

Кантуреева М.Л., Сеньковская А.Л., Болысбек М.А.

**ПРОЕКТИРОВАНИЕ ПРОДУКЦИОННОЙ БАЗЫ ЗНАНИЙ
НА ОСНОВЕ РАЗДЕЛЕНИЯ НА ИНФОРМАЦИОННЫЕ ЕДИНИЦЫ
ВЫБРАННОЙ ПРЕДМЕТНОЙ ОБЛАСТИ**

М.А. Kantureeva, A.A. Senkovskaya, M.A. Bolysbek

**DESIGN OF KNOWLEDGE PRODUCTION BASE BASED ON THE SEPARATION
OF UNITS OF SELECTED INFORMATION SUBJECT AREA**

УДК: 1339:001/859

В этой работе описывается процедура формализации проектирования продукционной базы знаний, на основе разделения на информационных единиц выбранной предметной области. При этом указывается некоторые особенности, связанные обработкой этих знаний.

In this work formalization procedure is designed of the production of knowledge base, on the basis of division of information units is described of the chosen subject domain. It is specified some features that connected by processing of this knowledge.

В большинстве современных систем управления очень остро встает вопрос о придании им свойства интеллектуальности, то есть реализации в них каких-то моделей принятия решений естественным интеллектом (человеком). Одной из существенных особенностей таких систем является наличие внутри системы модели внешнего мира, которая и называется внутренним представлением. В процедуре внутреннего представления центральное место занимает создание базы знания (БЗ) выбранной предметной области. БЗ в общем случае содержит сведения, которые отражают закономерности, существующие в предметной области (ПО), и позволяют, как выводить новые факты, имеющие место в данном состоянии ПО, так и прогнозировать потенциально возможные состояния. Из-за сложности формализации различных свойств и поведения ПО, практика показала [1,2,3], что описание этих объектов в виде знаний, легче производить в виде информационных сущностей. В этом случае ПО разделяется на глобальные информационности (или единицы ГИЕ) и на локальные информационные единицы (ЛИЕ). Кроме того, причинно-следственные связи этих единиц представляют знание о ПО [1,4,5].

Информация, с которой будем иметь дело в ЭВМ, разделяется на процедурную и декларативную. Процедурная информация описана в программах, которые выполняются в процессе решения задач, декларативная информация - в данных, с которыми эти программы работают [5]. С развитием информационных структур для представления данных, появились способы описания данных в виде векторов и матриц, возникли списочные структуры, иерархические структуры и т. д. Появление базы данных (БД) знаменовало собой еще один шаг на пути организации работы с декларативной информацией. В базах данных могут одновременно храниться большие объемы

информации, а специальные средства, образующие систему управления базами данных (СУБД), позволяют эффективно манипулировать данными, при необходимости извлекать их из БД и записывать в нужном порядке в базу. По мере развития исследований в области искусственного интеллекта (ИС) возникла концепция знаний, которые объединили в себе многие черты процедурной и декларативной информации. Особенности знаний, и его отличие от данных описан в работе [5]. Мы будем придерживаться определений, приведенных в этой работе. Перечисляется пять особенностей информационных единиц (ИЕ): внутренняя интерпретируемость, структурированность, связность, семантическая метрика и активность определяют ту грань, за которой данные превращаются в знания, а базы данных перерастают в базы знаний (БЗ).

Таким образом ИЕ предметной области находятся в определенных отношениях друг с другом (ассоциациях), которые также можно рассматривать как ИЕ и включать в предметную область. Между ИЕ наблюдаются различные отношения подобия. Совокупность подобных ИЕ составляет класс ИЕ, являющейся новой ИЕ предметной области. Отношения между ИЕ, в совокупности с самими ИЕ составляют предметную область. Языки, предназначенные для описания предметных областей, называется методом представления знаний [5].

Одним из таких методов является продукционный язык представления знания. С помощью этого метода процедурная информация явно выделяется и описывается иными средствами, чем декларативная информация, и делается вывод на знаниях. Ниже приведено описание продукционного метода, представляющего предметную область в виде следственно-причинных связей.

Пусть задана совокупность множеств $M_j, j = 1, m$, которые назовем признаками (атрибутами) предметной области, зависящими от фиксированных параметров $\lambda(i), i = 1, 2, \dots$, а их элементы - значениями соответствующих атрибутов [6].

Предметная область. Под ней будем понимать многомерный вектор, компонентами которого являются ИЕ. Так, если некоторую описываемую предметную область обозначить как S_i , а атрибуты соответственно через $\alpha_{ji}(\lambda), i = \overline{1, p}; j = \overline{1, m}$, то предметный объект $S_i = \{S_{ji}(\alpha_{ji}(\lambda)), \dots, S_{mi}(\alpha_{mi}(\lambda))\}$ будет допустимым, если $\alpha_{ji}(\lambda) \in M_j$. Очевидно, что набор значений атрибутов при

фиксированном значении $\lambda(i)$, $S_j(\lambda) = \{a_{ij}(\lambda), \dots, \lambda_{\pi}(\lambda)\}$ будет ИЕ предметной области.

Базы знаний. Пусть заданы подмножества $Y = \{a_1, \dots, a_n\}$: $Y \in M_j$; $\Lambda' = \{\Lambda'_1, \dots, \Lambda_n\}$. Из них с помощью алгоритма А образуем множества 2 следующим образом:

- все конечно упорядоченные подмножества с элементами из Y и принадлежит Σ_2 ;

- если $S'_1, 8^n \in \Sigma_2$, то $S', \Lambda', S' \in \Sigma_2$ тогда Σ_2 будет продукционной базой знания.

Класс знаний. Пусть для каждого класса K_j задан набор ГИЕ $V_1^s, \dots, V_{n(j)}^s$, и операций $Q^s, Q^{s_{q(j)}}$, применение которых к значениям ГИЕ K_j состоит только из тех, которые получаются из V_1^s , с применением к ним операций $Q^s; u = \overline{1, k(j)}; t = \overline{1, q(j)}$.

Информационная структуризация предметной области.

Описание $I^*(K_j(S_{ij})) = \{V_1^j, \dots, V_{k(j)}^j, Q_1^j, \dots, Q_{q(j)}^j\}$ класса K_j называется информационно-структурным описанием S_{ij} , набор называется $B^j u = \{B_1^j, \dots, B^j_{k(j)}\}$ глобальной ИЕ в $I^*(K_j(S_{ij}))$, а $b_c^{uj} = \{b_1^{uj}, \dots, b_{q(i)}^{uj}\}_c$ - локальной ИЕ в $I^*(K_j(S_{ij}))$.

Ядро продукционного знания. Если с помощью операции $Q_1^j, \dots, Q_{q(j)}^j$ соединить глобальные и локальные порождающие ИЕ, то получится ядро элементарных продукционных знаний. Если $I^*(K_j(S_{ij}))$ описать набором формул $\{\Phi_k^j\} \in \Sigma$ в виде

$$Q; I^*(\Phi_k^j) = (B_u^j \rightarrow \sum_{c=1}^{q(i)} b_c^{uj})_k; N$$

где $B_u^j \in \{B_1^j, \dots, B_{k(j)}^j\}$, то $b_c^{uj} \in \{b_1^{uj}, \dots, b_{q(i)}^{uj}\}_c$

будет базой продукционного знания, а $(B_1^j \rightarrow b_{q(i)}^{uj})$ - ядром продукционного знание. Знак следования « \rightarrow » в $A \rightarrow B$ определен здесь лишь для случая $A=1$ стандартным образом: $A=1$, то $B=1$, т.е. это и есть условие активизации информационной базы.

Здесь k - номер продукции, j - номер класса ИЕ, и u - номер ГИЕ, c - номер ЛИЕ. Элемент Q характеризует сферу применения продукции. Такие сферы легко выделяются в когнитивных структурах человека. Наши знания как бы «разложены по палочкам». На одной «палочке» хранятся знания о том, как надо готовить пищу, на другой - как добраться на работу и т.п. Разделение знаний на отдельные сферы позволяет экономить время на поиск нужных знаний. Такое же разделение на сферы в БЗ системы целесообразно и при использовании для представления знаний продукционных моделей. Элемент N описывает подусловия продукции. Они актуализируется только в том случае, если ядро продукции реализовалось. Подусловия продукция описывают действия и

процедуры, которые необходимо выполнить после реализации

$\sum_{c=1}^{q(i)} b_c^{uj}$. Например, в клинике после диагностики болезни, надо осуществить процедуру лечения. Выполнение N может происходить не сразу после реализации ядро продукции.

Продукция $\{\Phi_k^j\}$ может соответствовать процедуре нахождения закономерностей по эмпирическим данным. На основании просмотра и анализа данных выдвигается гипотеза о наличии закономерностей и, убедившись в их приемлимости и достаточной обоснованности, записывают их в базу знаний. Эта процедура называется нахождения значений ГИЕ и ЛИЕ [1,3,7].

В этом случае объективность БЗ зависит от степени профессионализма эксперта. Но с другой стороны выделение этих ИЕ можно осуществлять алгоритмическим методом. В этом случая из множества M , выделяется такие атрибуты, которые имеют глобальные свойства в M , и атрибуты, имеющие локальные свойства в M . А такие атрибуты всегда связаны между собой причинно - следственным образом [1]. Конечно, если будет выполняться принцип допустимости для S_i .

Надо отметить, что в продукционном БЗ в вид X , легко можно вести семантическую метрику, это вопрос особо исследования.

Следующая теорема будет выражать корреляционную особенность продукции $\{\Phi_k^j\}$.

Теорема 1. Информационные единицы, составляющие V^3 , будет иметь глобальные свойства если корреляционное взаимоотношение с другими информационными единицами максимальное в Q .

Доказательство. Корреляционные взаимоотношения ИЕ выражают поведенческую влияния этих единиц друг на друга. Если положительная коррелированность сильная, это говорит, что изменение ИЕ повлияет изменению других ИЕ в сторону увеличения. Если отрицательная коррелированность сильная, то увеличением одной ИЕ приводит к уменьшению другой ИЕ. Из этого следует, что ИЕ, имеющие максимальное число коррелированных взаимоотношений в Q с другими ИЕ, будет иметь максимальные причинные свойства. Это и есть глобальное свойства ИЕ.

Из этих соображений следует, что ИЕ, составляющими которых является b_c^{uj} , имеют такие же свойства только в классе K_j . Информационные единицы, имеющие такие свойства, называются локальными информационными единицами.

Также следует отметить что, локальные и глобальные свойства ИЕ можно определить следующим образом [1].

Допустим нам задан ПО S :

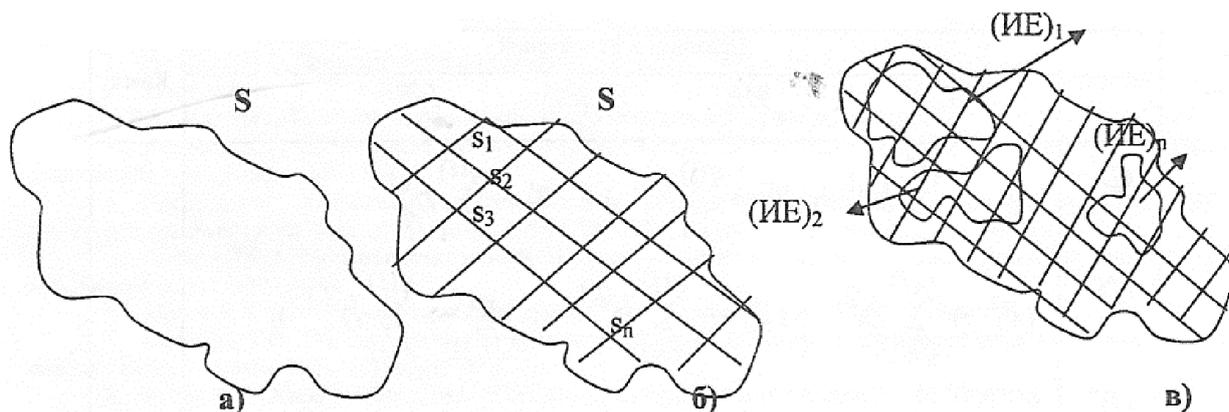


Рис. 1. Предметный объект (а), ПО представленный в совокупности признаков (б), ПО представленный в совокупности информационных единиц (в).

Каждая S описывается совокупностью признаков $\{a_y\}$ и эти данные сводятся в таблицу.

Таблица 1

Объект	Признаки и их значений	Класс
	a_1, a_2, \dots, a_m	
S1	$a_{11}, a_{12}, \dots, a_{1m}$	K1
S2	$a_{21}, a_{22}, \dots, a_{2m}$	K2
....
S2	$a_{n1}, a_{n2}, \dots, a_{nm}$	K1

На основе обобщения по признакам 8, при этом учитывая наличие в T_{mn}^s статических закономерностей в его проявлениях, представляется в виде информационных единиц (ИЕ). Такое представление производится следующим образом. Из информационных единиц выделяются ИЕ, имеющие глобальные свойства (B_{-i}), и локальные свойства $\Gamma_{сч}$ а эти ИЕ объединяются с помощью импликации « \rightarrow ». Эта процедура осуществляется на основе теоремы 1. В результате 8 (рис1.(в)) будет представлен в следующем виде:

$$\sum_Q : \Phi_{ik}^j = (\Gamma ИЕ)_k^j \rightarrow (\ЛИЕ)_c^{u,j}$$

Σ_Q – будет продукционной БЗ. Эту процедура можно семантически представить как:

$$БЗ: S \rightarrow T_{mn}^s \rightarrow \Sigma_Q \tag{2}$$

Если информационные единицы задаются в количественном виде, продукционную БЗ можно строить с помощью формулы (1). Если информационные единицы задаются в семантическом виде, то формулу (1) надо представить в следующем виде:

$$Q; I^*(\Phi_k) = (B_{iu}^j \rightarrow \prod_{c=1}^{q(i)} b_{c,u,j}^{u,j})_k; N \tag{3}$$

Тогда ГИЕ и ЛИЕ определяется семантическим методом. В продукционной БЗ обработка осуществляется процедурой сравнение и принятием решений [8]. Для этого БЗ, представленную по формулам (1) и (3), надо представить в табличной форме [9]. В этом случае по строкам ставится продукции, по столбцам номер ИЕ, на пересечении строк и столбцов ставится ядро продукции.

Таблица 2

Объект	Признаки и их значений	Класс
	ИЕ 1, ИЕ 2, ..., ИЕ q	
Φ_1^1	$(B_1^1 \rightarrow \sum_{c=1}^{q(i)} b_{c,1,1}^{1,1})_1 (B_1^1 \rightarrow \sum_{c=1}^{q(i)} b_{c,2,1}^{1,1})_2 \dots (B_1^1 \rightarrow \sum_{c=1}^{q(i)} b_{c,q,1}^{1,1})_q$	K1
Φ_1^2	$(B_1^2 \rightarrow \sum_{c=1}^{q(i)} b_{c,1,2}^{2,1})_1 (B_1^2 \rightarrow \sum_{c=1}^{q(i)} b_{c,2,2}^{2,1})_2 \dots (B_1^2 \rightarrow \sum_{c=1}^{q(i)} b_{c,q,2}^{2,1})_q$	K2
Φ_1^n	$(B_1^n \rightarrow \sum_{c=1}^{q(i)} b_{c,1,n}^{n,1})_1 (B_1^n \rightarrow \sum_{c=1}^{q(i)} b_{c,2,n}^{n,1})_2 \dots (B_1^n \rightarrow \sum_{c=1}^{q(i)} b_{c,q,n}^{n,1})_q$	K1

Φ_2^1	$(B_2^1 \rightarrow \sum_{c=1}^{q(i)} b_1^{1,2})_1 (B_2^1 \rightarrow \sum_{c=1}^{q(i)} b_2^{1,2})_2 \dots (B_2^1 \rightarrow \sum_{c=1}^{q(i)} b_q^{1,2})_q$	K ₂
Φ_2^2	$(B_2^2 \rightarrow \sum_{c=1}^{q(i)} b_1^{2,2})_1 (B_2^2 \rightarrow \sum_{c=1}^{q(i)} b_2^{2,2})_2 \dots (B_2^2 \rightarrow \sum_{c=1}^{q(i)} b_q^{2,2})_q$	
$\Phi_2^{n'}$	$(B_2^{n'} \rightarrow \sum_{c=1}^{q(i)} b_1^{1,2})_1 (B_2^{n'} \rightarrow \sum_{c=1}^{q(i)} b_2^{1,2})_2 \dots (B_2^{n'} \rightarrow \sum_{c=1}^{q(i)} b_q^{1,2})_q$	
Φ_n^1	$(B_n^1 \rightarrow \sum_{c=1}^{q(i)} b_1^{1,n})_1 (B_n^1 \rightarrow \sum_{c=1}^{q(i)} b_2^{1,n})_2 \dots (B_n^1 \rightarrow \sum_{c=1}^{q(i)} b_q^{1,n})_q$	K ₁
Φ_n^2	$(B_n^2 \rightarrow \sum_{c=1}^{q(i)} b_1^{2,n})_1 (B_n^2 \rightarrow \sum_{c=1}^{q(i)} b_2^{2,n})_2 \dots (B_n^2 \rightarrow \sum_{c=1}^{q(i)} b_q^{2,n})_q$	
Φ_n^m	$(B_n^m \rightarrow \sum_{c=1}^{q(i)} b_1^{m,n})_1 (B_n^m \rightarrow \sum_{c=1}^{q(i)} b_2^{m,n})_2 \dots (B_n^m \rightarrow \sum_{c=1}^{q(i)} b_q^{m,n})_q$	

Σ_α будет продукционной БЗ для предметной области На Ед очень просто реализовать процедуру сравнения.

Если определить на Σ_α семантическую метрику, характеризующую близость $\{\Phi_k^i\}$, то процедуру приобретения новых знаний можно осуществить алгоритмами геспознавания [10].

Резюмируя работу, следует подчеркнуть некоторые принципиальные особенности продукционных :n!нии. Термин «продукция» принадлежит американскому логик Э. Посту [11]. В понимании Поста в Еяегсве продукции выступила только та ее часть, которую теперь называют ядром. Запись «ЕСЛИ А, 70 В» трактовалась как оператор замены цепочки А цепочкой В в некотором входном слове, т.е. продукции были тем, что позже стали называть подстановками и использовать при описании различных тс-чнений понятия алгоритма. В предлагаемой работе продукция представляется не только заменой А на Е. еще раскрывается семантические свойства А и В с помощью «причинно-следственных» связей. Такой ШЯДОД, расширить практические возможности продукционных моделей знаний [1,2,9].

Литература:

1. Рустамов Н.Т. Прикладное распознавание. Туркестан. 1999, с. 84.
2. Темирбеков А.Н., Рустамов Н.Т., Сейтикеримова Д.С., - Рустамов Б.К. К вопросу создания базы знаний для хронических заболеваний. Вестник МКТУ, № 1, 2007, с. 15-21.
3. Ларичев О.И., Мечитов А.И., Мошкович Е.М., Фуремс Е.М. Выявление экспертных знаний (процедуры и ранящий). - М.: Наука. 1989 -128 с.
4. Поспелов И.Г., Поспелова Л.Л. Динамическое описание систем продукций и проверка непротиворечивого продукционных экспертных систем //Изв. АН СССР, Техн. Кибернетика, 1987, №1. - с. 184-192.
5. Поспелов Д.А. Представление знаний. Искусственный интеллект. Кн.2. Модели и методы: Справочник /— М.: Радио и связь, 1990. -304 с.
6. Мейер Д. Теория реляционных базы данных: Пер. с англ. - М.: Мир, 1987. -608 с.
7. Dietterich T.G. The methodology of knowledge layers for inducing descriptions of sequentially orderd events. Master's thesis. – Urbana: Univ. of llinois, 1979.
8. Кузин Л.Т. Основы кибернетики. С2. Основы кибернетических моделей. - М.: Энергия.
9. Рустамов Н. Т., Темирбеков А. Н., Сейдикеримова Д. С. Табличный метод организаций знаний. Вестник МКТУ, Туркестан, №1, 2007 с.14-17.
10. Журавлев Ю.И. Об алгоритмическом подходе к решению задач распознавания и классификации // Про- 5юсы кибернетики. - М.: Наука, 1978. -Вып. 33, -с 5-68.
11. Pospelov D. Models of human communication: dialogue with computer//Int J. Genaral Systems. Methodology Applications Educatuion. - 1986. - Vol. 12, № 4 - P. 333-338.

Рецензент: д.т.н., профессор Усенов К.Ж.