$

Следствие 1. Замыкание частого множества признаков тоже является частым, так как совпадают их поддержки.

Определение 1.6.3 Пусть дан формальный контекст $K = (G, M, I)$. Множество признаков $XC \subseteq M$ называется *частым замкнутым множеством признаков*, если $\text{supp}(XC) \geq \text{minsupp}$ и $XC = XC''$.

Утверждение 2. Для всякого контекста $K = (G, M, I)$ и любых $X, Y \subseteq M$ при $X \subseteq Y$ верно неравенство

$$\delta(Y) \leq \delta(X). \quad (9)$$

То есть, поддержка подмножества не меньше поддержки надмножества. То есть поддержка, как и отображения Галуа – функции антимонотонны.

Следствия из утверждения 2:

Следствие 1. Все подмножества частого множества признаков являются частыми.

Следствие 2. Все надмножества множества признаков, не являющегося частым, не частые.

Следствие 2 позволяет создавать алгоритмы, использующие поуровневый поиск частых множеств признаков. А следствие 1 помогает сократить число шагов такого поиска, т.е. не рассматривать надмножества множеств признаков, не являющихся частыми. Алгоритмы поиска частых множеств признаков, использующие эти два свойства, называются поуровневыми.

Поддержкой $\text{supp}(X \Rightarrow Y)$ ассоциативного правила $X \Rightarrow Y$ в $K = (G, M, I)$ называется величина отношения числа объектов, у которых есть признаки $X \cup Y$, к общему числу объектов, в этом контексте.

$$\text{supp}(X \Rightarrow Y) = \text{supp}(X \cup Y) = |(X \cup Y)'|/|G|, \quad (10)$$

Достоверность $\gamma(X \Rightarrow Y)$ ассоциативного правила $X \Rightarrow Y$ в контекста K находится как отношение числа объектов, обладающих всеми

признаками из $X \cup Y$, к числу объектов, обладающие только признаками X :

$$\gamma(X \Rightarrow Y) = |(X \cup Y)'|/|X'|.$$

Достоверность ассоциативного правила через поддержку:

$$\gamma(X \Rightarrow Y) = \text{supp}(X \Rightarrow Y)/\text{supp}(X) = \text{supp}(X \cup Y)/\text{supp}(X).$$

Если $\gamma(X \Rightarrow Y) = 1$, то такое ассоциативное правило будем называть *строгим ассоциативным правилом*. Таким образом, строгое ассоциативное правило — правило с достоверностью 1 и любой ненулевой поддержкой.

1.7. Свойства строгих ассоциативных правил

Известен критерий наличия в контексте строгого ассоциативного правила [6]. Приведём его с доказательством.

Утверждение 5. Достоверность ассоциативного правила $X \Rightarrow Y$ относительно контекста $K = (G, M, I)$ равна 1 тогда и только тогда, когда $X' \subseteq Y'$ (или $Y \subseteq X''$).

Заметим, что утверждение 5 выполняется так же для $X \Rightarrow X''$ во всяком контексте K и при любом $X \subseteq M$. Рассмотрим некоторые частные случаи утверждения 5, важные с точки зрения устранения избыточности в множестве строгих ассоциативных правил.

Случай 1: Ассоциативные правила вида $X \Rightarrow Y$ при любых $Y \subseteq X \subseteq M$. В силу антимонотонности 6 отображения «'» при $Y \subseteq X$ справедливо включение $X' \subseteq Y'$. Условие утверждения 5 выполняется, поэтому $\gamma(X \Rightarrow Y) = 1$. Таким образом, если в ассоциативном правиле заключение является подмножеством посылки, то такое правило имеет достоверность 1 в любом контексте K с поддержкой $\delta(X)$. Подобные строгие ассоциативные правила не несут в себе информации о существенных отношениях между множествами признаков X и Y , поэтому их следует считать тривиальными и не принимать во внимание. В частности, ассоциативные правила вида $\emptyset \Rightarrow \emptyset$, $M \Rightarrow Y$ и $X \Rightarrow \emptyset$ при любых $X, Y \subseteq M$ относятся к тривиальным строгим правилам.

Случай 2: Ассоциативные правила вида $\emptyset \Rightarrow Y$ при $\emptyset \neq Y \subseteq M$.

Для всякого контекста $K = (G, M, I)$ и $Y \subseteq M$ всегда $Y' \subseteq G$. Кроме того, $\psi(\emptyset) = \emptyset' = G$ и $\delta(\emptyset) = 1$. Поэтому для правила $\emptyset \Rightarrow Y$ имеем

$$\gamma(\emptyset \Rightarrow Y) = \delta(\emptyset \cup Y) / \delta(\emptyset) = \delta(Y).$$

Исходя из утверждения 5, равенство $\gamma(\emptyset \Rightarrow Y) = 1$ имеет место тогда и только тогда, когда $\emptyset' \subseteq Y'$. Поскольку $\psi(\emptyset) = \emptyset' = G$, то $\emptyset' \subseteq Y'$ верно лишь при $G = Y'$. Таким образом, строгое ассоциативное правило $\emptyset \Rightarrow Y$ при $Y = \emptyset$ имеет поддержку $\delta(\emptyset) = 1$ и отражает наличие жёсткого ограничения на контекст K : все объекты, представленные в этом контексте, обязательно обладают множеством признаков Y .

Рассмотрим строгое ассоциативное правило $X \Rightarrow Y$. В силу (11) его поддержка всегда совпадает с поддержкой его посылки: $\delta(X \Rightarrow Y) = \delta(X)$. Если $\delta(X) \geq \delta_0$, то также $\delta(X \Rightarrow Y) \geq \delta_0$. Если после какого-либо изменения правила $X \Rightarrow Y$ результирующее правило имеет поддержку не менее $\delta(X)$, то говорят, что такое изменение сохраняет поддержку исходного правила. Докажем свойства строгих ассоциативных правил, которые позволяют из одних строгих ассоциативных правил вывести другие строгие ассоциативные правила (с сохранением или без сохранения поддержки).

Лемма 1. Пусть в контексте $K = (G, M, I)$ множество $X \subseteq M$ имеет поддержку $\delta(X) \geq \delta_0$. Тогда для контекста K при любом $Y \subseteq X$ всегда справедливо строгое ассоциативное правило $X \Rightarrow Y$ с поддержкой $\delta(X \Rightarrow Y) \geq \delta_0$.

Лемма 2. Если для контекста $K = (G, M, I)$ справедливо строгое ассоциативное правило $X \Rightarrow Y$ с поддержкой $\delta(X)$, то при любом $Z \subseteq M$ для этого контекста также верно строгое ассоциативное правило $X \cup Z \Rightarrow Y$ с поддержкой $\delta(X \cup Z) \leq \delta(X)$.

Лемма 2 утверждает, что строгое ассоциативное правило сохраняется при пополнении посылки. Поддержку этого правила может уменьшиться.

Следствие 1. Если для контекста $K = (G, M, I)$ справедливо строгое ассоциативное правило $X \Rightarrow Y$ с поддержкой $\delta(X)$, то при любом $Z \subseteq Y$ для этого контекста также справедливо строгое ассоциативное правило $X \cup Z \Rightarrow Y$ с поддержкой $\delta(X)$.

Лемма 3. Пусть в контексте $K = (G, M, I)$ множество $X \subseteq M$ имеет поддержку $\delta(X) \geq \delta_0$. Если в контексте K существуют строгие ассоциативные правила $X \Rightarrow Y$ и $X \Rightarrow Z$, то для этого контекста также

справедливо строгое ассоциативное правило $X \Rightarrow Y \cup Z$ с поддержкой $\delta(X) \geq \delta_0$.

Легко убедиться, что свойства рефлексивности и пополнения невыполнимы для произвольных (δ_0, γ_0) - ассоциативных правил. Однако следующее свойство, называемое *проективностью*, выполняется для любых (δ_0, γ_0) -ассоциативных правил.

Лемма 4. Если в контексте $K = (G, M, I)$ существует (δ_0, γ_0) - ассоциативное правило $X \Rightarrow Y$, то при любых $Z \subseteq Y$ и $Y \neq \emptyset$ для этого контекста также верно (δ_0, γ_0) - ассоциативное правило $X \Rightarrow Z$.

Лемма 4 утверждает, что следствие всякого (δ_0, γ_0) - ассоциативного правила можно разложить до его подмножества, сохраняя при этом поддержку и достоверность в заданных границах. Для строгих ассоциативных правил леммы 3 и 4 констатируют равноценность различных эквивалентных форм записи этих правил.

Следствие 2. Представление строгого ассоциативного правила $X \Rightarrow Y \cup Z$ эквивалентно его представлению в виде двух строгих ассоциативных правил $X \Rightarrow Y$ и $X \Rightarrow Z$, при этом $\delta(X \Rightarrow Y \cup Z) = \delta(X \Rightarrow Y) = \delta(X \Rightarrow Z) = \delta(X)$.

Лемма 5. Если в контексте $K = (G, M, I)$ существуют строгие ассоциативные правила $X \Rightarrow Y$ и $Y \Rightarrow W$ и $\delta(X) \geq \delta_0$, то какими бы ни были $X, Y, W \subseteq M$, для этого контекста также справедливо строгое ассоциативное правило $X \Rightarrow W$ с поддержкой $\delta(X) \geq \delta_0$.

Следующая лемма обобщает лемму 5 и определяет свойство, называемое *псевдотранзитивностью* строгих ассоциативных правил.

Лемма 6. Если в контексте $K = (G, M, I)$ существуют строгие ассоциативные правила $X \Rightarrow Y$ и $Y \cup Z \Rightarrow W$, то какими бы ни были $X, Y, Z, W \subseteq M$, для контекста также справедливо строгое ассоциативное правило $X \cup Z \Rightarrow W$ с поддержкой $\delta(X \cup W \cup Z) \leq \delta(X)$.

Леммы 1–6 позволяют сформулировать следующую теорему.

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

A_1 . *Рефлексивность*: $X \Rightarrow X$.

A₂. Пополнение: если $X \Rightarrow Y$, то $X \cup Z \Rightarrow Y$.

A₃. Аддитивность: если $X \Rightarrow Y$ и $X \Rightarrow Z$, то $X \Rightarrow Y \cup Z$.

A₄. Проективность: если $X \Rightarrow Y$ и $Z \subseteq Y$, то $X \Rightarrow Z$.

A₅. Транзитивность: если $X \Rightarrow Y$ и $Y \Rightarrow W$, то $X \Rightarrow W$.

A₆. Псевдотранзитивность: если $X \Rightarrow Y$ и $Y \cup Z \Rightarrow W$, то $X \cup Z \Rightarrow W$.

Указанные в теореме 1 свойства A_1 – A_6 позволяют из некоторого множества строгих ассоциативных правил вывести многие другие строгие ассоциативные правила без дополнительного сканирования контекста. Выводимости, подобные A_1 – A_6 , справедливы и для функциональных зависимостей, имеющих место в теории реляционных баз данных, где их принято называть аксиомами Амстронга. Из доказанных лемм следует, что выводимости A_1 , A_3 , A_4 , A_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.

1.8.Алгоритм *CLOSE* и *M-CLOSE*

Алгоритм *CLOSE* - это итеративный алгоритм для извлечения минимальных генераторов и частых замкнутых наборов элементов по уровням. Вовремя k -ой итерации, рассматривается список возможных k -генераторов, вычисляются их замыкания и поддержка, с помощью данных и нечастые генераторы удаляются. Частые генераторы далее используются при создании списка $(k+1)$ - генераторов. Замыкание частых генераторов — это частое замыкание множества признаков, а поддержка генератора так же является поддержкой его замыкания.

Вовремя k -ой итерации, обрабатывается множество FC_k . Каждый элемент этого множества состоит из: k - генератора, его замыкания и их поддержек. Сначала алгоритм инициализирует 1-элементные генераторы в FC_1 списком из 1 признака, а затем производятся остальные итерации. На итерации под номером k выполняются:

1. Вычисляется замыкание всех возможных k -генераторов и их поддержка. Количество объектов, находящихся в генератора определяет его поддержку, а пересечение генераторов определяет его замыкание. Каждый объект обрабатывается один раз, а значит этот шаг требует только один проход по выборке.
2. Нечастые k -генераторы, то есть генераторы с поддержкой меньше, чем Minsupp удаляются из FCk .
3. Множество возможных $(k+1)$ - элементных генераторов создается при помощи соединения k -генераторов следующим образом:
 - а) Два k - элементных генератора в FCk , у которых одинаковый элемент с индексом $k-1$ соединяются, для того чтобы создать возможный $(k+1)$ - элементный генератора. Например, 3-генераторы $\{ABC\}$ и $\{ABD\}$ будут соединены и будет получен 4-генератор $\{ABCD\}$.
 - б) Возможный $(k+1)$ - генератор, который не является частым или не минимальный удаляются.
 - с) Также, удаляются те $(k+1)$ - элементный генераторы, замыкание которых уже посчитано. Такой генератор легко найти, так как он включён в замыкание

Алгоритм останавливается, когда не возможно создать новые возможные генераторы. Тогда, каждое множество FCk содержит частые k -элементные генераторы, их замыкание и их поддержку.

Но при построении минимаксного базиса строгих ассоциативных правил с помощью алгоритма *Close* могут содержаться избыточные строгие ассоциативные правила. Алгоритм *MClose* с помощью выводов A_1, A_3, A_4, A_5 расширяет возможности алгоритма *Close*, он в процессе построения минимаксного базиса распознаёт избыточные строгие ассоциативные правила и устраняет их.

Определение 5. Избыточные строгие ассоциативные правила. Пусть AR — множество строгих ассоциативных правил. Будем говорить, что строгое ассоциативное правило $X \Rightarrow Y$ логически следует из множества AR , если оно может быть выведено из AR с помощью свойств A_1, A_3, A_4, A_5 .

$$\text{AR} \models X \Rightarrow Y.$$

Строгое ассоциативное правило $X \Rightarrow Y$ назовём *избыточным* в AR , если его можно вывести:

$$AR \setminus \{X \Rightarrow Y\} \models X \Rightarrow Y. \quad (15)$$

Множество строгих ассоциативных правил неизбыточное, если оно не содержит избыточных строгих ассоциативных правил (согласно определению 5).

Суть алгоритма *MClose* состоит в том, что генерируются минимаксных строгих ассоциативных правил с помощью алгоритма *Close* и устраняются среди них избыточные. Выявление избыточного строгого ассоциативного правила в AR основано на проверке логического следования (15). Алгоритм такой проверки использует понятие замыкания множества признаков относительно множества AR . Замыканием множества $X \subseteq M$ относительно AR (обозначается X^+) называется множество всех признаков $m \in M$, таких, что верно логическое следование $AR \models X \Rightarrow m$. Заметим, что неизменно $X^+ \subseteq M$. Из свойств A_1, A_3, A_4 вытекает справедливость следующего критерия: логическое следование $AR \models X \Rightarrow Y$ имеет место тогда и только тогда, когда $Y \subseteq X^+$. Отсюда

$$AR \models X \Rightarrow X^+, \quad AR \models X \Rightarrow X^+ \setminus X.$$

Чтобы убедиться в справедливости (15), достаточно вычислить X^+ относительно $AR \setminus \{X \Rightarrow Y\}$ и проверить включение $Y \subseteq X^+$. Если $Y \subseteq X^+$, то строгое ассоциативное правило $X \Rightarrow Y$ избыточно в AR , иначе оно не является избыточным.

Алгоритм вычисления X^+ целиком базируется на выводимостях A_1, A_3, A_4, A_5 и сводится к выполнению следующих действий. Сначала полагается $X^+ = X$. Далее осуществляется просмотр правил из AR и пополнение X^+ по следующему принципу: если для правила $Y \Rightarrow Z \in AR$ верно включение $Y \subseteq X^+$, то множество Z добавляется к X^+ . Этот процесс повторяется до тех пор, пока изменяется X^+ . Поскольку множества M и AR конечные, то процесс вычисления X^+ конечен.

Так как построение X^+ не требует доступа к контексту, то время выполнения намного меньше чем получения частых замкнутых множеств признаков. Для 1- генераторов считаем $X^+ = X''$. Для k - генераторов будем находить замыкание X^+ , как конкатенация замыканий X^+ $(k-1)$ - генераторов которые образовали соответствующий k генератор.

Для того чтобы исключить избыточные ассоциативные правила из AR , необходимо проверить для всех не 1-элементных генераторов, если $X^+ = X''$, то это минимаксное ассоциативное правило является избыточным, и оно исключается из AR .

Таким образом, по построению результирующее множество AR состоит из минимаксных строгих ассоциативных правил и является неизбыточным. Заметим, что оперативное удаление избыточных правил сдерживает рост мощности AR и снижает время выполнения алгоритма.

1.9. Программная реализация алгоритма MClose и результаты

Псевдокод алгоритм MClose:

```

1:  begin
2:   $AR \leftarrow \emptyset$ 
3:   $k \leftarrow 1$ 
4:  while  $p_k \neq \emptyset$ 
5:    Gen-Closure ( $p_k$ )
6:    if  $p_k \neq p_k''$ 
7:       $p_k^+ \leftarrow SX(p_k)$ 
8:    end if
9:     $AR \leftarrow (p_k \Rightarrow p_k'' \setminus p_k)$ 
10:   Gen-Generator ( $k + 1$ )
11:    $k \leftarrow k + 1$ 
12: end while
13: Non-Redundancy ( $AR$ )
14: end

```

Процедура Gen-Closure формирует замыкание генераторов, а Gen-Generator формирует $(k+1)$ - генераторы. Процедура SX осуществляет построение замыкания p_k^+ относительно AR . Процедура Non-Redundancy, устраняет в AR избыточные строгие ассоциативные правила после завершения его генерации.

Алгоритм MClose программно реализован на языке программирования Python в среде разработки JetBrains PyCharm 2019.1.1 x64. Исходными данными для него служат контекст $K = (G, M, I)$.

Входные данные. Входные данные считываются из текстового файла input.txt

Пример входных данных:

minsupp :	0							
G\M	A	B	C	D	E	F	G	
0	1	0	1	1	0	0	1	
1	0	1	1	0	1	0	0	
2	1	1	1	0	1	1	0	
3	0	1	0	0	1	0	1	
4	1	1	1	0	1	1	0	
5	0	1	1	0	1	0	0	

Выходные данные. Результатом работы алгоритма MClose являются все выявленные частые замкнутые множества признаков и сжатый строгий базис. Они записываются в файл (output.xlsx). Пример выходных данных на рисунках 1 и 2.

Рис. 1. Таблица выходных данных

X	K	X+	X'	X''	3	Л.С.
A	+	AC	024	AC	-	-
B	+	BE	12345	BE	-	-
C	+	C	01245	C	+	-
D	+	ACDG	0	ACDG	-	-
E	+	BE	12345	BE	-	-
F	+	ABCEF	24	ABCEF	-	-
G	+	G	03	G	+	-
AB	+	ABCE	24	ABCEF	-	-
AC	-	AC	024	AC	+	+
AD	-	ACDG	0	ACDG	-	+
AE	+	ABCE	24	ABCEF	-	-
AF	-	ABCEF	24	ABCEF	-	+
AG	+	ACG	0	ACDG	-	-
BC	+	BCE	1245	BCE	-	+
BD	+	ABCDEG	∅	∅	-	-
BE	-	BE	12345	BE	+	+
BF	-	ABCEF	24	ABCEF	-	+
BG	+	BEG	3	BEG	-	+
CD	-	ACDG	0	ACDG	-	+
CE	+	BCE	1245	BCE	-	+
CF	-	ABCEF	24	ABCEF	-	+
CG	+	CG	0	ACDG	-	-
DE	+	ABCDEG	∅	∅	-	-
DF	+	ABCDEF	∅	∅	-	-
DG	-	ACDG	0	ACDG	-	+
EF	-	ABCEF	24	ABCEF	-	+
EG	+	BEG	3	BEG	-	+
FG	+	ABCEFG	∅	∅	-	-
ABE	-	ABCE	24	ABCEF	-	-
ABG	+	ABCEG	∅	∅	-	-
AEG	+	ABCEG	∅	∅	-	-
BCD	-	ABCDEG	∅	∅	-	-
BCG	+	BCEG	∅	∅	-	-
BDG	-	ABCDEG	∅	∅	-	-
CEG	+	BCEG	∅	∅	-	-
DEF	-	ABCDEF	∅	∅	-	-

Рис.2. Базис строгих избыточных ассоциативных правил

X ->	X''		X ->	X''\X
A ->	AC		A ->	C
B ->	BE		B ->	E
C ->	C			
D ->	ACDG		D ->	ACG
E ->	BE		E ->	B
F ->	ABCEF		F ->	ABCE
G ->	G			
AB ->	ABCEF		AB ->	CEF
AE ->	ABCEF		AE ->	BCF
AG ->	ACDG		AG ->	CD
CG ->	ACDG		CG ->	AD

Заключение

Список литературы

1. Витяев Е. Е., Демин А. В., Пономарев Д. К. Вероятностное обобщение формальных понятий // Программирование. 2012. № 5. С. 18–34.
2. Городецкий В. И., Самойлов В. В. Ассоциативный и причинный анализ и ассоциативные байесовские сети // Тр. СПИИРАН. 2009. Вып. 9. С. 13–65.
3. Кузнецов С. О. Автоматическое обучение на основе анализа формальных понятий // Автоматика и телемеханика. 2001. № 10. С. 3–27.
4. Биркгоф Г., Барти Т. Современная прикладная алгебра. СПб.: Лань, 2005. 400 с.
5. Гуров С. И. Булевы алгебры, упорядоченные множества, решетки: определения, свойства, примеры. М.: Книжный дом ³/₄ЛИБРОКОМ_г, 2013. 352 с.
6. Ganter B. and Wille R. Formal Concept Analyses: Mathematical Foundations. Springer Science and Business Media, 2012. 314 p.

7. Pasquier N., Bastide Y., Taoui R., and Lakhal L. Generating a condensed representation for association rules // J. Intelligent Inform. Systems. 2005. V. 24. No. 1. P. 29–60.
8. Майер Д. Теория реляционных баз данных. М.: Мир, 1987. 608 с.
9. Батура Т. В. Модели и методы анализа компьютерных социальных сетей // Программные продукты и системы. 2013. № 3. С. 130–137.
10. Платонов В. В., Семенов П. О. Методы сокращения размерности в системах обнаружения сетевых атак // Проблемы информационной безопасности. Компьютерные системы. 2012. № 3. С. 40–45.
11. Payaraja M. and Meyyappan T. Mining medical data to identify frequent diseases using Apriori algorithm // Pattern Recognition, Informatics and Mobile Engineering (PRIME), IEEE, 2013. P. 194–199.
12. Duquenne V. and Obiedkov S. A. Attribute-incremental construction of the canonical implication basis // Ann. Math. Artif. Intelligence. 2007. V. 49. No. 1. P. 77–99.
13. Rudolph S. Some notes on pseudo-closed sets // LNCS. 2007. V. 4390. P. 151–165.
14. Zaki M. J. and Hsiao C.-J. Efficient algorithms for mining closed item sets and their lattice structure // IEEE Trans. Knowledge Data Eng. 2005. V. 17. No. 4. P. 462–478.
15. Uno T., Asai T., Uchida Y., and Arimura H. An efficient algorithm for enumerating closed patterns in transaction databases // LNCS. 2004. V. 3245. P. 16–31.
16. Zhang C. and Zhang S. Association Rules Mining. Springer, 2002. 240 p.
17. Geng L. and Hamilton H. J. Interestingness measures for data mining: a survey // ACM Computing Surveys. 2006. V. 38. No. 3. Article 9.

Приложение

Рис. 1. Таблица выходных данных

X	K	X+	X'	X''	3	Л.С.
A	+	AC	024	AC	-	-
B	+	BE	12345	BE	-	-
C	+	C	01245	C	+	-
D	+	ACDG	0	ACDG	-	-
E	+	BE	12345	BE	-	-
F	+	ABCEF	24	ABCEF	-	-
G	+	G	03	G	+	-
AB	+	ABCE	24	ABCEF	-	-
AC	-	AC	024	AC	+	+
AD	-	ACDG	0	ACDG	-	+
AE	+	ABCE	24	ABCEF	-	-
AF	-	ABCEF	24	ABCEF	-	+
AG	+	ACG	0	ACDG	-	-
BC	+	BCE	1245	BCE	-	+
BD	+	ABCDEG	∅	∅	-	-
BE	-	BE	12345	BE	+	+
BF	-	ABCEF	24	ABCEF	-	+
BG	+	BEG	3	BEG	-	+
CD	-	ACDG	0	ACDG	-	+
CE	+	BCE	1245	BCE	-	+
CF	-	ABCEF	24	ABCEF	-	+
CG	+	CG	0	ACDG	-	-
DE	+	ABCDEG	∅	∅	-	-
DF	+	ABCDEF	∅	∅	-	-
DG	-	ACDG	0	ACDG	-	+
EF	-	ABCEF	24	ABCEF	-	+
EG	+	BEG	3	BEG	-	+
FG	+	ABCEFG	∅	∅	-	-
ABE	-	ABCE	24	ABCEF	-	-
ABG	+	ABCEG	∅	∅	-	-
AEG	+	ABCEG	∅	∅	-	-
BCD	-	ABCDEG	∅	∅	-	-
BCG	+	BCEG	∅	∅	-	-
BDG	-	ABCDEG	∅	∅	-	-
CEG	+	BCEG	∅	∅	-	-
DEF	-	ABCDEF	∅	∅	-	-

Рис.2. Базис строгих избыточных ассоциативных правил

X ->	X''		X ->	X''\X
A ->	AC		A ->	C
B ->	BE		B ->	E
C ->	C			
D ->	ACDG		D ->	ACG
E ->	BE		E ->	B
F ->	ABCEF		F ->	ABCE
G ->	G			
AB ->	ABCEF		AB ->	CEF
AE ->	ABCEF		AE ->	BCF
AG ->	ACDG		AG ->	CD
CG ->	ACDG		CG ->	AD