УДК 681.3.00:007

Д.Э. Ситников[1], О.А. Рябов[2], Е.В. Титова[1], О.А. Романенко[3]

[1]*Харьковская государственная академия культуры, Харьков, Украина*
[2]*Национальный институт современной промышленной науки и технологии, Цукуба, Япония*
[3]*ООО Укргазтехкомплекс, Харьков, Украина*

МЕТОД НАХОЖДЕНИЯ АППРОКСИМАЦИЙ ПРИБЛИЖЕННЫХ МНОЖЕСТВ И ПОСТРОЕНИЯ ЛОГИЧЕСКИХ ПРАВИЛ НА ОСНОВЕ АЛГЕБРЫ КОНЕЧНЫХ ПРЕДИКАТОВ

Математическая теория приближенных множеств, позволяющая получать модели в форме логических правил является мощным аппаратом интеллектуального анализа данных. Статья посвящена описанию алгебраического метода нахождения верхней и нижней аппроксимаций приближенных множеств. Данный метод использует только булевы операции над бинарными строками, что делает его весьма эффективным с вычислительной точки зрения.

интеллектуальный анализ данных, приближенное множество, неразличимые отношения, верхняя и нижняя аппроксимация, конечные предикаты, логические правила

Введение

Постановка проблемы. Интеллектуальный анализ данных (Data Mining) направлен на обнаружение в "сырых" данных определенных закономерностей (знаний) с помощью методов и алгоритмов искусственного интеллекта. Для представления полученных знаний служат модели, виды которых зависят от методов их создания. Представленная около 10 лет назад польским ученым Павлаком теория приближенных множеств [1, 2] представляет собой замечательную формальную базу для построения Data Mining моделей в форме логических правил. Полученные правила позволяют классифицировать новые объекты (записи в базе данных, элементы множества) на основании имеющихся характеристик. В отличие от классической теории множеств, в которой множество однозначно определяется набором элементов, о которых не требуется дополнительная информация, теория приближенных множеств (rough sets) утверждает, что для определения множества необходима некоторая информация об элементах. Причем с разными элементами может ассоциироваться одинаковая информация, т.е. такие элементы могут быть "неразличимы" с точки зрения имеющейся информации. Таким образом, одним из основных понятий данной теории является понятие "неразличимых" отношений. Для определения приближенного множества используются так называемые верхняя и нижняя аппроксимации. Нахождение верхней и нижней аппроксимаций приближенного множества представляет собой определенную проблему и является достаточно трудоемким процессом, т.к. требует многократного просмотра базы данных. Безусловно, решением этой проблемы была бы возможность определения приближенных аппроксимаций алгебраическим путем, который существенно сокращает количество операций в процессе генерации логических правил.

Анализ последних исследований и публикаций. В теории приближенных множеств, предложенной Павлаком [1, 2] утверждается, что некоторые объекты (элементы) на основании имеющейся информации не могут быть однозначно отнесены ни к некоторому множеству, ни к его дополнению. Например, в медицинской базе данных могут существовать записи о пациентах, имеющих одинаковые симптомы, однако часть из них имеет некоторое заболевание, часть – нет. Какое из решений должно быть принято, когда база данных содержит противоречивую информацию и как классифицировать нового пациента с теми же симптомами? В данном случае мы имеем дело с приближенным множеством, которое может быть определено с помощью верхней и нижней аппроксимаций.

Классические определения нижней и верхней аппроксимаций были введены для описания топологических свойств приближенных множеств [1 – 3]. Эти определения базировались на отношении "неразличимости", которое рассматривалось как отношение эквивалентности.

Пусть X – некоторое множество, являющееся подмножеством универсума U и I – отношение "неразличимости". Нижней аппроксимацией X является:

$$I_*(X) = \{x \in U : I(x) \subseteq X\},$$

а верхней аппроксимацией:

$$I^*(X) = \{x \in U : I(x) \cap X \neq 0\},$$

где I(x) определяет множество объектов, "неразличимых" с x (с точки зрения имеющейся информации) [1].

В дальнейшем были предложены более общие определения приближенных аппроксимаций, большинство из которых основаны на менее строгих типах отношений [4]. В частности, некоторые авторы рассматривают отношение толерантности (рефлексивное и симметричное) как основу аппроксимаций, что позволяет выразить более слабые формы "неразличимости" [5]. В некоторых статьях предлагаются интересные определения приближенных аппроксимаций, основанные на отношении подобия [6].

Авторы вышеуказанных работ используют классический топологический подход к определению аппроксимаций приближенных множеств, т.е. рассматривают "неразличимые" отношения, определяют их специфические свойства и после этого вводят понятия нижней и верхней аппроксимаций. В работе [7] был предложен иной подход к определению аппроксимаций с использованием конечных предикатов и без использования понятия "неразличимого" отношения.

Целью статьи является разработка метода нахождения верхней и нижней аппроксимаций приближенных множеств с использованием конечных предикатов и булевых операций, что позволит быстро определять аппроксимации множества и на их основе генерировать логические правила.

Алгебраический подход к определению приближенных множеств

Классическая теория приближенных множеств тесно связана с понятием "неразличимости". Эта концепция рассматривает приближенные множества, некоторые элементы которые не могут быть "различимы" в терминах некоего отношения. В данной работе предлагается иной подход к концепции неразличимости.

Предположим, что имеющаяся информация об элементах множества представляется с помощью их свойств. Такой подход позволяет рассматривать любые отношения, определённые на множестве свойств элементов, пар элементов, упорядоченных множеств элементов и т.д. К этим отношениям могут применяться различные операции для получения в итоге новых отношений. Таким образом, любая информация об элементах множества может быть представлена в форме отношений. Главная проблема формулируется следующим образом: как описать некоторое отношение между элементами множества в терминах существующих отношений, т.е. каким образом представить новую информацию в терминах имеющейся информации. Иногда возможно точно представить данное отношение в терминах имеющихся отношений, но чаще это невозможно. Таким образом, возникает необходимость рассматривать некоторые аппроксимации для описания отношений. Продемонстрируем вышесказанное на простом примере.

Пусть на элементах $a_1, a_2, a_3, a_4, a_5, a_6$ определены два множества: $A=\{a_1, a_2, a_3, a_4\}$ и $B=\{a_2, a_4, a_5\}$ и множество $X=\{a_2, a_4\}$. Что можно сказать о множестве X в терминах множеств A и B? В данном случае можно утверждать, что элемент принадлежит X тогда и только тогда, когда он принадлежит одновременно A и B. Другими словами, X является пересечением двух множеств: $X = A \cap B$.

Пусть теперь $X=\{a_2, a_3, a_4\}$. В этом случае, мы не можем получить множество X из множеств A и B, используя операции пересечения, объединения и отрицания и будем вынуждены рассматривать множества, приближающиеся к X. Например, можно утверждать, что если элемент принадлежит пересечению A и B, то он однозначно принадлежит X, т.е. $\{a_2, a_4\}$ является нижней аппроксимацией X.

Рассмотрим теперь обобщенный подход к описанию элементов одного множества в терминах других множеств.

Предположим, существует конечное непустое множество объектов $U=\{a_1, a_2, \ldots, a_n\}$, называемое универсумом. Рассмотрим также множество унарных предикатов (функций принимающих одно из значений из множества $\{0,1\}$) определённых на U: $P_1(t), P_2(t), \ldots, P_k(t)$ и называемых координатами.

Предикаты $P_1(t), P_2(t), \ldots, P_k(t)$ могут интерпретироваться как характеристические функции для некоторых свойств объектов универсума. В этом случае объект a_i имеет свойство P_j тогда и только тогда, когда соответствующий предикат обращается в единицу: $P_j(a_i)=1$. Следуя базовым концепциям теории приближенных множеств, нам необходимо описать некоторое множество $X \subseteq U$ в терминах координат. Поскольку существует однозначное соответствие между всеми предикатами, определёнными на U и всеми подмножествами U, то вместо $X \subseteq U$ мы можем рассматривать предикат X(t) который равен 1 тогда и только тогда, когда $t \in X$. Таким образом мы должны дать описание некоторого предиката X(t) в терминах предикатов $P_1(t), P_2(t), \ldots, P_k(t)$.

В [7] было предложено рассматривать некоторый логический или алгебраический язык (язык аппроксимации), позволяющий выявить связь между X(t) и $P_1(t), P_2(t), \ldots, P_k(t)$. Будем считать, что язык аппроксимации состоит из унарных предикатов $P_1(t), P_2(t), \ldots, P_k(t)$ и булевых операций. Следует отметить, что в общем случае язык аппроксимации может включать другие типы предикатов и операций.

Рассмотрим множество Ф всех возможных формул, которые строятся с помощью булевых операций конъюнкции, дизъюнкции и отрицания, применяемых к предикатам $P_1(t), P_2(t), \ldots, P_k(t)$. Например: $(P_1 \wedge P_2 \wedge \overline{P_3}) \vee P_4$, $(P_1 \vee (P_2 \wedge P_3)) \wedge P_4$, $\overline{(P_1 \wedge P_2 \wedge \ldots \wedge P_k)}$ и т.д. Множество предикатов, получаемое применением формул из множества Ф, обозначим Λ. Следует отметить, что различные формулы могут определять один и тот же предикат.

Например, пусть U={a_1, a_2}; $P_1(a_1)=1$, $P_1(a_2)=1$, $P_2(a_1)=1$, $P_2(a_2)=0$. Формулы $\varphi = P_1 \wedge P_2$ и $\pi=P_1$ дадут нам одинаковый результат – предикат P_1, хотя не являются логически эквивалентными и не могут быть получены одна из другой путем эквивалентных преобразований.

Таким образом, если предикат X(t) принадлежит Λ, это означает что X(t) может быть выражен в терминах координат и множество X, соответствующее предикату X(t), может быть названо "точным" (четким) в терминах данных координат. Если предикат X(t) не принадлежит к Λ, этот предикат не может быть описан в терминах координат, и мы можем описать его приблизительно.

Также с точки зрения предложенного в [7] подхода, можно дать следующие определения верхней и нижней аппроксимации для X(t), а также определения точной верхней и точной нижней аппроксимаций:

Определение 1. Если A(t)∈Λ и ∀t∈U, A(t)→X(t), тогда мы говорим что A(t) есть нижняя аппроксимация X(t).

Определение 2. Если B(t)∈Λ и ∀t∈U, X(t)→B(t), тогда мы говорим что B(t) есть верхняя аппроксимация X(t).

Определение 3. Если предикат $I_*(t)$ есть нижняя аппроксимация предиката X(t) и для любых аппроксимаций A(t) этого предиката ∀t∈U, $A(t)\to I_*(t)$, то мы говорим, что $I_*(t)$ есть точная нижняя аппроксимация X(t).

Определение 4. Если предикат $I^*(t)$ есть верхняя аппроксимация предиката X(t) и для любых аппроксимаций B(t) этого предиката ∀t∈U, $I^*(t)\to B(t)$, то мы говорим что $I^*(t)$ есть точная верхняя аппроксимация X(t).

Свойства аппроксимаций

1. Множество нижних аппроксимаций не пусто для любых X(t). Это следует из того, что предикат $0 = P_1 \wedge \overline{P_1}$ всегда является нижней аппроксимацией для X(t).

2. Множество верхних аппроксимаций не пустое для любых X(t). Это следует из того, что предикат $1 = P_1 \vee \overline{P_1}$ всегда является верхней аппроксимацией для X(t).

3. Для любого предиката X(t) не может существовать более чем одна точная нижняя аппроксимация.

Предположим, что для предиката X(t) существуют две точные нижние аппроксимации $I_{*_1}(t)$ и $I_{*_2}(t)$. Рассмотрим предикат: $I_*(t) = I_{*_1}(t) \vee I_{*_2}(t)$. Очевидно, что $I_*(t)$ является нижней аппроксимацией, т.к. $I_{*_1}(t) \to I_*(t)$ и $I_{*_2}(t) \to I_*(t)$ ∀t∈U. С другой стороны $I_{*_1}(t)$ и $I_{*_2}(t)$ – точные нижние аппроксимации для X(t).

∀t∈U $I_*(t) \to I_{*_1}(t)$ и $I_*(t) \to I_{*_2}(t)$. Следовательно, $I_*(t) = I_{*_1}(t) = I_{*_2}(t)$.

4. Для любого предиката X(t) существует как минимум одна точная нижняя аппроксимация.

Рассмотрим предикат $I_*(t)$, являющийся дизъюнкцией всех нижних аппроксимаций X(t). Очевидно, что для любой нижней аппроксимации A(t) выполняется следующее: ∀t∈U $A(t) \to I_*(t)$. Таким образом, $I_*(t)$ является точной нижней аппроксимацией.

5. Для любого предиката X(t) существует одна единственная точная нижняя аппроксимация. Это следует из свойств 3 и 4.

6. Для любого предиката X(t) не может существовать более чем одна точная верхняя аппроксимация.

Предположим, что для предиката X(t) существуют две точные верхние аппроксимации $I_1^*(t)$ и $I_2^*(t)$. Рассмотрим предикат: $I^*(t) = I_1^*(t) \wedge I_2^*(t)$. Очевидно, что $I^*(t)$ является верхней аппроксимацией, т.к. $I^*(t) \to I_1^*(t)$ и $I^*(t) \to I_2^*(t)$ ∀t∈U. С другой стороны $I_1^*(t)$ и $I_2^*(t)$ – точные верхние аппроксимации для X(t).

∀t∈U $I_*(t) \to I_{*1}(t)$ и $I_*(t) \to I_{*2}(t)$. Следовательно, $I_*(t) = I_{*1}(t) = I_{*2}(t)$.

7. Для любого предиката X(t) существует как минимум одна точная верхняя аппроксимация.

Рассмотрим предикат $I^*(t)$, являющийся конъюнкцией всех верхних аппроксимаций X(t). Очевидно, что для любой верхней аппроксимации B(t) выполняется следующее: ∀t∈U $I^*(t) \to B(t)$. Таким образом, $I^*(t)$ является точной верхней аппроксимацией.

8. Для любого предиката X(t) существует одна единственная точная верхняя аппроксимация. Это следует из свойств 6 и 7.

Сравнение с классическими определениями аппроксимаций

Сравним классическое определение нижней и верхней аппроксимаций с определениями, введенными в данной статье. В классической теории приближенных множеств отношение "неразличимости" (эквивалентности) составляет математическую основу приближенных аппроксимаций. Предположим, дано отношение эквивалентности I, используемое для построения аппроксимаций множества X [2]. Отношение I определяет множество классов эквивалентности, где каждому классу может быть поставлен в соответствие предикат, принимающий значение 1, если элемент множества принадлежит этому классу, и значение 0 для всех других элементов универсума. Таким образом, имеется множество предикатов $P_1(t), P_2(t), ..., P_k(t)$, удовлетворяющих следующим условиям, которые соответствуют известным свойствам классов эквивалентности:

$\exists t\ P_i(t)$, i=1,2,...,k, (1)

$\forall t \in U\ \ P_1(t) \vee P_2(t) \vee ... \vee P_k(t)$ (2)

$\forall t \in U\ \ (\overline{(P_i(t) \wedge P_j(t))}$; i,j=1,2,...,k; i≠j; (3)

Таким образом, из классического определения приближенных аппроксимаций может быть получено следующее:

1. Нижняя аппроксимация множества X является объединением классов эквивалентности подмножеств X.

2. Верхней аппроксимацией множества X является объединение всех классов эквивалентности, которые имеют непустое пересечение с X.

Используя аппарат математической логики, мы можем переформулировать эти определения следующим образом:

Определение 5. Нижней аппроксимацией для X(t) является дизъюнкция всех предикатов $P_i(t)$, для которых $\forall t \in U\ P_i(t) \to X(t)$.

Определение 6. Верхней аппроксимацией для X(t) является дизъюнкция всех предикатов $P_i(t)$, для которых $\exists t \in U\ P_i(t) \wedge X(t)$.

Покажем, что определения 5 и 6 эквивалентны соответственно определениям 3 и 4.

Рассмотрим произвольный предикат X(t), $t \in U$. При построении любой булевой формулы из предикатов $P_1(t), P_2(t), ..., P_k(t)$ мы получаем дизъюнкцию, составленную из некоторого числа этих предикатов ($P_i(t) \wedge P_j(t) \equiv 0$, i≠j; операция отрицания, примененная к дизъюнкции некоторого числа предикатов $P_i(t)$ дает дизъюнкцию оставшихся предикатов (согласно (2) и (3)).

Таким образом:

- любая нижняя аппроксимация X(t) является дизъюнкцией некоторого числа предикатов $P_i(t)$, для которых $\forall t \in U\ P_i(t) \to X(t)$.
- точная нижняя аппроксимация для X(t) является дизъюнкцией всех нижних аппроксимаций, полученных согласно определению 1 (свойство 4).
- дизъюнкция D всех предикатов $P_i(t)$, для которых $\exists t \in U\ P_i(t) \wedge X(t)$ является верней аппроксимацией для X(t).
- любая другая верхняя аппроксимация F является дизъюнкцией некоторого числа предикатов $P_i(t)$, которые обязательно входят в D.

Таким образом, мы показали, что определения 3 и 5, а также 4 и 6 эквивалентны между собой и мы можем описывать приближенные множества с помощью аппроксимаций в терминах алгебры предикатов.

Метод нахождения точных аппроксимаций предиката

Перебор всех возможных булевых формул для получения точных аппроксимаций, безусловно, является достаточно трудоемкой процедурой.

В данной статье предлагается метод, позволяющий получать аппроксимации предиката без перебора всех формул.

Пусть требуется описать предикат X в терминах координат $P_1(t), P_2(t), ..., P_k(t)$.

Таблица 1

Унарные предикаты и предикат X, определённые на универсуме (в общем виде)

Элементы универсума / Координаты	a_1	a_2	…	a_n
P_1	δ_{11}	δ_{12}	…	δ_{1n}
P_2	δ_{21}	δ_{22}	…	δ_{2n}
…	…	…	…	…
P_k	δ_{k1}	δ_{k1}	…	δ_{kn}
Предикат X	λ_1	λ_2	…	λ_n

В табл.. 1: $\delta_{ij}, \lambda_i \in \{0,1\}$;

если $\delta_{ij}=1$, то $P_j(a_i)=1$;

если $\delta_{ij}=0$, то $P_j(a_i)=0$;

если $\lambda_i=1$, то $X(a_i)=1$;

если $\lambda_i=0$, то $X(a_i)=0$.

Для нахождения точной верхней аппроксимации X рассмотрим те столбцы таблицы, на которых X обращается в 1 и запишем соответствующую дизъюнктивную нормальную форму.

Для нахождения точной нижней аппроксимации X рассмотрим те столбцы таблицы, на которых предикат X обращается в 0 и запишем соответствующую конъюнктивную нормальную форму.

Продемонстрируем вышесказанное на простом примере.

Таблица 2

Пример координат и предиката X, определённых на универсуме

Элементы универсума / Координаты	a_1	a_2	a_3	a_4	a_5
P_1	1	0	0	1	0
P_2	0	0	1	1	1
P_3	0	1	0	0	0
Предикат X	0	1	0	1	1

Для предиката X получаем следующую формулу верней аппроксимации:

$$I^* = \left(\overline{P_1} \wedge \overline{P_2} \wedge P_3\right) \vee \left(P_1 \wedge P_2 \wedge \overline{P_3}\right) \vee \left(\overline{P_1} \wedge P_2 \wedge \overline{P_3}\right). \quad (4)$$

Также согласно предложенному методу получим формулу нижней аппроксимации:

$$I_* = \left(\overline{P_1} \vee P_2 \vee P_3\right) \wedge \left(P_1 \vee \overline{P_2} \vee P_3\right). \quad (5)$$

Из формул (4) и (5) получаем результат, приведенный в табл. 3. В общем случае предикаты I^* и I_* могут быть получены следующим образом:

Таблица 3

Верхняя и нижняя аппроксимации предиката X

Элементы универсума / Координаты	a_1	a_2	a_3	a_4	a_5
P_1	1	0	0	1	0
P_2	0	0	1	1	1
P_3	0	1	0	0	0
Предикат X	0	1	0	1	1
I^*	0	1	1	1	1
I_*	0	1	0	1	0

$$I^* = (\lambda_1 \wedge P_1 * \delta_{11} \wedge P_2 * \delta_{21} \wedge \ldots \wedge P_k * \delta_{k1}) \vee \vee(\lambda_2 \wedge P_1 * \delta_{12} \wedge P_2 * \delta_{22} \wedge \ldots \wedge P_k * \delta_{k2}) \vee \ldots \ldots \vee (\lambda_n \wedge P_1 * \delta_{1n} \wedge P_2 * \delta_{2n} \wedge \ldots \wedge P_k * \delta_{kn}). \quad (6)$$

$$I_* = (\lambda_1 \vee P_1 * (1-\delta_{11}) \vee P_2 * (1-\delta_{21}) \vee \ldots \ldots \vee P_k * (1-\delta_{k1})) \wedge (\lambda_2 \vee P_1 * (1-\delta_{12}) \vee \vee P_2 * (1-\delta_{22}) \vee \ldots \vee P_k * (1-\delta_{k2})) \vee \ldots \ldots \vee (\lambda_n \vee P_1 * (1-\delta_{1n}) \vee P_2 * (1-\delta_{2n}) \vee \ldots \ldots \vee P_k * (1-\delta_{kn})). \quad (7)$$

где $P*\delta=P$, если $\delta=1$ и $P*\delta=\overline{P}$, если $\delta=0$ для всех предикатов P.

Покажем, что предикаты I^* и I_* являются точной верхней и нижней аппроксимациями. Очевидно, что предикат I^* – верхняя аппроксимация X согласно определению 2. Если удалить из формулы (6) любую конъюнкцию, где λ_i=1, то результирующая формула не будет являться формулой для верхней аппроксимации, т.к. предикат I^* будет обращаться в 0 на том столбце, где X обращается в 1 (предполагается, одинаковые конъюнкции удаляются). Например, если удалить $\overline{P_1} \wedge P_2 \wedge \overline{P_3}$ из (4), то $I^*(a_5)=0$, в то время как $X(a_5)$=1. Это означает, что аппроксимация I^* не может быть улучшена и является, таким образом точной верхней аппроксимацией. Предикат I_* очевидно является нижней аппроксимацией X согласно определению 1. Если удалить из (7) любую дизъюнкцию, где λ_i=0, то результирующая формула не будет являться формулой для нижней аппроксимации, т.к. предикат I_* будет обращаться в 1 на том столбце, где X обращается в 0 (предполагается, одинаковые дизъюнкции удаляются). Например, если удалить $P_1 \vee \overline{P_2} \vee P_3$ из (5), то $I_*(a_3)=1$, в то время как $X(a_3)$=0. Это означает, что аппроксимация I_* не может быть улучшена и является, таким образом, точной нижней аппроксимацией.

Следует отметить следующий момент. Одной из проблем Data Mining является большой объем анализируемых данных. Предлагаемый метод позволяет находить нижнюю и верхнюю аппроксимации множества за один просмотр таблицы (базы) данных, что делает его весьма привлекательным с вычислительной точки зрения.

Логические правила, получаемые на основе аппроксимаций приближенного множества

Рассмотрим пример генерации логических правил с помощью формул (4) и (5). Используя традиционную терминологию теории нечётких множеств, можно сказать, что правила, основанные на точной верхней аппроксимации **могут** существовать для анализируемого набора (базы) данных, а правила, основанные на точной нижней аппроксимации, **должны** выполняться для анализируемого набора (базы) данных.

Преобразуем выражение в правой части формулы (5) и получим:

$$I_* = P_3 \vee \left(\overline{P_1} \vee P_2\right) \wedge \left(P_1 \vee \overline{P_2}\right) = = P_3 \vee (P_1 \vee P_2) \wedge \left(\overline{P_1} \vee \overline{P_2}\right).$$

Исходя из полученной формулы можно сформулировать следующие точные правила:

A. Элемент принадлежит множеству X если:

a) он обладает свойством P_3 (P_3=1);

ИЛИ

b) он обладает свойствами P_1 и P_2;

ИЛИ

c) он не обладает ни свойством P_1, ни свойством P_2.

Это правило утверждает, что если хотя бы одно из условий a), b) и c) выполняется, то элемент принадлежит множеству X.

Обратимся теперь к формуле верхней аппроксимации (4) и упростим выражение в правой части.

$$I^* = \left(\overline{P_1} \wedge \overline{P_2} \wedge P_3\right) \vee \left(P_1 \wedge \overline{P_1}\right) \vee \left(P_2 \wedge \overline{P_3}\right) = = \left(\overline{P_1} \wedge \overline{P_2} \wedge P_3\right) \vee \left(P_2 \wedge \overline{P_3}\right).$$

Полученное выражение позволяет сформулировать следующие приблизительные (те, которые могут выполняться на анализируемом наборе данных) правила:

B. Элемент может принадлежать множеству X, если он не обладает ни свойством P_1, ни свойством P_2 И обладает свойством P_3 (P_3=1);

C. Элемент может принадлежать множеству X, если он обладает свойством P_2 и не обладает свойством P_3.

Отметим, что правило B можно исключить из набора полученных правил, т.к. согласно правилу A, если элемент обладает свойством P_3, то он обязательно принадлежит множеству X.

Таким образом, было показано, что логические правила позволяют на основе имеющейся информации об элементах дать ответ на вопрос, принадлежит или нет элемент данному множеству. Также следует отметить, что подобные расчёты (оперирование булевыми переменными) являются очень быстрыми и позволяют получать результирующие правила даже для большого числа элементов и характеристических функций.

Выводы

В данной статье представлен алгебраический метод нахождения аппроксимаций приближенных множеств, основной идеей которого является описание одного множества в терминах других множеств. Данный подход может применяться к иным логическим и алгебраическим структурам, описывающим приближенные множества. Например, отказываясь от применения операции отрицания или добавляя иные операции над предикатами, можно получить новые интересные свойства аппроксимаций.

Понятия точной верхней и точной нижней аппроксимаций были введены для того, чтобы подчеркнуть тот факт, что для произвольного множества может существовать несколько различных аппроксимаций. Однако среди существующих аппроксимаций возможно выбрать такие, которые не могут быть улучшены в терминах языка аппроксимации. Таким аппроксимации наиболее точно описывают приближенное множество.

Логические правила, получаемые с помощью точных аппроксимаций могут быть использованы для классификации новых элементов, т.е.для определения принадлежит или нет элемент множеству, что является классической задачей Data Mining.

Также следует отметить, что предлагаемый метод использует только операции сравнения и булевы операции, что делает процесс поиска логических правил очень быстрым с вычислительной точки зрения.

Представляется перспективным определение наиболее важных и интересных свойств (предикатов), участвующих в классификации элементов, а также рассмотрение случая, когда характеристические функции могут принимать значения из множества возможных (небинарный случай).

Список литературы

1. *Pawlak Z. Rough sets // International Journal of Computer and Information Sciences. – 1982. – №11. – P. 341-356.*
2. *Pawlak Z. Vagueness and uncertainty: a rough set perspective // Computational Intelligence. – 1995. – Vol. 11 (issue 2). – P. 227-232.*
3. *Pawlak Z. Rough sets, rough relations and rough functions // Fundamenta Informaticae. – 1996. – Vol. 27, № 2/3. – P. 103-108.*
4. *Yao, Y.; Wong, S. Generalization of rough sets using relationships between attribute values // Proc. of the 2nd Annual Joint Conference on Information Sciences. – 1995. – P. 30-33.*
5. *Polkowski, L.; Skowron, A.; Zytkow, J. Tolerance based rough sets // Lin, T; Wildberger, A.; (eds.), Soft computing: Rough Sets, Fuzzy Logic, Neural Networks, Uncertainty Management, Simulation councils. – San Diego, 1995. – P. 55-58.*
6. *Slowinski R.; Vanderpooten D. Similarity relation as a bases for rough approximations // Wong, P.P.; (ed.), Advances in Machine Intelligence and Soft-Computing. – Bookwrights, Raleigh, NC, 1997. – Vol. IV. – P. 17-33.*
7. *Sitnikov D., Ryabov O. An algebraic approach to defining rough set approximations and generating logic rules // Zanasi, A.; Ebecken, N.; Brebbia, C.; (eds), Data Mining V. – Malaga, Spain, 2004. – P. 179-188.*

Поступила в редколлегию 9.04.2007

Рецензент: д-р техн. наук, доцент И.В. Гребенник, Харьковский национальный университет радиоэлектроники, Харьков.