Вероятностные понятия в формальных контекстах*

Е.Е.Витяев[1] А.В.Демин[2] Д.К.Пономарев[2]
vityaev@math.nsc.ru alexandredemin@yandex.ru ponom@iis.nsk.su

[1]Институт Математики им. С.Л.Соболева СО РАН
[2]Институт Систем Информатики им. А.П.Ершова СО РАН

Аннотация

В данной работе мы обобщаем ключевые определения из метода «формального анализа понятий» с помощью идей семантического вероятностного вывода. Показано, что в стандартных ограничениях вводимые нами объекты полностью соответствуют исходным определениям в «формальном анализе понятий». С практической точки зрения представлен способ, позволяющий восстанавливать понятия в формальных контекстах в условиях шума на данных.

1 Введение

Представим, что исследователю необходимо классифицировать конечное число некоторых наблюдаемых объектов по n признакам. Наблюдения производятся в серии экспериментов, в каждом из которых устанавливается, имеет ли объект тот или иной признак. Фактически результат каждого эксперимента можно представить в виде таблицы, строки которой помечены именами наблюдаемых объектов, столбцы – названиями признаков, и каждая клетка с координатой (i, j) заполнена в том и только том случае, если i-ый объект имеет признак j. На результатах одного эксперимента логично классифицировать объекты следующим образом – выделить в группы те объекты, которые имеют общий набор признаков, и никакой другой объект ровно этим же набором признаков не обладает. Известно, что получаемые таким образом пары <группа объектов, набор признаков> можно естественным образом упорядочить и представить в некотором наглядном виде, что является предметом широко распространенного метода «формального анализа понятий» [4, 5] (Formal Concept Analysis, в дальнейшем сокращенно FCA). Однако представим, что нам известны результаты не только одного эксперимента, но целой серии экспериментов, и для построения классификации объектов мы хотели бы привлечь всю совокупность полученных данных. При этом надо понимать, что для объекта в некотором числе экспериментов может быть установлено наличие определенного признака, а в оставшейся части наблюдений у объекта этот признак может отсутствовать. Чтобы учесть эту нечеткость при построении классификации объектов на основе серии экспериментальных данных мы используем метод семантического вероятностного вывода, представленный в работах [1, 11, 12]. В данной статье мы обобщим стандартное используемое в FCA понятие истинности импликации на данных с помощью некоторой оценки истинности, основанной на вероятностной мере. Далее мы определим аналог классификационной единицы, возникающей в методе FCA, с помощью неподвижных точек импликаций, истинных на данных с учетом введенной оценки. Нам не известны работы, в которых бы предлагался аналогичный вероятностный подход. Например, в работе [3]

*Работа выполнена при поддержке гранта Президента РФ (МК-2037.2011.9), гранта РФФИ (11-07-00560а) и Интеграционных проектов СО РАН (гранты № 47, 111, 119).

основные объекты «формального анализа понятий» переформулируются в терминах вероятностной логики, однако при этом само их определение в рамках FCA не обобщается.

В данной статье мы преследуем цель установить связи между «формальным анализом понятий» и методом семантического вероятностного вывода. Научным вкладом данной работы является обобщение ключевых понятий метода FCA в рамках семантического вероятностного вывода.

2 Предварительные определения

Начнем с основных определений и результатов «формального анализа понятий».

Определение 1. *Формальным контекстом называется тройка* (G, M, I), *где* G *и* M – *некоторые множества, а* $I \subseteq G \times M$ – *некоторое отношение между элементами* G *и* M. *Элементы* G *называются объектами контекста а элементы* M – *атрибутами контекста. Формальный контекст назовем конечным, если* G *и* M *являются конечными множествами.*

Далее для краткости мы опускаем слово «формальный» и будем называть тройки (G, M, I), указанные в определении, контекстами. Любой контекст можно представить в виде таблицы, аналогично тому, как было замечено во введении. Если (G, M, I) – контекст, то на подмножествах $A \subseteq G$, $B \subseteq M$ определим операцию $'$ следующим образом:

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

Если $g \in G$, то обозначение g' будет служить сокращением для множества $\{g\}'$.

Определение 2. *Понятием в контексте* (G, M, I) *называется пара* (A, B), *где* $A \subseteq G$, $B \subseteq M$, $A' = B$ *и* $B' = A$. *При этом множество* A *называется объемом, а* B – *содержанием понятия* (A, B).

Фактически «понятие» является классификационной единицей, группирующей объекты и атрибуты контекста.

Следующий простой факт будет многократно использоваться в доказательствах основных утверждений статьи:

Лемма 1. *Для любого контекста* (G, M, I) *и множеств* $B_1, B_2 \subseteq M$ *верно:*

1. $B_1 \subseteq B_2 \Longrightarrow B_2' \subseteq B_1'$
2. $B_1 \subseteq B_1''$.

Определение 3. *Определим (частичное) упорядочение* $\leqslant$ *понятий контекста следующим образом: если* (A_1, B_1) *и* (A_2, B_2) – *понятия в некотором контексте, то полагаем* $(A_1, B_1) \leqslant (A_2, B_2)$, *если* $A_1 \subseteq A_2$ *(или, что эквивалентно по лемме 1, если* $B_2 \subseteq B_1$*).*

Теорема. *Отношение* $\leqslant$ *порождает на множестве понятий контекста полную решетку, в которой инфимум и супремум подмножеств определяются, соответственно, следующим образом:*

$$\bigwedge_{j \in J} (A_j, B_j) = (\bigcap_{j \in J} A_j,\ (\bigcup_{j \in J} B_j)'')$$

$$\bigvee_{j \in J} (A_j, B_j) = ((\bigcup_{j \in J} A_j)'',\ \bigcap_{j \in J} B_j).$$

I	m_1	m_2	m_3	m_4
g_1	×	×	×	
g_2	×	×		
g_3			×	
g_4				×

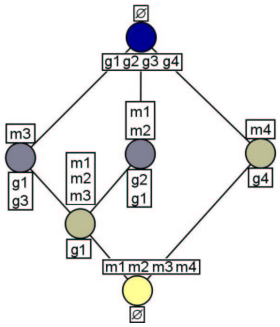

Рис. 1: Контекст и соответствующая ему решетка понятий.

Пример 1. *Рассмотрим конечный контекст* $K = (\{g_1, g_2, g_3, g_4\}, \{m_1, m_2, m_3, m_4\}, I)$, *представленный в табличном виде на рисунке 1. Решетка всех понятий в контексте* K *представлена на этом же рисунке, каждый элемент решетки помечен множеством объектов и атрибутов, являющихся, соответственно, объемом и содержанием понятия.*

Процедуры вычисления полной решетки понятий по заданному конечному контексту [9, 10] являются одними из ключевых алгоритмов в методе FCA . Фактически, они производят построение классификации объектов контекста в соответствии с указанными для них атрибутами и позволяют найти все существующие классы.

Если задан некоторый контекст $K = (G, M, I)$, то можно говорить об истинности на K утверждений следующего вида: «все объекты, обладающие атрибутами $B_1 \subseteq M$, имеют также множество атрибутов $B_2 \subseteq M$». Поскольку все свойства контекста в определенном смысле симметричны относительно множеств G и M, то аналогичные утверждения можно сформулировать о подмножествах G: «все атрибуты, имеющие своими объектами $A_1 \subseteq G$, также имеют своими объектами и $A_2 \subseteq G$». Без ограничения общности мы будем рассматривать утверждения лишь первого вида. Фактически, они определяют монотонный оператор, импликацию, на булевой алгебре подмножеств M. Понятно, что если контекст K конечен, то множество всех таких утверждений, истинных на K, также конечно. Формализуем понятие импликации, истинной на контексте, с помощью определений главы 2.3 в [4].

Определение 4. *Импликацией на некотором множестве* M *назовем упорядоченную пару подмножеств* $A, B \subseteq M$, *обозначаемую как* $A \to B$. *Множество* A *назовем посылкой, а* B *– заключением импликации* $A \to B$. *Скажем, что множество* $T \subseteq M$ *удовлетворяет импликации* $A \to B$, *если* $A \not\subseteq T$ *или* $B \subseteq T$. *Семейство подмножеств* M *удовлетворяет импликации, если каждое из множеств этого семейства удовлетворяет ей.*

Если $K = (G, M, I)$ *– некоторый контекст, то импликация* $A \to B$ *истинна на* K *(обозначение* $K \models A \to B$*), если* $A, B \subseteq M$ *и семейство множеств* $\{g' \mid g \in G\}$ *удовлетворяет* $A \to B$.

Скажем, что посылка импликации $A \to B$ *ложна на* K, *если не существует* $g \in G$ *такого, что* $A \subseteq g'$. *Назовем импликацию* $A \to B$ *тавтологией, если* $B \subseteq A$.

Для контекста $K = (G, M, I)$ будем обозначать множество всех импликаций на M, которые истинны на контексте K, через $Imp(K)$. Легко проверить, что подмножеством $Imp(K)$ является множество тавтологий, а так же множество тех импликаций, посылка которых ложна на K. Для обозначения того, что множество или семейство множеств

удовлетворяет некоторой импликации, будем использовать тот же символ $\models$, когда это не приводит к путанице.

Любое множество импликаций L на множестве M порождает монотонный оператор $f_L : 2^M \to 2^M$, определяемый следующим образом:

$$f_L(X) = X \cup \{B \mid A \to B \in L,\ A \subseteq X\}.$$

Ясно, что для любого $X \subseteq M$ верно $f_L(X) = X \Leftrightarrow X \models L$.

Замечание 1. *Пусть L – семейство импликаций на некотором множестве M. Тогда для любого $X \subseteq M$ существует минимальное $Y \subseteq M$ такое, что $X \subseteq Y$ и $f_L(Y) = Y$.*

Доказательство. Рассмотрим очевидный индуктивный процесс построения расширений заданного множества $X \subseteq M$. Положим начальное множество $X_0 = X$. Если построено множество X_i, то полагаем $X_{i+1} = f_L(X_i)$. Тогда искомое $Y = \bigcup_{i\in\omega} X_i$. □

Таким образом, любое семейство импликаций L на множестве M определяет оператор $\bar{f}_L : 2^M \to 2^M$, который для каждого $X \subseteq M$ дает минимальное $Y \subseteq M$, удовлетворяющее условиям замечания. Очевидно, что для любого множества $X \subseteq M$ верно $f_L(X) = X \Leftrightarrow \bar{f}_L(X) = X$.

Замечание 2. *Если $K = (G, M, I)$ – контекст, $A \to B$ – импликация на M, то $K \models A \to B \Leftrightarrow \forall m \in B\ (K \models A \to \{m\})$.*

В дальнейшем мы будем рассматривать импликации только вида $A \to \{m\}$ и использовать обозначение $A \to m$ для таких импликаций.

Если K – контекст, то для каждой импликации $A \to m \in Imp(K)$ найдется множество $\{A_0 \to m \in Imp(K) \mid A_0 \subseteq A$ и для любого множества $A_1 \subseteq A$ из $A_1 \subset A_0$ следует $A_1 \to m \notin Imp(K)\}$. Для контекста K обозначим через $MinImp(K)$ множество всех импликаций вида $A_0 \to m$, истинных на K, в которых множество A_0 минимально в указанном смысле. Отметим, что определение такого рода импликаций является вариантом определения импликации как закона в [1, 12, 11].

Далее приведем доказательство несколько измененного Предложения 20 из [4], которое является ключевым в данной статье.

Предложение 1. *Пусть $K = (G, M, I)$ – контекст, $T \subseteq Imp(K)$ – множество тавтологий на M, а $F \subseteq Imp(K)$ – множество импликаций, посылки которых ложны на K. Тогда для любого множества $B \subseteq M$ выполняется следующее:*

1. $f_{MinImp(K)\setminus T}(B) = B \Leftrightarrow B'' = B$;
2. *если $B' \neq \varnothing$, то $f_{MinImp(K)\setminus\{F\cup T\}}(B) = B \Leftrightarrow B'' = B$.*

Доказательство. Покажем сначала, что для любого подмножества $B \subseteq M$ верно $f_{Imp(K)}(B) = B$ тогда и только тогда, когда $f_{MinImp(K)}(B) = B$. Действительно, если $f_{Imp(K)}(B) \supset B$ для некоторого B, то (с учетом замечания 2) существует импликация $A \to m \in Imp(K)$ такая, что $A \subseteq B$, но $m \notin B$. Тогда найдется импликация $A_0 \to m \in MinImp(K)$, где $A_0 \subseteq A$ и поэтому $A_0 \subseteq B$, $m \notin B$, $f_{MinImp(K)}(B) \supset B$ – противоречие. В обратную сторону утверждение очевидно, поскольку $MinImp(K) \subseteq Imp(K)$.

Аналогичным образом нетрудно проверить, что $f_{MinImp(K)\setminus L}(B) = B \Leftrightarrow f_{Imp(K)\setminus L}(B) = B$, где $L = T$, либо $L = F \cup T$. Это следует из того, что для любой импликации $A \to m$ на M и любого подмножества $A_0 \subseteq A$ условие $A \to m \notin T$, очевидно, влечет $A_0 \to m \notin T$, а из условия $A' \neq \varnothing$ следует $A'_0 \neq \varnothing$ по лемме 1. Поэтому далее докажем утверждения 1 и 2 данного предложения относительно множества $Imp(K)$, а не $MinImp(K)$.

1.$\Leftarrow$: Пусть $B'' = B$, $A_1 \to A_2 \in Imp(K) \setminus T$ и $A_1 \subseteq B$. Покажем, что тогда $A_2 \subseteq B$. Действительно, для любого $g \in B'$ выполняется $g' \supseteq A_2$, поскольку по лемме 1, $g' \supseteq B'' = B$, а импликация $A_1 \to A_2$ истинна на K. Из этого $\bigcap\{g' \mid g \in B'\} \supseteq A_2$. С другой стороны, $\bigcap\{g' \mid g \in B'\} = B''$, но так как $B'' = B$, получаем $B \supseteq A_2$.

1.$\Rightarrow$: По лемме 1 в любом случае имеем $B'' \supseteq B$, поэтому предположим, что $f_{Imp(K)\setminus T}(B) = B$, но $B'' \not\subseteq B$. Тогда $B \not\models B \to B'' \notin T$ и для того, чтобы прийти к противоречию достаточно показать, что $B \to B'' \in Imp(K)$.

а) Если $B' = \varnothing$, то это, очевидно, имеет место, поскольку в данном случае не существует $g \in G$ такого, что $B \subseteq g'$, т.е. посылка импликации ложна на K.

б) Пусть $B' \neq \varnothing$, надо показать, что $\forall g \in G$ $(B \subseteq g' \Rightarrow B'' \subseteq g')$. Ясно, что $\forall g \in G$ $(B \subseteq g' \Leftrightarrow g \in B')$ и, по лемме 1, $\forall g \in B'$ $(B'' \subseteq g')$. Таким образом, если $B \subseteq g'$ для некоторого $g \in G$, то $B'' \subseteq g'$, то есть $B \to B'' \in Imp(K)$. Более того, $B \to B'' \in Imp(K) \setminus F$, поскольку $B' \neq \varnothing$.

2. Достаточность следует из доказательства пункта 1, поскольку ясно, что если $f_{MinImp(K)\setminus T}(B) = B$, то и $f_{MinImp(K)\setminus\{F\cup T\}}(B) = B$. Необходимость доказывается пунктом б) выше. $\square$

Для любого контекста $K = (G, M, I)$ из определения 2, очевидно, следует, что подмножество $B \subseteq M$ является содержанием некоторого понятия в контексте K тогда и только тогда, когда $B'' = B$. Таким образом, как только задан контекст $K = (G, M, I)$, мы имеем множество $Imp(K)$ всех импликаций, истинных на K, и неподвижные точки оператора $f_{MinImp(K)\setminus T} : 2^M \to 2^M$ совпадают с содержаниями понятий контекста K. Если же среди импликаций из $MinImp(K) \setminus T$, не рассматривать множество тех импликаций F, посылка которых ложна на K, то неподвижные точки оператора $f_{MinImp(K)\setminus\{F\cup T\}} : 2^M \to 2^M$ совпадают с содержаниями понятий контекста K за исключением единственного понятия $(\varnothing, M)$. В силу того, что для любого $B \subseteq M$ условие $B'' \neq M$, очевидно, влечет $B' \neq \varnothing$.

3 Вероятностные понятия на классе контекстов

Выше было определено понятие истинности импликации на некотором отдельно взятом контексте. Опишем, как обобщить данное понятие с помощью оценки истинности импликации на классе контекстов. Перейдем к изложению идей метода семантического вероятностного вывода в применении к FCA. В рамках данного метода, представленного в [1, 11, 12], закономерности на данных, в частности, импликации формализуются в виде универсальных формул языка логики первого порядка счетной сигнатуры, состоящей из предикатов и констант. Таким образом, стандартное понятие импликации, определенное в [4], оказывается более узким, чем понятие закономерности на данных, рассматриваемое в семантическом вероятностном выводе (отметим, что в статьях по FCA также изучались импликации, далеко выходящие за рамки определений из [4]). Однако для того, чтобы показать применимость данного метода для «анализа формальных понятий», нам будет удобнее не выходить за рамки стандартно используемых алгебраических определений. Поэтому далее мы представим некоторое ограничение метода семантического вероятностного вывода в терминах, близких к используемым в FCA.

Определение 5. *Классом контекстов над множествами G и M назовем семейство $\mathcal{K} = \{(G, M, I_j)\}_{j\in J\neq\varnothing}$, где для каждого $j \in J$ тройка (G, M, I_j) является контекстом. Будем использовать обозначение $\mathcal{K}(G, M)$ для класса контекстов $\mathcal{K}$ над множествами G и M. Вероятностной моделью первого типа назовем пару $\mathcal{M} = (\mathcal{K}(G, M), \rho)$, где $G \neq \varnothing$ и ρ – вероятностная мера на множестве $\mathcal{K}$, удовлетворяющая условию:* $\forall\ S_1, S_2 \subseteq G \times M\ \ \forall\ (G, M, I) \in \mathcal{K}$

$$(S_1 \not\subseteq I \text{ или } S_2 \subseteq I) \Longleftrightarrow \rho(\{(G, M, I_j) \mid S_1 \cup S_2 \subseteq I_j\}) = \rho(\{(G, M, I_j) \mid S_1 \subseteq I_j\}).$$

Если $S \subseteq G \times M$, то вероятностью множества S на модели $\mathcal{M}$ назовем значение функции $\nu_{\mathcal{M}}(S) = \rho(\{(G, M, I) \in \mathcal{K} \mid S \subseteq I\})$.

В рамках данного раздела для краткости мы будем называть пары $(\mathcal{K}(G, M), \rho)$ из определения выше *вероятностными моделями* или просто *моделями*.

Пусть $\mathcal{M} = (\mathcal{K}(G, M), \rho)$ – вероятностная модель и $A \to m$ – некоторая импликация на множестве M. Подстановочным случаем импликации $A \to m$ на модели $\mathcal{M}$ назовем пару $\langle g, A \to m\rangle$, где $g \in G$. Вероятностью пары $\langle g, A \to m\rangle$ на модели $\mathcal{M}$ назовем значение функции

$$\mu_{\mathcal{M}}(\langle g, A \to m\rangle) = \begin{cases} \frac{\nu_{\mathcal{M}}(\ S \cup \{<g,m>\}\ )}{\nu_{\mathcal{M}}(S)}, & \text{если } \nu_{\mathcal{M}}(S) \neq 0, \text{ где } S = \{< g, a > \mid a \in A\} \\ \text{не определено} & \text{в противном случае} \end{cases}$$

Вероятностью импликации $A \to m$ на модели $\mathcal{M}$ назовем значение функции

$$\eta_{\mathcal{M}}(A \to m) = \begin{cases} \text{не определено}, & \text{если } \forall g \in G\ \mu_{\mathcal{M}}(\langle g, A \to m\rangle) \text{ не определено} \\ \inf_{g \in G} \mu_{\mathcal{M}}(\langle g, A \to m\rangle) & \text{в противном случае} \end{cases}$$

Замечание 3. *Пусть $\mathcal{M} = (\mathcal{K}(G, M), \rho)$ – вероятностная модель, $A \to m$ – некоторая импликация на множестве M, вероятность которой на модели $\mathcal{M}$ определена. Тогда $\eta_{\mathcal{M}}(A \to m) = 1$ в том и только том случае, если $\forall K \in \mathcal{K}\ (A \to m \in Imp(K))$.*

Доказательство. $\Rightarrow$: Условие $\eta_{\mathcal{M}}(A \to m) = 1$ означает, что для каждого $g \in G$ значение $\mu_{\mathcal{M}}(\langle g, A \to m\rangle)$ либо не определено, либо равно 1. Поэтому для каждого $g \in G$ и каждого контекста $K \in \mathcal{K}$, по определению функций $\mu_{\mathcal{M}}$ и ρ, выполняется $A \not\subseteq g'$ или $m \in g'$. А это непосредственно означает, что $\forall K \in \mathcal{K}\ (A \to m \in Imp(K))$.

$\Leftarrow$: Предположим, что $\eta_{\mathcal{M}}(A \to m) < 1$. Тогда найдется $g \in G$ такое, что значение $\mu_{\mathcal{M}}(\langle g, A \to m\rangle)$ определено и также строго меньше единицы. В свою очередь, тогда существует $K \in \mathcal{K}$, в котором $A \subseteq g'$, но $m \notin g'$, а это означает, что $A \to m \notin Imp(K)$. $\square$

Определение 6. *Пусть $\mathcal{M} = (\mathcal{K}(G, M), \rho)$ – вероятностная модель, $imp(M)$ – множество тех импликаций на M, вероятность которых на $\mathcal{M}$ определена. Вероятностными закономерностями (вероятностными законами [1, 11, 12]) на $\mathcal{M}$ назовем импликации $A \to m \in imp(M)$, для которых выполняется следующее:*

- $\eta_{\mathcal{M}}(A \to m) \neq 0$;
- *если $A_0 \to m \in imp(M)$ и $A_0 \subset A$, то $\eta_{\mathcal{M}}(A_0 \to m) < \eta_{\mathcal{M}}(A \to m)$.*

Импликацию $A \to m \in imp(M)$ назовем максимально специфичной вероятностной закономерностью (максимально специфичным законом [1, 11]) на $\mathcal{M}$, если она есть вероятностная закономерность на $\mathcal{M}$, $A \neq \{m\}$, и не существует такой вероятностной закономерности $A_0 \to m$ на $\mathcal{M}$, что $A \subset A_0$ и $A_0 \to m$ не тавтология.

Замечание 4. *Если импликация является максимально специфичной вероятностной закономерностью на модели $\mathcal{M}$, то она не тавтология.*

Определение 7. *Пусть $\mathcal{M} = (\mathcal{K}(G, M), \rho)$ – вероятностная модель, $S(\mathcal{M})$ – множество всех максимально специфичных вероятностных закономерностей на $\mathcal{M}$. Импликацию $A \to m \in S(\mathcal{M})$ назовем сильнейшей вероятностной закономерностью на $\mathcal{M}$, если значение ее вероятности на $\mathcal{M}$ максимально среди всех импликаций $B \to m \in S(\mathcal{M})$.*

Будем использовать обозначение $D(\mathcal{M})$ для множества всех сильнейших вероятностных закономерностей на модели $\mathcal{M}$.

Отметим, что в виду достаточно произвольного задания функции ρ в определении вероятностной модели, ничем не гарантируется существование максимума в смысле определения 7 и, таким образом, существование самих сильнейших вероятностных закономерностей. Однако далее будет приведен способ задания вероятностной модели (на основе конечного класса конечных контекстов), при котором возникает широкий класс моделей, гарантирующих наличие таких импликаций. Отметим, что в общем случае для произвольного m может существовать несколько импликаций вида $A \to m$, являющихся сильнейшими вероятностными закономерностями.

Неформально каждую импликацию на вероятностной модели стоит рассматривать как «предсказание» с некоторой оценкой истинности того, что каждый объект, имеющий набор атрибутов из посылки, будет иметь также атрибут из заключения импликации. Как и в «формальном анализе понятий» (напомним предложение 1), импликации в методе семантического вероятностного вывода непосредственно связаны с процессом выделения объектов и атрибутов в классификационные единицы. Если данные представлены в виде некоторого класса контекстов $\mathcal{K}$, то от того, какого рода импликации выделить среди всех возможных импликаций на вероятностной модели $(\mathcal{K}, \rho)$, зависит формирование самих классов на основе предоставленных данных. Минимальные импликации в смысле множества $MinImp(K)$, а также вероятностные закономерности, максимально специфичные и сильнейшие вероятностные закономерности являются адаптацией соответствующих вероятностных определений из [1, 11, 12, 15] применительно к методу FCA. Такого рода импликации обладают рядом теоретических и практических полезных свойств, которые обосновывают их использование:

- множество всех минимальных импликаций, истинных на каждом контексте из некоторого класса $\mathcal{K}$, дает в определенном смысле аксиоматизацию этого класса контекстов – из них семантически выводится импликативная теория класса контекстов, ограниченная на импликации с неложной посылкой [1, 12] (аналог теоремы Duquenne, Guigues о базисе импликаций [6]);

- вероятностная закономерность исключает возможность «предсказания» атрибута в ее заключении некоторым собственным подмножеством атрибутов посылки с большей (либо равной) вероятностью, чем у самой закономерности; вместе с требованием максимальной специфичности на практике это приводит к группированию атрибутов в меньшие классы, с наибольшей вероятностью [2];

- в [1, 11] доказано, что если в импликациях допускается негативная информация, то максимально специфические вероятностные закономерности образуют непротиворечивое множество утверждений (не возникает ситуации одновременного «предсказания» наличия атрибута и его отсутствия);

- сильнейшие вероятностные закономерности приводят к отнесению атрибута к тому классу атрибутов, который «предсказывает» его с максимальной вероятностью; при этом не исключается ситуация, когда один и тот же атрибут может входить в разные классы [14];

- разработана программная система Discovery, реализующая обнаружение упомянутых типов импликаций на табличных данных и построение соответствующих им классов объекты-признаки. Эта система успешно применялась для решения целого ряда прикладных задач [7, 1, 15].

Определение 8. *Пусть $\mathcal{M} = (\mathcal{K}(G, M), \rho)$ – вероятностная модель первого типа. Вероятностным понятием контекста $(G, M, I) \in \mathcal{K}$ в модели $\mathcal{M}$ назовем пару множеств (A, B), удовлетворяющих следующим условиям:*

- $A \subseteq G$, $B \subseteq M$,
- $f_{D(\mathcal{M})}(B) = B$,
- $\exists E \subseteq B$ ($\bar{f}_{D(\mathcal{M})}(E) = B$ *и* $E \neq \varnothing \neq E'$),
- $A = \bigcup\{E' \mid \varnothing \neq E \subseteq B,\ \bar{f}_{D(\mathcal{M})}(E) = B\}$,

где $'$ *– операция в рамках контекста* (G, M, I). *При этом множество* A *назовем объемом, а* B *– содержанием вероятностного понятия* (A, B).

Таким образом, как только задана вероятностная модель $\mathcal{M} = (\mathcal{K}(G, M), \rho)$, множество неподвижных точек оператора $f_{D(\mathcal{M})}$ ограничивает множество всех возможных вероятностных понятий контекстов класса $\mathcal{K}$ в модели $\mathcal{M}$.

Теорема 1. *Рассмотрим контекст* $K = (\varnothing \neq G, M, I)$, *и вероятностную модель* $\mathcal{M} = (\{K\}, \rho)$. *Тогда для любых непустых подмножеств* $A \subseteq G$ *и* $B \subseteq M$ *пара* (A, B) *является понятием в контексте* K *тогда и только тогда, когда* (A, B) *является вероятностным понятием контекста* K *в модели* $\mathcal{M}$.

Доказательство. Пусть $S \subseteq Imp(K)$ – множество, состоящее из всех тавтологий на M и всех импликаций, посылки которых ложны на контексте K. Покажем, что $MinImp(K) \setminus S = D(\mathcal{M})$.

$\subseteq$: Рассмотрим произвольную импликацию $A \to m \in MinImp(K) \setminus S$. В силу определения модели $\mathcal{M}$, для любого подмножества $S \subseteq G \times M$ имеем $\rho(S) = 0 \Leftrightarrow S \not\subseteq I$. Так как посылка A не ложна на K, из этого получаем, что вероятность импликации $A \to m$ на модели $\mathcal{M}$ определена, и тогда по замечанию 3 имеем $\eta_{\mathcal{M}}(A \to m) = 1$. В силу минимальности посылки A, любая импликация $A_0 \to m$, где $A_0 \subset A$, уже не истинна на K. Кроме того, поскольку посылка A не ложна на K, то и A_0 не ложна на K. По замечанию 3 тогда $\eta_{\mathcal{M}}(A_0 \to m) = 0$ и, таким образом, импликация $A \to m$ является вероятностной закономерностью на $\mathcal{M}$. С учетом $\eta_{\mathcal{M}}(A \to m) = 1$ и $m \notin A$ из этого получаем, что $A \to m \in D(\mathcal{M})$.

$\supseteq$: По заданию модели $\mathcal{M}$ имеем $\forall S \subseteq G \times M$ $\rho(S) \in \{0, 1\}$, следовательно, для любой импликации $A \to m \in D(\mathcal{M})$ из определения вероятностной закономерности имеем $\eta_{\mathcal{M}}(A \to m) = 1$. По определению функции $\mu_{\mathcal{M}}$ тогда получаем, что посылка A не ложна на K, и поэтому с учетом замечаний 3 и 4 имеем $A \to m \in Imp(K) \setminus S$. Предположим, что найдется импликация $A_0 \to m \in Imp(K)$ такая, что $A_0 \subset A$. Тогда $A_0 \to m \in Imp(K) \setminus S$, $\eta_{\mathcal{M}}(A_0 \to m) = 1$ и приходим к противоречию с тем, что $A \to m$ – вероятностная закономерность на $\mathcal{M}$. Следовательно, $A \to m \in MinImp(K) \setminus S$.

Пусть (A, B) – вероятностное понятие контекста K в модели $\mathcal{M}$. Покажем, что (A, B) является понятием в контексте K – для этого достаточно проверить, что $A' = B$ и $B' = A$. Рассмотрим множество $\mathcal{C} = \{E \subseteq B \mid \bar{f}_{D(\mathcal{M})}(E) = B,\ E \neq \varnothing \neq E'\}$, по определению вероятностного понятия оно не пусто. Для каждого $E \in \mathcal{C}$, вследствие $\bar{f}_{D(\mathcal{M})}(E) = B$ и доказанного выше, найдется импликация $E \to B \in Imp(K)$, поэтому $B' \neq \varnothing$ и с учетом $f_{D(\mathcal{M})}(B) = B$ по пункту 2 предложения 1 получаем, что $B'' = B$. Кроме того, из $E \to B \in Imp(K)$ следует, что для каждого $g \in E'$ верно $g' \supseteq B$. Это означает, что для каждого $g \in \bigcup\{E' \mid E \in \mathcal{C}\} = A$ выполнено $g' \supseteq B$ и поэтому $A \subseteq B'$. С другой стороны, для каждого $E \in \mathcal{C}$ из условия $E \subseteq B$ имеем $B' \subseteq E'$, поэтому $B' \subseteq \bigcup\{E' \mid E \in \mathcal{C}\} = A$. Таким образом, имеем $A = B'$, что вместе с $B'' = B$ дает $A' = B$.

Пусть (A, B) – понятие в контексте K, причем множества A и B не пустые. Проверим, что (A, B) является вероятностным понятием контекста K в модели $\mathcal{M}$. Действительно, так как $A \neq \varnothing$ и $B' = A$, имеем $B' \neq \varnothing$ и, поскольку верно $B'' = B$, по пункту 2 предложения 1, в силу доказанного выше, получаем $f_{D(\mathcal{M})}(B) = B$. Остается проверить, что $A = \bigcup\{E' \mid E \in \mathcal{C}\}$, где $\mathcal{C} = \{E \subseteq B \mid E \neq \varnothing,\ \bar{f}_{D(\mathcal{M})}(E) = B\}$, так как, очевидно,

$B \in \mathcal{C}$. Имеем $\bigcup\{E' \mid E \in \mathcal{C}\} \supseteq B' = A$. Наоборот, если $g \in \bigcup\{E' \mid E \in \mathcal{C}\}$, то найдется $E \in \mathcal{C}$ такое, что $g \in E'$ и, таким образом, $g' \supseteq E$. В силу $\bar{f}_{D(\mathcal{M})}(E) = B$, имеем $E \to B \in Imp(K)$, поэтому $g' \supseteq B$ и значит, $g \in B' = A$. Таким образом, все условия в определении вероятностного понятия выполнены. □

Пусть $\mathcal{K} = \{(\varnothing \neq G, M, I_j)\}_{j \in J \neq \varnothing}$ – конечный класс, состоящий из конечных контекстов. Укажем естественный способ определить вероятностную модель $(\mathcal{K}, \rho)$ на классе $\mathcal{K}$. Для каждого контекста $K \in \mathcal{K}$ положим $\rho(\{K\}) = 1/|J|$, а для подмножества $\mathcal{C} \subseteq \mathcal{K}$ определим $\rho(\mathcal{C}) = \sum_{K \in \mathcal{C}} \rho(\{K\})$.

Тогда ρ – дискретная вероятностная мера на $\mathcal{K}$ и для каждого $S \subseteq G \times M$ выполняется $\nu_{\mathcal{M}}(S) = |\tilde{J}|/|J|$, где $\tilde{J}$ – наибольшее подмножество J, удовлетворяющее условию $\forall j \in \tilde{J}\ (S \subseteq I_j)$. Нетрудно проверить, что тогда $(\mathcal{K}, \rho)$, действительно, является вероятностной моделью. Определенную таким образом модель $\mathcal{M}$ назовем *частотной вероятностной моделью*.

Проиллюстрируем на примере введенные нами определения.

Пример 2. *Пусть даны множества* $G = \{g_1, g_2\}$ *и* $M = \{m_1, m_2, m_3\}$. *Рассмотрим класс* $\mathcal{K} = \{(G, M, I_j)\}_{j \in \{1,2,3\}}$, *состоящий из трех контекстов, приведенных ниже в табличном виде:*

I_1	m_1	m_2	m_3
g_1	×		×
g_2		×	

I_2	m_1	m_2	m_3
g_1		×	×
g_2	×	×	

I_3	m_1	m_2	m_3
g_1	×		×
g_2	×	×	

Тогда пары $(\{g_1\}, \{m_1, m_2, m_3\})$ *и* $(\{g_1, g_2\}, \{m_1, m_2\})$ *являются единственными вероятностными понятиями контекста* (G, M, I_1) *в частотной вероятностной модели* $\mathcal{M} = (\mathcal{K}, \rho)$.

Доказательство. Вероятностная мера ρ однозначно определяет значение $\eta_{\mathcal{M}}(A \to m)$ для каждой импликации $A \to m$ на множестве M. В таблицах ниже приведем значения вероятности всех возможных импликаций вида $A \to m$ на $\mathcal{M}$, не являющихся тавтологиями.

$A \to m$	$\eta_{\mathcal{M}}(A \to m)$
$\{\varnothing\} \to m_1$	$2/3$
$m_2 \to m_1$	0
$m_3 \to m_1$	$2/3$
$m_2, m_3 \to m_1$	0
$\{\varnothing\} \to m_2$	$1/3$
$m_1 \to m_2$	0

$A \to m$	$\eta_{\mathcal{M}}(A \to m)$
$m_3 \to m_2$	$1/3$
$m_1, m_3 \to m_2$	0
$\varnothing \to m_3$	0
$m_1 \to m_3$	0
$m_2 \to m_3$	0
$m_1, m_2 \to m_3$	0

Посылки импликаций, которые составляют множество $D(\mathcal{M})$ всех сильнейших вероятностных закономерностей на $\mathcal{M}$, отмечены фигурными скобками. Приведем пример вычисления вероятности одной из импликаций в таблице выше:

$\eta_{\mathcal{M}}(m_3 \to m_1) = inf_{g \in G}\ \mu_{\mathcal{M}}(\langle g, m_3 \to m_1 \rangle) = inf_{g \in G} \frac{\nu_{\mathcal{M}}(\{<g,m_3>,<g,m_1>\})}{\nu_{\mathcal{M}}(\{<g,m_3>\})} =$
$= \frac{\nu_{\mathcal{M}}(\{<g_1,m_3>,<g_1,m_1>\})}{\nu_{\mathcal{M}}(\{<g_1,m_3>\})} = 2/3$, поскольку значение $\mu_{\mathcal{M}}(\langle g_2, m_3 \to m_1 \rangle)$ не определено, вследствие $\nu_{\mathcal{M}}(\{< g_2, m_3 >\}) = 0$. Отметим, что импликация $m_3 \to m_1$ не является вероятностной закономерностью, поскольку существует импликация $\varnothing \to m_1$ с таким же значением вероятности на $\mathcal{M}$.

Приведем значение оператора $f_{D(\mathcal{M})}$ на подмножествах $B \subseteq M$:

$B \subseteq M$	$f_{D(\mathcal{M})}(B)$
m_1	m_1, m_2
m_2	m_1, m_2
m_3	m_1, m_2, m_3
m_1, m_2	m_1, m_2

$B \subseteq M$	$f_{D(\mathcal{M})}(B)$
m_1, m_3	m_1, m_2, m_3
m_2, m_3	m_1, m_2, m_3
m_1, m_2, m_3	m_1, m_2, m_3
$\varnothing$	m_1, m_2

Очевидно, ровно два подмножества $B \subseteq M$ удовлетворяют условию $f_{D(\mathcal{M})}(B) = B$, а именно $\{m_1, m_2\}$ и $\{m_1, m_2, m_3\}$. Наконец,

$$\bigcup\{E' \mid \varnothing \neq E \subseteq \{m_1, m_2\},\ \bar{f}_{D(\mathcal{M})}(E) = \{m_1, m_2\}\} = \{g_1, g_2\},$$

$$\bigcup\{E' \mid \varnothing \neq E \subseteq \{m_1, m_2, m_3\},\ \bar{f}_{D(\mathcal{M})}(E) = \{m_1, m_2, m_3\}\} = \{g_1\}.$$

Единственное подмножество $E \subseteq \{m_1, m_2, m_3\}$, удовлетворяющее условиям в определении вероятностного понятия, это множество $\{m_3\}$ и для него имеем $\{m_3\}' = g_1$.

Таким образом, $(\{g_1\}, \{m_1, m_2, m_3\})$ и $(\{g_1, g_2\}, \{m_1, m_2\})$ являются единственными вероятностными понятиями контекста (G, M, I_1) в модели $\mathcal{M}$. □

4 Вероятностные понятия на одном контексте

В разделе 3 было рассмотрено понятие вероятностной модели первого типа, определенной на классе контекстов. Фактически, любой класс контекстов $\mathcal{K}$, на котором можно ввести вероятностную меру, порождает некоторое множество вероятностных моделей и посредством этого определяет возможные семейства импликаций, являющиеся сильнейшими вероятностными закономерностями. С помощью таких семейств импликаций осуществлялось «предсказание» атрибутов у объектов в произвольном выбранном контексте из класса $\mathcal{K}$. Аналогично описанному подходу можно выделить сильнейшие вероятностные закономерности на основе лишь одного заданного формального контекста. Для этого нам потребуется только незначительно изменить определение 5 вероятностной модели.

Определение 9. *Вероятностной моделью второго типа (вероятностным контекстом) назовем пару $\mathcal{M} = (K, \rho)$, где $K = (G, M, I)$ – контекст и ρ – вероятностная мера на множестве G, удовлетворяющая свойству*

$$\forall\ B, C \subseteq M\ (B' \subseteq C' \Leftrightarrow \rho((B \cup C)') = \rho(B')).$$

Если $B \to m$ – некоторая импликация на множестве M, то ее вероятностью на модели $\mathcal{M}$ назовем значение функции

$$\eta_{\mathcal{M}}(B \to m) = \begin{cases} \frac{\rho((B \cup \{m\})')}{\rho(B')}, & \text{если } \rho(B') \neq 0 \\ \text{не определено} & \text{в противном случае} \end{cases}$$

В данном разделе для краткости будем называть пары (K, ρ) из определения выше *вероятностными моделями* или просто *моделями*.

Если $K = (\varnothing \neq G, M, I)$ – конечный контекст, то назовем модель $\mathcal{M} = (K, \rho)$ *частотной вероятностной моделью*, если для каждого $g \in G$ имеем $\rho(\{g\}) = 1/|G|$ и для каждого подмножества $A \subseteq G$ $\rho(A) = \sum_{g \in A} \rho(\{g\})$. Таким образом, $\forall B \subseteq M\ (\rho(B') = |B'|/|G|)$. Подчеркнем, что $\mathcal{M}$, действительно, является моделью, т.к. для любых подмножеств $B, C \subseteq M$ верно $B' \subseteq C' \Leftrightarrow (B \cup C)' = B' \Leftrightarrow |(B \cup C)'| = |B'|$.

Замечание 5. *Для любой вероятностной модели $\mathcal{M} = (K, \rho)$, где $K = (G, M, I)$, и любой импликации $B \to m$ на множестве M имеем $\eta_{\mathcal{M}}(B \to m) = 1$ тогда и только тогда, когда $B \to m \in Imp(K)$ и $B' \neq \varnothing$ (где $'$ – операция в рамках контекста K).*

Доказательство. Если $\eta_{\mathcal{M}}(B \to m) = 1$, то $\rho(B') \neq \varnothing$ и, следовательно, $B' \neq \varnothing$, т.е. посылка B не ложна на K. С другой стороны, это условие означает, что $\rho((B \cup \{m\})') = \rho(B')$, поэтому имеем $B' \subseteq \{m\}'$, что равносильно $B \to m \in Imp(K)$. Этот же аргумент доказывает утверждение в обратную сторону. □

Определим понятия вероятностной закономерности, максимально специфичной вероятностной закономерности и сильнейшей вероятностной закономерности на модели второго типа полностью аналогично определениям 6 и 7. Будем использовать то же обозначение $D(\mathcal{M})$ для множества всех сильнейших вероятностных закономерностей на модели второго типа $\mathcal{M}$, что и в разделе 3.

Предложение 2. *Пусть $\mathcal{M} = (K, \rho)$ – вероятностная модель, где $K = (G, M, I)$, а $S \subseteq Imp(K)$ – множество, состоящее из всех тавтологий на M и всех импликаций, посылка которых ложна на K. Тогда выполнено $MinImp(K) \setminus S \subseteq D(\mathcal{M})$.*

Доказательство. Действительно, для каждой импликации $B \to m \in MinImp(K) \setminus S$ верно $B' \neq \varnothing$, поэтому, в силу замечания 5, имеем $\eta_{\mathcal{M}}(B \to m) = 1$. Условие максимальности значения вероятности для импликации $B \to m$ на модели $\mathcal{M}$ выполнено и, очевидно, не может существовать вероятностной закономерности $B_1 \to m$ на $\mathcal{M}$ такой, что $B \subset B_1$. Кроме того, импликация $B \to m$ сама является вероятностной закономерностью, так как из условия $B \to m \in MinImp(K) \setminus S$ и замечания 5 для любого множества $B_0 \subset B$ имеем $\eta_{\mathcal{M}}(B_0 \to m) < 1$. Таким образом, все условия в определении сильнейшей вероятностной закономерности выполнены и $B \to m \in D(\mathcal{M})$. □

Определение 10. *Пусть $\mathcal{M} = (K, \rho)$ – вероятностная модель второго типа, где $K = (G, M, I)$. Вероятностным понятием в модели $\mathcal{M}$ (понятием вероятностного контекста $\mathcal{M}$) назовем пару множеств (A, B), удовлетворяющих условиям определения 8.*

Пусть $\mathcal{M} = (K, \rho)$ – вероятностная модель, где $K = (G, M, I)$. Рассмотрим контекст $\overline{K} = (G, M, \bar{I})$, где $\bar{I} = \{< g, m > \mid \ g \in G,\ m \in \bar{f}_{D(\mathcal{M})}(g')\}$, $'$ - операция в рамках контекста K. Другими словами, $I \subseteq \bar{I}$ и отношение $\bar{I}$ получается из исходного I добавлением пар $< g, m >$, «предсказываемых» семейством импликаций $D(\mathcal{M})$. Для пояснения взаимосвязи понятий в контексте K и вероятностных понятий в модели $\mathcal{M}$ важно отметить, что *следующее утверждение не верно в обе стороны:*

для любых непустых подмножеств $A \subseteq G$ и $B \subseteq M$ пара (A, B) является вероятностным понятием в модели $\mathcal{M}$ тогда и только тогда, когда (A, B) является понятием в контексте $\overline{K}$.

Для подтверждения этого достаточно рассмотреть любой из приведенных ниже контекстов $K_1 = (\{g_1, g_2\}, \{m_1\},\ I_1)$, $K_2 = (\{g_1, g_2\}, \{m_1, m_2, m_3\},\ I_2)$ и соответствующие им частотные вероятностные модели $\mathcal{M}_1 = (K_1, \rho_1)$ и $\mathcal{M}_2 = (K_2, \rho_2)$.

I_1	m_1
g_1	×
g_2	

I_2	m_1	m_2	m_3
g_1	×		
g_2		×	×

Для этих моделей имеем $D(\mathcal{M}_1) = \{\varnothing \to m_1\}$ и $D(\mathcal{M}_2) = \{\varnothing \to m_1,\ \{m_2\} \to m_3, \{m_3\} \to m_2\}$. Поэтому множество всех вероятностных понятий в модели $\mathcal{M}_1$ состоит ровно из одного понятия $(\{g_1\}, \{m_1\})$, а множество $\{\ (\{g_1\}, \{m_1\}), (\{g_2\}, \{m_1, m_2, m_3\})\ \}$ – есть все вероятностные понятия в модели $\mathcal{M}_2$.

Легко убедиться в том, что для каждого $j = 1, 2$ контекст $\overline{K}_j$ получается из K_j определением $\bar{I}_j = I_j \cup \{< g_2, m_1 >\}$. Остается лишь подчеркнуть, что множество всех понятий в контексте $\overline{K}_1$ состоит ровно из одной пары $(\{g_1, g_2\}, \{m_1\})$, а множество

$\{\ (\{g_1, g_2\}, \{m_1\}),\ (\{g_2\}, \{m_1, m_2, m_3\})\ \}$, соответственно, есть все понятия в контексте $\overline{K}_2$.

Тем не менее, можно гарантировать следующее свойство, характеризующее взаимосвязь понятий в контексте K и вероятностных понятий в модели $\mathcal{M} = (K, \rho)$:

Теорема 2. *Для любой вероятностной модели $\mathcal{M} = (K, \rho)$, где $K = (G, M, I)$, выполняется следующее:*

1. *если (A, B) – понятие в контексте K, причем $A \neq \varnothing \neq B$, то существует вероятностное понятие (A_1, B_1) в модели $\mathcal{M}$ такое, что $A \subseteq A_1$ и $B \subseteq B_1$;*

2. *если (A_1, B_1) – вероятностное понятие в модели $\mathcal{M}$, то найдется понятие (A, B) в контексте K такое, что $\varnothing \neq A \subseteq A_1$ и $\varnothing \neq B \subseteq B_1$. Более того, множество A_1 является объединением объемов некоторых таких понятий.*

Доказательство. 1. Пусть $S \subseteq Imp(K)$ множество, состоящее из всех тавтологий на M и всех импликаций, посылка которых ложна на K. Поскольку (A, B) – понятие в контексте K, имеем $B'' = B$, $B' = A \neq \varnothing$ и по предложению 1 тогда получаем, что $f_{MinImp(K)\setminus S}(B) = B$.

По предложению 2 имеет место включение $MinImp(K) \setminus S \subseteq D(\mathcal{M})$. Кроме того, для любых семейств импликаций L_1 и L_2 на множестве M и любого подмножества $B \subseteq M$ из $L_1 \subseteq L_2$ следует $\bar{f}_{L_1}(B) \subseteq \bar{f}_{L_2}(B)$, поэтому $B \subseteq \bar{f}_{D(\mathcal{M})}(B)$. Обозначим $B_1 = \bar{f}_{D(\mathcal{M})}(B)$, $\mathcal{C} = \{E \subseteq B_1 \mid \bar{f}_{D(\mathcal{M})}(E) = B_1,\ E \neq \varnothing \neq E'\}$ и $A_1 = \cup\{E' \mid E \in \mathcal{C}\}$. Ясно, что $\bar{f}_{D(\mathcal{M})}(B_1) = B_1$. Кроме того, заметим, что $B \in \mathcal{C}$, $A = B'$ и $B' \subseteq A_1$. Поэтому имеем $A \subseteq A_1$ и, таким образом, (A_1, B_1) – искомое вероятностное понятие в модели $\mathcal{M}$.

2. Рассмотрим множество $\mathcal{C} = \{E \subseteq B_1 \mid \bar{f}_{D(\mathcal{M})}(E) = B_1,\ E \neq \varnothing \neq E'\}$ и произвольное $E \in \mathcal{C}$. Имеем $MinImp(K) \setminus S \subseteq D(\mathcal{M})$, поэтому $\bar{f}_{MinImp(K)\setminus S}(E) \subseteq B_1$. Обозначим $B = \bar{f}_{MinImp(K)\setminus S}(E)$; ясно, что $\bar{f}_{MinImp(K)\setminus S}(B) = B$. Кроме того, из $E \neq \varnothing \neq E'$ следует $B \neq \varnothing \neq B'$, поэтому по предложению 1 получаем $B'' = B$. С другой стороны, $E \subseteq B$, поэтому $E' \supseteq B'$ и $A_1 = \cup\{E' \mid E \in \mathcal{C}\} \supseteq B'$. Получаем, что (B', B) – искомое понятие в контексте K.

Теперь заметим, что условие $B = \bar{f}_{MinImp(K)\setminus S}(E)$ влечет $E \to B \in Imp(K)$, что эквивалентно $E' \subseteq B'$; поэтому получаем $E' = B'$. В силу произвольного выбора множества $E \in \mathcal{C}$ и того, что $A_1 = \cup\{E' \mid E \in \mathcal{C}\}$, заключаем, что множество A_1 является объединением объемов некоторых понятий (A, B) в контексте K, у которых $\varnothing \neq B \subseteq B_1$. □

Далее приведем схемы вычислительных процедур поиска вероятностных закономерностей и вероятностных понятий для заданной частотной вероятностной модели $\mathcal{M} = (K, \rho)$, где $K = (G, M, I)$ и $\forall m \in M\ (\{m\}' \neq \varnothing)$.

Пусть $S \subseteq Imp(K)$ – множество, состоящее из всех тавтологий на M и всех импликаций, посылка которых ложна на K. Отметим, что для заданного контекста K мощность множества $MinImp(K) \setminus S$ может экспоненциально зависеть от значения $|G| \times |M|$. Это следует из теоремы 1 в [8], где приводится пример построения такого контекста. По предложению 2 имеем $MinImp(K) \setminus S \subseteq D(\mathcal{M})$, а множество всех вероятностных закономерностей на $\mathcal{M}$ по определению содержит $D(\mathcal{M})$. По этой причине процедура поиска вероятностных закономерностей реализуется с использованием эвристики.

Введем несколько вспомогательных обозначений. Длиной импликации $A \to m$ на множестве M будем называть мощность ее посылки, т.е. множества A, и использовать обозначение $len(A \to m)$. Скажем, что импликация $A_2 \to m$ является уточнением импликации $A_1 \to m$, если $A_2 = A_1 \cup \{n\}$, где $n \in M \setminus A_1$. Если L – некоторое семейство

импликаций, то через $Spec(L)$ будем обозначать множество всех возможных уточнений импликаций из L.

Вычислительная процедура обнаружения вероятностных закономерностей основана на идеях метода семантического вероятностного вывода. Ее суть заключается в последовательном уточнении импликаций с проверкой выполнимости условий для вероятностной закономерности. По существу реализуется направленный перебор импликаций, позволяющий существенно сократить пространство их поиска. Сокращение перебора достигается за счет использования эвристики, которая заключается в том, что, начиная с момента, когда длина посылки импликаций достигает некоторой заданной величины, называемой глубиной базового перебора, начинается последовательное уточнение только тех импликаций, которые являются вероятностными закономерностями.

Для простоты опишем вычислительную процедуру поиска вероятностных закономерностей вида $A \to m$ на модели $\mathcal{M}$ для некоторого выбранного атрибута $m \in M$. Кроме указанной вероятностной модели $\mathcal{M}$ и элемента $m \in M$, входным параметром данной процедуры также является глубина базового перебораd $(1 \leqslant d \leqslant |M|)$. Выходными данными является множество найденных вероятностных закономерностей на модели$\mathcal{M}$ с элементом m в заключении.

На шаге $k = 0$ генерируется множество $imp(\mathcal{M})_{(k)}$ импликаций, состоящее из одной импликации нулевой длины, имеющей вид $R = \varnothing \to m$. Импликация R проходит проверку на выполнение условий для вероятностной закономерности, сформулированных в определении 6. Обозначим множество всех вероятностных закономерностей, обнаруженных на k-ом шаге вычислительной процедуры через $REG^{(k)}_{\mathcal{M}}(m)$. Если R является вероятностной закономерностью, то $REG^{(0)}_{\mathcal{M}}(m) = \{R\}$. В противном случае $REG^{(0)}_{\mathcal{M}}(m) = \varnothing$ и на выход процедуры выдается пустое множество. Действительно, в этом случае $\eta_{\mathcal{M}}(\varnothing \to m) = 0$ и поэтому, в силу определения модели $\mathcal{M}$, вероятность любой импликации $B \to m$ либо не определена, либо равна нулю на $\mathcal{M}$, а это означает, что ни одна из таких импликаций не может быть вероятностной закономерностью на модели$\mathcal{M}$.

На шаге $1 \leqslant k \leqslant d$ формируется множество $imp(\mathcal{M})_{(k)}$ всех уточнений всех импликаций, построенных на предыдущем шаге, вероятность которых определена, но не равна нулю, либо единице: $imp(\mathcal{M})_{(k)} = Spec(\{R \mid R \in imp(\mathcal{M})_{(k-1)},\ 0 < \eta_{\mathcal{M}}(R) < 1\})$, каждая импликация в этом множестве имеет длину k. Все импликации из множества $imp(\mathcal{M})_{(k)}$ проходят проверку на выполнение условий для вероятностных закономерностей, и формируется множество $REG^{(k)}_{\mathcal{M}}(m)$.

На шаге $d < k \leqslant |M|$ генерируется множество $imp(\mathcal{M})_{(k)}$ всех уточнений всех вероятностных закономерностей, обнаруженных на предыдущем шаге, вероятность которых строго меньше единицы: $imp(\mathcal{M})_{(k)} = Spec(\{R \mid R \in REG^{(k-1)}_{\mathcal{M}}(m),\ \eta_{\mathcal{M}}(R) < 1\})$. Все полученные импликации из проходят проверку на выполнение условий для вероятностных закономерностей, и формируется множество $REG^{(k)}_{\mathcal{M}}(m)$. Исполнение вычислительной процедуры завершается либо на шаге$k = |M|$, либо если на очередном шаге $d < k < |M|$ не будет обнаружено ни одной вероятностной закономерности, т.е. когда$REG^{(k)}_{\mathcal{M}}(m) = \varnothing$. Искомое множество вероятностных закономерностей для атрибутаm будет равно объединению $\bigcup_k REG^{(k)}_{\mathcal{M}}(m)$; оно выдается на выход.

Чтобы из полученного множества импликаций выбрать семейство сильнейших вероятностных закономерностей (по отношению к входным параметрам процедуры), достаточно впрямую проверить условия определения 7.

Шаги $k \leqslant d$ называются базовым перебором, а шаги $k > d$ – дополнительным перебором. Как показывают эксперименты, для большого числа практических задач достаточно использовать глубину базового перебора $d \leqslant 3$. Отметим, что на практике проверка выполнимости неравенств, указанных в определении 6, осуществляется с уче-

том статистического критерия Фишера (точный критерий независимости Фишера для таблиц сопряженности), который применяется с некоторым (выбранным пользователем) доверительным уровнем α.

Пусть L – некоторое непустое множество вероятностных закономерностей на модели $\mathcal{M}$. Отметим, что если L является выходными данными приведенной выше процедуры для глубины базового перебора $d = |M|$, то $L = D(\mathcal{M})$.

Опишем итеративную процедуру нахождения вероятностных понятий в модели $\mathcal{M}$ относительно семейства импликаций L.
На шаге $k = 1$ генерируется множество $C^{(1)} = \{\bar{f}_L(A \cup \{m\}) \mid A \to m \in L\}$.

На шаге $k > 1$ в случае если $C^{(k-1)} = \varnothing$, на выход процедуры выдается список обнаруженных вероятностных понятий. В противном случае для каждого $B \in C^{(k-1)}$, рассматривая семейство импликаций $L_B = \{A \to m \in L \mid A \subseteq B\}$, вычисляем множество $A = \{g \in G \mid g' \cap B \neq \varnothing,\ f_{L_B}(g' \cap B) = B\}$. Если $A \neq \varnothing$, то пара (A, B) добавляется в список найденных вероятностных понятий. Далее генерируется множество $C^{(k)} = \{\bar{f}_L(B \cup C) \mid B, C \in C^{(k-1)},\ \bar{f}_L(B \cup C) \notin C^{(k-1)}\}$ и производится переход на следующий шаг итерации. Описание процедуры закончено.

Пример 3. Рассмотрим контексты K_1 и K_2, представленные на рисунке 2. Понятиями с непустыми объемом и содержанием в контексте K_1 являются пары $(\{g_1, \ldots, g_{20}\}, \{m_1, \ldots, m_5\})$ и $(\{g_{21}, \ldots, g_{40}\}, \{m_6, \ldots, m_{10}\})$. Контекст K_2 был получен из K_1 добавлением случайного шума. Задача заключалась в том, чтобы на зашумленном контексте восстановить исходные понятия. В соответствии с описанными алгоритмами, на частотной модели $\mathcal{M} = (K_2, \rho)$ было вычислено множество сильнейших вероятностных закономерностей, оно оказалось составленным из 22 импликаций. Множество вероятностных понятий в модели $\mathcal{M}$ совпало с множеством понятий в исходном контексте K_1, имеющих непустой объем и содержание.

5 Заключение

Из определений 5 и 9 легко заметить, что с теоретической точки зрения разделение между вероятностными моделями первого и второго типа весьма условное. В частности, для любой модели $\mathcal{M}_2 = (K, \rho_2)$ второго типа, где $K = (\varnothing \neq G, M, I)$, можно определить модель первого типа $\mathcal{M}_1 = (\mathcal{K}, \rho_1)$ так, чтобы $D(\mathcal{M}_1) = D(\mathcal{M}_2)$. В самом деле, достаточно положить $\mathcal{K} = \{K_g \mid g \in G,\ K_g = (\{h\}, M, I_g),\ I_g = \{< h, m > \mid < g, m > \in I\}\}$ и определить $\forall\ \mathcal{C} \subseteq \mathcal{K}\ \rho_1(\mathcal{C}) = \rho_2(\{g \mid K_g \in \mathcal{C}\})$. Тогда для каждой импликации $B \to m$ на множестве M имеем $\eta_{\mathcal{M}_1}(B \to m) = \mu_{\mathcal{M}_1}(\langle h, B \to m \rangle) = \frac{\rho_1(\{K_g \mid \{<h,n> \mid n \in B \cup \{m\}\} \subseteq I_g\})}{\rho_1(\{K_g \mid \{<h,n> \mid n \in B\} \subseteq I_g\})} = \frac{\rho_2((B \cup \{m\})')}{\rho_2(B')}$ и поэтому $\eta_{\mathcal{M}_1}(B \to m) = \eta_{\mathcal{M}_2}(B \to m)$, что, очевидно, влечет $D(\mathcal{M}_1) = D(\mathcal{M}_2)$. Тем не менее, на практике оказывается важным разделение между анализом данных, представленных в виде класса контекстов, и анализ данных на основе лишь одного заданного контекста. В первом случае речь идет о задаче классификации объектов, наблюдаемых в серии экспериментов, в каждом из которых устанавливается, имеет ли объект тот или иной атрибут. Во втором случае классификация объектов строится на основе одного контекста, который представляет всю совокупность экспериментальных данных, полученных об этих объектах. Контекст однозначно определяет наличие того или иного атрибута у объекта, и метод FCA позволяет строить четкую классификацию объектов на основе заданного контекста. В свою очередь, выявление вероятностных закономерностей на модели, определенной на заданном контексте, позволяет получать устойчивые к шумам классификационные единицы.

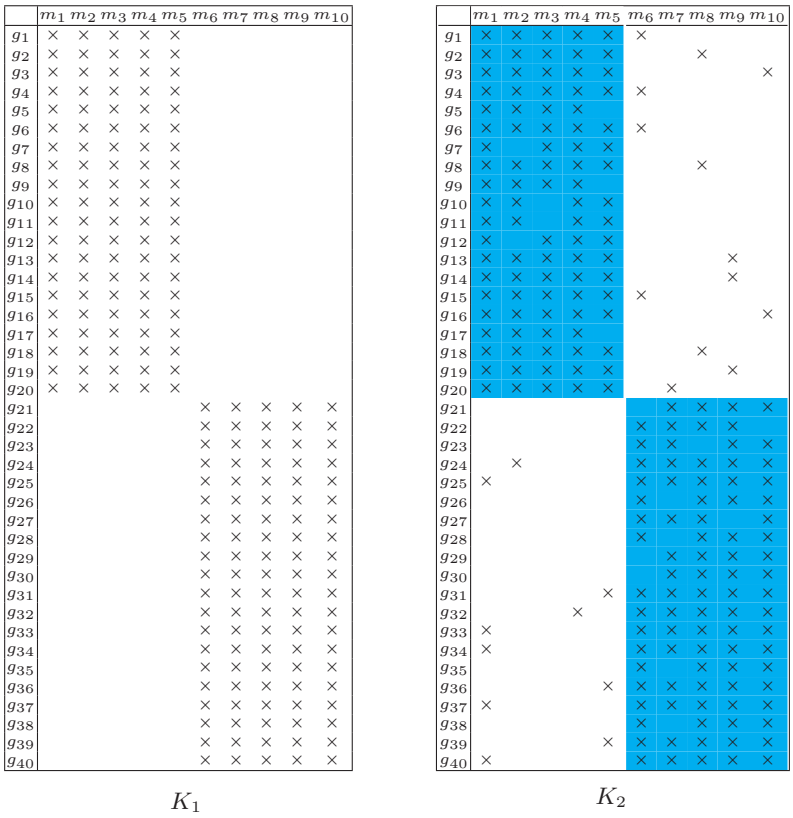

Рис. 2: Восстановленные понятия в зашумленном контексте.

Пример 3 показывает, что некоторый уровень шума не меняет множество понятий в контексте, а именно: множество понятий с непустыми объемом и содержанием в исходном заданном контексте совпадает с множеством вероятностных понятий нового контекста, полученного из исходного внесением шума. Существуют шумы (см. формальное определение в [1]), любой уровень которых не меняет множества понятий в контексте - множества понятий и вероятностных понятий для него совпадают. Такие шумы называются сохраняющими [1]. Это ставит проблему характеризации такого рода шумов.

В определениях импликаций и закономерностей, рассмотренных в данной статье, не используется понятие отрицания. Поэтому, в частности, теорема 2 звучит слабее, чем это можно было ожидать. Это связано с тем, что в самих основах метода «формального анализа понятий» не используется отрицание, и в данной статье мы хотели получить наиболее просто излагаемое обобщение основных определений данного метода. Расширение метода FCA в рамках рассмотренных идей позволит формализовать понятия «естественной классификации» и «идеализации», как определяется в [1, 13].

Семантический вероятностный вывод, лежащий в основе формализации вероятностных понятий, разработан для логики первого порядка и может обнаруживать достаточно сложные закономерности по сравнению с теми, которые были рассмотрены в данной работе. Более того, в реляционном подходе, изложенном в монографиях [1, 7], аргументируется, что использование языка первого порядка для формализации закономерностей принципиально важно для возможности в полном объеме анализировать информацию, содержащуюся в данных. Некоторые примеры таких закономерностей приведены на вебсайте [15] по адресу

http://math.nsc.ru/AP/ScientificDiscovery/pages/Examples_of_rules.html

Список литературы

[1] Витяев Е.Е. Компьютерное познание. — Новосибирский Гос. Ун-т, 293 с., 2006.

[2] Смердов С.О., Витяев Е.Е. Синтез логики, вероятности и обучения: формализация предсказания. // Сибирские электронные математические известия — 2009 — N 6 — С.340–365.

[3] Deogun J., Jiang L., Xie Y., Raghavan V. Probability logic modeling of knowledge discovery in databases. // Тез. ISMIS'2003 / Springer Verlag — С. 402–407, 2003.

[4] Ganter B., Wille R. Formal Concept Analysis: Mathematical Foundations. — Springer Verlag, 1999.

[5] Formal Concept Analysis: Foundations and Applications. / Edited by B. Ganter, G. Stumme, R.Wille — Springer Verlag, 2005.

[6] Guigues J.– L., Duquenne V. Families minimales d'implications informatives resultant d'un tableau de données binaires. // Math. Sci. Humaines — 1986 — N 95 — С.5–18.

[7] Kovalerchuk B., Vityaev E. Data Mining in Finance: Advances in Relational and Hybrid methods. — Kluwer, 2000.

[8] Kuznetsov S. On the intractability of computing the Duquenne–Guigues base. // J. UCS — 2004 — N 10 (8) — С. 927–933.

[9] Kuznetsov S., Obiedkov S. Comparing performance of algorithms for generating concept lattices // J. Exp. Theor. Artif. Intell. — 2002 — N 14 (2–3) — С. 189–216.

[10] Merwe D., Obiedkov S., Kourie D. AddIntent: A New Incremental Algorithm for Constructing Concept Lattices. // Тез. ICFCA'2004 / Springer Verlag — С. 372–385, 2004.

[11] Vityaev E. The logic of prediction. // Тез. The 9th Asian Logic Conference / World Scientific — С.263–276, 2006.

[12] Vityaev E., Kovalerchuk B. Empirical theories discovery based on the Measurement Theory. // Mind and Machine — 2005 — N 14 (4) — С. 551–573.

[13] Vityaev E., Lapardin K., Khomicheva I., Proskura A. Transcription factor binding site recognition by regularity matrices based on the natural classification method. // Intelligent Data Analysis. Спец. выпуск: «New Methods in Bioinformatics» под ред. Е. Витяева и Н. Колчанова. — 2008 — N 12 (5) — С. 495–512.

[14] Vityaev E., Smerdov S. New definition of prediction without logical inference // Тез. IASTED International Conference on Computational Intelligence (CI 2009) — С.48–54, 2009.

[15] Scientific Discovery website: http://math.nsc.ru/AP/ScientificDiscovery/