

Использование экспертных систем для интеллектуального анализа данных

А.И. Леженко, И.А. Кузнецов, С.К. Кузнецов

Аннотация. Статья посвящена использованию экспертных систем (ЭС) для интеллектуальной обработки данных (ИАД) в приложениях информационно-телекоммуникационных технологий. Сделан общий обзор ЭС и методов теории нечетких множеств и нечеткой логики, а далее с использованием ЭС FuzzyCLIPS рассмотрен конкретный способ построения процесса ИАД для исследования рентабельности коммерческого центра обработки данных (ЦОД) на основе факторного анализа.

Ключевые слова: интеллектуальный анализ данных, экспертные системы, нечеткие множества, нечеткая логика, факторный анализ, FuzzyCLIPS, поддержка принятия решений, информационно-телекоммуникационные технологии, хранилища данных, центры обработки данных

Введение

Приложения информационных и телекоммуникационных технологий (ПИТТ) как системы коллективного пользования становятся постоянно пополняющимся хранилищем и концентратором интеллектуального потенциала их пользователей. ПИТТ являются той областью деятельности, цель которой – объединение ее возможностей и способностей человека, позволяющее достичь поставленных целей более эффективно, чем при использовании этих возможностей раздельно. На практике человек или коллектив, используя ИТТ, свой интеллект и знания, в конкретной обстановке должен найти приемлемое решение, приводящее к оптимальному или эффективному результату. Одним из математических аппаратов для решения подобных задач является теория выбора и принятия решений (ТВПР) [1].

Методы ТВПР широко используются в компьютерных экспертных системах (ЭС) как самостоятельном разделе систем и методов искусственного интеллекта (ИИ). Использование ЭС в различных областях деятельности уже сегодня дало значительный эффект и привело к уникальным результатам [2, 7].

ЭС применяются для управления процессами переработки информации. Они незаменимы особенно тогда, когда информация из одной подсистемы передается в другую. С помощью задания правил и механизмов логического вывода ЭС может выходную информацию из одной подсистемы подготовить в необходимом виде для обработки последующей подсистемой.

Ниже приведен обзор ЭС.

1. Оболочка CLIPS и связанные системы

CLIPS (Язык С, Интегрированная Производственная Система) – OVS-подобная производственная система, использующая вывод от фактов к цели, написанная на языке С в Космическом центре Джонсона NASA. Механизм логического вывода CLIPS включает сопровождение, динамическое добавление правил и настраиваемые стратегии разрешения противоречий. CLIPS легко встраивается в другие прикладные программы. CLIPS включает объектно-ориентированный язык, названный COOL (объектно-ориентированный язык CLIPS), который прямо интегрирован с механизмом логического вывода. CLIPS выполняется на многих платформах, включая IBM PC.

DYNACLIPS (Динамические утилиты CLIPS) включает доску объявлений, механизм динамического обмена знаниями и инструментальные средства для CLIPS. Она построена как набор библиотек, который может быть связан с CLIPS. Для связи с другими интеллектуальными средствами используется доска объявлений.

FuzzyCLIPS – версия CLIPS, оболочка экспертной системы, основанная на правилах, используется для представления и управления нечеткими фактами и правилами. В дополнение к функциональным возможностям CLIPS, FuzzyCLIPS может иметь дело с точными, нечеткими (или неточными) знаниями, сложными рассуждениями, которые можно свободно смешивать в правилах и фактах экспертной системы. Система использует две базисных концепции о неточности: нечеткость и неопределенность.

WxCLIPS снабжает CLIPS с нечетким представлением знаний простым графическим внешним интерфейсом.

2. Другие оболочки экспертных систем

OVS5 содержит механизмы представления знаний и управления. Хотя эта система обеспечивает основные потребности инженерии знаний, она не ориентирована на конкретные стратегии решения задач или схемы представления знаний. Система разрешает программисту использовать символы и представлять отношения между символами, однако эти символы и отношения не имеют заранее определенных значений. Последние полностью определяются порождающими правилами, которые пишет программист. Механизм управления интерпретатора OVS5 представляет собой простой цикл, называемый «циклом распознавания», детали которого пользователь разрабатывает сам в соответствии со своими потребностями.

BABYLON – среда для разработки для экспертных систем. Она включает фреймы, модели данных, прологоподобный логический формализм и язык для написания диагностических прикладных программ. Она написана на Лиспе и переносима на широкий диапазон аппаратных платформ.

MIKE (Микроинтерпретатор для Инженерии Знаний) – это полная свободная и переносимая программная среда, разработанная для целей обучения в Открытом университете Великобритании. Она включает прямые и обратные правила вывода «от цели к фактам» с определяемыми пользователем стратегиями разрешения противоречий и фреймовый язык представления знаний с наследственностью и «демонами» плюс определенные пользователем стратегии наследования. Правила вывода автоматически снабжаются объяснениями «как»; пользователь может сформировать объяснения «почему». Порядок применения правил в процессе трассировки и выполнения может отображаться графически на дисплее. MIKE формирует ядро курса по инженерии знаний Открытого университета, он написан на консервативном и переносимом подмножестве Пролога. MIKE включает RETE-алгоритм для быстрого поиска вперед, систему сопровождения, обработки неопределенности и гипотетических миров. Он имеет интерфейс, который управляется с помощью меню, а также инструментальные средства для создания и просмотра фреймов.

WindExS (Экспертная Система под Windows) – полнофункциональная экспертная система, использует вывод от фактов к цели, работает на базе Windows. Ее модульная архитектура позволяет пользователю заменять модули так, как это требуется для расширения возможностей системы. WindExS содержит процессор правил на естественном языке, механизм логического вывода, диспетчер файлов, интерфейс пользователя, администратор сообщений и модули базы знаний. Она поддерживает вывод от фактов к цели и графическое представление базы знаний.

RT-EXVERT – экспертная система общего назначения, что позволяет программистам интегрировать правила экспертной системы в прикладные программы на языке C или C++. В состав RT-EXVERT входит транслятор правил, который компилирует правила в код C, и библиотека, содержащая механизм выполнения правил. RT-EXVERT под DOS работает с трансляторами Borland Turbo C, Borland C++ и Microsoft C/C++. Лицензионная версия программы используется в области образования,

исследований и хобби. Прикладные программы, созданные с помощью RT-EXVERT, не лицензированы для коммерческих целей. Профессиональные издания пригодны для коммерческих прикладных программ, использующих Windows и Unix среды.

3. Пример использования экспертной системы FuzzyCLIPS

В данной работе мы будем использовать продукт FuzzyCLIPS версии 6.04A [3]. Оболочка этой экспертной системы основана на правилах и используется для представления и управления нечеткими фактами и правилами.

В работе [4] мы провели факторный анализ для получения интегральных оценок и аналитической информации о функционировании центра обработки данных (ЦОД) ручным

способом с карандашом, а второй – автоматический, с использованием экспертных систем, – выполним в этой работе.

Результаты факторного анализа параметров функционирования ЦОД, приведенных в Табл. 1 отражены в Табл. 2. Анализ, осуществленный посредством метода главных компонент и процедуры варимаксного вращения, выявил 4 фактора, каждый из которых может быть представлен как нечеткое множество (НМ).

Анализ этих факторов выполним с помощью ЭС FuzzyCLIPS.

Для каждой переменной $V_{i=1,31}$ и фактора $F_{j=1,4}$ в качестве меры неопределенности $\mu(F_j, V_i)$ возьмем модуль факторного веса переменной V_i .

Обозначим полученные НМ: *fact1*, *fact2*, *fact3*, *fact4*; загрузим их в FuzzyCLIPS в виде группы (group).

Табл. 1. Параметры функционирования ЦОД

V1	Затраты на строительство / аренду и подготовку помещения
V2	Затраты на обслуживание и поддержку
V3	Затраты на систему электроснабжения
V4	Затраты на систему холодоснабжения
V5	Затраты на систему телекоммуникаций
V6	Затраты на систему безопасности
V7	Прочие затраты
V8	Программное обеспечение
V9	Подключение к каналам связи
V10	Подключение к интернет
V11	Организация каналов связи
V12	Предоставление VPN каналов
V13	Подключение к удалённым дата-центрам
V14	Резервный узел корпоративной информационной сети
V15	Резервный офис клиента
V16	Резервное копирование и хранение данных
V17	Размещение сервера
V18	Аренда сервера
V19	Размещение стойки
V20	Аренда стойки
V21	SaaS: использование лицензионного ПО от Microsoft
V22	Разработка высокопроизводительных информационных систем
V23	Комплексное администрирование
V24	Конфигурирование стандартных опций в установленном программном обеспечении
V25	Площадь
V26	Число стойкомест
V27	Потребляемая электроэнергия
V28	Курс доллара
V29	Курс евро
V30	Индекс РТС
V31	Индекс ММВБ

Табл. 2. Повернутая матрица компонент

	Component			
	1	2	3	4
V14	.972	.184	.091	.100
V11	.961	.102	-.053	.198
V25	.925	.318	.181	.098
V2	.925	.318	.181	.098
V24	.925	.318	.181	.098
V26	.925	.318	.181	.098
V3	.925	.318	.181	.098
V5	.925	.318	.181	.098
V7	.925	.318	.181	.098
V4	.925	.318	.181	.098
V18	.925	.318	.181	.098
V20	.925	.318	.181	.098
V6	.925	.318	.181	.098
V21	.925	.318	.181	.098
V16	.925	.318	.181	.098
V1	.925	.318	.181	.098
V15	.925	.318	.181	.098
V8	.925	.318	.181	.098
V10	.904	.375	.115	.083
V13	.894	.288	.010	.162
V9	.811	.264	.385	.266
V23	.711	.613	.282	.069
V27	.711	.613	.282	.069
V19	.535	.762	.095	.058
V22	.432	.735	.474	.058
V12	.675	.705	-.033	.046
V28	.178	.034	.944	.062
V31	-.241	-.245	-.671	-.599
V30	-.242	-.242	-.664	-.607
V17	.004	-.086	.174	.810
V29	.436	.489	-.225	.564

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 11 iterations.

С помощью специальной команды имеется возможность графического отображения переменных НМ, введенных в ЭС FuzzyCLIPS. Для переменной *fact1* результат выполнения команды представлен на рисунке. Подобная возможность отображения, заложенная в ЭС FuzzyCLIPS, позволяет повысить уровень контроля данных и улучшить сервис по их обработке и анализу.

Следующим этапом процесса работы с ЭС FuzzyCLIPS является построение правил обработки НМ для логического вывода.

Приведем основу математического аппарата для построения правил [5, 6].

Правила определяются соединением:

$$R = F_a \cap F_c,$$

где F_a – НМ предшествующей модели,

F_c – НМ результатов вывода.

Отношение R вычисляется по формуле:

$$\mu_R(u, v) = \min(\mu_{F_a}(u), \mu_{F_c}(v)), \quad \forall (u, v) \in U \cap V.$$

Другие алгоритмы для формирования этого отношения изложены в работе [7]. Вывод основан на правилах композиционного вывода, описание которого приведено ниже.

$$F_c = F_a \bar{U} R.$$

F_c – это НМ, определяющее выводы, рассчитываемое следующим образом:

$$\mu_{F_c}(v) = \max_{u \in U} (\min(\mu_{F_a}(u), R(u, v))),$$

что может быть упрощено:

$$\mu_{F_c}(v) = \min(z, \mu_{F_c}(v)),$$

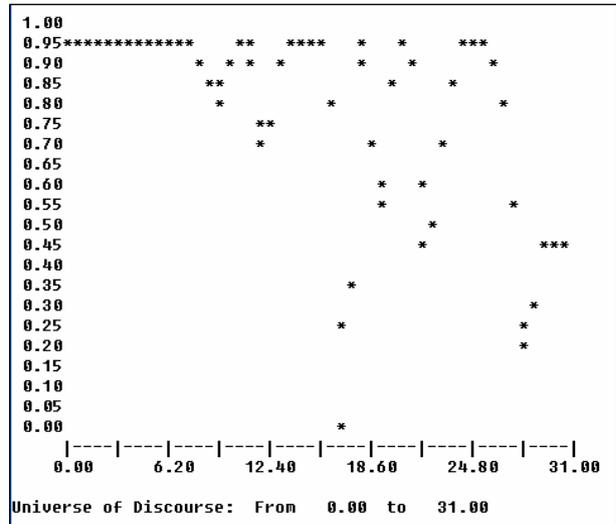
где $z = \max(\min(\mu_{F_a}(u), \mu_{F_a}(U)))$.

Кроме простых правил вывода существуют комплексные правила вывода, когда в цепочке вывода используются несколько НМ.

Вывод осуществляется с помощью алгоритма, который подробно описан в работе [7]. Таким образом, имеется развитый аппарат построения простых и комплексных правил обработки НМ.

Для дополнения алгоритмов обработки НМ и построения правил вывода предлагается способ обработки НМ, суть которого изложена ниже.

В нечеткой логике введены эквиваленты операций обычной логики “И”, “ИЛИ”, “НЕ”, выполняемые по следующим правилам [8]:



Графическое представление НМ переменной *fact1*

$$p1 \text{ И } p2 = \min(p1, p2);$$

$$p1 \text{ ИЛИ } p2 = \max(p1, p2);$$

$$\text{НЕ } p1 = 1 - p1.$$

Кроме того, для уточнения правил Шортлиффом введена формула вычисления «меры доверия» (МД) [9]:

$$МД(h:L1,L2) = МД(h:L1) + МД(h:L2) * (1 - МД(h:L1)).$$

Эта формула уточняет МД в пользу предложения *h*, если у него кроме свидетельства *L1* имеется также свидетельство *L2*. Данная формула может быть использована для построения правил логического вывода. А цепочку правил вывода предлагается строить согласно убыванию факторных весов.

В работе [4] был выполнен ручной анализ результатов факторного анализа жизнедеятельности ЦОД. Первый фактор (НМ *fact1*) объединяет финансовые показатели как по доходам, так и по расходам; назовем его “internal financial indices”. Второй фактор (НМ *fact2*) определяет ключевые показатели по доходности ЦОД (“profitability”). Третий фактор (НМ *fact3*) определяет зависимость ключевых финансовых показателей деятельности ЦОД от российских биржевых индексов (“external financial indices”). Четвертый фактор (НМ *fact4*) определяет зависимость внутренних финансовых показателей от внешних (“correlation between internal and external financial indices”).

Напишем простейшее правило обработки сформированных НМ. Данное правило обрабатывает НМ и выдает данные о том, какие показатели обладают большими, средними и малыми факторными весами. ЭС FuzzyCLIPS имеет очень развитый аппарат построения правил для обработки НМ, с помощью которых можно строить даже функции управления и регулирования.

Вот описание правила, например, для НМ *fact1*:

```
; Define the fuzzy linguistic variable needed --
1 inputs (group)
;
(load "factor")
;
((defrule simple-fuzzy-group fact1)
(get-U fuzzy fact1)
(do i=1,31)
(if u > 0.7) then
(printout "high internal financial indices")
(get fuzzy-template-i)
(printout I)
(else if U > 0,3)
(printout "middle internal financial indices")
(get fuzzy-template-i)
(printout I)
(else)
printout "low internal financial indices"
(get fuzzy-template-i)
(printout I)
)
```

Необходимо подчеркнуть, что при работе с реальными БД корпоративных информационных систем, с которыми работают опытные менеджеры, их опыт автоматически становится доступным для начинающих менеджеров.

Заключение

Таким образом, в рамках предложенной методики ИАД сделано следующее: на базе результатов факторного анализа (в качестве примера выбран анализ параметров рентабельности коммерческого ЦОД) были сформирова-

ны простейшие правила ИАД и введены в ЭС FuzzyCLIPS. Результаты дальнейших исследований по данной методике могут быть использованы для поддержки принятия управленческих решений о направлениях развития с целью повышения рентабельности не только ЦОД, но и промышленных предприятий, а также во многих других сферах человеческой деятельности.

Литература

1. Макаров И.М. Теория выбора и принятия решений. - М.: Наука, 1982.
2. Амосов А.В., Данилов И.Г., Леженко А.И. Вопросы применения средств и методов искусственного интеллекта в САПР // ЭВМ в проектировании и производстве. Л.: Машиностроение, 1989. С. 86-101.
3. Orcharad R. A. FuzzyCLIPS Version 6.04A User's Guide. Integrated Reasoning, Institute for Information Technology, National Research Council Canada, 1998.
4. Леженко А.И., Кузнецов И.А. Способы и методы интеллектуальной обработки информационных потоков в приложениях информационных и телекоммуникационных технологий // Труды Международной научно-практической конференции «Передовые информационные технологии, средства и системы автоматизации и их внедрение на российских предприятиях», Москва, 4-8 апреля 2011.
5. Tzi-cker Chiueh: Optimization of Fuzzy Logic Inference Architecture. IEEE Computer 25(5): 6771 (1992)
6. Mizumoto M., Fukami S. and Tanaka K. Some methods of fuzzy reasoning. In Advances in Fuzzy Set Theory Applications (eds. Gupta M. M., Ragade R. K. and Yager R. R.). North-Holland, Amsterdam, 1979.
7. Lezhenko A. On the Methods of Formalization of the Information Trends and Data Processing in Integrated Information and Telecommunication Technologies. // 8th Multi-Conference on Systemic, Cybernetics and Informatics (SCI 2004), July 21-25 2004 in Orlando, Florida, USA.
8. Zaden E.A. Fuzzy sets // Information and Control.-1965/-V8.
9. Buchanan B.G. and Shortliffe E.H. (1984, eds.). Rule-based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project. Addison-Wesley.

Леженко Анатолий Иванович. Главный инженер АНО «Институт физико-технической информатики» (АНО ИФТИ). Окончил Новосибирский государственный университет в 1972 году. Доктор физико-математических наук. Автор более 50 печатных работ. Область научных интересов: информационные и телекоммуникационные технологии, АСУ, АСУ ТП, САПР. E-mail: lezhenko@gmail.com

Кузнецов Илья Анатольевич. Ведущий инженер «Центра информационных технологий и систем органов исполнительной власти» (ФГНУ ЦИТиС). Московский государственный университет им. М.В. Ломоносова в 2003 году. Автор 4 печатных работ. Область научных интересов: интеллектуальный анализ данных; имитационное моделирование; разработка ПО, основанная на моделях. E-mail: i441@mail.ru

Кузнецов Сергей Константинович. Научный сотрудник Института проблем управления РАН им. Трапезникова. Окончил Тульский политехнический институт в 1976 году. Автор 5 печатных работ. Область научных интересов: информационные технологии, разработка ПО на VB.NET. E-mail: skuznetsov@bk.ru