

АНАЛИЗ ЭФФЕКТИВНОСТИ ЛОГИЧЕСКИХ ФОРМ В ИНТЕЛЛЕКТУАЛЬНОЙ БАЗЕ ДАННЫХ

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

В работе рассматриваются способы представления и анализа субъективной надежности логических форм в интеллектуальной базе данных (ИБД).

Поскольку оболочка является системой, обрабатывающей суждения, то можно говорить, что все её команды сравнивают одну величину с другой и, в зависимости от результатов сравнения, осуществляют определённые действия, а в результате действия имеется суждение или оценивание. Однако, суждение не последняя логическая форма представления знаний. В ИБД обрабатываются несколько логических форм: *определения, аксиомы, суждения, умозаключения, метатеоремы, эпитеоремы*. Переход от одной логической формы к другой – *действие*. Совокупность действий в ИБД представлена операцией. *Операция* – элемент схемы интегрированного вывода. Понятию в ИБД соответствует определение в режиме естественного языка СНАТ, операции над понятиями – команды генератора [1]. Аксиомы DS_i содержат соотношения экономических показателей, которые представляются и обрабатываются генератором в режиме CALC [1]. Суждение «*неопределённое*» определения: определить что-нибудь – значит, исчерпать все его признаки, для суждения этого не требуется. Суждения в ИБД создаются группой команд BUILD, а для представления субъективной надежности суждений (которую не способна выразить байесовская вероятность) в ИБД используются алгебры коэффициентов уверенности $f(CF)$: связи КОМБ обрабатываются интегрированным выводом, связи И/ИЛИ – генератором. $CF[X-O]_i$ определяются с помощью базовой $CF(X_i)$, которая “закрывается” в подмножестве X_i (рис. 1).

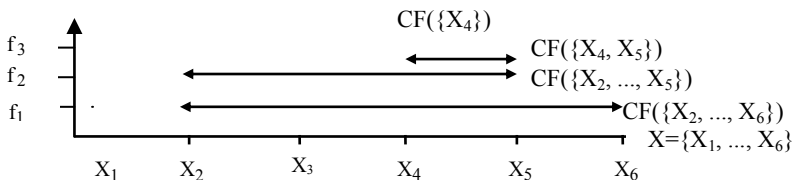


Рис. 1. Образ базовой уверенности $CF(X_i)$

Пусть X – посылка $R_{na}(H_2)$, отражающая группировку фактов по DS_i , а X_i ($i=1,2,\dots$) её подмножества, тогда базовый коэффициент уверенности при значениях переменных $E.CFJO$, $E.CFCO$, $E.CFVA$ по умол-

чанию, как в стандартных оболочках, определяется следующим образом:

LET [X_i-O]=TRUE CF(X_i),

а пересчитывается с помощью функции

$$f(\text{CF}) = \begin{cases} \text{E.UNKN}, & 0 \leq \text{CF}(X_i) \leq 20; \\ \text{E.CFJO}, & 20 < \text{CF}(X_i) \leq 100. \end{cases} \quad (1)$$

Нижнюю границу CF можно определить следующим образом:

$$\text{CF}_*(X_i) = \sum_{X_j \subseteq X_i} \text{CF}(X_j). \quad (2)$$

С другой стороны, верхняя граница CF определяется как

$$\text{CF}^*(X_i) = 100 - \text{CF}_*(\overline{X_i}) = 100 - \sum_{X_j \subseteq \overline{X_i}} \text{CF}(X_j). \quad (3)$$

Правило комбинации CF. Пусть CF₁ и CF₂ базовые для гипотезы, предложенной экспертом, тогда новый базовый CF можно представить как

$$\text{CF}(X_k) = \frac{\sum_{X_{1i} \cap X_{2j} = X_k} \text{CF}_1(X_{1i}) * \text{CF}_2(X_{2j})}{100 - \sum_{X_{1i} \cap X_{2j} = \emptyset} \text{CF}_1(X_{1i}) * \text{CF}_2(X_{2j})} (X_k \neq \emptyset). \quad (4)$$

Числитель в этой формуле означает, что произведение соответствующих базовых CF распределяется на пересечение X_{1i} и X_{2j}, как указано на рис. 2, а знаменатель нормирует это произведение,

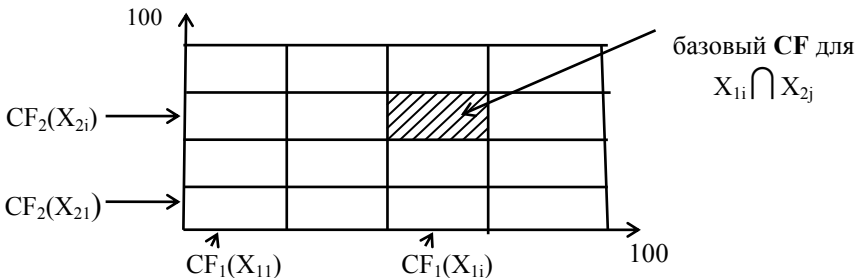


Рис.2. Базовая уверенность

Логические формы умозаключения, метатеоремы и эпитеоремы предполагают обоснование своей точки зрения (метатеоремы и эпитеоремы требуют доказательства). Умозаключения, метатеоремы и эпитеоремы на этапе прототипирования ИБД описываются в режиме BUILD и обрабатываются с помощью интерфейса CLINC библиотекой функций на C⁺⁺ [3]. При этом метатеоремы и эпитеоремы рассматриваются и как операции над низшими логическими формами и как логические формы высокого уровня, обрабатываемые интегрированным выводом. Модель эксперта в ИБД представлена промежу-

точными аналитическими таблицами (генератор ИНТЕР-ЭКСПЕРТ позволяет использовать одновременно 15 таблиц в MS DOS и 20 таблиц в UNIX). В логиках генератора ИНТЕР-ЭКСПЕРТ невыразимы суждения, использующие многозначные отображения, играющие центральную роль в индуктивном обобщении умозаключений. Противоречия в наборе правил возникают не из самих «противоречивых понятий», а из свойств, которые им приписываются. Значительная часть информации, необходимой для оценки рациональности логики эксперта не может быть выражена в генераторе ИНТЕР-ЭКСПЕРТ по двум причинам:

- 1) неявное представление стратегии рассуждения;
- 2) отсутствие в оболочке средств представления и обработки метазависимостей, позволяющих обосновать ход рассуждений экспертов с точки зрения фундаментальных положений процесса капиталообразования.

Поскольку в ИБД конкретная процедура установления связи $X \Rightarrow_{\mathbf{h}} \mathbf{O}$ или

её опровержения системой словаря справочника базы метазависимостей (СССБМЗ) зависит от последовательности использования наборов правил интегрированным выводом, каждая такая последовательность образует некоторую стратегию вывода. Стратегия вывода ИБД описывается как способ обхода узлов И/ИЛИ-графов. Метаправила ИБД подчиняются схеме:

правило : : = (IF <предпосылка> THEN <действие> (ELSE <действие>))

предпосылка : : = (AND {условие }⁺)

условие : : = (OR {<условие >}⁺) / (<предикат> <INT > <вектор CF >)

< INT > : : = (<объект> <атрибут> <значение - CF >)

действие : : = { <результат >}⁺

результат : : = (<элемент > < f(CF) >) .

Определение 1. Любое непротиворечивое совместное использование эвристик $\mathbf{H}_1 \in \mathbf{ЭБД}$ имеет $\mathbf{R}_{\text{на}}(\mathbf{H}_2)$, где ЭБД – база данных экспериментов [2].

Определение 2. $\mathbf{R}_{\text{на}}(\mathbf{H}_2)$ полон, если $\mathbf{H}_1(\mathbf{DS}_i) \cap \mathbf{H}_1(\mathbf{DS}_j) \neq \emptyset$ и логический вывод каждого элемента правил может быть достигнут.

Определение 3. $\mathbf{R}_{\text{на}}(\mathbf{H}_2)$ корректен, если ни одна продукция в $\mathbf{R}_{\text{на}}(\mathbf{H}_1)$ не содержит прямой рекурсии.

Определение 4. $\mathbf{R}_{\text{на}}(\mathbf{H}_2)$ противоречив, если в нем имеются такие продукции, в левых частях которых одинаковые условия порождают разные действия в правых частях.

Проверка названных требований в ИБД автоматизирована, если использовать не более 6 условий в левой части правил.

Определение 5. Отображение $\mathbf{H}_2 : \mathbf{R}_{\text{на}}(\mathbf{H}_1) \cup \mathbf{R}_{\text{на}}(\mathbf{H}'_0) \xrightarrow{\mathbf{M}(\mathbf{h})} \mathbf{R}_{\text{на}}(\mathbf{H}'_2)$ является моделью интерпретации, если для связей И, ИЛИ выполняются условия:

$$(R_{na}(H_1) \cup R_{na}(H'_0)) \subseteq R_{na}(H'_2); \quad (5)$$

$$\text{если } h \in R_{na}(H_2), \text{ то } CF^*(H'_2(h)) = CF(h); \quad (6)$$

$$\text{если } [X-O] \in R_{na}(H'_0), \text{ то } CF^*(H'_2[X-O]) \xrightarrow{f(CF)} CF(h) \pm CF(X-O); \quad (7)$$

$$\text{если } \forall h_1, \dots, \forall h_i \forall M(h_1), \dots, M(h_i), \text{ то } CF(h, M(h)) \xrightarrow{f(CF)} CF(M(h), h), \quad (8)$$

где $M(h)$ – многозначная зависимость.

Определение 6. Схема F_{in} применима для $R_{na}(H_2)$ только тогда, когда $\forall f_i(R_{na}(H_{1i}))$ такая, что $\forall H_{1i} \in \text{ЭБД}$, из $\exists \sum_1^n R_{na}(H_{1i}) \subseteq f(R_{na}(H_1))$

$$\text{следует } \exists \sum_1^n R_{na}(H_{2i}) \subseteq f(\Delta R_{na}(H_1)). \quad (9)$$

При этом формирование аналитических таблиц в ИБД требует:

- определить примитивные операции, рассматриваемые как единицы иерархии схем интегрированного вывода, которые могут быть объяснены (в качестве такой примитивной операции предложено рассматривать вызов набора правил);

- добавить в интегрированный вывод средства сохранения цепочки вызывавшихся наборов правил, т.е. истории выполнения;

- представление структуры управления вызовом правил в виде И/ИЛИ дерева целей (для понимания истории выполнения);

- для исследования предыдущего, текущего и потенциального состояния реорганизовать функциональную схему генератора ИНТЕР-ЭКСПЕРТ.

Представление в ИБД возможных стратегий рассуждения экспертов метатеоремами позволяет через стратегии выбора определять функциональную ориентацию интерпретации цепочки вызывавшихся наборов правил. Интерпретатор выстраивает общую схему поиска действий, а стратегии управляют деталями поиска. При этом $f(CF)$ является средством проверки эффективности наборов правил сформулированных экспертом. Поэтому вместо проблемы поиска у когнитолога часто возникает новая проблема: до какого уровня абстрагировать метатеоремы.

Теорема 1. Если H_1 – механизм целеполагания, а h – суждение, которое по определенным правилам вывода можно вывести из H_1 , то $\forall DS_i \exists INT(M(h) \in R_{na}(H'_1))$ и $H_1|_h$ определяется с помощью стратегии абстрагирования

$$H'_0 = f \{ [X-O]_h \cup R_{na}(H_1) \}, h \in R_{na}(H_1). \quad (10)$$

Доказательство. Пусть CF_{X_i} – CF для посылки X_i , CF_{h_i} – CF, присвоенный переменной IS в действии правила $X \rightarrow O$ при посылке X_i .

Предположим, что получен CF_{X_j} для посылки X_j , $CF_{h_j} - CF$, присвоенный переменной IS в действии правила $X \rightarrow O$ при посылке X_j . Тогда $CF_i(IS)$ и $CF_j(IS)$ определим подтверждающими формулами в два этапа:

1) в пределах $R_{na}(H_1)$: как $f(CF_{X_i}, CF_{h_i})$, $f(CF_{X_j}, CF_{h_j})$ соответственно:

$$\text{MIN}(CF_{X_i}, CF_{h_i}), \text{MIN}(CF_{X_j}, CF_{h_j}); \quad (11)$$

$$(CF_{X_i} * CF_{h_i})/100, (CF_{X_j} * CF_{h_j})/100; \quad (12)$$

$$(\text{MIN}(CF_{X_i}, CF_{h_i}) + (CF_{X_i} * CF_{h_i})/100)/2 \text{ и} \\ (\text{MIN}(CF_{X_j}, CF_{h_j}) + (CF_{X_j} * CF_{h_j})/100)/2; \quad (13)$$

$$((CF_{X_i} * CF_{h_i})/100) * (2 - \text{MAX}(CF_{X_i}, CF_{h_i})/100) \text{ и} \\ ((CF_{X_j} * CF_{h_j})/100) * (2 - \text{MAX}((CF_{X_j}, CF_{h_j})/100)); \quad (14)$$

$CF_i(IS)$ по формулам (13) и (14) всегда дают значения между экстремальными, полученными по подтверждающим формулам (11) и (12);

2) по $f_i [X-O]$ (этот этап пропускается, если было включено хотя бы одно правило, дающее в результате определенное значение IS):

$$\text{MAX}(CF_{X_i}, CF_{h_i}), \text{MAX}(CF_{X_j}, CF_{h_j}); \quad (15)$$

$$(CF_{X_i} + CF_{h_i}) - (CF_{X_i} * CF_{h_i})/100, (CF_{X_j} + CF_{h_j}) - (CF_{X_j} * CF_{h_j})/100; \quad (16)$$

$$\text{MAX}(CF_{X_i}, CF_{h_i}) + ((CF_{X_i} + CF_{h_i}) - (CF_{X_i} * CF_{h_i})/100)/2; \quad (17)$$

$$\text{MAX}(CF_{X_j}, CF_{h_j}) + \text{MIN}(CF_{X_j}, CF_{h_j}) * (1 - CF_{X_j}/100) * (1 - CF_{h_j}/100). \quad (18)$$

Используя механизм трассировки, выявим отличия в действиях схем вывода. Рассмотрение правил приостанавливается, если $\exists CF_{X_i} > E.UNKN$, но если $\forall CF_{X_i} < E.UNKN$, то $f(CF_{X_j}) = E.CFVA$ и для соотношений (11), (12), (13), (14), (15) абстрагируем H_1 и h :

$$H'_1 = \{X \cup O \cup R_{na}(M(h))\}. \quad (19)$$

После этого анализируем $CF_i(IS)$ и $CF_j(IS)$ по формулам:

$$CF_i(IS) + CF_j(IS) - CF_i(IS) * CF_j(IS)/100; \quad (20)$$

$$\text{MAX}(CF_i(IS), CF_j(IS)), \quad (21)$$

$$((CF_i(IS) + CF_j(IS) - CF_i(IS) * CF_j(IS)/100) + \text{MAX}(CF_i(IS), CF_j(IS))), \quad (22)$$

$$\text{MAX}(CF_i(IS), CF_j(IS)) + \text{MIN}(CF_i(IS), CF_j(IS)) * \\ * (1 - CF_i(IS)/100) * (1 - CF_j(IS)/100). \quad (23)$$

В случае получения из независимых доказательств более двух базовых CF_{X_i} условный коэффициент неопределенности можно представить путем последовательного применения формулы (4). В итоге схема вывода определяет CF для $H'_0 = f \{ [X-O]_n \cup R_{na}(H_1) \}$. Построение доказательства обрывается, если эксперт не может указать, какая из альтернатив верна. Остаётся определить способ вывода $H_1 | - h$ из $H'_1 | - M(h)$, при котором сохраняется каждый вывод, обусловленный (10). Если начать с $X \cup O \cup R_{na} | - O \cup R_{na}$ и проверить вывод $X(h) \cup O(M(h)) \cup R_{na}(H_1)$, то получим

$$X(h) \cup O(M(h)) \cup R_{na}(H_1) | - O(M(h)) \cup R_{na}(H'_1), \quad (24)$$

где $O \cup R_{na}$ - пропозициональное абстрагирование выводимого суждения $O(M(h)) \cup R_{na}(H'_1)$, поэтому $O \cup R_{na}, O | - R_{na}$. Таким образом, отправляясь от дерева вывода $H'_1 | - M(h)$ можно получить дерево вывода $H_1 | - h$ и

$$X(h) \cup R_{na}(H_1), O(M(h)) | - R_{na}(H'_0). \quad (25)$$

Формализм доказательства теоремы 1 может быть дополнен «семантическими» рассмотрениями. При этом $f(CF) = E.CFVA$ рассматривается как:

- метод соединения CF посылки с CF действия (первый символ);
- способ соединения CF наборов правил в CF результирующей модели.

Для каждого типа перехода СССБМЗ создаётся словарь ключевых слов WORDMAN [2]. Каждый элемент словаря снабжен уникальным идентификатором, что обеспечивает декомпозицию образа динамической ситуации DS_i на составляющие и выявление интервалов допустимости.

При формировании модели интерпретации базы данных с неполной информацией (БДНИ) для DS_i и модели перехода $DS_i \rightarrow DS_j$ строится И/ИЛИ дерево целей, и выполняются выбранные правила. Система интерпретируемых доменов может образовывать иерархию, если используется обратный логический вывод с механизмом возвратов. И/ИЛИ дерево целей строится на основе $[X_i - O_i]$, фактов БДНИ и целей в следующей последовательности:

- 1) определяются правила для достижения поставленной цели

$$R_{na}(H_2) : R_{na}(H_1) \xrightarrow{X-O} R_{na}(H'_0); \quad (26)$$

- 2) выбираются $R_{na}(H_2)$, левые части которых уже находятся в БДНИ;

- 3) заносятся правые части этих правил в БДНИ и исключаются из дальнейшего рассмотрения.

Поиск решения происходит как выбор пути следования по И/ИЛИ дереву вниз и слева - направо (в глубину). Для поддержки механизма прерываний, позволяющего переоценивать правила и изменять текущую цель, необходимо:

- 1) в *простом* условии $R_{na}(H_2)$ указывать метку, чтобы:
 - соответствие устанавливал эксперт;
 - описанное значение было достаточно полным;
 - был указатель специальных составляющих наборов правил [1];
- 2) в *сложном* условии $R_{na}(H_2)$ указывать отношения между $[X_i - O_i]$, соответствующими нескольким $R_{na}(H_1)$, что позволяет включать в процесс аргументации анализ техники мышления эксперта [3];
- 3) использовать значение по умолчанию в качестве переменной вместо логических ограничителей. Наличие в ИБД схем вывода позволяет протестировать влияние набора правил на значение переменной IS и общую стратегию абстрагирования (10). При этом каждому шагу вывода соответствует пересчет значений коэффициентов уверенности, обусловленный соотношением на множестве суждений: $f(CF)$. Пересчет значений коэффициентов уверенности как бы сопровождает процесс вывода, ведущийся на точных высказываниях. Соединив множество $[R_{na}(H_1)]_i$, получим дерево состояний для DS_i .

Таким образом, при уточнении схемы вывода гипотезы, рассмотренные способы представления и анализа субъективной надежности логических форм, позволяют учитывать в ИБД все релевантные этой гипотезе свидетельства (с последующим объединением значений CF) и предотвращают многократное влияние CF одних и тех же свидетельств.

ЛИТЕРАТУРА

1. Экспертные системы (ИНТЕР-ЭКСПЕРТ, версия 2). Часть 2. – Калинин : НПО “ЦЕНТРПРОГРАММСИСТЕМ”, 1992. – 144 с.
2. Новожилова Т.Н. Особенности функциональной схемы обобщения процесса капиталообразования в интеллектуальной базе данных // Системи обробки інформації. – Харків: НАНУ, ПАНМ, ХВУ. – 2000. – Вип.3(9). – С. 8 - 15.
3. Новожилова Т.Н. Анализ техники мышления в интеллектуальной базе данных // Системи обробки інформації. – Харків: НАНУ, ПАНМ, ХВУ. – 2000. – Вип.4(10). – С. 107 - 114.

Поступила в редколлегию 20.10.2000