

КЛАССИФИКАЦИЯ ОБЪЕКТОВ В ИНТЕЛЛЕКТУАЛЬНОЙ БАЗЕ ДАННЫХ

к.т.н. Т.Н. Новожилова
(представил д.т.н., проф. Г.Г. Асеев)

Предложены идеи классификации наборов правил для автоматизации процесса приобретения знаний в интеллектуальной базе данных (ИБД).

Задачи анализа экспертных знаний в среде приобретения знаний предполагают построение алгебры классов объектов в пределах базового набора концепций объектно-ориентированного проектирования (ООП). Поскольку набор основных понятий ООП не содержит подходящего понятия для представления результата запроса, то обычно расширяют базовый набор понятий концепцией множества объектов [1]. В этом случае результатом запроса будет подмножество объектов-экземпляров класса, что автоматически исключает возможность наличия в языке запросов средств, аналогичных реляционному оператору соединения. Продукционные языки представления знаний (ЯПЗ) также не гарантируют классификации каждого объекта знаний эксперта. Существенное отличие ЯПЗ ИБД - поддержание алгебры классов наборов правил [2,3]. Объектами классификации в ИБД являются логические формы ЯПЗ, формализуемые для выбранной экспертом системы исчисления динамических ситуаций (**DS**). По заданной фактами **DS_i** экспертом выдвигается объясняющая их гипотеза и для этой гипотезы формируется утверждение

$$\mathbf{R}_{na}(\mathbf{h}) : \mathbf{DS}_i \xrightarrow{\Delta \mathbf{DS}_i} \mathbf{DS}_j. \quad (1)$$

Для доказательства (1) экспертом выбирается стандартный сценарий действий ИБД или предлагается новый. В любом случае сценарий оценивается в ИБД соответствующей алгеброй коэффициентов уверенности (**f(CF)**). Обработка метатеорем вывода включает интерпретацию сценария для реализации **R_{na}(h)** и преобразование кортежа наборов правил от цели к фактам, используя манипуляции с дедуктивной компонентой постреляционной оболочки [1], в которой частично поддерживается структурная часть ООП:

$$\left[\mathbf{R}_{na}(\mathbf{h}) \rightarrow \mathbf{R}_{na}(\mathbf{H}_1) \cdot \mathbf{RSS} \rightarrow \left\{ \mathbf{R}_{na} \left[\begin{array}{cc} \mathbf{h} & \mathbf{H}_2 \\ \mathbf{X} & \mathbf{O} \end{array} \right] \rightarrow \mathbf{R}_{na}(\mathbf{H}_2) \right\} \rightarrow \mathbf{R}(\mathbf{H}'_0) \right], \quad (2)$$

где **H₁**, **H₂**, **H'₀** - классы наборов правил; **R_{na}**, **R** - типы наборов правил; **na** - индекс типа. При обработке (2) выявляются ошибочные гипотезы в механизме целеполагания **H₁**. Утверждению (1) эксперта присваивается неопределен-

ный класс \mathbf{H}_2 и формируется $\mathbf{R}_{na}(\mathbf{H}_2, r_1, \dots, r_n; h_1, \dots, h_n)$, где перечисление предположений можно ограничить до уровня объектов. Системы классификации и типизации объектов ИБД позволяют для заданного отображения $\mathbf{DS}_i \xrightarrow{\Delta \mathbf{DS}_i} \mathbf{DS}_j$ получать несколько $\mathbf{R}_{na}(\mathbf{h})$ в одном кортеже. Средства связи с семантикой позволяют установить однозначное соответствие между опытом эксперта и его представлением в ЯПЗ ИБД с помощью приписывания свойств каждому элементу набора правил и использования причинных отношений как логических правил. Таким образом, цель обработки (1) - получение схемы вывода $\mathbf{f}(\mathbf{R}_{na})$, объясняющей целевую \mathbf{DS}_j , задаваемую как

$$\mathbf{DS}_j = \langle \mathbf{h}, \mathbf{DS}_i \oplus \Delta \mathbf{DS}_i, \mathbf{IS} + \Delta \mathbf{IS}, \mathbf{R}_{na} \oplus \Delta \mathbf{R}_{na} \rangle, \quad (3)$$

где $\Delta \mathbf{R}_{na}(\mathbf{H}_1, r_1, \dots, r_n; h_1, \dots, h_n; \mathbf{q}(\mathbf{h})_1, \dots, \mathbf{q}(\mathbf{h})_n)$ - общее выражение изменения вида операций вносимых в \mathbf{R}_{na} : (r_1, \dots, r_n) - операторы над \mathbf{H}_1 ; (h_1, \dots, h_n) - элементы \mathbf{H}_1 ; $\mathbf{q}(\mathbf{h})$ - переход.

Определение 1. Основой для формулирования запроса в ИБД служит кортеж наборов правил

$$\mathbf{R}_{na}(\mathbf{H}_2) \rightarrow \mathbf{R}_{na}(\mathbf{H}'_0) \rightarrow \mathbf{R}_{na} \begin{bmatrix} \mathbf{h} & \mathbf{H}_1 \\ \mathbf{X} & \mathbf{O} \end{bmatrix} \rightarrow \Delta \mathbf{R}_{na}, \quad (4)$$

который для $\forall \mathbf{DS}_i$ определяет $\mathbf{R}_{na}(\mathbf{H}_1)$ и метод представления кортежа - $\mathbf{f}(\mathbf{R}_{na})$. Таким образом, ЯПЗ ИБД характеризуется взаимопроникновением наборов правил и процедур. Запросу к набору правил соответствует операция паросочетания. В качестве операции отбора правил рассматривается $\mathbf{f}(\Delta \mathbf{R}_{na})$ - некоторое расширение принципа преобразования наборов правил [1].

Определение 2. Если для вывода \mathbf{H}'_{0j} по $\mathbf{R}_{na}(\mathbf{h})$ из \mathbf{H}'_{0i} заданы: механизм формализации и метод получения $\mathbf{R}_{na}(\mathbf{h})$, \mathbf{H}_1 и $\mathbf{f}(\mathbf{R}_{na})$, то отображение $\mathbf{H}'_{0j} \xrightarrow{\mathbf{R}_{na}(\mathbf{h})} \mathbf{H}'_{0i}$ можно рассматривать как подобие по $\mathbf{R}_{na}(\mathbf{h})$:

$$\mathbf{f}(\mathbf{R}_{na}(\mathbf{h})) : \mathbf{H}'_{0i} \xrightarrow{\mathbf{H}_1} \mathbf{H}'_{0j}. \quad (5)$$

Определение 3. Если $\mathbf{R}_{na}(\mathbf{H}_2)$ унифицируем с одним из сомножителей $\mathbf{R}_{na}(\mathbf{h})$, представленного конъюнкцией гипотез и аксиом, то $\mathbf{R}_{na}(\mathbf{H}_2)$ связывается с $\mathbf{R}_{na}(\mathbf{h})$ операцией паросочетания.

Используя схожесть между системами аксиом \mathbf{DS} , можно преобразовать $\mathbf{R}_{na}(\mathbf{h})$ одной из систем в кортеж наборов правил для другой системы аксиом и сделать вывод о том, что этот набор правил есть сценарий действий.

Определение 4. Два набора правил $\mathbf{R}_{na}(\mathbf{H}'_{0j})$ и $\mathbf{R}_{na}(\mathbf{h})$, где $\mathbf{H}'_{0j} \in \mathbf{DS}_j$ паросочетательны тогда и только тогда, когда $\exists \Delta \mathbf{R}_{na}$ такой, что при подстановке значения индикатора устойчивости $\mathbf{IS}_j = \mathbf{IS}_i + \Delta \mathbf{IS}_i$ в \mathbf{DS}_j :

$$\lim_{\Delta \mathbf{R}_{na} \rightarrow r} (\mathbf{R}_{na}(\mathbf{h}) \oplus \Delta \mathbf{R}_{na}) = \mathbf{R}_{na}(\mathbf{H}'_{0j}). \quad (6)$$

Определение 5. $R_{na}(H'_0)$ противоречив, если в составе $R_{na}(h)$, $R_{na}(H_1)$ и $R_{na}(H_2)$ отсутствуют правила, соответствующие обозначениям $R_{na}(h)$ или возникают несоответствие условий.

Пусть экспертом предложена $h \in M(H_1, H'_0, H_2)$. Если $\exists H_1$, для которого $\exists H_1 \rightarrow H_2$, определяемое схемой связывания $f(R_{ni})$, где каждое значение аргумента типизировано согласно схеме обобщения $F_{ni}(DS_j)$ [2]:

$$F_{ni}(DS_j) = \lim_{\Delta R_{na} \rightarrow r} \sum f(R_{ni}), \quad (7)$$

$$\text{где } f(R_{ni}) = \lim_{h \rightarrow \Delta R_{na}} (f(R_{na}(h)) : H'_{0i} \xrightarrow{H_1} H'_{0j}). \quad (8)$$

Если H_1 подтверждается в H_2 , то

$$M(h) : H_1 \oplus H_2 \xrightarrow{h} H'_0, \quad (9)$$

отображение (8) назовём обобщением по h , а $M(h)$ – классом обобщения.

Классическое определение отношения семантического следования также непригодно для формализации модифицируемых рассуждений, поскольку с чисто синтаксической точки зрения построение немонотонной системы вывода ИБД вызывает необходимость ослабления дедуктивных свойств оболочки [1] и введения понятия «устойчивое множество рассуждений».

Определение 6. Последовательность $R_{na}(H_2)_1, R_{na}(H_2)_2, \dots, R_{na}(H_2)_n$ сходится по h к устойчивому множеству утверждений, если

$$\exists \lim_{h \rightarrow r} f[CF(R_{na}(H_2, h))] > E.UNKN, \quad (10)$$

где $E.UNKN$ – переменная среды, определяющая порог неизвестного.

В немонотонной логике ИБД в сценарии включаются наборы правил, присутствующие во всех выводимых устойчивых множествах рассуждений, (устанавливающих существование для данного набора правил устойчивого выполнимого множества предположений (10)) включающие рассуждения:

$$1) \text{ модифицируемые из-за неопределённости, для которых } f(CF(H_2, h)) \leq E.UNKN; \quad (11)$$

2) модифицируемые экспертом из-за неприемлемого сценария перехода.

Определение 7. Пусть $M_k(h) \in M(H_1, H'_0, H_2)$ ($k=1,2$) два класса обобщения. Рассмотрим два механизма целеполагания

$$R_{na}(H_{11}) \in R_{TX}(H_1), R_{na}(H_{12}) \in R_{TX}(H_1), \quad (12)$$

где $R_{TX}(H_1)$ – общий набор метаправил [3], и две аргументации:

$$M_1(h_1) : R_{na}(H_{12}) \xrightarrow{h_1} R_{na}(H'_0); \quad (13)$$

$$M_2(h_2) : R_{na}(H_{12}) \xrightarrow{h_2} R_{na}(H'_0). \quad (14)$$

Если для $M_k(h) = M_1(h_1) \oplus M_2(h_2)$, $\exists r : X'O \xrightarrow{h} XO$, $r \in R_{na}(H_1)$ и

$$CF(M_k(h)) \geq CF(M_1(h_1), M_2(h_2)), \quad (15)$$

то

$$\lim_{\mathbf{h} \rightarrow \mathbf{r}} [\mathbf{R}_{\text{на}}(\mathbf{h}) : \text{DS}_i \xrightarrow{\mathbf{M}_k(\mathbf{h})} \text{DS}_j] = \mathbf{R}_{\text{ТХ}}(\mathbf{H}_1). \quad (16)$$

Определение 8. Последовательность $\mathbf{R}_{\text{на}}(\mathbf{H}_2)_1, \mathbf{R}_{\text{на}}(\mathbf{H}_2)_2, \dots, \mathbf{R}_{\text{на}}(\mathbf{H}_2)_n$ сходится к $\mathbf{R}_{\text{на}}(\mathbf{H}'_0)$ по $\mathbf{R}_{\text{на}}(\mathbf{h})$, если $\exists \mathbf{h} \in \mathbf{H}_1$, для которой

$$\mathbf{f}(\text{CF}(\mathbf{H}'_0, \mathbf{H}_2)) \geq \mathbf{f}(\text{CF}(\mathbf{R}_{\text{на}}(\mathbf{h}))) > \mathbf{E.UNKN}, \quad (17)$$

выражение (15) является критерием сходимости для $\mathbf{R}_{\text{на}}(\mathbf{H}_2)_1, \dots, \mathbf{R}_{\text{на}}(\mathbf{H}_2)_n$.

Определение 9. Последовательность $\mathbf{R}_{\text{на}}(\mathbf{H}_2)_1, \mathbf{R}_{\text{на}}(\mathbf{H}_2)_2, \dots, \mathbf{R}_{\text{на}}(\mathbf{H}_2)_n$ сходится к $\mathbf{R}_{\text{на}}(\mathbf{H}'_0)$ по схеме вывода, если $\exists \mathbf{f}(\mathbf{R}_{\text{на}})$ такая, что

$$\lim_{\Delta \mathbf{R}_{\text{на}} \rightarrow \mathbf{r}} \left[\mathbf{R}_{\text{на}}(\mathbf{H}_2)_i \xrightarrow{\Delta \mathbf{R}_{\text{на}}} \mathbf{R}_{\text{на}}(\mathbf{H}'_0) \right] = \mathbf{f}(\mathbf{R}_{\text{на}}). \quad (18)$$

Теорема определения класса. Чтобы существовало отображение

$$\mathbf{R}_{\text{на}}(\mathbf{H}_2)_i \xrightarrow{\Delta \mathbf{R}_{\text{на}}} \mathbf{R}_{\text{на}}(\mathbf{H}'_0) \quad (19)$$

для наборов правил неопределённого класса $\mathbf{R}_{\text{на}}(\mathbf{H}_2)_i$, необходимо и достаточно, чтобы $\exists \mathbf{h} \in \mathbf{H}_1$, для которой

$$\mathbf{f}(\text{CF}(\mathbf{H}_2, \mathbf{h})) \geq \mathbf{f}(\text{CF}(\mathbf{H}_1)) > \mathbf{E.UNKN}. \quad (20)$$

Доказательство. Чтобы существовало отображение (19) по определению 6, последовательность $\mathbf{R}_{\text{на}}(\mathbf{H}_2)_1, \dots, \mathbf{R}_{\text{на}}(\mathbf{H}_2)_n$ должна сходиться к устойчивому множеству рассуждений. Следовательно, необходимо, чтобы $\exists \mathbf{h} \in \mathbf{H}_1$ для которого $\exists \lim_{\mathbf{h} \rightarrow \mathbf{r}} \mathbf{f}[\text{CF}(\mathbf{R}_{\text{на}}(\mathbf{H}_2, \mathbf{h}))] > \mathbf{E.UNKN}$. (21)

Соблюдение условия (21) вызывает необходимость ограничения

$$\mathbf{f}(\text{CF}(\mathbf{H}_2, \mathbf{h})) > \mathbf{E.UNKN}. \quad (22)$$

Чтобы существовало отображение (19), последовательность $\mathbf{R}_{\text{на}}(\mathbf{H}_2)_1, \dots, \mathbf{R}_{\text{на}}(\mathbf{H}_2)_n$ должна сходиться и по схеме вывода должен существовать

$$\lim_{\mathbf{h} \rightarrow \mathbf{r}} \mathbf{f}[\text{CF}(\mathbf{R}_{\text{на}}(\mathbf{H}_1))] > \mathbf{E.UNKN}. \quad (23)$$

Соблюдение условия (23) вызывает необходимость ограничения

$$\mathbf{f}(\text{CF}(\mathbf{H}_1)) > \mathbf{E.UNKN}, \quad (24)$$

следовательно, необходимость условия (20) доказана.

Если последовательность $\mathbf{R}_{\text{на}}(\mathbf{H}_2)_1, \dots, \mathbf{R}_{\text{на}}(\mathbf{H}_2)_n$ сходится по \mathbf{h} и по $\mathbf{f}(\mathbf{R}_{\text{на}})$, то она должна сходиться по $\mathbf{R}_{\text{на}}(\mathbf{h})$:

$$\mathbf{f}(\text{CF}(\mathbf{R}_{\text{на}}(\mathbf{h}))) > \mathbf{E.UNKN}. \quad (25)$$

Чтобы существовало отображение (19), достаточно с помощью $\mathbf{f}(\mathbf{R}_{\text{на}})$, используя преобразование (3), произвести частичный отбор наборов правил:

$$\left[R_{na}(\mathbf{h}) \rightarrow R_{na}(\mathbf{H}_1) \cdot RSS \rightarrow R_{na} \begin{bmatrix} \mathbf{h} & \mathbf{H}'_0 \\ \mathbf{X} & \mathbf{O} \end{bmatrix} \rightarrow R_{na}(\mathbf{H}_2) \right]. \quad (26)$$

Для существования (26) достаточно справедливости (20). Теорема доказана.

Следствие 1. Пусть сформированы $\{R_{na}(\mathbf{H}_{2i}, \mathbf{h})\}$, для которых $\sum_{i \in \mathbf{N}} f[\mathbf{CF}(R_{na}(\mathbf{H}_{2i}, \mathbf{h}))] = 100$. Тогда согласно теореме определения класса существует хотя бы один механизм целеполагания \mathbf{H}_1 такой, что

$$f(\mathbf{CF}(\mathbf{H}_1)) > E.UNKN. \quad (27)$$

Таким образом, $f(R_{na})$ определяют допустимые паросочетания в кортежах наборов правил и позволяют рассматривать исчисление \mathbf{DS} как дерево состояний, а $f(\Delta R_{na})$ предоставляют возможность объявления дополнительных атрибутов и методов для индексируемых объектов: конкретный индексируемый объект-представитель класса может обладать типом, являющимся подтипом типа класса. Поскольку с такими атрибутами не работают стандартные методы класса [1], для индексируемого объекта в ЯПЗ ИБД определяются $f(R_{ni})$, в которых дополнительные атрибуты уже доступны. Для реализации исключительных атрибутов и методов требуется развитие в ЯПЗ ИБД техники позднего связывания (для неопределённых схем вывода), поддерживаемой $f(R_{ni})$, которые включают средства конструирования значений-кортежей, множеств и списков. При этом элементами значений-множеств могут являться объекты (простая выборка), либо значения-кортежи с элементами-объектами разных классов. В совокупности эти особенности ЯПЗ ИБД позволяют формулировать запросы над несколькими классами наборов правил (специфическое соединение, порождающее не новые объекты, а кортежи из существующих объектов), а также применять вложенные подзапросы.

ЛИТЕРАТУРА

1. Экспертные системы (ИНТЕР-ЭКСПЕРТ, версия 2). Руководство пользователя. Ч. 3. – Калинин : НПО “ЦЕНТРПРОГРАММСИСТЕМ”, 1992. – 156 с.
2. Новожилова Т.Н. Анализ техники мышления в интеллектуальной базе данных // Системи обробки інформації. – Харків: НАНУ, ПАНМ, ХВУ. – 2000. – Вип.4(10). – С. 107 - 114.
3. 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.

Поступила в редколлегию 19.03.2001