

# Об одном подходе к объединению продукционных правил, полученных из различных источников знаний<sup>1</sup>

**Аннотация.** Обсуждается проблема объединения продукционных правил, полученных из источников знаний различной типологии (эксперты, проблемно-ориентированные тексты, электронные носители в виде баз данных), для построения полных и непротиворечивых баз знаний в интегрированных экспертных системах. Описываются особенности распределенного приобретения знаний в рамках задачно-ориентированной методологии построения интегрированных экспертных систем и созданной на её основе автоматизированной технологии, включающей инструментарий нового поколения – программный комплекс АТ-ТЕХНОЛОГИЯ. Рассматриваются вопросы применения теории мультимножеств для объединения продукционных правил, полученных из различных источников знаний.

**Ключевые слова:** интегрированные экспертные системы, ИЭС, распределенное приобретение знаний, база знаний, объединение наборов продукционных правил, продукционные правила, теория мультимножеств.

## Введение

Приобретение знаний является центральной проблемой, возникающей при разработке современных интеллектуальных систем, в частности, традиционных экспертных систем и более сложных – интегрированных экспертных систем (ИЭС), обладающих масштабируемой архитектурой и расширяемой функциональностью [1, 2]. Однако, несмотря на значительное число исследований и разработок в этой важнейшей области искусственного интеллекта, например [1, 3-6], вопросы практического использования традиционных методов приобретения знаний и создания технологии автоматизированного приобретения знаний по-прежнему являются актуальной проблемой.

Данная работа является продолжением обсуждения результатов, представленных в [7], поэтому напомним кратко, что в центре внимания авторов находится проблема приобретения знаний в контексте разработки и практического использования прикладных ИЭС, созданных на основе задачно-ориентированной методологии (ЗОМ) [1] и инструментального комплекса АТ-ТЕХНОЛОГИЯ.

Опыт разработки систем подобного класса сложности [1, 2, 7, 8], в которых существенно возрастают объемы противоречивых и «зашумленных» данных и знаний, выявил острую необходимость автоматизации труда экспертов, являющихся основным источником знаний, а также создания как специальных программных средств приобретения знаний из дополнительных источников знаний (текстов на естественных языках и электронных информационных ресурсов в виде баз данных), так и соответствующих средств интеграции знаний.

Однако в настоящее время, несмотря на бурное развитие таких технологий как Text Mining, Data Mining и Knowledge Discovery in Databases, практически отсутствуют исследования в области создания инструментальных средств и технологий распределенного приобретения знаний из источников различной типологии. Таким образом, возникает необходимость создания новой автоматизированной технологии приобретения и интеграции знаний, распределенных по различным источникам, в связи с чем в [7] авторами было введено понятие «распределенного приобретения знаний».

<sup>1</sup> Работа выполнена при финансовой поддержке РФФИ (проект 09-01-00638)

Следует отметить, что в соответствии с концептуальными основами ЗОМ построения ИЭС неотъемлемой частью данной методологии является ЗОМ приобретения знаний, представляющая собой совокупность комбинированного метода приобретения знаний (КМПЗ) и технологии его использования на различных стадиях жизненного цикла построения ИЭС и веб-ИЭС [1, 9]. В рамках базового КМПЗ и средств его реализации рассматривался так называемый *локальный вариант* приобретения знаний, но при переходе к веб-версии комплекса АТ-ТЕХНОЛОГИЯ реализован *распределенный вариант* автоматизированного приобретения знаний на основе КМПЗ, обеспечивающий в рамках клиент-серверной архитектуры интеграцию различных типов источников знаний, учет их географической распределенности, а также возможность работы с группами удаленных источников знаний [1, 7, 9].

В контексте КМПЗ, реализованного в рамках ЗОМ, источники знаний 1-го и 2-го типа [1, 5, 9] рассматриваются как *совмещенные*, поскольку в КМПЗ существует совокупность хорошо апробированных технологических процедур, позволяющих дополнять информацию, полученную от эксперта/экспертов, за счет информации, выявленной из проблемно-ориентированных ЕЯ-текстов (в данном случае – это обработка протоколов интервьюирования экспертов, сбор лексики инженера по знаниям / системного аналитика, анализ сигнальных лексем во входных ЕЯ-текстах и др.).

Поэтому основной акцент *распределенного* приобретения знаний смещается в сторону интеграции с информацией, полученной из БД как источника знаний 3-го типа с целью автоматизированного построения максимально полных и непротиворечивых моделей проблемной области (ПрО).

Данная работа посвящена теоретическим аспектам объединения (интеграции) знаний, полученных из источников различной типологии, в частности, рассматриваются вопросы применения теории множеств для объединения продукционных правил.

## 1. Общая характеристика распределенного варианта КМПЗ

Детальное описание базовой КМПЗ содержится в [1], поэтому ниже приводится в общем виде теоретико-множественная модель расширенного для источников знаний различной

типологии КМПЗ с учетом распределенного варианта приобретения знаний:

$$M_{км} = \langle N^-, S^-, F^-, K, Z \rangle,$$

где  $N^- = (M_{мртз}, I, T, E)$ ;

$M_{мртз}$  – совокупность моделей решения типовых задач (МРТЗ) диагностики, проектирования, планирования, обучения и управления;

$I = (I^{aem}, I^{epyn})$  – знания, полученные от экспертов путем применения МРТЗ, где  $I^{aem}$  – знания, полученные от автономного эксперта, а  $I^{epyn}$  – знания от групп экспертов;

$E$  – знания, приобретенные из БД;

$T = (Q, D, W)$  – знания, выявленные в результате лингвистической обработки ЕЯ-текстов, где  $Q$  – знания, получаемые в результате «понимания» ЕЯ-текстов, описывающих решаемую проблему;  $D = \{D_i\}$  – знания, приобретенные от эксперта/экспертов;  $W$  – лексические знания, получаемые на этапе анализа системных требований пользователя.

Компонент  $S^- = \{S^-_{m_j}\}$ ,  $m=1, \dots, m_m$  – есть множество *структурированных описаний ПрО* в виде *поля знаний* (ПЗ) [1], выполняющего важную функцию в процессе структурирования полученной от эксперта информации, обеспечивая единое внутреннее представление и унификацию основных понятий и отношений ПрО, выявленных из различных источников знаний.

Соответственно, с учетом особенностей распределенного приобретения знаний обобщенную модель ПЗ можно представить в виде:

$$S^-_m = \langle IS_m, SS_m, O_m, R_m \rangle,$$

где  $IS_m$  – порядковый номер структурированного описания проблемной области;

$SS_m = \langle T_{SS}, I_{SS} \rangle$  – источник, откуда получено описание, где  $T_{SS}$  – тип источника (эксперт, группа экспертов, БД, текст),  $I_{SS}$  – идентификатор источника знаний;

$O_m = \{O_{mj}\}$ ,  $j=1, \dots, n$  – множество объектов;

$R_m = \{R_{mk}\}$ ,  $k=1, \dots, p$  – множество правил.

Компонент  $K$  – множество процедур конвертации сформированного ПЗ в форматы языков представления знаний различных инструментальных средств (зарегистрированных в комплексе АТ-ТЕХНОЛОГИЯ), а  $Z$  – фрагменты БЗ в форматах языков представления знаний.

Компонент  $F^-$  определяет множество процедур внутреннего малого жизненного цикла формирования фрагментов базы знаний (БЗ) на основе ПЗ. Распределенный вариант приобре-

тения знаний на основе КМПЗ позволяет строить более *полные* и *непротиворечивые* БЗ за счет использования как географически удаленных источников знаний, так и наличия групп источников знаний различной типологии. При переходе от локального варианта приобретения знаний к распределенному, компонент  $F^-$  представляется в виде:

$$F^- = F_{БАЗ} \cup F_{РАСПР},$$

где  $F_{БАЗ}$  – множество базовых процедур;

$F_{РАСПР} = \{F_{P1}, \dots, F_{P4}\}$  – множество процедур, добавленных при переходе от локального варианта приобретения знаний к распределенному,

$F_{P1}$  – процедуры группового приобретения знаний;

$F_{P2}$  – процедуры построения хранилищ данных;

$F_{P3}$  – процедуры распределенного приобретения знаний из БД;

$F_{P4}$  – процедуры объединения фрагментов ПЗ, полученных из источников разной типологии.

Объединение фрагментов ПЗ, полученных из источников знаний различной типологии является одной из сложнейших задач при построении БЗ ИЭС.

Решение проблемы объединения и группировки, продукционных правил рассматривается в работах [15-18].

В [15] предлагается подход, основанный на иерархии атрибутов, которая формируется либо с помощью алгоритмов приобретения знаний, либо при помощи эксперта, далее эксперт строит срез по полученному дереву. В результате получаются группы правил, причем каждую группу определяет агрегированное правило, на основе которого эксперт может определить нужна данная группа или нет.

В [16, 17] предложены подходы, которые рассматривают ассоциативные правила, где могут присутствовать как числовые, так и строковые атрибуты в посылке правила и строковые атрибуты в заключении правила. Существует также подход, основанный на подсчете *меры схожести* между утверждениями правил [18].

В рамках КМПЗ для объединения фрагментов ПЗ, полученных из источников различной типологии реализован метод, основанный на применении теории мультимножеств и вычислении меры схожести посылок и заключений правил. Множество процедур объединения фрагментов ПЗ при этом представлено процедурой объединения типов атрибутов, процеду-

рой объединения объектов и непосредственно процедурой объединения продукционных правил [1, 7, 8]. Рассмотрим последовательно эти процессы.

## 2. Автоматизированное объединение объектов, полученных из различных источников знаний

Решение задачи объединения фрагментов ПЗ, полученных из источников знаний различной типологии, существенно затрудняется задачей *объединения объектов*, описанных в каждом фрагменте ПЗ. Данная проблема в рамках КМПЗ решается в два этапа: на первом этапе производится объединение типов атрибутов, на втором осуществляется объединение самих объектов.

В каждом фрагменте ПЗ описан свой набор типов, которые могут совпадать по именам, типам и списку возможных значений. Сопоставление типов атрибутов можно разбить на два шага.

*Шаг 1.* Производится сравнение имен атрибутов. В случае отсутствия прямого совпадения для каждого имени атрибута из словаря синонимов получается список синонимичных значений. Производится попарное сравнение. Вне зависимости от обнаружения совпадений имен типов осуществляется переход на шаг 2.

*Шаг 2.* Рассматриваются два варианта:

- Типы являются числовыми. В этом случае дальнейшие действия над типами определяются экспертом.

- Типы являются строковыми и имеют конечный набор значений. В этом случае дальнейшее сравнение производится по списку возможных значений типов. Обозначим через  $T_A$  не пустое множество возможных значений типа  $A$ , через  $T_B$  – не пустое множество возможных значений типа  $B$ . Тогда возможны следующие варианты:

- $T_A \cap T_B = \emptyset$ , в этом случае осуществляется сравнение возможных значений по словарю синонимов, после чего шаг 2 повторяется;

- $T_A \neq T_B$ ,  $T_A \cap T_B \neq \emptyset$ , т.е. возможные значения типов совпадают частично. Не совпадающие возможные значения также подвергаются сравнению по словарю синонимов;

- $T_A = T_B$ , т.е. возможные значения типов совпадают полностью. В таком случае при совпадении имен типов или их синонимов типы  $A$  и  $B$  автоматически приравниваются и объединяются;

–  $T_B \subset T_A$ , в таком случае при совпадении имен типов либо их синонимов возможно автоматическое замещение типа  $B$  типом  $A$ .

Для реализации второго этапа объединения объектов применяется метод, основанный на теории мультимножеств [10-12].

Рассмотрим два объекта  $A_m$  и  $A_k$ , полученных из источников  $I_1$  и  $I_2$  соответственно. Будем говорить, что объекты возможно объединить, либо заменить объект  $A_m$  объектом  $A_k$  в следующих случаях:  $Name_{O_{A_m}} = Name_{O_{A_k}}$ , т.е. совпадают имена объектов;  $Синоним(Name_{O_{A_m}}) = Name_{O_{A_k}}$ , т.е. синоним имени объекта  $A_m$  совпадает с именем объекта  $A_k$ ;  $Name_{O_{A_m}} = Синоним(Name_{O_{A_k}})$ , т.е. имя объекта  $A_m$  совпадает с синонимом имени объекта  $A_k$ ;  $Синоним(Name_{O_{A_m}}) = Синоним(Name_{O_{A_k}})$ , т.е. имеет место совпадение синонимов имен объектов.

Рассмотрим совокупность  $A = \{A_1, \dots, A_k\}$  объектов, которые описываются  $m$  атрибутами  $Q_1, \dots, Q_m$ , имеющими конечное число значений. В случае распределенного приобретения знаний каждый объект  $A_i$  получен из  $k$  источников и описывается  $m$  атрибутами. В таком случае при объединении всех фрагментов ПЗ объект  $A_i$  представляется в  $m$ -мерном пространстве  $Q$  уже не одной точкой  $q_i$ , а группой из  $k$  точек. Тогда удобно представить объект в виде мультимножества [10-12]: пусть количество фрагментов ПЗ равно  $k$ , суммарное количество всех уникальных атрибутов равно  $m$ ; вводится обобщенная шкала признаков – множество  $G = \{Q_1, \dots, Q_m\}$ , состоящая из  $m$  групп признаков, и объект  $A_i$  представляется в символическом виде:

$$A_i = \{k_{A_i}(q_1^1) \circ q_1^1, \dots, k_{A_i}(q_1^{h_1}) \circ q_1^{h_1}, \dots, k_{A_i}(q_m^1) \circ q_m^1, \dots, k_{A_i}(q_m^{h_m}) \circ q_m^{h_m}\},$$

где число  $k_{A_i}(q_s^{es})$  указывает сколько раз атрибут  $q_s^{es} \in Q_s$  встречается в описании объекта  $A_i$ , знак  $\circ$  обозначает кратность вхождения атрибута  $q_s^{es}$ . Например, при получении объекта  $A_i$  из нескольких источников знаний число  $k_{A_i}(q_s^{es})$  равно числу источников, в которых объекту  $A_i$  присвоен атрибут  $q_s^{es}$  по критерию  $Q_s$ .

Объекты группируются в мультимножества по именам либо синонимам имен. Объекты можно объединить либо заменить объект  $A_m$  объектом  $A_k$  в случае совпадения имен объектов, совпадения синонима имени объекта  $A_m$  с именем объекта  $A_k$ , а также в случае совпадения синонимов имен объектов.

В случае, если объединение объектов по словарю синонимов не может гарантировать приемлемых результатов необходимо воспользоваться подсчетом расстояния между мультимножествами. Между двумя объектами, представленными в виде мультимножеств, рассматривается метрическое пространство [10]:  $d_{1p}(A, B) = [m(A \Delta B)]^{1/p}$ , где  $p$  – целое число,  $m$  – мера мультимножества, действительная неотрицательная функция, заданная на алгебре мультимножеств  $L(Z)$ , причем  $m(A) = \sum_i w_i k_A(x_i)$ ,  $w_i > 0$ . В таком случае  $d_{1p}(A, B)$  можно представить в виде:

$$d_{ip}(A, B) = \left( \sum_{x_i \in G} w_i |k_A(x_i) - k_B(x_i)| \right)^{1/p}.$$

Метрика  $d_{1p}(A, B)$  аналогична мере схожести Хемминга [13] для мультимножеств. Для объединения нескольких фрагментов ПЗ в один необходимо привести множество многомерных точек, образующих мультимножество к единственной многомерной точке. В случае, если сделать это невозможно (расстояние между точками слишком велико, а также не совпадают имена, либо синонимы имен объектов), каждая точка выделяется в виде отдельного многомерного объекта. На рисунке ниже на примере двумерного пространства, приведена последовательность преобразований мультимножеств  $A$  и  $B$ , а также разбиение мультимножества  $B$  на две многомерные точки.

Рассмотрим две точки мультимножества  $O_m$  и  $O_k$ . Для подсчета расстояния между парой точек из облака точек многомерного множества воспользуемся Евклидовым расстоянием:

$$d_{O_m O_k} = \sqrt{\sum_{i=1}^n (x_{O_m i} - x_{O_k i})^2},$$

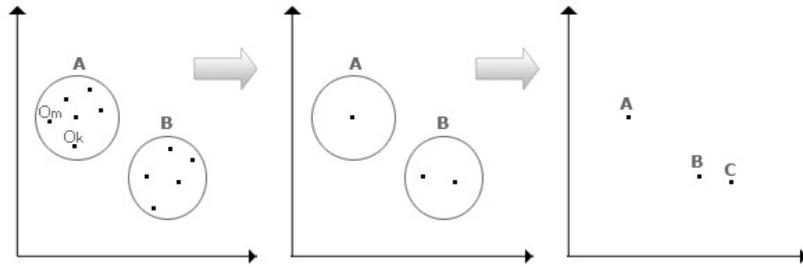
где  $O_m$  и  $O_k$  – объекты в  $n$ -мерном пространстве,  $i$  – порядковый номер признака,  $x_{O_m i}$  и  $x_{O_k i}$  – координаты объектов  $O_m$  и  $O_k$  по  $i$ -ому атрибуту, причем

$$x_O = \begin{cases} 0, & \text{в случае отсутствия атрибута в объекте} \\ 1, & \text{в случае присутствия атрибута в объекте} \end{cases}$$

В таком случае квадрат геометрического расстояния:

$$(d_{O_m O_k})^2 = \sum_{i=1}^n (x_{O_m i} - x_{O_k i})^2$$

есть количество атрибутов, различных для объектов  $O_m$  и  $O_k$ .



Приведение множеств многомерных точек к единой точке и разбиение мультимножества на множество многомерных точек

Обозначим  $A_m$  – не пустое множество атрибутов объекта  $O_m$ , а  $A_k$  – не пустое множество атрибутов объекта  $O_k$ . Объекты  $O_m$  и  $O_k$  имеют одинаковые имена, либо совпадающие синонимы имен. Рассмотрим следующие состояния:

- $(d_{O_k O_m})^2 = 0$ , т.е.  $A_k = A_m$ , в этом случае будем называть объекты  $A_k$  и  $A_m$  идентичными, возможно их автоматическое объединение;

- $(d_{O_k O_m})^2 = |A_k| + |A_m|$  и  $A_k \cap A_m = \emptyset$ , в этом случае, либо допущена ошибка при сопоставлении имен и синонимов имен объектов, либо имена объектов полностью идентичны, но объекты различны. Примером данной ситуации могут послужить два объекта: «Ручка» (дверная) с атрибутами *материал (металл, пластик), конструкция (раздельная, ручка-кнопка, ручка-скоба)*, и «Ручка» (шариковая) с атрибутами *цвет (красный, синий, черный), тип (шариковая, роллер)*. Для устранения данного совпадения необходимо уточнение экспертом, либо инженером по знаниям имени объекта. В качестве решения проблемы в приведенном примере может послужить переименование объектов в «Ручка\_дверная» и «Ручка\_шариковая», а все найденные примеры такого рода сохраняются в списке исключений;

- $(d_{O_k O_m})^2 > 0$  и  $A_k \neq A_m$ , в этом случае возможны следующие состояния:

а)  $A_k \subset A_m$ , при таком условии возможна автоматическая замена объекта  $A_k$  объектом  $A_m$ , т.к. объект  $A_m$  является более детально описанным;

б)  $A_k \not\subset A_m$  и  $A_k \cap A_m \neq \emptyset$  при таких условиях необходима ручная обработка объектов экспертом, либо инженером по знаниям. Скорее всего, объекты, описание которых получено из источников различной типологии, описаны также с точки зрения применения в различных примерах ПрО. В этом случае возможно как объединение объектов с частичным или полным объединением атрибутов, так и замена выбранных

атрибутов, либо декомпозиция объектов. Все перечисленные операции производятся экспертом, либо инженером по знаниям.

При каждом объединении  $k$  объектов производится переформирование всех правил, в посылках или заключениях которых участвуют по отдельности рассматриваемые объекты. Вместо  $k$  различных объектов в утверждениях каждого правила используется объединенный либо замененный (оставшийся один из  $k$ ) объект.

Следующим после завершения объединения объектов этапом объединения фрагментов ПЗ, полученных из источников знаний различной типологии, является этап объединения наборов продукционных правил.

### 3. Автоматизированное объединение наборов продукционных правил, полученных из различных источников знаний

В качестве анализируемой структуры для эффективного и быстрого объединения наборов правил в ЗОМ используются расширенные таблицы решений (РТР) [14], представляющие собой набор строк и столбцов, где каждая ячейка строки хранит данные о вхождении и параметрах вхождения *утверждения*, характеризующегося заголовком строки в конкретное *правило*. Каждая ячейка РТР разбита на две части: одна – для посылок правил, а другая – для заключений правил.

Сначала РТР пуста, а по мере рассмотрения правил, входящих в состав ПЗ, она пополняется новыми строками, однозначно идентифицирующимися парой «объект - атрибут объекта». Правила представляются в РТР ее столбцами. В каждую ячейку РТР записывается «тип» утверждения, он может принимать следующие значения: 0 - утверждение отсутствует в рассматри-

ваемом правиле; 1 – утверждение присутствует в рассматриваемом правиле. Для каждого рассматриваемого правила предусмотрены два столбца: наличие утверждения в посылке правила и наличие утверждения в заключении правила;

Применение РТР упрощает и позволяет в значительной степени автоматизировать анализ наборов правил, полученных из различных источников. Построение и анализ РТР являются лишь промежуточными этапами объединения наборов правил, полученных из различных источников.

Для объединения двух наборов правил в единый используется анализ РТР, который сводится к подсчету совпадающих атрибутов, участвующих в правилах  $R_i$  и  $R_k$ , а также общего количества атрибутов, участвующих в данных правилах. Далее отдельно для левой и правой частей правил подсчитывается мера сходства Хемминга [13] ( $\mu_{ik}^{NL}$  и  $\mu_{ik}^{NR}$ ):  $\mu_{ik}^N = n_{ik}/N$ , где  $n_{ik}$  – есть число совпадающих признаков у образцов  $R_i$  и  $R_k$ ,  $\mu_{ik}^{NL}$  – есть отношение количества совпавших атрибутов правых частей правил  $R_i$  и  $R_k$  к количеству всех атрибутов, участвующих в правых частях правил.

Далее формируется таблица мер схожести правил, которая имеет число строк и столбцов равное суммарному числу правил, находящихся в сравниваемых наборах правил. На первом этапе работы алгоритма создается пустая таблица, каждому столбцу и строке которой присваивается имя (номер) рассматриваемого правила. Как в столбцах, так и в строках таблицы находятся все правила, составляющие оба сравниваемых набора. На пересечении каждого столбца и строки таблицы имеется две ячейки, одна предназначена для хранения меры схожести посылок, другая – для хранения меры схожести заключений. В каждую ячейку соответственно записываются правая и левая меры схожести пересекающихся правил (пересекающейся строки и столбца). Для вычисления каждой меры схожести проводится анализ РТР:

- выбирается первая незаполненная строка таблицы мер схожести;
- в РТР выбирается столбец, номер (имя) которого равен номеру текущей строки таблицы мер схожести;
- проводится пошаговое сравнение с каждым столбцом РТР, вычисляются меры схожести посылок и заключений пар правил;

- меры схожести посылок и заключений записываются в соответствующие ячейки таблицы мер схожести;

- по окончании анализа РТР и заполнения таблицы мер схожести полученный результат сохраняется для дальнейшей обработки.

Очевидно, что главная диагональ такой таблицы будет представлена единицами, а сама таблица симметрична относительно главной диагонали, что позволяет хранить только верхнюю ее половину.

Перед началом работы процедуры объединения правил для определения последовательности их вывода устанавливается *контрольная зона* мер схожести. В каждой строке производится анализ ячеек, содержащих соответствующие меры схожести *посылки* и меры схожести *заключения* в заданный интервал, пары правил, образующие пересечение столбца и строки таблицы мер схожести помещаются в список удовлетворяющих заданным условиям и могут быть выведены для дальнейшего анализа инженером по знаниям.

По окончании работы полученный фрагмент ПЗ подвергается завершающей обработке – все объекты и правила собираются в единый XML файл и проходят заключительные процедуры: верификацию [14] и финальную перенумерацию.

#### 4. Экспериментальное программное исследование распределенного варианта КМПЗ

В настоящее время проводится экспериментальная апробация предложенных алгоритмов и разработанных программных средств для задач медицинской диагностики. Кроме того, для определения географического местонахождения IP-адресов, в задачах контроля радиационных дозовых нагрузок персонала АЭС РФ, для чего были использованы фрагменты 10 БД атомных электростанций (Балаковской, Белоярской, Билибинской, Калининской, Кольской, Курской, Ленинградской, Нововоронежской, Ростовской и Смоленской). Объем контрольных выборок, полученных из каждого фрагмента БД, составил от 2500 до 4000 записей. Суммарное время работы реализованных методов и алгоритмов распределенного приобретения знаний на десяти контрольных выборках (общий объем 35072 записи) составило 2 мин. 6 с., а суммарное количество

правил, выведенных из десяти контрольных выборок, составило 10350 продукционных правил.

Все используемые для проверки работы фрагменты БД имели одинаковую структуру, как следствие, контрольные выборки имели одинаковые наборы полей, а соответствующие атрибуты - одинаковые имена, типы и списки возможных значений. Таким образом, объединение всех десяти фрагментов ПЗ, полученных из различных БД, было осуществлено в полностью автоматическом режиме и заняло менее 10 мин. Объем итогового ПЗ составил 1674 правила.

Анализ экспериментальных данных, полученных при создании ряда БЗ, показал, что *распределенный* вариант использования БД в качестве дополнительного источника знаний способен пополнить объем разрабатываемых БЗ на 12-25%.

## Заключение

Предложенные методы и алгоритмы объединения наборов продукционных правил способны значительно сократить временные и финансовые затраты при разработке БЗ. При этом в среднем все разрабатываемые БЗ при использовании БД в качестве дополнительного источника знаний были пополнены на 15%, что свидетельствует об эффективности применения методов и инструментальных программных средств для автоматизированного построения БЗ ИЭС.

## Литература

1. Рыбина Г.В. Теория и технология построения интегрированных экспертных систем. – М.: Научтехлитиздат. 2008. 482 с.
2. Рыбина Г.В. Интегрированные экспертные системы. Приборы и системы. Управление, контроль, диагностика. 2011. №4. С. 22-47.
3. Осипов Г.С. Лекции по искусственному интеллекту. М.: КРАСАНД. 2009. 272 с.
4. Частиков А.П., Гаврилова Т.А. Белов Д.Л. Разработка экспертных систем. Среда CLIPS. СПб.: БХВ. Петербург. 2003. 608 с.
5. Рыбина Г.В. Основы построения интеллектуальных систем. М.: Финансы и статистика. ИНФРА-М. 2010. 432 с.
6. Люггер Дж. Ф. Искусственный интеллект: стратегии и методы решения сложных проблем. М.: Издательский Дом «Вильямс». 2003. 864 с.
7. Рыбина Г.В., Дейнеко А.О. Распределенное приобретение знаний для автоматизированного построения интегрированных экспертных систем. Искусственный интеллект и принятие решений. 2010. №4. С. 56-62.
8. Рыбина Г.В., Дейнеко А.О., Нистратов О.В. Автоматизированное построение полных и непротиворечивых баз знаний в интегрированных экспертных системах. Приборы и системы. Управление. Контроль. Диагностика. 2010. №5. С. 21-24.
9. Рыбина Г.В. Концептуальные основы задачно-ориентированной методологии построения интегрированных экспертных систем Приборы и системы. Управление, контроль, диагностика. 2011. №5. С. 33-51.
10. Петровский А.Б. Новые классы метрических пространств измеримых множеств и мультимножеств в кластерном анализе. Методы поддержки принятия решений. Труды ИСА РАН. т. 12. М.: Едиториал УРСС. 2001. С. 54-67.
11. Петровский А.Б. Метрические пространства мультимножеств. Доклады Академии наук. 1995, Т.344, №2, С.175-177.
12. Петровский А.Б. Пространства множеств и мультимножеств. М.: Едиториал УРСС. 2003. 248 с.
13. Загоруйко Н.Г. Прикладные методы анализа данных и знаний. Новосибирск. Издательство института математики. 1999. 17с.
14. Рыбина Г.В., Смирнов В.В. Методы и алгоритмы верификации баз знаний в интегрированных экспертных системах. Новости искусственного интеллекта. №3. 2005. С. 7-19.
15. Adomavicius G., Tuzhilin A. Expert-Driven Validation of Rule-Based User Models in Personalization Applications. Data Mining and Knowledge Discovery Journal. vol. 5. 2001.
16. Lent B., Swami A., Widom J. Clustering association rules. In Proceedings of the 13th International Conference on Data Engineering. 1997. p. 220-231.
17. Wang K., Tay S., Liu B. Interestingness-based interval merger for numeric association rules. In proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining. 1998.
18. Загоруйко Н.Г. Прикладные методы анализа данных и знаний. Новосибирск: Издательство института математики. 1999. 17с.

**Рыбина Галина Валентиновна.** Профессор кафедры кибернетики Национального исследовательского ядерного университета «МИФИ» (НИЯУ МИФИ). Окончила Московский инженерно-физический институт (государственный университет) в 1971 году. Доктор технических наук, профессор. Лауреат премии Президента РФ в области образования. Автор свыше 400 печатных работ. Области интересов: интеллектуальные системы и технологии, статические, динамические и интегрированные экспертные системы, интеллектуальные диалоговые системы, многоагентные системы, инструментальные средства.

**Дейнеко Александр Олегович.** Аспирант кафедры кибернетики Национального исследовательского ядерного университета «МИФИ» (НИЯУ МИФИ). Окончил Московский инженерно-физический институт (государственный университет) в 2008 году. Автор 8 печатных работ. Области интересов: интеллектуальные системы и технологии, интегрированные экспертные системы, инструментальные средства, приобретение знаний, технологии Data Mining и KDD.