АНАЛИЗ ДАННЫХ В ЭКСПЕРТНЫХ СИСТЕМАХ
(Вычислительные системы)
1986 год Выпуск 117

УДК 519.812.2.+519.7

МЕРЫ РАССТОЯНИЯ В ПРОСТРАНСТВЕ ЗНАНИЙ

Н.Г.Загоруйко, М.Б.Бушуев

Введение

Одной из основных задач блока гомеостата экспертных систем [1] является автоматическая структуризация базы знаний. Будем рассматривать базу знаний как систему, состоящую из таких элементов, как продукции (правила вывода вида "если А, то В") и высказывания, констатирующие факты или их совокупности ("А и В"). Условно назовем единичным знанием элемент базы знаний предметной области.

В большинстве экспертных систем база знаний представляет собой просто неупорядоченный набор продукций, такой способ хранения знаний часто называют "школьной доской" ("school board"). Он может себя оправдывать только в экспертных системах с базой знаний небольшого размера. В больших системах знаний, насчитывающих десятки тысяч продукций, остро встает проблема структуризации знаний. При хранении информации в неструктурированном виде с ростом объема базы знаний значительно усложняются все операции манипулирования с системой продукции. Особенно это касается операции поиска по образцу ("matching "), которая является основой дедуктивной системы логического вывода. В большинстве экспертных систем количество продукций обычно не превышает нескольких сотен, а в наиболее развитых - нескольких тысяч. Для того чтобы резко увеличить объемы базы знаний, необходимо изменить способ хранения знаний, структурировать ее.

В первом приближении базу знаний системы можно структурировать выделением групп знаний, в некотором смысле "похожих" друг на друга. После этого процедура рассуждения, логического вывода в систему может выглядеть следующим образом. Поставленная задача разбивается на подзадачи так, что каждой выделенной подзадаче соответствует группа (таксон) знаний системы, а затем уже в нем ищется решение этой подзадачи.

Для применения на практике такого подхода необходимо определить меру "похожести", "близости" знаний друг от друга.

I. Расстояние между формулами классической логики

Наиболее часто эксперты формулируют свои суждения в виде высказываний двух типов:

I) констатации отдельных фактов (простые высказывания) или их совокупности: A = "зимой холодно", $B \wedge C$ "(сейчас лето) и (температура +25°C)";

2) продукции, т.е. высказывания вида "если A, то B". При этом условий может быть несколько, а следствие одно: "если (почва сухая) и (сейчас лето), то (необходим полив)".

Один из способов записи таких высказываний - это их представление в виде формул исчисления высказываний с использованием логических связок $\wedge$ - конъюнкции, $\to$ - импликации и $\neg$ - отрицания: I) $A, B \wedge C$; 2) $D \wedge C \to E$, где заглавными латинскими буквами обозначены простые высказывания.

ОПРЕДЕЛЕНИЕ. Назовем носителем M формулы исчисления высказываний множество простых высказываний или их отрицаний, используемых при ее построении: $M(B \wedge C) = \{B, C\}$; $M(D \wedge C \to E) = \{D, C, E\}$.

Пусть база знаний системы состоит из K формул исчисления высказываний не обязательно вида I или 2: $\Sigma = \{A_1, A_2, \ldots, A_k\}$.

ОПРЕДЕЛЕНИЕ. Назовем носителем системы знаний $M(\Sigma)$ - объединение носителей формул, входящих в Σ:

$$M(\Sigma) = M(A_1, A_2, \ldots, A_k) = \bigcup_{i=1}^{k} M(A_i).$$

Мощность этого множества обозначим $m = card(M(\Sigma))$. Система знаний Σ фиксирует некоторый гипотетический мир $W(M(\Sigma))$, объекты которого описываются с помощью всех возможных конъюнкций простых высказываний из носителя $M(\Sigma)$ и их отрицаний. Всего таких объектов в $W(M(\Sigma))$, т.е. его мощность $card(W(M(\Sigma))) = 2^m$.

Из логики известно [2], что любая формула исчисления высказываний может быть представлена в виде совершенной дизъюнктивной нормальной формы (СДНФ), т.е. дизъюнкции конъюнктов K_j, где каждый конъюнкт K_j описывает объект из мира $W(M(\Sigma))$.

Будем представлять формулу исчисления высказываний Φ вектором φ длины 2^m, состоящим из нулей и единиц, следующим образом:

$\varphi_j = 1$, если K_j входит в СДНФ формулы Φ;

$\varphi_j = 0$, если K_j не входит в СДНФ формулы Φ.

Таким образом, вектор φ представляет характеристику объектов мира $W(M(\Sigma))$, на которых формула Φ истинна ($\varphi_j = 1$) или ложна ($\varphi_j = 0$).

ОПРЕДЕЛЕНИЕ. Вектор φ называется представлением Φ в мире $W(M(\Sigma))$.

Естественно было бы измерять различие формул Φ и Ψ количеством объектов, на которых эти формулы принимают разные значения. Для этой цели можно взять нормированное хэммингово расстояние между векторами φ и ϕ, представляющими формулы Φ и Ψ в мире $W(M(\Sigma))$.

ОПРЕДЕЛЕНИЕ. Расстоянием между формулами Φ и Ψ в мире $W(M(\Sigma))$ назовем

$$h(\Phi,\Psi) = \frac{\mathrm{card}\{\, i \in \{1, \ldots, 2^m\} \mid \varphi_i \cdot \phi_i = 0 \text{ и } \varphi_i + \phi_i = 1 \}}{2^m}.$$

Рассмотрим свойства меры h.

I. Для любых формул исчисления высказываний Φ и Ψ выполнено:

$h(\Phi,\Phi) = 0$,

$h(\Phi,\neg\Phi) = 1$,

$h(\Phi,\Psi) = 1 - h(\Phi,\neg\Psi) = h(\neg\Phi,\neg\Psi)$

2. Имеет место симметричность $h(\Phi,\Psi) = h(\Psi,\Phi)$.

3. Свойство неравенства треугольника:

$h(\Phi,\Psi) \le h(\Phi,\Upsilon) + h(\Upsilon,\Psi)$ для всех Υ,

$h(\Phi,\neg\Phi) = h(\Phi,\Upsilon) + h(\Upsilon,\neg\Phi)$ для всех Υ.

4. $h(\Phi,\Psi) = h(\Phi\wedge\Psi,\ \Phi\vee\Psi)$.

Доказательство свойств I–3 очевидно и следует из свойств хэммингова расстояния.

Докажем свойство 4. Вектор k, представляющий конъюнкцию двух формул, равен $k_i = \min(\varphi_i, \phi_i)$ для $i = 1, 2, \ldots, 2^m$.

Вектор d, представляющий дизъюнкцию двух формул, равен $d_i = \max(\varphi_i, \phi_i)$ для $i = 1, 2, \ldots, 2^m$.

Таким образом, если в формулах Φ и Ψ на i-м месте стояли разные значения, то $k_i = 0$, а $d_i = 1$, т.е. на i-м объекте формулы $\Phi\wedge\Psi$ и $\Phi\vee\Psi$ тоже принимают различные значения. Этим доказывается свойство 4.

Введенное нами расстояние рассматривается во всем мире. Очень часто требуется подсчитать расстояние между формулами, носители которых составляют небольшое подмножество мира $W(M(\Sigma))$. Естественно встает вопрос: нельзя ли упростить вычисление расстояния h, т.е. считать его не среди всех объектов $W(M(\Sigma))$, а среди некоторого подмножества $W(M(\Sigma))$? Ответ дает

ТЕОРЕМА I (о сужении). Расстояние между формулами Φ и Ψ во всем мире $W(M(\Sigma))$ равно расстоянию между ними в сужении этого мира: $h_{W(M(\Sigma))}(\Phi,\Psi) = h_{W(M(\Phi)\cup M(\Psi))}(\Phi,\Psi)$.

ДОКАЗАТЕЛЬСТВО. Для доказательства теоремы нам нужна

ЛЕММА (о расширении представления). Пусть φ - представление Φ в мире $W(M(\Phi))$, а φ' - представление Φ в мире $W(M(\Phi)\cup C)$, если C не принадлежит $M(\Phi)$, то $\varphi' = (\varphi,\varphi)$.

ДОКАЗАТЕЛЬСТВО леммы. При добавлении к $M(\Phi)$ еще одного простого высказывания C, принимающего два значения, количество объектов в мире $W(M(\Phi)\cup C)$ увеличится в два раза, так как к каждой конъюнкции, описывающей объект, добавится C или $\neg C$. Зафиксируем некоторый порядок в множестве $M(\Phi)$ и будем кодировать объекты мира $W(M(\Phi))$ вектором O длины $card(M(\Phi))$:

$O_i = 1$, если на объекте выполнено высказывание под номером i из $M(\Phi)$:

$O_i = 0$, если выполнено отрицание высказывания.

Упорядочим лексикографически векторы O^k, описывающие объекты мира $W(M(\Phi))$. Мир $W(M(\Phi))$ описывается набором векторов, представляющие собой записи двоичных чисел от $2^{card(M(\Phi))-1}-1$ до 0, упорядоченных по убыванию. Мир $W(M(\Phi)\cup C)$ будет описываться следующим образом: слева к векторам, описывающим объекты $W(M(\Phi))$, приписывается 1, обозначающая объекты, на которых высказывание C истинно, снизу к этим векторам приписываются векторы мира $W(M(\Phi))$ с нулем впереди, соответствующие объектам мира $W(M(\Phi)\cup C)$, на которых C ложно. Таким образом, мир $W(M(\Phi)\cup C)$ описывается двоичными числами от $2^{card(M(\Phi))}-1$ до 0, упорядоченными по убыванию.

При таком способе расширения мира к объектам, на которых были истинны Φ и C, добавятся объекты, на которых Φ истинна и C ложно, т.е. представлением Φ в мире $W(M(\Phi)\cup C)$ будет вектор $\varphi' = (\varphi,\varphi)$. Лемма доказана.

Пусть φ', ψ' – представления Φ, Ψ в мире $W(M(\Phi) \cup M(\Psi))$ длиной 2 в степени m_1, а размерность мира $W(M(\Sigma))$ равна m. Применяя $(m-m_1)$ раз лемму о расширении, получаем векторы φ, ψ, представляющие Φ, Ψ в мире $W(M(\Sigma))$: $\varphi = (\varphi', \dots, \varphi')$, $\psi = (\psi', \dots, \psi')$, длина каждого 2^m.

Пусть расстояние между формулами в мире $h_{W(M(\Phi) \cup M(\Psi))} = h_1$, т.е. количество объектов, на которых формулы принимают различные значения, $L = h_1 2^{m_1}$. Тогда расстояние в мире $W(M(\Sigma))$

$$h_{W(M(\Sigma))}(\Phi,\Psi) = \frac{L\, 2^{(m-m_1)}}{2^m} = h_1.$$

Теорема доказана.

Для подсчета расстояния между формулами, пересечение носителей которых пусто, используется

ТЕОРЕМА 2. Если пересечение носителей формул Φ и Ψ пусто, то $h(\Phi,\Psi) = t_\Phi + t_\Psi - 2t_\Phi \cdot t_\Psi$, где $t_\Phi(t_\Psi)$ – доля объектов, на которых формула $\Phi(\Psi)$ истинна в мире $W(M(\Phi))[W(M(\Psi))]$.

ДОКАЗАТЕЛЬСТВО. По теореме о сужении можем сосчитать расстояние между Φ и Ψ в мире $W(M(\Phi) \cup M(\Psi))$. Пусть φ – представление Φ в этом мире, длина φ равна 2^m; φ' – представление Φ в мире $W(M(\Phi))$, $L_1 = \mathrm{card}(M(\Phi))$, ψ' – представление Ψ в мире $W(M(\Psi))$, $L_2 = \mathrm{card}(M(\Psi))$. По лемме о расширении представлением Φ в мире $W(M(\Phi) \cup M(\Psi))$ будет вектор $\varphi = (\varphi', \dots, \varphi')$, где φ' повторено 2^{L_2} раз. Применим описание объектов мира из леммы о расширении, при этом объекты мира $W(M(\Phi) \cup M(\Psi))$ будут располагаться следующим образом:

$$\begin{bmatrix} 1 \dots 1\; 1\; A \\ 1 \dots 1\; 0\; A \\ \dots\dots \\ 0 \dots 0\; 0\; A \end{bmatrix}, \tag{1}$$

где матрица A размерности 2^{L_1} на L_1 описывает мир $W(M(\Phi))$:

$$A = \begin{bmatrix} 1 & 1 & \ldots & 1 & 1 \\ 1 & 1 & \ldots & 1 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \ldots & 0 & 1 \\ 0 & 0 & \ldots & 0 & 0 \end{bmatrix},$$

а $\mathbb{1}$, $\mathbb{0}$ – столбцы единиц и нулей длины 2^{L_1}. Их количество L_2. В связи с тем, что пересечение носителей $M(\Phi)$ и $M(\Psi)$ пусто, вектор ψ, представляющий Ψ в мире $W(M(\Phi) \cup M(\Psi))$, зависит только от выделенных значений $\mathbb{1}, \mathbb{0}$ в левой части матрицы (I) и состоит из 2 в степени L_2 единичных ($\mathbb{1}$) или нулевых ($\mathbb{0}$) векторов длины 2 в степени L_1, т.е. если $\psi' = (10\ldots01)$, то $\psi = (\mathbb{1}, \mathbb{0}, \ldots, \mathbb{0}, \mathbb{1})$.

Найдем теперь расстояние h между $\varphi = (\varphi', \varphi', \ldots, \varphi', \varphi')$ и $\psi = (\mathbb{1}, \mathbb{0}, \ldots, \mathbb{0}, \mathbb{1})$.

Количество объектов, на которых Φ и Ψ принимают разные значения, равно количеству нулей в векторе φ', умноженному на количество $\mathbb{1}$ в ψ, плюс количество единиц в векторе φ', умноженному на количество $\mathbb{0}$ в ψ: $L = T_\Phi \cdot F_\Psi + T_\Psi \cdot F_\Phi$, где T, F означают количество объектов, на которых формулы истинны (T) или ложны (F) в своем мире:

$$h(\Phi,\Psi) = \frac{L}{2^{L_1+L_2}} = \frac{T_\Phi}{2^{L_1}} \cdot \frac{F_\Psi}{2^{L_2}} + \frac{T_\Psi}{2^{L_2}} \cdot \frac{F_\Phi}{2^{L_1}} = t_\Phi \cdot f_\Psi + t_\Psi \cdot f_\Phi =$$

$$= t_\Phi \cdot (1 - t_\Psi) + t_\Psi \cdot (1 - t_\Phi) = t_\Phi + t_\Psi - 2 t_\Phi \cdot t_\Psi,$$

где доли t_Φ, t_Ψ могут принимать значения из интервала $[0, I]$. Теорема 2 доказана.

СЛЕДСТВИЕ. Пусть C – простое высказывание, Φ – формула исчисления высказываний. Если C не принадлежит $M(\Phi)$, то $h(\Phi, C) = \frac{1}{2}$.

ДОКАЗАТЕЛЬСТВО. Доля объектов, на которых истинно C в мире $W(M(C))$, равна I/2. По теореме 2, расстояние между Φ и C при этом равно I/2.

Еще две возможности аналитического вычисления расстояния для наиболее часто встречающихся в экспертных системах типов высказываний I и 2 дает

ТЕОРЕМА 3.

а) Пусть Φ,Ψ - высказывания типа 1, т.е. конъюнкции простых высказываний, тогда

$$h(\Phi\,\Psi) = \frac{1}{2^{l_1}} + \frac{1}{2^{l_2}} - \frac{\rho}{2^{n-1}}, \tag{2}$$

где $n = card(M(\Phi) \cup M(\Psi))$, $l_1 = card(M(\Phi))$, $l_2 = card(M(\Psi))$, $\rho = 1$, если в $M(\Phi)$ содержатся высказывания, отрицания которых входят в $M(\Psi)$; $\rho = 0$, если в $M(\Phi)$ не содержатся высказывания, отрицания которых входят в $M(\Psi)$.

б) Пусть Φ,Ψ - высказывания второго типа, тогда расстояние между ними вычисляется по (2).

ДОКАЗАТЕЛЬСТВО. По теореме о сужении, будем рассматривать расстояние между формулами Φ и Ψ в мире $W(M(\Phi) \cup M(\Psi))$. Пусть объекты этого мира упорядочены так же, как и в лемме о расширении, а φ, ϕ - представление формул Φ, Ψ в мире $W(M(\Phi) \cup M(\Psi))$.

Докажем "а". Известно [2], что конъюнкция истинна, когда истинны все составляющие ее простые высказывания. Это выполняется только на одном объекте в мире, составленном из этих простых высказываний.

I) Пусть в $M(\Phi)$ не содержатся высказывания, отрицания которых входят в $M(\Psi)$, т.е. формулы вида $\Phi = A \wedge B \wedge C$; $\Psi = B \wedge C \wedge D$. В мире $W(M(\Phi) \cup M(\Psi))$ формула Φ истинна на $1\cdot 2^{(n-l_2)}$ объектах, а Ψ - на $1\cdot 2^{(n-l_1)}$ объектах. При этом только на одном объекте Φ,Ψ истинны одновременно (на объекте $A \wedge B \wedge C \wedge D$). Таким образом, количество объектов, на которых формулы принимают разные значения,

$$L = 2^{(n-l_2)} + 2^{(n-l_1)} - 2,$$

а расстояние

$$h = \frac{L}{2^n} = \frac{1}{2^{l_1}} + \frac{1}{2^{l_2}} - \frac{1}{2^{n-1}}.$$

2) Пусть в $M(\Phi)$ содержатся высказывания, отрицания которых входят в $M(\Psi)$, т.е. формулы вида $\Phi = A \wedge B \wedge C$; $\Psi = B \wedge \neg C \wedge D$. Тогда формулы Φ, Ψ ни на одном объекте одновременно истинны быть не могут и расстояние

$$h(\Phi,\Psi) = \frac{2^{(n-l_2)} + 2^{(n-l_1)}}{2^n} = \frac{1}{2^{l_1}} + \frac{1}{2^{l_2}}.$$

Первая часть теоремы ("а") доказана. Доказательство "б" аналогично.

Импликация ложна, когда посылки истинны, а следствие ложно [2], т.е. импликация ложна только на одном объекте в мире из простых высказываний, используемых в этой импликации. Таким образом, в мире $W(M(\Phi) \cup M(\Psi))$ импликация Φ ложна на $2^{(n-l_2)}$ объектах, а Ψ – на $2^{(n-l_1)}$ объектах. В первом случае Φ, Ψ одновременно ложны только на одном объекте, если следствия импликаций Φ, Ψ совпадают. Во втором случае таких объектов нет. Следовательно, как и для первой части теоремы 3, получаем формулу (2). Теорема доказана.

Таким образом, расстояние, введенное нами для формул классической логики, вполне соответствует интуитивным представлениям эксперта и позволяет решить поставленную во введении задачу группировки высказываний в базе знаний. При этом аналитические формулы, доказанные в теоремах 2 и 3, позволяют упростить вычисление расстояний для основных типов высказываний в экспертной системе, что особенно важно при больших размерностях системы.

3. Расстояние между распределениями

Выше в анализе участвовали простые высказывания. В реальных базах знаний элементы высказывания могут иметь более сложный вид. Например, одна продукция может выглядеть так: "Если влаги в почве более 50 мм, а азота от 100 до 150 кг/га, то...", а другая так: "Если влаги в почве от 40 до 100 мм, а азота от 120 до 200 кг/га, то...". Возникает проблема оценки меры "близости" или меры "расстояния" между такими продукциями.

Если два эксперта одной и той же характеристике x_j приписывают одинаковый диапазон значений, то естественно считать, что расстояние между мнениями этих экспертов равно нулю. Если же один из них считает, что $x_j = x_{j,max}$, а второй, что $x_j = x_{j,min}$, то расстояние между их мнениями максимально и можно считать его равным I. Эксперты могут указывать диапазоны значений x_j и вообще любые распределения допустимых значений в пределах от $x_{j,min}$ до $x_{j,max}$. Следовательно, задача оценки расстояния между мнениями экспертов сводится к задаче поиска расстояния между двумя распределениями.

Пусть характеристика x имеет конечное число (n) градаций. Если первый эксперт указывает в качестве допустимых градаций x_1, $x_2, \ldots, x_\alpha, \ldots, x_A$ с вероятностями $P_\alpha \left(\sum_{\alpha=1}^{A} P_\alpha = 1 \right)$, а второй эксперт допускает градации $x_1, x_2, \ldots, x_\beta, \ldots, x_B$ с вероятностями P_β, то расстояние Z между этими распределениями можно определить так:

$$Z = \frac{\sum_{\alpha=1}^{A} \sum_{\beta=1}^{B} |x_\alpha - x_\beta| \cdot (P_\alpha + P_\beta)}{\sum_{\alpha=1}^{A} \sum_{\beta=1}^{B} (P_\alpha + P_\beta)}$$

Дополнительно следует выделить случаи, когда распределения хотя бы частично перекрываются. Если перекрытия нет, то будем считать, что доля неперекрывающихся распределений $W = 1$, а если гистограммы распределения полностью совпадают, то $W = 0$.

Далее будем считать, что при одинаковом расстоянии между математическими ожиданиями двух распределений "похожесть" этих распределений тем меньше, чем меньше неопределенность высказываний экспертов. Это можно учесть с помощью величины

$$H = \frac{H_{max} - H_{12}}{H_{max}},$$

где $H_{max} = \ln n$, $H_{12} = \frac{H_1 + H_2}{2}$. Здесь

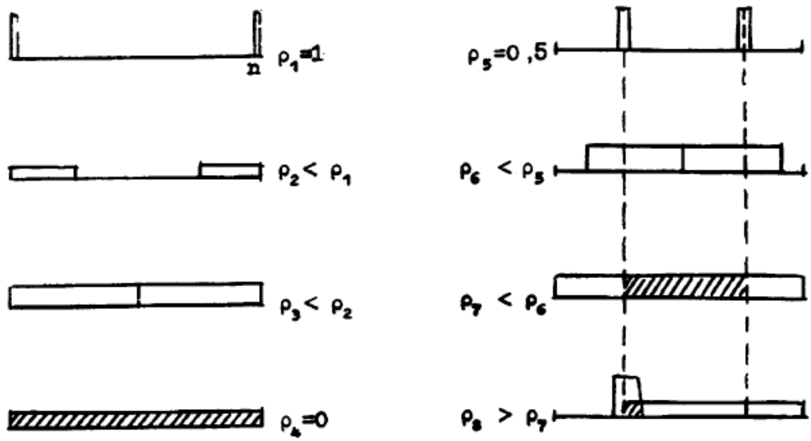
$$H_1 = \sum_{\alpha=1}^{A} P_\alpha \cdot \ln P_\alpha, \quad H_2 = \sum_{\beta=1}^{B} P_\beta \cdot \ln P_\beta .$$

С учетом сказанного общая мера расстояния между двумя распределениями ρ определяется так:

$$\rho = Z \cdot W \cdot H, \qquad (3)$$

где $0 \le \rho \le 1$.

На рисунке приведены примеры распределений и расстояний между ними.



Аналогичные меры можно сконструировать и для порядковых, и для номинальных шкал.

Данные в шкале порядка могут быть представлены такими высказываниями: "влажность высокая", "влажность не выше средней".

Набор возможных значений качественной переменной часто фиксирован в виде списка типа: "очень мало", "мало", "средне", "много", "очень много". Список значений может быть и более кратким (например, "мало", "средне", "много"), и более детальным ("очень-очень мало", "очень мало", "ниже среднего", "средне" и т.д.). Если различаемых упорядоченных состояний n, то все они могут быть занумерованы числами от I до n, после чего расстояние между высказываниями в шкале порядка можно определять по той же формуле (3), что и для сильной шкалы.

Рассмотрим теперь шкалу наименований. Пусть количество имен различных объектов, которые будут встречаться в высказываниях экспертов, зафиксировано и равно n (например, в высказываниях о сортах семян будут фигурировать n сортов). Если первый эксперт указывает A имен x_α, с их вероятностями P_α, а второй – B имен x_β со своими вероятностями P_β, то расстояние ρ между такими распределениями снова можно считать по той же формуле (3) с той лишь поправкой, что расстояние $|x_\alpha - x_\beta|$ равно 0, если имя $x_\alpha = x_\beta$, и равно 1, если $x_\alpha \neq x_\beta$.

Так, например, если в наборе из 8 сортов один указывает с равной вероятностью два сорта – первый и пятый, а второй – с равной вероятностью четыре сорта - второй, третий, пятый и восьмой, то расстояния между высказываниями этих экспертов будет равно $\frac{21}{64}$, так как

$$Z = \frac{7}{8}, \quad W = \frac{3}{4}, \quad H_1 = 1, \; H_2 = 2, \; H_{12} = 1.5,$$

$$H_{max} = 3, \quad H = \frac{1}{2} \quad \text{и} \quad \rho = \frac{7}{8} \cdot \frac{3}{4} \cdot \frac{1}{2} = \frac{21}{64}.$$

Если в импликации, сформулированной одним экспертом, содержится характеристика x_j, а другой эксперт об этой характеристике ничего не говорит, то часто можно интерпретировать этот факт, как то, что второй эксперт не отдает предпочтения одних значений x_j перед другими: он либо не знает влияния x_j на целевую функцию x_0, либо считает все значения x_j равноценными. На этом основании можно считать, что второй эксперт задает (по умолчанию) равномерное распределение в диапазоне от $x_{j,min}$ до $x_{j,max}$. При таком предположении появляется возможность считать расстояние между двумя импликациями и в том случае, когда использованные двумя экспертами характеристики не совпадают частично или даже полностью.

Если в двух сравниваемых импликациях содержатся высказывания об m разных характеристиках x_j, $j=1,2,\ldots m$, то после вычисления расстояний ρ_j для каждой из этих характеристик общее расстояние R между двумя импликациями можно определить так:

$$R = \frac{\sum_{j=1}^{m} \gamma_j \cdot \rho_j}{\sum_{j=1}^{m} \gamma_j},$$

где γ_j - весовой коэффициент, характеризующий относительную "информативность", "важность" характеристики x_j. В простейшем случае можно предположить, что все $\gamma_j = 1$.

ПРИМЕР. Два эксперта высказываются о влажности (L), которая измеряется порядковой шкалой с четырьмя значениями (нулевая, низкая, средняя, высокая), о количестве азотных удобрений — диапазон изменений которого от 0 до 320 кг/га, а величина градации 40кг/га, о типе почвы (G), число которых 4, и видах культур-предшественников (B), число которых 8.

На вопрос, каковы условия для получения хорошего урожая пшеницы, эксперты дали такие ответы:

1: (L = средняя) $\wedge$ (N = 160–200) $\wedge$ (G = чернозем),

2: (L = средняя $\vee$ высокая) $\wedge$ (N = 200–280) $\wedge$ (B = бобовые $\vee$ травяные).

Различие мнений о влажности равно $\rho_1 = \frac{1}{12}\cdot\frac{1}{2}\cdot\frac{3}{4} = 0{,}03$, об азоте $\rho_2 = \frac{3}{16}\cdot 1\cdot\frac{3}{4} = \frac{9}{64} = 0{,}14$, о типе почв $\rho_3 = \frac{3}{4}\cdot\frac{3}{4}\cdot\frac{1}{2} = 0{,}28$, о виде предшественника $\rho_4 = \frac{3}{4}\cdot\frac{1}{2}\cdot\frac{3}{4} = 0{,}28$. Общее расстояние при $\gamma_j = 1$

$$R = \frac{0{,}03 + 0{,}14 + 0{,}28 + 0{,}28}{4} = 0{,}18 .$$

Авторы выражают благодарность Л.Я.Савельеву за очень полезные дискуссии о мерах близости между распределениями.

Литература

1. ЗАГОРУЙКО Н.Г. Экспертные системы и распознавание образов. –Настоящий сборник, с. 3–10.

2. ЧЕРЧ А. Введение в математическую логику. Т.1. – М.: ИЛ, 1960.

Поступила в ред.-изд.отд.
21 июля 1986 года