

## АЛГЕБРАИЧЕСКИЕ СВОЙСТВА ОПЕРАЦИЙ ИНТЕЛЛЕКТУАЛЬНОЙ БАЗЫ ДАННЫХ

к.т.н. Т.Н. Новожилова  
(представила д.т.н., проф. Р.Т. Волколупова)

Рассмотрены механизмы определения связей между наборами правил разных классов в интеллектуальной базе данных (ИБД).

Сложность определения алгебраических свойств операций между наборами правил при организации системы словаря справочника базы метазависимостей (СССБМЗ) связана с расширяемостью набора метаправил: каждый новый кортеж вводит собственную алгебру, для которой в СССРМЗ может не быть эпитеоремы [1]; в процессе приобретения знаний трудно с первого раза записать метатеорему кортежем наборов правил [2].

Поскольку для управления процессом приобретения знаний требуются сложные структуры данных и условия выборки, отличные от реляционных, в языке представления знаний (ЯПЗ) ИБД не соблюдается замкнутость логических форм. Разрабатываемый ЯПЗ для задач экспертной классификации частично основывается на семантике включения, т.е. суперкласс как кортеж наборов правил включает наборы правил различных классов.

Операция паросочетания в ИБД позволяет отложить классификацию набора правил (связывание) [3]. В этом случае в ИБД можно не выводить схему запроса, а использовать эпитеорему СССРМЗ и на этой основе изменять критерии выбора схемы связывания.

Если СССРМЗ не содержит информации о возможной коммутативности двух операций кортежа, то необходимо *формальное определение алгебраических свойств операций набора правил при его разработке*. Операндами являются наборы правил, а результатом, как правило, является набор правил неопределённого класса  $R_{na}(H_2, r_1, \dots, r_n; h_1, \dots, h_n)$ , где  $\{r_1, \dots, r_n\}$  - операторы над механизмом целеполагания  $H_1$ . Классифицированный набор правил рассматривается как параметризованный

$$R_{na}(H'_0) = [X_1 \rightarrow O_1, X_2 \rightarrow O_2, \dots, X_i \rightarrow O_i, M], \quad (1)$$

где  $M$  – множество преобразований над  $[X_i \rightarrow O_i]$ . Все  $[X_i \rightarrow O_i]$ , входящие в группировку фактов по ситуациям  $R_{na}(H'_0)$ , представляются множеством значений  $\{h_1, h_2, \dots, h_i\}$ .

Среди операций ЯПЗ ИБД над классами наборов правил различаются: теоретико-множественные операции и операция селекции, формирующие из наборов правил неопределённого класса, приобретаемых у эксперта, логиче-

ские формы ЯПЗ ИБД со схемой вывода из СССБМЗ; операции проекции и соединения, формирующие логические формы ЯПЗ ИБД со схемой вывода, которая в общем случае не описывалась статически в составе схем вывода ИБД.

Теоретико-множественные операции определяются для наборов правил, согласованных с точностью до типов, и класс результата помещается в решетку классов схемы ИБД в соответствии с семантикой включения. Таким образом, операции ИБД формируют не только значение кортежа (метатеорему), но и тип этого значения (эпитеорему).

**Определение 1.** При объединении двух наборов правил по  $\mathbf{h}$  получается набор, включающий все кортежи, входящие хотя бы в один из наборов правил:

$$\{\exists \mathbf{h} \mid \mathbf{M}(\mathbf{h}) \cup \mathbf{R}_{\text{на}}(\mathbf{H}_1) = \mathbf{R}_{\text{на}}(\mathbf{H}_2)\}, \quad (2)$$

если 
$$\text{CF} \left[ \mathbf{R}_{\text{на}}(\mathbf{H}_1) \cup \mathbf{R}_{\text{на}}(\mathbf{H}'_0) \xrightarrow{\mathbf{M}(\mathbf{h})} \mathbf{R}_{\text{на}}(\mathbf{H}'_2) \right] > \mathbf{E.UNKN}, \quad (3)$$

где  $\mathbf{R}_{\text{на}}(\mathbf{H}'_2)$  - модель интерпретации,  $\mathbf{E.UNKN}$  - порог неизвестного [2].

Как и в теории множеств, в ЯПЗ ИБД операция объединения осмыслена для любых двух наборов правил. Результатом операции будет множество разнотипных кортежей, т.е. набор правил неопределённого класса. Такая операция объединения является бессмысленной в реляционной алгебре, исходящей из требований замкнутости.

**Определение 2.** Операция пересечения двух наборов правил по  $\mathbf{h}$  производит набор правил, включающий все кортежи  $\{\mathbf{r}_1, \dots, \mathbf{r}_n\}$ , входящие в оба набора:

$$\{\exists \mathbf{h} \mid \mathbf{R}_{\text{на}}(\mathbf{H}_2) \cap \mathbf{R}_{\text{на}}(\mathbf{h}) = \mathbf{R}_{\text{на}}(\mathbf{h})\}, \quad (4)$$

если 
$$\mathbf{R}_{\text{на}}(\mathbf{h}) \in \mathbf{R}_{\text{на}}(\mathbf{H}_1) \cap \mathbf{R}_{\text{на}}(\mathbf{H}_2) \neq \emptyset; \quad (5)$$

$$\mathbf{R}_{\text{на}}(\mathbf{h}) \subseteq \mathbf{R}_{\text{на}}(\mathbf{H}_2). \quad (6)$$

**Определение 3.** При соединении двух наборов правил по  $\mathbf{h}$  образуется набор правил неопределённого класса

$$\{\exists \mathbf{h} \mid \mathbf{R}_{\text{на}}(\mathbf{H}_1) \oplus \mathbf{R}_{\text{на}}(\mathbf{h}) = \mathbf{R}_{\text{на}}(\mathbf{H}_2)\}. \quad (7)$$

**Определение 4.** Ограничением набора правил по  $\mathbf{r}$  является набор правил, удовлетворяющий условию

$$\{\exists \mathbf{h} \mid \mathbf{R}_{\text{на}}(\mathbf{H}_2) \xrightarrow{\mathbf{r}} \mathbf{R}_{\text{на}}(\mathbf{h})\}, \quad (8)$$

если 
$$\exists \mathbf{M}(\mathbf{h}) : \mathbf{H}_2 \xrightarrow{\mathbf{h}} \mathbf{H}_1; \quad (9)$$

$$\{\mathbf{M}_2(\mathbf{h}) \mid \exists \mathbf{M}_1(\mathbf{h}) \in (\mathbf{R}_{\text{на}}(\mathbf{r}) \cap \mathbf{R}_{\text{на}}(\mathbf{h}))\}. \quad (10)$$

**Определение 5.** Разность двух наборов правил по  $\mathbf{h}$  включает все кортежи, входящие в первый набор, такие, что ни один из них не входит во второй набор

$$\{\exists \mathbf{h} \mid \mathbf{R}_{\text{на}}(\mathbf{H}_2) \setminus \mathbf{R}_{\text{на}}(\mathbf{h}) = \mathbf{R}_{\text{на}}(\mathbf{H}_1)\}. \quad (11)$$

**Определение 6.** Переименование  $\mathbf{R}_{\text{на}}(\mathbf{H}_0)$  в набор правил с той же структурой производится изменением атрибутов в соответствии с требованиями логической формы

$$\{\exists h \mid R_{na}(H_0) = R_{na}(H_2)\}. \quad (12)$$

**Определение 7.** Операция присваивания позволяет сохранить результат логического вывода в логической форме определённого класса

$$\{\exists h \mid R_{na}(H_2) = R_{na}(h)\}, \quad (13)$$

если

$$\lim_{h \rightarrow r} \sum R_{na}(H_2) > E.UNKN. \quad (14)$$

В этом случае класс, которому присваивается результат алгебраического выражения, должен быть определен в схеме СУБМЗ ИБД заранее.

**Определение 8.**  $R_{TX}(H_1 \oplus H_2, M(h))$  является обобщением  $H_1$  и  $H_2$ , если справедливо равенство

$$\begin{aligned} & \lim_{r \Rightarrow h} (R_{na}(H_1) \cup R_{na}(H_2)) \cap M(r) = \\ & = (R_{na}(H_1) \cap M(h)) \cup ((R_{na}(H_2) \cap M(h))). \end{aligned} \quad (15)$$

Любую эвристику можно включить в операцию обобщения, если  $\left\{ \forall h \mid \exists M(h) \in [R_{na}(H_1) \cup R_{na}(H_2)] \xrightarrow{M(h)} R_{TX}(H_1 \oplus H_2, M(h)) \right\}$ . (16)

В отличие от обобщения, операция ассоциации связывает набор правил неопределённого класса и  $R_{na}(h)$  в кортеж вывода:

$$\left\{ R_{na} \left[ \begin{array}{cc} h & H_2 \\ X & O \end{array} \right] \rightarrow R_{na}(H_2) \right\}. \quad (17)$$

Другая запись  $\Delta R_{na} : R_{na}(h) \xrightarrow{[X-O]} R_{na}(H_2)$ . (18)

**Определение 9.** Операция селекции формирует класс, являющийся подклассом последовательности  $R_{na}(H_2)_1, R_{na}(H_2)_2, \dots, R_{na}(H_2)_n$ :

$$\sigma_h(R_{na}(H_2) \xrightarrow{r} R_{na}(H_1)) = \{R_{na}(H_2) \cap M(h)\}, \quad (19)$$

если

$$\exists \lim_{h \rightarrow r} CF(R_{na}(H_2) \cap M(h)) \geq E.UNKN. \quad (20)$$

**Определение 10.** При проекции  $R_{na}(H_2)$  на заданный  $\{r_1, \dots, r_n\}$  получается набор правил, кортежи которого производятся путем взятия соответствующих значений из кортежей  $\{r_1, \dots, r_n\}, \{h_1, \dots, h_n\}$

$$\pi_h(H'_0) : R_{na}(H_2) \xrightarrow{M(h)} R_{na}(H_1), \quad (21)$$

если  $f(CF(H'_0)) = f(CF[R_{na}(H_1)] \oplus \left\{ \bigcup_i R_{na}(H_{2i}) \right\})$ . (22)

**Определение 11.** Набор правил  $R_{na}(H'_0)$ , кортежи которого являются сцеплением кортежей  $R_{na}(H_1)$  и  $R_{na}(h)$ , получается через произведение

$$R_{na}(H_1) \otimes R_{na}(h) = R_{na}(H'_0). \quad (23)$$

Если наборы правил формируются как связи между другими наборами правил, то (23) выражает агрегацию [2].

Операцию соединения классов можно определить с использованием операций произведения и пересечения:

$$\mathbf{R}_{na}(\mathbf{H}_1) \otimes \mathbf{R}_{na}(\mathbf{h}) = \mathbf{R}_{na}(\mathbf{H}'_0); \quad (24)$$

$$\{\mathbf{h} \mid \exists \mathbf{M}(\mathbf{h}) \in (\mathbf{R}_{na}(\mathbf{H}'_0) \cap \mathbf{R}_{na}(\mathbf{h})) = \mathbf{R}_{na}(\mathbf{H}_2)\}. \quad (25)$$

Для обобщения по выделенной структуре отношения в ИБД вводятся нормальные формы (НФ)  $\mathbf{R}_{na}(\mathbf{H}_2)$ .

1. В 1НФ в кортеже набора правил  $\mathbf{R}_{na}(\mathbf{H}_2)$  устраняются повторяющиеся  $\mathbf{r}_i$ , т.е. производится выявление неясных  $\mathbf{h}_i$ , "замаскированных" под  $\mathbf{r}_i$ ,

$$\forall \mathbf{r}_i \neq \mathbf{M}_1(\mathbf{h}) \in \mathbf{M}. \quad (26)$$

2. Во 2НФ в кортеже набора правил  $\mathbf{R}_{na}(\mathbf{H}_2)$  устраняются  $\mathbf{r}_i$ , зависящие только от части устойчивого множества рассуждений  $\mathbf{R}_{na}(\mathbf{H}_2)$ , определяющей  $\mathbf{M}_1(\mathbf{h}) = \{\mathbf{X}_0 - \mathbf{O}_0\}$  с точностью до схемы вывода [3].

3. В 3НФ в  $\mathbf{R}_{na}(\mathbf{H}_2)$  устраняются  $\mathbf{r}_i$ , зависящие от  $\mathbf{M}_2(\mathbf{h}) \in \mathbf{M}$ , не входящих в устойчивое множество рассуждений,  $\forall \mathbf{r}_i$  зависящие от  $\mathbf{M}_2(\mathbf{h})$  - основа отдельной логической формы, согласованной с точностью до класса [3].

Таким образом, если результатом операции ЯПЗ ИБД является  $\mathbf{R}_{na}(\mathbf{H}_2)$ , то можно образовывать логические формы, в которых вместо операнда находится вложенное логическое выражение. Во всех конфликтных случаях предполагается операция переименования класса.

**Теорема.**  $\mathbf{H}_1$  и  $\mathbf{H}'_0$  будут сопоставимыми логическими формами в том и только в том случае, если  $\exists \mathbf{M}(\mathbf{h}) \in \{\mathbf{R}_{na}(\mathbf{H}_2)\}$ , такая, что

$$\lim_{\mathbf{h} \rightarrow \mathbf{r}} \bigcup_{\mathbf{h}} \mathbf{M}(\mathbf{h}) = \mathbf{R}_{na}(\mathbf{H}_1). \quad (27)$$

*Доказательство.* Очевидно, что теорема будет доказана, если  $\mathbf{R}_{na}(\mathbf{H}_2)$  полученный из

$$\{\exists \mathbf{h} \mid \mathbf{R}_{na}(\mathbf{H}_1) \oplus \mathbf{R}_{na}(\mathbf{h}) = \mathbf{R}_{na}(\mathbf{H}_2)\}, \quad (28)$$

будет классифицирован как  $\mathbf{R}_{na}(\mathbf{H}'_0)$ . Для этого операцию паросочетания необходимо выразить через рассмотренные выше операции ЯПЗ ИБД, объединенных схемой вывода

$$\left[ \mathbf{R}_{na}(\mathbf{h}) \rightarrow \mathbf{R}_{na}(\mathbf{H}_1) \rightarrow \left\{ \mathbf{R}_{na} \left[ \begin{array}{cc} \mathbf{h} & \mathbf{H}_2 \\ \mathbf{X} & \mathbf{O} \end{array} \right] \rightarrow \mathbf{R}_{na}(\mathbf{H}_2) \right\} \rightarrow \mathbf{R}(\mathbf{H}'_0) \right]. \quad (29)$$

Поскольку два набора правил  $\mathbf{R}_{na}(\mathbf{H}'_{0i})$  и  $\mathbf{R}_{na}(\mathbf{h})$  паросочетательны тогда и только тогда [3], когда  $\exists \Delta \mathbf{R}_{na}$  такой, что

$$\mathbf{R}_{na}(\mathbf{H}'_{0i}) = \mathbf{R}_{na}(\mathbf{h}) \& \Delta \mathbf{R}_{na}, \quad (30)$$

то подставив в (29) вместо вложенного логического выражения (17),  $\Delta \mathbf{R}_{na}$ , и заменив в (30) операцию паросочетания соединением, устанавливаем соответствие между (29) и (30):

$$\lim_{\mathbf{R}_{na}(\mathbf{H}_1) \rightarrow \mathbf{R}_{na}(\mathbf{h})} \mathbf{R}_{na}(\mathbf{H}_1) \oplus \mathbf{R}_{na}(\mathbf{H}_2) = \mathbf{R}(\mathbf{H}'_0). \quad (31)$$

Выведенное в [2] отображение

$$R_{na}(H_2) : R_{na}(h) \xrightarrow{|X-O|} R_{na}(H'_0), \quad (32)$$

перепишем следующим образом:

$$\pi_h(H'_0) : R_{na}(H_2) \xrightarrow{R_{na}(h)} R_{na}(H_1). \quad (33)$$

Очевидно, при подстановке  $R_{na}(h)=M(h)$ , соблюдение

$$M(h) : H_1 \xrightarrow{h} H'_0, \quad (34)$$

невозможно без соблюдения

$$\lim_{h \rightarrow r} \bigcup_h M(h) = R_{na}(H_1).$$

**Следствие.** Переопределения атрибутов механизма целеполагания необходимо и достаточно для моделирования любой из пяти основных логических форм с помощью операций ИБД [2].

Таким образом, возможность переопределения атрибутов в наборе правил, формируемом экспертом, увеличивает гибкость, но порождает дополнительную проблему: при компиляции могут быть неизвестны структура и схема вывода набора правил, хотя его класс известен. Для этого в ЯПЗ ИБД применяется интерпретационный режим с уточнением семантики набора правил. Если внешняя семантика операций ИБД на данном этапе прототипирования определяется однозначно, то отображение выполнения операции проекции на внутренней структуре метатеорем пока проблематично. Поэтому предлагается отдельно определять эпитеорему, а затем метатеоремы с данной схемой. В настоящее время ведутся активные теоретические поиски подходов к организации языка запросов к ИБД, но практически реализованного языка запросов пока нет.

## ЛИТЕРАТУРА

1. Новожилова Т.Н. Особенности функциональной схемы обобщения процесса капиталообразования в интеллектуальной базе данных // Системи обробки інформації. – Харків: НАНУ, ПАНМ, ХВУ. – 2000 – Вип.1(7). – С. 8 - 15.
2. Новожилова Т.Н. Анализ эффективности логических форм в интеллектуальной базе данных // Системи обробки інформації. – ХФВ “Транспорт України”, 2001. – Вип.1(11). – С. 22 – 28.
3. Новожилова Т.Н. Классификация объектов в интеллектуальной базе данных // Системи обробки інформації. – ХФВ “Транспорт України”, 2001. – Вип. 3(13). – С. 141 - 145.

*Поступила в редколлегию 06.06.2001*