УДК 517.9

ПОИСК МАКСИМАЛЬНЫХ ИНТЕРВАЛОВ ОБЛАСТИ ПРИЗНАКОВОГО ПРОСТРАНСТВА В ЛЕКСИКОГРАФИЧЕСКОМ ПОРЯДКЕ

© Ильченко А.В.

Таврический национальный университет им. В.И. Вернадского
факультет математики и информатики
пр-т Вернадского, 4, г. Симферополь, 95007, Украина
e-mail: *ilch@crimea.edu*

Abstract. The properties of characteristic vector families for intervals of the feature space are under investigation. The search algorithms for the closed characteristic vectors and vectors generating the maximum intervals of the specified feature space region are considered in the paper.

Введение

Алгоритмы и методы кластерного анализа во многом составляют основу приложений анализа данных. Рост объёма данных, подлежащих обработке, предъявляет ряд требований к используемым алгоритмам кластеризации: возможность находить кластеры в пространствах большой размерности, наглядность, лёгкость интерпретации полученных результатов, отсутствие необходимости приведения исходных данных к какому-либо каноническому виду и так далее.

В некоторых случаях выполнение многих из этих требований могут обеспечить алгоритмы и методы анализа формальных понятий [1, 2, 6], сеточные методы.

Целью статьи является рассмотрение понятия индикаторного вектора интервала признакового пространства, изучение свойств семейства таких векторов, построение и обоснование алгоритма поиска замкнутых индикаторных векторов и индикаторных векторов, порождающих максимальные интервалы заданной области признакового пространства.

1. Максимальные интервалы области признакового пространства

$\{A_0, A_1, \dots, A_{n-1},\}$ – семейство множеств, каждое из которых конечно и линейно упорядоченно. Прямое произведение множеств семейства A определяет n-мерное признаковое пространство S:

$$S = A_0 \times A_1 \times \dots \times A_{n-1}.$$

Иногда на обозначение A_j ссылаются как на измерение (атрибут, домен) признакового пространства. Поскольку множество значений по каждому A_j линейно упорядоченно, можно говорить о положительном и отрицательном направлениях по соответствующему измерению.

$n_i = |A_i|$, $i = 0, 1, \dots, n-1$ – мощность домена A_i.

Элементы домена A_i можно перенумеровать и, в дальнейшем, всегда считать, что $A_i = \{0, 1, \ldots, n-1\}$.

Пусть $H \subseteq S$ – некоторая область признакового пространства S. Интервал I, содержащийся в области H, называется максимальным интервалом области H, если не существует его собственного надинтервала, содержащегося в области H.

Множество всех максимальных интервалов области H называется сокращенной интервальной структурой этой области [4]. Каждую область H признакового пространства S можно представить в виде объединения интервалов ее сокращенной интервальной структуры.

Удобно точки области H называть единичными точками (1-точки) так, как если бы была задана индикаторная функция этой области, а интервалы области H называть 1-интервалами. Те точки признакового пространства, которые не содержатся в области H, будем называть 0-точками признакового пространства.

2. Векторное представление интервалов

У пространства S, размерность которого n, имеется $2n$ направлений. В каждом измерении по два направления: отрицательное и положительное. Любой интервал может быть получен из интервала, представляющего пространство S, удалением некоторого числа значений из доменов, по соответствующим направлениям.

Для описания интервала будет использоваться вектор длины $2n$:

$$A = (a_0, a_1, \ldots, a_{2i}, a_{2i+1}, \ldots, a_{2n-2}, a_{2n-1}) \tag{1}$$

Компоненты такого вектора рассматриваются парами. Каждая пара соответствует одному из измерений признакового пространства. Каждая компонента такой пары отвечает одному из направлений соответствующего измерения.

Пара a_{2i}, a_{2i+1} соответствует i-му измерению.

Компонента a_{2i} отвечает отрицательному направлению i-го измерения.

Компонента a_{2i+1} - положительному направлению.

Значение каждой компоненты такой пары - число значений, удалённых из домена A_i по соответствующему направлению.

Такой вектор будем называть индикаторным вектором интервала или векторным описанием интервала.

Если вектор A является индикаторным вектором интервала I, то удобно говорить, что вектор A порождает интервал I (обозначение $A(I)$), а интервал I порождается вектором A (обозначение $I(A)$).

Когда вектор A является описанием непустого интервала, то для каждого измерения признакового пространства сумма компонент по этому измерению меньше мощности домена значений этого измерения:

$$a_{2i} + a_{2i+1} < n_i, \; i = 0, 1, \ldots, n-1. \tag{2}$$

Если для какого-либо из измерений $a_{2i} + a_{2i+1} \geq n_i$, это означает, что такой вектор является описанием пустого подмножества.

Примечание 1. Для нас представляют интерес непустые интервалы. Поэтому вектора-описания, в которых хотя бы для одного измерения выполняется условие

$a_{2i} + a_{2i+1} \geq n_i$, исключаются из рассмотрения. Если всё же потребуется использовать пустое подмножество, то в качестве индикаторного вектора будет использоваться вектор, в котором значение каждой компоненты равно мощности домена по соответствующему измерению признакового пространства, то есть вектор, в котором

$$a_{2i} = a_{2i+1} = n_i, \; i = 0, 1, \ldots, n-1. \tag{3}$$

3. Лексикографический порядок на множестве индикаторных векторов

Определение 1. Пусть A, B – индикаторные вектора. Вектор A лексически предшествует вектору B, если существует направление k такое, что $a_k < b_k$, а по всем направлениям, предшествующим направлению k, значения компонент векторов A и B совпадают:

$$A \lesssim B \Leftrightarrow \exists k(a_k < b_k, \forall j < k(a_j = b_j)). \tag{4}$$

Для сокращения записи будет использоваться следующее обозначение:

$$A \lesssim_k B \Leftrightarrow a_k < b_k, \forall j < k(a_j = b_j). \tag{5}$$

Тогда, определение лексического порядка запишется в виде

$$A \lesssim B \Leftrightarrow \exists k(A \lesssim_k B). \tag{6}$$

Удобно индикаторные вектора перебирать в лексическом порядке, начиная с вектора $(0, 0, \ldots, 0, 0)$, который является индикаторным вектором признакового пространства S.

Алгоритм вычисления вектора, непосредственно лексически следующим за текущим вектором, основан на следующих соображениях.

Лексический порядок – порядок линейный. Поэтому вектора, расположенные в лексическом порядке, можно пронумеровать числами $0, 1, \ldots$, начиная с вектора $(0, 0, \ldots, 0, 0)$. Тогда переход от текущего вектора, к вектору непосредственно лексически следующим за текущим, осуществляется прибавлением единицы к номеру текущего вектора и записи векторного представления полученного числа. То есть, индикаторный вектор можно рассматривать как позиционную запись неотрицательного целого числа.

Отличия от обычного позиционного представления чисел заключаются в следующем:

1. Для каждой пары позиций, соответствующих одному из измерений, используется своё основание, которое совпадает с мощностью домена этого измерения.
2. Для каждой пары позиций, соответствующих одному из измерений, должно выполняться ограничение (2), что позволяет исключить из рассмотрения пустые интервалы.

4. Вычисление непосредственно лексически следующего индикаторного вектора

Псевдокод алгоритма.

Вход. A – текущий индикаторный вектор;
Выход. $A+1$ – индикаторный вектор, непосредственно лексически следующий за вектором A.
Метод.

$i = n - 1$
while $i \geq 0$
    $a_{2i+1} = a_{2i+1} + 1$
    if $a_{2i} + a_{2i+1} < n_i$ **then return** A
    $a_{2i+1} = 0$
    $a_{2i} = a_{2i} + 1$
    if $a_{2i} < n_i$ **then return** A
    $a_{2i} = 0$
    $i = i - 1$
end while
return A

Приведенный псевдокод основан на соображениях предшествующего раздела.

В дальнейшем, вычисление вектора, непосредственно лексически следующего за текущим вектором A, будет обозначаться $A+1$.

5. Замкнутые интервалы признакового пространства

Пусть $A \subseteq S$ – некоторое подмножество точек признакового пространства.

Обозначим $[A]$ наименьший интервал пространства S, содержащий подмножество точек A (интервальное замыкание, интервальная оболочка множества A).

Операция интервального замыкания обладает обычными свойствами операции замыкания [5]:

$$A \subseteq B \Rightarrow [A] \subseteq [B], \tag{7}$$

$$A \subseteq [A], \tag{8}$$

$$[[A]] = [A]. \tag{9}$$

Обозначим $I(S)$ – семейство всех интервалов признакового пространства S.

$I \in I(S)$ – фиксированный интервал из этого семейства.

$I \cap H$ – подмножество точек области H, содержащихся в интервале I (1-точки пространства S, содержащиеся в интервале I).

$[I \cap H]$ – наименьший интервал пространства S, содержащий подмножество $I \cap H$ (интервальное замыкание подмножества 1-точек пространства S, содержащихся в интервале I).

Определение 2. Интервал $I \in I(S)$ называется замкнутым (относительно области H)), если $[I \cap H] = I$.

Утверждение 1. *Каждый 1-интервал является замкнутым интервалом.*

Доказательство. Пусть $I \in I(S)$ и $I \subseteq H$.

Так как $I \subseteq H$, то $I \cap H = I$ и, следовательно, $[I \cap H] = [I]$.

Это, с учётом очевидного равенства $[I] = I$, позволяет сделать вывод, что $[I \cap H] = I$. □

Примечание 2. В силу утверждения 1, максимальные 1-интервалы имеет смысл искать среди замкнутых интервалов.

Примечание 3. Замкнутые интервалы интересны и сами по себе. Например, в анализе формальных понятий [1], [2] их можно рассматриваться как содержания «интервальных» формальных понятий. В задачах распознавания образов замкнутые интервалы могут быть использованы для построения различных гипотез [6].

6. Решётка интервалов и решётка индикаторных векторов

Для семейства $I(S)$ выполняется свойство замыкания [3] относительно операции пересечения. Поэтому $I(S)$, упорядоченное отношением включения $\subseteq$, является полной решёткой.

Обозначим $A(S)$ семейство всех индикаторных векторов признакового пространства S.

Между множествами $I(S)$ и $A(S)$ существует взаимно-однозначное соответствие (с учётом примечания 1):

$I(A) \leftrightarrow A(I)$.

На векторах семейства $A(S)$ следующим образом определяется отношение предшествование $\leq$:

$$A \leq B \Leftrightarrow I(B) \subseteq I(A). \qquad (10)$$

Так определённое отношение предшествования является отношением порядка, двойственным к порядку, определяемому отношением включения $\subseteq$ на множестве $I(S)$. Поэтому, множество $A(S)$, упорядоченное отношением $\leq$, является двойственным к решётке $\langle I(S), \subseteq \rangle$ и, следовательно, является полной решёткой.

С использованием введенного отношения предшествования, определяются отношения строгого $<$ и непосредственного $\prec$ предшествования.

$A < B \Leftrightarrow A \leq B$ и $A \neq B$.

$A \prec B \Leftrightarrow A < B$ и $\bar{\exists} C \in A(S) : A < C < B$.

Утверждение 2. *Пусть* $A, B \in A(S)$.

$$A \leq B \Leftrightarrow a_j \leq b_j,\ j = 0, 1, \ldots, 2n - 1. \qquad (11)$$

Доказательство. $A \leq B \Leftrightarrow I(B) \subseteq I(A)$

$\Leftrightarrow$ по каждому из $2n$ направлений признакового пространства, в интервале $I(A)$ удалено не больше значений, чем удалено по соответствующим направлениям в интервале $I(B)$

$\Leftrightarrow a_j \leq b_j,\ j = 0, 1, \ldots, 2n - 1.$ □

Из утверждения 2 следует

Утверждение 3. *Решёточный порядок $\leq$ на множестве индикаторных векторов с сохранением порядка вложен в лексический порядок $\lesssim$:*

$$A \leq B \Rightarrow A \lesssim B. \quad (12)$$

Так же, из утверждения 2 следует

Утверждение 4.

$$A \prec B \Leftrightarrow \exists! k(a_k + 1 = b_k, \forall j \neq k(a_j = b_j)). \quad (13)$$

7. Оператор замыкания на решетке индикаторных векторов

Пусть $H \subseteq S$ – область пространства S. На множестве $A(S)$ определим оператор $''$, который каждому индикаторному вектору A ставит в соответствие индикаторный вектор A'', порождающий наименьший интервал, содержащий те же 1-точки, что и интервал, порождаемый исходной вектором A.

Результат действия оператора $''$, удобно представить в виде следующей последовательности шагов:

$A \Rightarrow I(A) \Rightarrow I(A) \cap H \Rightarrow [I(A) \cap H] \Rightarrow A''$.

В силу своего определения, оператор $''$' обладает следующими свойствами, аналогичными свойствам (7) – (9):

$$A \leq B \Rightarrow A'' \leq B'', \quad (14)$$

$$A \leq A'', \quad (15)$$

$$(A'')'' = A'', \quad (16)$$

и, следовательно, является оператором замыкания [3] на множестве $A(S)$.

Вектор A'' называется замыканием вектора A.

Определение 3. Индикаторный вектор A называется замкнутым, если он совпадает со своим замыканием, то есть, если $A'' = A$.

В силу определения оператора замыкания, именно замкнутые индикаторные вектора порождают интересующие нас замкнутые интервалы.

Если у вектора A его компоненты обозначены как a_j, $j = 0, 1, \ldots, 2n-1$, то соответствующие компоненты вектора A'' будут обозначаться a''_j, $j = 0, 1, \ldots, 2n-1$.

В силу монотонности оператора замыкания (14) и утверждения 2, компоненты векторов A и A'' удовлетворяют условию

$$a_j \leq a''_j, \ j = 0, 1, \ldots, 2n-1. \quad (17)$$

Пусть $A \in A(S)$ и $A \lesssim_m A+1$. В этом случае

$A = (a_0, \ldots, a_m, a_{m+1}, \ldots, a_{2n-1})$ и $A + 1 = (a_0, \ldots, a_m + 1, 0, \ldots, 0)$, где

$$m = max\{j : a_j \neq 0, \ j = 0, 1, \ldots, 2n-1, \} \quad (18)$$

то есть, m – наибольшее направление признакового пространства, по которому компонента a_j вектора $A+1$ отлична от нуля.

Обозначим M – подмножество всех векторов, которым вектор A предшествует по компоненте m:

$$M = \{D \in A(S) \mid A \lesssim_m D\}. \tag{19}$$

Любой вектор $C \in A(S)$ такой, что $A \lesssim_k C$ и $k < m$ находится лексически дальше от вектора A, чем каждый из векторов подмножества M. Поэтому, именно среди векторов подмножества M следует искать замкнутый вектор, лексически ближайший к вектору A. И только, если в подмножестве M искомого вектора не окажется, переходить к векторам, которым вектор A лексически предшествует по направлениям, меньшим направления m.

В множестве $A(S)$, упорядоченном отношением лексического порядка, подмножество M является интервалом. Обозначим этот интервал $[B, B{+}{+}]$. Нижняя граница этого интервала B - вектор, непосредственно лексически следующий за вектором A, то есть, $B = A+1$. Вектор $B{+}{+}$ определяется по вектору $A+1$ и соответствующему измерению m (18).

Утверждение 5. $\forall A \in A(S)\ :\ A \lesssim (A+1)''$.

Утверждение 6. *Пусть $A \in A(S)$. Если $A \lesssim_m A+1$, D - замкнутый вектор такой, что $A \lesssim_m D$, то $(A+1)'' \leq D$ и, следовательно, $(A+1)'' \lesssim D$.*

Утверждение 7. *Пусть $A \in A(S)$. Если $A \lesssim_m A+1$, D - замкнутый вектор такой, что $A \lesssim_m D$, то $A \lesssim_m (A+1)''$.*

Из утверджений 6 и 7 вытекают следующие утверждения.

Утверждение 8. *Если $A \lesssim_m A+1$ и $A \lesssim_m (A+1)''$, то $(A+1)''$ - лексически наименьший замкнутый вектор среди замкнутых индикаторных векторов, лексически следующих за вектором A.*

Утверждение 9. *Если $(A+1)''$ не является лексически наименьшим замкнутым вектором среди замкнутых индикаторных векторов, лексически следующих за вектором A, то не выполняется условие $A \lesssim_m (A+1)''$.*

Последнее утверждение позволяет сформулировать условие проверки, является ли вектор $(A+1)''$ искомым индикаторным вектором.

Утверждение 10. *Пусть*

$$k = min\{j \mid a''_j \neq a_j,\ j = 0, 1, \ldots, 2n-1\}, \tag{20}$$

то есть, k – наименьшее направление признакового пространства, по которому компоненты векторов A и $(A+1)''$ не совпадают. Тогда, если $k < m$, то $(A+1)''$ не является искомым вектором.

Таким образом, если выполняются условия утверждения 10, интервал $[B, B{+}{+}]$, где $B = A+1$, не содержит замкнутых векторов и его элементы следует пропустить. Для этого следует перейти к вектору, непосредственно лексически следующему за вектором $B{+}{+}$.

Интервал $[B, B{+}{+}]$ удобно называть исключаемым интервалом незамкнутых векторов.

8. Алгоритм поиска лексически следующего замкнутого индикаторного вектора

Алгоритм $lexicalNextClose(A)$ – алгоритм поиска лексически наименьшего замкнутого индикаторного вектора, следующего за вектором A.

Псевдокод алгоритма $lexicalNextClose(A)$.

Вход. A – текущий замкнутый индикаторный вектор;
Выход. C – лексически наименьший замкнутый индикаторный вектор, следующий за вектором A.
Метод.

1) $B = A + 1$
2) $C = A''$
3) **repeat**
4) $\quad m = max\{j \mid b_j \neq 0,\ j = 0, 1, \ldots, 2n - 1\}$
5) $\quad k = min\{j \mid b_j \neq c_j,\ j = 0, 1, \ldots, 2n - 1\}$
6) $\quad$ **if** $k < m$ **then begin**
7) $\qquad B = B + +$
8) $\qquad B = B + 1$
9) $\qquad C = B''$
10) $\quad$ **end**
11) **until** $k < m$
12) **return** C

Пояснения к псевдокоду алгоритма $lexicalNextClose(A)$.

Шаг 1.	B – вектор, непосредственно лексически следующий за A.
Шаг 2.	C – замыкание вектора B.
Шаги 3 - 11.	Цикл, осуществляющий вычисление искомого вектора.
Шаг 4.	m – наибольшее измерение признакового пространства, по которому компонента вектора B отлична от нуля.
Шаг 5.	k – наименьшее измерение признакового пространства, по которому компоненты векторов B и C не совпадают.
Шаг 6.	Проверка, является ли вычисленное C искомым вектором. Если условие выполняется, то – не является.
Шаг 7.	Пропуск лексического интервала незамкнутых векторов, следующих за текущим вектором B.
Шаг 8.	Замена текущего значения вектора B на вектор, непосредственно лексически следующий за вектором $B + +$.
Шаг 9.	C – замыкание вектора B.
Шаг 11.	Проверка, найден ли искомый вектор.
Шаг 12.	Возврат найденного вектора C.

9. Исключаемый интервал

Пусть $A \in A(S)$ – 1-вектор, $I(A)$ – порождаемый им интервал.

$$A^{\triangle}_{\leq} = \{D \in A(S) \mid A \leq D\}. \tag{21}$$

Элементы подмножества $A^{\triangle}_{\leq}$ – это 1-вектора, порождающие подинтервалы интервала $I(A)$.

$$A^{\triangle}_{\lesssim} = \{D \in A(S) \mid A \lesssim D\}. \tag{22}$$

Элементы подмножества $A^{\triangle}_{\lesssim}$ – все вектора, лексически следующие за вектором A.

В силу линейности лексического порядка $\lesssim$, подмножество $A^{\triangle}_{\lesssim}$ – цепь в упорядоченном множестве $\langle A(S), \lesssim \rangle$, начинающаяся с вектора A.

Рассмотрим на подмножестве $A^{\triangle}_{\lesssim}$ подцепочку максимальной длины, начальный элемент которой вектор A, и состоящую из элементов подмножества $A^{\triangle}_{\leq}$, непосредственно лексически следующих друг за другом.

Такая подцепочка образует некоторый интервал $[A, B]$ подмножества $A^{\triangle}_{\lesssim}$. Вектор A – нижняя грань этого интервала. Вектор B - это вектор из подмножества $A^{\triangle}_{\leq}$ такой, что непосредственно лексически следующий за ним вектор (вектор $B+1$) уже не является элементом подмножества $A^{\triangle}_{\leq}$.

1-интервалы, порождаемые векторами из интервала $[A, B]$, не являются максимальными (за исключением, может быть, вектора A), и, следовательно, могут быть исключены из рассмотрения (пропущены). Таким образом, от вектора A можно сразу перейти к вектору B.

Интервал $[A, B]$ будем называть исключаемым интервалом 1-векторов.

Вычисление вектора B по заданному вектору A осуществляется следующим образом.

Пусть $A = (a_0, \ldots, a_{2n-1})$ и

$$m = max\{i \mid a_{2i} \neq 0 \vee a_{2i+1} \neq 0,\ i = 0, 1, \ldots, n-1\}. \tag{23}$$

то есть, m - наибольшее измерение признакового пространства S, в котором хотя бы по одному из направлений есть отличная от нуля компонента.

Различаются два случая: $a_{2m+1} = 0$ и $a_{2m+1} \neq 0$.

Каждый из этих случаев рассмотрим отдельно.

Случай 1). $a_{2m+1} = 0$.

Здесь вектора A и B имеют следующий вид:

$$A = (a_0, \ldots, a_{2m-1}, a_{2m}, 0, \ldots, 0), \tag{24}$$

$$B = (a_0, \ldots, a_{2m-1}, b_{2m}, b_{2m+1}, \ldots, b_{2n-1}), \tag{25}$$

где

$$\begin{array}{lllllll} b_{2m} & = & (n_m - 1), & b_{2m+1} & = & 0, \\ b_{2m+2} & = & (n_{m+1} - 1), & b_{2m+3} & = & 0, \\ \ldots & & & & & \\ b_{2n-2} & = & (n_{n-1} - 1), & b_{2n-1} & = & 0. \end{array} \tag{26}$$

Для компонент векторов A и B выполняются условия покомпонентного предшествования:

$a_0 = b_0, \ldots, a_{2m-1} = b_{2m-1}, a_{2m} \le b_{2m}, \ldots, a_{2n-1} \le b_{2n-1}$.

Отсюда, в силу (11), $A \le B$ и, следовательно, $B \in A^{\triangle}_{\le}$.

Кроме того, так как $A \le B$, то $A \lesssim B$, а значит, $[A, B]$ – интервал в $A^{\triangle}_{\lesssim}$.

Каждый вектор $D \in [A, B]$ имеет вид:

$D = (a_0, \ldots, a_{2m-1}, d_{2m}, d_{2m+1}, \ldots, d_{2n-1})$,

и для его компонент выполняются условия

$a_j \le d_j, \; j = 0, 1, \ldots, 2n-1$.

Поэтому $A \le D$ и, следовательно, $D \in A^{\triangle}_{\le}$.

Таким образом, $[A, B] \subseteq A^{\triangle}_{\le}$.

Покажем, что вектор, непосредственно лексически следующий за вектором B, не принадлежит подмножеству $A^{\triangle}_{\le}$.

Пусть $C = B{+}1$ – вектор, непосредственно лексически следующий за вектором B.

В векторе C компонента $c_{2m} = 0$. Так как $a_{2m} \ne 0$, то $c_{2m} \not\le a_{2m}$ и, следовательно, $A \not\le C$, то есть, $C \notin A^{\triangle}_{\le}$.

Случай 2). $a_{2m+1} \ne 0$.

Здесь вектора A и B имеют следующий вид:

$$A = (a_0, \ldots, a_{2m-1}, a_{2m}, a_{2m+1}, 0, \ldots, 0), \tag{27}$$

$$B = (a_0, \ldots, a_{2m-1}, a_{2m}, b_{2m+1}, \ldots, b_{2n-1}), \tag{28}$$

где

$$\begin{array}{lcllcl} b_{2m} & = & a_{2m}, & b_{2m+1} & = & (n_m - 1) - a_{2m}, \\ b_{2m+2} & = & (n_{m+1} - 1), & b_{2m+3} & = & 0, \\ \ldots & & & & & \\ b_{2n-2} & = & (n_{n-1} - 1), & b_{2n-1} & = & 0. \end{array} \tag{26}$$

Для этого случая обоснование почти полностью повторяет предыдущий вариант, за исключением того, что в векторе C теперь компонента $c_{2m+1} = 0$.

Таким образом, по заданному 1-вектору A, используя (26) или (29), можно вычислить исключаемый интервал 1-векторов $[A, B]$.

10. Расширение исключаемого интервала

A^- – вектор, непосредственно предшествующий вектору A по направлению $2m+1$.

Так как

$A = (a_0, \ldots, a_{2m-1}, a_{2m}, a_{2m+1}, 0, \ldots, 0)$, то, с учётом (13),

$A^- = (a_0, \ldots, a_{2m-1}, a_{2m}, a_{2m+1} - 1, 0, \ldots, 0)$.

$A^- \le A$ и, следовательно, $I(A) \subseteq I(A^-)$ и $A^- \lesssim A$.

Вектор A^- – это не 1-вектор. Если бы вектор A^- порождал 1-интервал, то $[A^-, B]$ был бы исключаемым интервалом 1-векторов, $A \in [A^-, B]$ и, поэтому, вектор A был бы исключён из рассмотрения.

Рассмотрим интервал J, определяемый следующим образом:

$$J = I(A^-) \setminus I(A). \tag{30}$$

Так как, $I \cap I(A) = \emptyset$ и $I(A^-) = J \cup I(A)$, то интервалы J и $I(A)$ порождают разбиение интервала A^-.

$I(A)$ – 1-интервал, $I(A^-)$ – не 1-интервал. Следовательно, именно интервал J, является интервалом, содержащим хотя бы одну 0-точку признакового пространства. Поэтому, любой интервал, содержащий интервал J, – не 1-интервал.

Индикаторный вектор интервала J:

$$A(J) = (a_0, \ldots, a_{2m-1}, (n_m - 1) - (a_{2m+1} - 1), a_{2m+1} - 1, 0, \ldots, 0). \tag{31}$$

Пусть $[A, B]$ – исключаемый интервал 1-векторов, порождаемый вектором A.

Рассмотрим следующие два шага.

Шаг 1).

$$B + 1 = (a_0, \ldots, a_{2m-1}, a_{2m} + 1, 0, \ldots, 0). \tag{32}$$

Так как, $B+1 \leq A(J)$, то $I(A(J)) \subseteq I(B+1)$ и, следовательно, $B+1$ – не 1-интервал.

$$C = (a_0, \ldots, a_{2m-1}, a_{2m} + 1, a_{2m+1} - 1, (n_{m+1} - 1), 0, \ldots, (n_{n-1} - 1), 0). \tag{33}$$

Так как, $C \leq A(J)$, то $I(A(J)) \subseteq I(C)$ и, следовательно, C – не 1-интервал.

Далее, $B + 1 \lesssim C$ и $\forall D \in [B + 1, C] \ : \ D \leq C$. Следовательно, $I(C) \subseteq I(D)$ и D – не 1-вектор. Поэтому, $[B+1, C]$ можно исключить из рассмотрения, тем самым расширив интервал $[A, B]$ до интервала $[A, C]$.

Шаг 2).

$$C + 1 = (a_0, \ldots, a_{2m-1}, a_{2m} + 1, a_{2m+1}, 0, 0, \ldots, 0, 0). \tag{34}$$

Так как, $A \leq C + 1$, то $I(C + 1) \subseteq I(A)$ и, следовательно, $[C + 1, (C + 1) + +] \subseteq A^{\triangle}_{\leq}$. Поэтому, интервал $[C+1, (C+1)++]$ можно исключить из рассмотрения, расширив интервал $[A, C]$ до интервала $[A, (C + 1) + +]$.

Обозначим $B = (C+1)++$ и расширенный таким образом исключаемый интервал снова запишем в виде $[A, B]$.

Шаги 1) и 2) повторять до тех пор, пока компоненты вектора B станут равными

$$b_{2m} \;=\; (n_m - 1) - a_{2m+1}, \quad b_{2m+1} \;=\; a_{2m+1}. \tag{35}$$

Примечание 4. Расширение исходного интервала $[A, B]$ проводилось только для случая $a_{2m+1} \neq 0$. Если $a_{2m+1} = 0$, то выражения (29) и (35) совпадут. Поэтому выражение (35) рассматривается как обобщение выражений (29) и (35).

Примечание 5. Вектор B, полученный в результате указанных вычислений будем обозначать $A + +m$, тем самым подчёркивая, что для его вычисления использован вектор A и информация по измерению m этого вектора (23).

11. Алгоритм поиска 1-векторов, порождающих максимальные 1-интервалы

Алгоритм $listMinAntichains$ – алгоритм поиска всех 1-векторов, порождающих максимальные 1-интервалы.

Псевдокод алгоритма $listMinAntichains$.

Вход. Нет;
Выход. B – список всех 1-векторов, порождающих максимальные 1-интервалы.
Метод.

1) **if** $(0,\ldots,0)$ – 1-вектор **return** $\{(0,\ldots,0)\}$
2) $B = \emptyset$
3) $A = (0,\ldots,0)''$
4) **repeat**
5)   **if** $(A-1)$ – 1-вектор **then begin**
6)   **if** $\overline{\exists} D \in B \,:\, D < A$ **then** $B = B \cup \{A\}$
7) $m = max\{i \mid a_{2i} \neq 0 \ \vee \ a_{2i+1} \neq 0,\ i = 0, 1, \ldots, n-1\}$
8) $A = A + +m$
9)   **end**
10) $A = lexicalNextClose(A)$
11) **until** $A \neq (0,\ldots,0)''$
12) **return** B

Пояснения к псевдокоду алгоритма $listMinAntichains$.

Шаг 1.	Если вектор, порождающий признаковое пространство, – 1-вектор, то он порождает единственный максимальный 1-интервал.
Шаг 2.	Изначально множество B искомых 1-векторов пусто.
Шаг 3.	A – текущий замкнутый вектор – замыкание вектора, порождающего признаковое пространство.
Шаги 4 - 11.	Цикл, осуществляющий вычисление искомого множества векторов.
Шаг 6.	Если текущий 1-вектор A – это максимальный 1-вектор, то включить его в множество B.
Шаг 7.	m – наибольшее измерение признакового пространства S, по которому хотя бы по одному из направлений есть отличная от нуля компонента.
Шаг 8.	Перейти к правой границе исключаемого интервала.
Шаг 10.	Найти замкнутый вектор, лексически следующий за текущим вектором A.

Заключение

В работе получены следующие результаты: Определено понятие интервала, замкнутого относительно заданной области признакового пространства и показано, что искомые максимальные интервалы являются замкнутыми. Ведено понятие индикаторного вектора интервала, как способа представления интервалов признакового пространства. На множестве таких векторов определён оператор замыкания, порождающий замкнутые индикаторные вектора, реализующие замкнутые интервалы. Предложен алгоритм поиска в лексикографическом порядке всех замкнутых индикаторных векторов. Предложен алгоритм поиска в лексикографическом порядке всех индикаторных векторов, порождающих максимальные интервалы области признакового пространства.

Список литературы

1. *Ganter, B., Wille, R.* Formal Concept Analysis – Mathematical Foundations. Springer-Verlag, Berlin., 1999.
2. *Gugisch, R.* Lattice Contexts – a Generalization in Formal Concept Analisys. Handuot to ICCS 2000, Darmstadt(2000) http://www.mathe2.uni-bayreuth.de/ralfg/papers/diplom.ps.gz.
3. *Биркгоф Г.* Теория решёток: Пер. с англ. – М.: Наука, 1984. – 568 с.
4. *Ильченко А.В.* Компактная компонентная и сокращённая интервальная структуры признакового пространства, порождаемые эмпирическими данными // Сб.Таврический вестник информатики и математики. – 2005. – №2. – С. 126-142.
5. *Колмогоров А.К., Фомин С.В.* Элементы теории функция и функционального анализа. – М.: Наука. Гл. ред. физ.-мат. Лит., 1989. – 624 с.
6. *Кузнецов С.О.* Автоматическое обучение на основа анализа формальных понятий // Автоматика и телемеханика. – 2001. – №10. – С. 3-27.

Статья поступила в редакцию 10.12.2010