ISSN 1691-5402
ISBN 978-9984-44-028-6
Environment. Technology. Resources
Proceedings of the 7th International Scientific and Practical Conference. Volume I1

ОСОБЕННОСТИ ПРЕДСТАВЛЕНИЯ ЗНАНИЙ В ТЕОРИИ ГРУБЫХ МНОЖЕСТВ
KNOWLEDGE REPRESENTING FEATURES IN ROUGH SETS THEORY

Олег Ужга-Ребров
Rēzeknes Augstskola, Atbrīvošanas aleja, 90, Rēzekne, LV 4600, Latvija
tālr. +(373)29472205, e-pasts: rebrovs@tvnet.lv

***Abstract**. Knowledge discovering and representing in information systems has evolved into an important area of research because of theoretical challenges and practical representing unknown knowledge in data. Rough set theory is a mathematical formalism for representing uncertainty that can be considered a special extension of the set theory. In this paper is analyzed knowledge representing problem and attribute signification's problem, based on rough sets approach.*

***Keywords**: information systems, knowledge representation, rough sets, significance of attributes.*

Введение

В настоящее время является общепризнанным, что все глобальные задачи обработки неопределенной информации могут быть разделены на три больших класса:
- обращение с неполными знаниями;
- комбинирование противоречивых частей информации;
- использование грубых (необработанных) массивов данных.

Для решения задач первого класса предложено большое число действенных методов. К их числу относятся различные экстраполяционные и интерполяционные методы, экспертное оценивание, интервальное оценивание, и другие. Наиболее известными подходами к решению задач второго класса являются теория Демпстера-Шейфера [2], различные её модификации, включая теорию Дезера-Смарандаке [3], и правила перераспределения конфликтов Дезера-Смарандаке. Теория грубых множеств предусмотрена для решения задач третьего класса. Автором теории является польский учёный З. Павлак. В систематическом виде основы теории изложены в [1]. Методы, предложенные в рамках этой теории, позволяют обрабатывать большие массивы неупорядоченных данных и на основе результатов такой обработки получать новые знания.

В настоящей статье даются краткие концептуальные основы теории грубых множеств, рассматриваются проблемы представления знаний в информационных системах на основе грубых множеств. Более подробно рассматривается задача оценивания значимости отдельных атрибутов в информационной системе и предлагается модификация существующего подхода к решению этой задачи.

1. Концептуальные основы теории грубых множеств

База знаний определяется как K = (U, R), где U – универсум элементов, R – отношение эквивалентности, на основе которого могут быть выделены классы эквивалентности (категории) элементов U, которые обозначаются через IND(R). В каждую категорию включаются элементы, которые обладают одинаковыми значениями классификационных признаков (атрибутов). Внутри каждой категории элементы считаются неразличимыми.

Пусть элементы универсума классифицированы по категориям на основе отношения эквивалентности R. Если мы возьмем целевое множество элементов

X $\in$ U, то относительно классификации IND(R) могут быть рассмотрены следующие ситуации:
1. Множество X является объединением некоторых категорий из IND(R). В этом случае множество X называется *R-точным*.
2. Множество X не может быть выражено как объединение некоторых категорий IND(R). В этом случае множество X называется *R-неточным* или *R-грубым*.
R-нижней аппроксимацией грубого множества X называется подмножество таких его элементов, которые могут быть классифицированы как принадлежащие X на основе классификации IND(R):

$$\underline{R}X = \cup\{Y \in IND(R) : Y \subseteq X\}. \tag{1}$$

R-верхней аппроксимацией грубого множества X называется подмножество таких его элементов, которые возможно принадлежат X:

$$\bar{R}X = \cup\{Y \in IND(R) : Y \cap X \neq \varnothing\}. \tag{2}$$

R-нижнюю аппроксимацию целевого множества X называют *R-положительной областью X*:

$$POS_R(X) = \underline{R}X. \tag{3}$$

Отрицательной областью X называется подмножество элементов универсума, которые с определённостью не принадлежат X:

$$NEG_R(X) = U - \bar{R}X. \tag{4}$$

Граничной областью X называется подмножество его элементов, которые принадлежат R-верхней аппроксимации, но не принадлежат R-нижней аппроксимации:

$$BN_R(X) = \bar{R}X - \underline{R}X. \tag{5}$$

Рассмотрим простой пример, иллюстрирующий выше введенные понятия.
Пример 1. Имеется база знаний K = (U, R), U = {x_1, x_2, … , x_{10}) – универсум элементов, R – отношение эквивалентности, на основе которого выделены следующие классы эквивалентности (категории) на U:

$$U / IND(R) = \{\{x_1, x_2\}, \{x_3, x_7, x_{10}\}, \{x_4\}, \{x_5\}, \{x_6\}, \{x_8\}, \{x_9\}\}.$$

Заданы целевые множества $X_1 = \{x_1, x_2, x_4, x_5\}, X_2 = \{x_1, x_2, x_3, x_4\}$. Необходимо определить аппроксимации, отрицательные и граничные области для этих множеств.
Решение.

$$\underline{R}X_1 = \{x_1, x_2, x_4, x_5\}, \bar{R}X_1 = \varnothing, NEG_R(X_1) = \{x_3, x_6, x_7, x_8, x_9, x_{10}\}, BN_R(X_1) = \varnothing.$$

$$\underline{R}X_2 = \{x_1, x_2, x_4\}, \bar{R}X_2 = \{x_1, x_2, x_3, x_4, x_7, x_{10}\}, NEG_R(X_2) = \{x_5, x_6, x_8, x_9\}, BN_R(X_2) = \{x_3, x_7, x_{10}\}.$$

Философия грубых множеств такова, что выделение релевантных категорий элементов на универсуме, характеризация целевых множеств и операции над этими множествами производятся только и только на основе существующих знаний. Эти знания включают в себя значения атрибутов и отношение (отношения) эквивалентности, на основе которого (которых) выделяются категории элементов. В практических приложениях грубых множеств часто возникает вопрос: нужны ли все существующие знания, чтобы решить поставленную задачу, или для этого достаточна только часть этих знаний?
Под *сокращением* знаний будем понимать процесс выделения той существенной части существующих знаний, которая достаточна для определения всех релевантных концепций. Результат такого процесса также будем называть *сокращением* знаний.
Введём следующие понятия и определения. Пусть **R** – множество отношений эквивалентности на заданном универсуме U и R $\in$ **R**. Отношение R является *несущественным* в **R**, если IND(**R**) = IND(**R** – {R}), в противном случае отношение R является *необходимым* в **R**. Множество отношений **R** называется *независимым*, если

каждое отношение R ∈ **R** является необходимым в **R**, в противном случае множество отношений **R** называется *зависимым*.

Пусть **Q** и **P** – множество отношений эквивалентности на U. **Q** ⊆ **P** называется *сокращением* **P**, если **Q** независимо и IND(**Q**) = IND(**P**).

Ядро знаний (*ядро* множества отношений эквивалентности) определяется как

$$CORE(\mathbf{P}) = \cap RED(\mathbf{P}), \tag{6}$$

где RED(**P**) – семейство всех сокращений **P**.

Расширим данные понятия и определения на два произвольных множества отношений эквивалентности на U.

Назовём $POS_{\mathbf{P}}(\mathbf{Q})$ **P**-*положительной областью* **Q**. Содержательно это понятие интерпретируется как множество категорий отношения **Q**, которые могут быть корректно определены на основе множества отношений эквивалентности **P**. Отношение R ∈ **P** является **Q**-несущественным в **P**, если

$$POS_{IND(\mathbf{P})}(IND(\mathbf{Q})) = POS_{IND(\mathbf{P}-\{R\})}IND(\mathbf{Q}), \tag{7}$$

в противном случае отношение R является **Q**-*необходимым* в **P**.

Если каждое отношение R ∈ **P** **Q**-необходимо, то **P** является **Q**-*независимым относительно* **Q**.

Множество отношений **S** ∈ **P** называется **Q**-*сокращением* **P**, если и только если **S** является **Q**-независимым подмножеством **P** и $POS_{\mathbf{S}}(\mathbf{Q}) = POS_{\mathbf{P}}(\mathbf{Q})$.

Множество всех **Q**-необходимых элементарных отношений в **P** называется **Q**-*ядром* **P**:

$$CORE_{\mathbf{Q}}(\mathbf{P}) = \cap RED_{\mathbf{Q}}(\mathbf{P}), \tag{8}$$

где $RED_{\mathbf{Q}}(\mathbf{P})$ множество всех **Q**-сокращений **P**.

Все концептуальные основы грубых множеств, рассмотренные в настоящем разделе, основаны на понятиях отношений эквивалентности (или множествах таких отношений). Если мы имеем массив исходных данных, гораздо более удобным представляется отказаться от отношений эквивалентности и непосредственно оперировать множествами значений атрибутов. В последующем разделе рассматриваются теоретические основы такого подхода и анализируются проблемы классификации элементов.

2. Проблемы представления знаний на основе грубых множеств

Корректное представление существующих знаний играет очень важную роль в различных областях науки, особенно в многочисленных направлениях искусственного интеллекта. Существует много методов представления знаний. Каждый метод служит для отражения существенных аспектов релевантных знаний. В настоящей работе рассматривается подход к представлению знаний на основе теории грубых множеств. Этот подход основан на главной концепции теории, которая заключается в том, что знания отражаются в разделении (классификации) элементов универсума. Такое разделение может пониматься, как семантика представления знаний. Однако, для того, чтобы успешно выявить семантику знаний, эти знания должны быть представлены в подходящей синтаксической форме. Такой формой является таблица данных, строки которой соответствуют элементам (объектам), а столбцы – признакам (атрибутам) этих элементов. В ячейке на пересечении i-й строки и j-го столбца отображается значение j-го атрибута для i-го элемента. В практических приложениях теории грубых множеств такие таблицы принято называть *информационными системами (ИС)*. Релевантные исходные данные в ИС могут быть получены различными путями. Их источниками могут быть измерения, наблюдения, субъективные оценки экспертов. В настоящей работе источники исходных данных остаются за рамками рассмотрения.

Введём формальные определения. *Информационная система* – это пара S = (U, A), где U – конечное множество (универсум) элементов, A – множество атрибутов. Каждый атрибут $a \in A$ может быть отображён функцией $a : U \rightarrow V_a$, где V_a – множество значений a, называемое *областью a*.
С каждым подмножеством атрибутов $B \subseteq A$ может быть связано бинарное отношение IND(B), называемое *отношением неразличимости*:

$$IND(B) = \{(x, y) \in U^2 : \forall a \in B, a(x) = a(y)\}. \tag{9}$$

Очевидно, что отношение IND(B) является отношением эквивалентности и

$$IND(B) = \cap_{a \in B} IND(a). \tag{10}$$

Отсюда следует, что между понятием базы знаний и соответствующей информационной системы существует соответствие один-к-одному. Все концептуальные основы теории грубых множеств, сформулированные в предыдущем разделе на основе отношений (множеств отношений) эквивалентности, могут быть сформулированы на основе атрибутов (множеств атрибутов).
Рассмотрим простой пример, иллюстрирующий понятия информационной системы и представления знаний.
Пример 2. В таблице 1 представлена условная информационная система.

Таблица 1.

Представление условной информационной системы

	Значения атрибутов				
Элементы	a_1	a_2	a_3	a_4	a_5
x_1	1	2	0	1	1
x_2	1	2	0	1	1
x_3	2	0	0	1	0
x_4	0	0	1	2	1
x_5	2	1	0	2	1
x_6	0	0	1	2	2
x_7	2	0	0	1	0
x_8	0	1	2	2	1
x_9	2	1	0	2	2
x_{10}	2	0	0	1	0

Необходимо произвести разделение элементов универсума по подмножествам атрибутов $\{a_1\}$, $\{a_1, a_2\}$, $\{a_1, a_2, a_3\}$, $\{a_1, a_2, a_3, a_4\}$, $\{a_1, a_2, a_3, a_4, a_5\}$.

Решение.

$$U / IND\{a_1\} = \{\{x_4, x_6, x_8\}, \{x_1, x_2\}, \{x_3, x_5, x_7, x_9, x_{10}\}\};$$

$$U / IND\{a_1, a_2\} = \{\{x_4, x_6\}, \{x_8\}, \{x_1, x_2\}, \{x_3, x_7, x_{10}\}, \{x_5, x_9\}\};$$

$$U / IND\{a_1, a_2, a_3\} = \{\{x_4, x_6\}, \{x_8\}, \{x_1, x_2\}, \{x_3, x_7, x_{10}\}, \{x_5, x_9\}\};$$

$$U / IND\{a_1, a_2, a_3, a_4\} = \{\{x_4, x_6\}, \{x_8\}, \{x_1, x_2\}, \{x_3, x_7, x_{10}\}, \{x_5, x_9\}\};$$

$$U / IND\{a_1, a_2, a_3, a_4, a_5\} = \{\{x_4\}, \{x_6\}, \{x_8\}, \{x_1, x_2\}, \{x_3, x_7, x_{10}\}, \{x_5\}, \{x_9\}\}.$$

Получены одинаковые разделения элементов на основе подмножеств атрибутов $\{a_1, a_2\}$, $\{a_1, a_2, a_3\}$, $\{a_1, a_2, a_3, a_4\}$. Поэтому подмножества атрибутов $\{a_1, a_2, a_3\}$, $\{a_1, a_2, a_3, a_4\}$ являются излишними. Достаточно произвести разделение универсума на основе подмножества атрибутов $\{a_1, a_2\}$. Однако это заключение имеет силу только для данного конкретного примера и ни в коем случае не может служить поводом для обобщений.

Пусть мы имеем два подмножества атрибутов B, C $\subset$ A, и на основе этих подмножеств произведено разделение элементов универсума. По аналогии с классификациями элементов на основе множеств отношений эквивалентности, рассмотренными в предыдущем разделе, нас может интересовать связь между этими разделениями. Мы можем выделить положительную область разделения C на основе разделения B: $POS_B(C)$. Содержательно эта область содержит те элементы разделения C, которые могут быть корректно классифицированы, используя разделение на основе подмножества B. Тогда степень зависимости между двумя разделениями З. Павлак [1] предложил оценивать по выражению

$$\gamma_B(C) = \frac{cardPOS_B(C)}{cardU}, \tag{11}$$

где card означает кардинальность соответствующего множества элементов.

Существенным аспектом представления знаний, содержащихся в информационной системе, является оценка значимости или весов атрибутов для разделений элементов универсума. Такие веса, как правило, назначаются на основе субъективных суждений экпертов-специалистов. Но философия грубых множеств заключается в том, что могут быть использованы только существующие знания, представленные в информационной системе. Поэтому З. Павлак [1] предложил следующую идею оценивания значимости атрибутов. Пусть имеются два разделения элементов универсума, выполненные на основе двух подмножеств атрибутов B, C $\subset$ A. Степень зависимости $\gamma_B(C)$ между этими разделениями может быть оценена по выражению (11). Теперь предположим, что из подмножества B удален один атрибут или некоторое подмножество атрибутов B′ и произведено разделение элементов на основе подмножества атрибутов B - B′. По выражению (11) может быть оценена степень зависимости $\gamma_{B-B'}(C)$. Тогда в качестве оценки значимости подмножества атрибутов B′ принимается разность

$$\gamma_B(C) - \gamma_{B-B'}(C). \tag{12}$$

Содержательно эта разность интерпретируется как степень изменения положительной области $POS_B(C)$ при переходе от подмножества атрибутов B к подмножеству B - B′.

Рассмотрим пример, иллюстрирующий методику оценки значимости атрибутов.

Пример 3. Примем за основу информационную систему, представленную в таблице 1. Необходимо:

1. Оценить важность атрибутов из подмножества C = $\{a_1, a_2, a_3\}$ для разделения элементов универсума по подмножеству атрибутов D = $\{a_4, a_5\}$.

2.Оценить важность атрибутов из подмножества D = $\{a_4, a_5\}$ для разделения элементов универсума по подмножеству атрибутов C = $\{a_1, a_2, a_3\}$.

Решение.

1. $U/IND(D) = \{\{x_3, x_7, x_{10}\}, \{x_1, x_2\}, \{x_4, x_5, x_8\}, \{x_6, x_9\}\};$

$U/IND(C) = \{\{x_4, x_6\}, \{x_8\}, \{x_1, x_2\}, \{x_3, x_7, x_{10}\}, \{x_5, x_9\}\};$

$POS_C(D) = \{x_1, x_2, x_3, x_7, x_{10}\};$

$$\gamma_C(D) = \frac{cardPOS_C(D)}{cardU} = \frac{5}{10} = 0,5.$$

$U/IND(C - \{a_1\}) = \{\{x_3, x_7, x_{10}\}, \{x_4, x_6\}, \{x_5, x_9\}, \{x_8\}, \{x_1, x_2\}\};$

$POS_{(C-\{a_1\})}(D) = \{x_1, x_2, x_3, x_7, x_{10}\};$

$$\gamma_{(C-\{a_1\})}(D) = \frac{cardPOS_{(C-\{a_1\})}(D)}{cardU} = \frac{5}{10} = 0,5; \quad \gamma_C(D) - \gamma_{(C-\{a_1\})}(D) = 0,5 - 0,5 = 0.$$

$U/IND(C - \{a_2\}) = \{\{x_4, x_6,\}, \{x_8\}, \{x_1, x_2\}, \{x_3, x_5, x_7, x_9, x_{10}\}\};$

$$POS_{(C-\{a_2\})}(D)=\{x_1,x_2\};$$

$$\gamma_{(C-\{a_2\})}(D)=\frac{cardPOS_{(C-\{a2\})}(D)}{cardU}=\frac{2}{10}=0,2;\quad \gamma_C(D)-\gamma_{(C-\{a_2\})}(D)=0,5-0,2=0,3.$$

$$U/IND(C-\{a_3\})=\{\{x_4,x_6\},\{x_8\},\{x_1,x_2\},\{x_3,x_7,x_{10}\},\{x_5,x_9\}\};$$

$$POS_{(C-\{a_3\})}(D)=\{x_1,x_2,x_3,x_7,x_{10}\};$$

$$\gamma_{(C-\{a_3\})}(D)=\frac{cardPOS_{(C-\{a_3\})}(D)}{cardU}=\frac{5}{10}=0,5;\quad \gamma_C(D)-\gamma_{(C-\{a_3\})}(D)=0,5-0,5=0.$$

Таким образом, наиболее важным для рассматриваемого разделения является атрибут a_2.

2. $POS_D(C)=\{x_1,x_2,x_3,x_{7,}x_{10}\};$

$$\gamma_D(C)=\frac{cardPOS_D(C)}{cardU}=\frac{5}{10}=0,5.$$

$$U/IND(D-\{a_5\})=\{\{x_3,x_7,x_{10}\},\{x_1,x_2,x_4,x_5,x_8\},\{x_6,x_9\}\};$$

$$POS_{(D-\{a_5\})}(C)=\{x_3,x_7,x_{10}\};$$

$$\gamma_{(D-\{a_5\})}(C)=\frac{card(D-\{a_5\})(C)}{cardU}=\frac{3}{10}=0,3;\quad \gamma_D(C)-\gamma_{(D-\{a_5\})}(C)=0,5-0,3=0,2.$$

$$U/IND(D-\{a_6\})=\{\{x_1,x_2,x_3,x_7,x_{10}\}\},\{x_4,x_5,x_6,x_8,x_9\}\};$$

$$POS_{(D-\{a_6\})}(C)=\varnothing;$$

$$\gamma_{(D-\{a_6\})}(C)=\frac{card(D-\{a_6\})(C)}{cardU}=\frac{0}{10}=0;\quad \gamma_D(C)-\gamma_{(D-\{a_6\})}(C)=0,5-0=0,5.$$

Отсюда следует, что оба атрибута, a_5 a_6, являются важными для разделения элементов универсума по подмножеству атрибутов C $=\{a_1, a_2, a_3\}$ на основе разделения по подмножеству атрибутов D = $\{a_5, a_6\}$. При этом, атрибут a_6 является наиболее важным. По своей сути оценка степени важности атрибутов, предложенная З.Павлаком, характеризует результаты одного разделения элементов относительно другого разделения. Рассмотрим несколько иную ситуацию. Предположим, что произведено разделение элементов универсума по всему множеству атрибутов и нам необходимо оценить важность отдельных атрибутов в таком разделении. Формально мы не можем использовать для этой цели оценку (12), поскольку у нас нет опорного разделения, относительно которого могут быть рассчитаны релевантные степени зависимости. Чтобы преодолеть этот недостаток, предлагается следующий подход. Предположим, что произведено разделение элементов универсума по всему множеству атрибутов A. Теперь предположим, что мы производим разделение элементов по множеству атрибутов B = A. Естественно, что мы получим точно такое же разделение элементов. Чисто формально можно рассчитать оценку $\gamma_B(A)$, которая будет равна 1. Затем, последовательно удаляя из множества B отдельные атрибуты, могут быть рассчитаны любые интересующие нас оценки степеней зависимости. Чтобы продемонстрировать предлагаемую методику, рассмотрим пример.

Пример 4. Примем за основу исходные данные примера 2. Необходимо оценить значимость всех отдельных атрибутов для разделения элементов универсума по всему заданному множеству атрибутов.

Решение. Очевидно, что

$$U/IND(B)=U/INFD(A)=\{\{x_4\},\{x_6\},\{x_8\},\{x_1,x_2\},\{x_3,x_7,x_{10}\},\{x_5\},\{x_9\}\} \text{ и } \gamma_B(A)=1.$$

$$U/IND(B-\{a_1\})=\{\{x_3,x_7,x_{10}\},\{x_4\},\{x_6\},\{x_5\},\{x_8\},\{x_9\},\{x_1,x_2\}\};$$

$$U/IND(B-\{a_1\})=U/IND(A),POS_{(B-\{a_1\})}(A)=U;$$

$$\gamma_{(B-\{a_1\})}(A)=1;\quad \gamma_B(A)-\gamma_{(B-\{a_1\})}(A)=1-1=0.$$
$$U/IND(B-\{a_2\})=\{\{x_4\},\{x_6\},\{x_8\},\{x_1,x_2\},\{x_3,x_7,x_{10}\},\{x_5\},\{x_9\}\};$$
$$U/IND(B-\{a_2\})=U/IND(A), POS_{(B-\{a_2\})}(A)=U;$$
$$\gamma_{(B-\{a_2\})}(A)=1;\quad \gamma_B(A)-\gamma_{(B-\{a_2\})}(A)=1-1=0.$$
$$U/IND(B-\{a_3\})=\{\{x_4\},\{x_6\},\{x_8\},\{x_1,x_2\},\{x_3,x_7,x_{10}\},\{x_5\},\{x_9\}\};$$
$$U/IND(B-\{a_3\})=U/IND(A), POS_{(B-\{a_3\})}(A)=U;$$
$$\gamma_{(B-\{a_3\})}(A)=1;\quad \gamma_B(A)-\gamma_{(B-\{a_3\})}(A)=1-1=0.$$
$$U/IND(B-\{a_4\})=\{\{x_4\},\{x_6\},\{x_8\},\{x_1,x_2\},\{x_3,x_7,x_{10}\},\{x_5\},\{x_9\}\};$$
$$U/IND(B-\{a_4\})=U/IND(A), POS_{(B-\{a_4\})}(A)=U;$$
$$\gamma_{(B-\{a_4\})}(A)=1;\quad \gamma_B(A)-\gamma_{(B-\{a_4\})}(A)=1-1=0.$$
$$U/IND(B-\{a_5\})=\{\{x_4,x_6\},\{x_8\},\{x_1,x_2\},\{x_3,x_7,x_{10}\},\{x_5,x_9\}\};$$
$$POS_{(B-\{a_5\})}(A)=\{x_1,x_2,x_3,x_7,x_8,x_{10}\};$$
$$\gamma_{(B-\{a_5\})}(B)=\frac{cardPOS_{(B-\{a_5\})}(A)}{cardU}=\frac{6}{10}=0,6;\quad \gamma_B(A)-\gamma_{(B-\{a_5\})}(A)=1-0,6=0,4.$$

Отсюда имеем, что наиболее важным для разделения заданного универсума элементов по всему множеству атрибутов является атрибут a_5.

Выводы

1. Теория грубых множеств обеспечивает действенные методы обработки больших массивов неупорядоченных исходных данных.
2. Все методы теории грубых множеств основываются исключительно на использовании существующих знаний, представленных в базе знаний.
3. Все операции, выполняемые в базе знаний на основе отношений эквивалентности, могут быть выполнены в информационной системе на основе разделений элементов по множествам релевантных атрибутов.
4. Важность конкретного атрибута для разделения элементов универсума не является абсолютным понятием. Она определяется только содержанием конкретной задачи.
5. Оценку важности атрибутов можно выполнить только на основе анализа релевантных разделений элементов.
6. Методика, предложенная в настоящей статье, позволяет выполнить оценку важности отдельных атрибутов (или любых подмножеств таких атрибутов) для разделения элементов универсума по всему множеству релевантных атрибутов.

Summary

In this paper is analyzed knowledge representing problem and attribute signification's problem, based on rough sets approach. Knowledge discovering and representing in information systems has evolved into an important area of research because of theoretical challenges and practical representing unknown knowledge in data. Rough set theory is a mathematical formalism for representing uncertainty that can be considered a special extension of the set theory.

This paper permit make following conclusions:

7. Rough sets theory provides effective methods for treatment with chaotic initial data.
8. All rough sets theory's methods are based exceptionally on knowledge in knowledge base.
9. All operations implementing in knowledge base on the base of equivalence relations can be implemented in information system on the base of elements' partitioning. This partitioning is implemented on the base of attribute sets.

10. Specific attribute significance is not absolute notion. Attributes' significance makes sense only in concrete problem.
11. Attributes' significance can be evaluated only on the base relevant elements' partitioning.
12. The method proposed in this paper allow evaluate separate attribute's significance (or any this attribute subsets) for elements' partitioning on the base of all relevant attribute.

Литература

1. Pawlak Z. Rough sets. Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, Dordrecht/Boston/London, 1991. 229 p.
2. Shafer G. A mathematical theory of evidence. Princeton University Press, 1976. 297 p.
3. Smarandache F. & Dezert J. (Editors). Advances and Applications of DSmT for Information Fusion. Volume 1, American Research Press, Rehobolth, 2004.
4. Tripathy H.K., Tripaty B.K., and Das P.K. An Intelligent Approach of Rough Set in Knowledge Discovery Databases. International Journal of Computer Science and Engineering, 2008. pp. 89-93.