

### **The approach to masters preparing educational program designing**

*Vladimir Isaakovich Freyman, Candidate of Technical Sciences, Associate professor, Perm National Research Polytechnical University*

*Efim Lyvovich Kon, Candidate of Technical Sciences, Professor, Perm National Research Polytechnical University*

*Alexandr Anatolyevich Yuzhakov, Doctor of Engineering, Professor, Head of the Department of AT, Perm national Research Polytechnical University*

*In this report the problems of moving High school education system to the Third generation Federal state educational standards are analyzed. The necessity of linking employers qualification requirements with science and technical direction development vector is pointed out. Approaches to the masters preparing educational programs designing, realized in the Perm National Research Polytechnical University, are dwelled on.*

*Keywords: competence-based approach, educational program, direction development vector, employers qualification requirements, functional-full basis.*

УДК 004.832.3

### **АРГУМЕНТАЦИЯ В ИНДУКТИВНОМ ФОРМИРОВАНИИ ПОНЯТИЙ**

*Вадим Николаевич Вагин, д-р. техн. наук, проф.,  
кафедра прикладной математики*

*Тел. 8(495)362-79-62, e-mail: vagin@aprrmat.ru*

*Марина Владимировна Фомина, доц., канд. техн. наук,  
кафедра вычислительной техники*

*Тел. 8(495)362-79-62, e-mail: vagin@aprrmat.ru*

*Национальный исследовательский университет*

*Московский энергетический институт*

*ФБГОУ ВПО «НИУ «МЭИ»*

*<http://mpei.ru>*

*В данной работе рассматриваются методы обработки неполной и противоречивой информации в таких подсистемах интеллектуальных систем поддержки принятия решений, как подсистема поиска решений и подсистема приобретения и накопления знаний. В подсистеме поиска решений основной акцент сделан на применение аргументации, с использованием степеней обоснования. В подсистеме приобретения и накопления знаний решается задача обобщения в условиях противоречивых и неполных данных. Для обобщения информации разработан алгоритм, основанный на теории приближенных множеств. Проведено исследование влияния шума на работу предложенного алгоритма обобщения. Представлены результаты программного моделирования.*

*Ключевые слова: Аргументация, пересматриваемые рассуждения, степени обоснования, приближенные множества, формирование понятий, обобщение, зашумленные данные.*

Работа выполнена при поддержке грантов  
РФФИ № 14-07-00862, 12-01-00589

**Введение**

В настоящее время весьма актуальной проблемой в области ИИ является проблема конструирования интеллектуальных систем поддержки принятия решений (ИСППР), ориентированных на открытые и динамические предметные области. К сожалению, большинство практических задач являются плохо формализованными, и могут содержать в себе противоречивую и недостоверную информацию. Причем, неопределенность может иметь не вероятностный характер. Классические методы логического вывода не применимы для решения таких задач. При поиске решения в таких задачах необходимо применение



**В.Н. Вагин**

методов правдоподобного вывода, позволяющих найти некоторое приемлемое (которое может и не быть оптимальным) решение. Основные подсистемы, входящие в состав ИСППР, были подробно рассмотрены в [1]. В данной работе предлагается определить те подсистемы в структуре ИСППР, которые требуют модификации для обработки противоречий и предложить такие модификации.

Прежде чем переходить к рассмотрению проблем, связанных с противоречивостью, отметим различные типы информации применяемой в ИСППР, и определим, что может приводить к противоречивости в базах знаний. Обычно в базах знаний ИСППР содержится информация следующих типов [2].

1) Объективная информация – информация, полученная из надежных источников, или которая может быть напрямую измерена или подтверждена.

2) Субъективная информация – информация, полученная из менее надежных источников. Это могут быть некоторые предположения, суждения, которые часто формулируются с помощью фраз «как правило», «обычно», «скорее всего».

3) Гипотетическая информация (используется для построения гипотез).

Если ИСППР использует субъективную или гипотетическую информацию, то есть, например, знания экспертов, необходимо учитывать возможность наличия противоречий в базах знаний.

Ключевыми частями ИСППР являются подсистема поиска решений и подсистема приобретения и накопления знаний.

Остановимся подробнее на том, как обрабатываются противоречия в процессе приобретения и накопления знаний и при поиске решений.

**Обработка противоречий в подсистеме поиска решений.**

Наличие противоречивой информации является весьма распространенным в ИСППР и ограничения, требующие ее отсутствия, довольно сильно сужают область применения всей системы. Одним из способов обработки конфликтных ситуаций и противоречивой информации является применение аппарата аргументации. Аргументация дает гораздо больше возможностей для моделирования правдоподобных рассуждений. Для работы с такой системой предлагается применять теорию аргументации. Существуют несколько формализаций теории аргументации. В данной работе будет рассмотрена системы аргументации, основанная на пересматриваемых рассуждениях, предложенных Джоном Поллоком [3]. Аргументация – процесс построения аргументов за и против некоторого предположения. Аргумент – пара, состоящая из множества посылок и заключения. Записывать такие пары будем в следующем виде  $p/X$ , где  $p$  – заключение, а  $X$  – множество посылок. Например, аргумент  $(p \& q) / \{ \sim a, b \}$  означает, что из посылок  $\sim a, b$  следует заключение  $p \& q$ . Ключевым моментом в любых системах пересматриваемых рассуждений является понятие конфликта. Будем рассматривать два типа конфликтов – опровержение



**М.В. Фомина**

(rebutting) и подрыв (undercut) [4]. Опровержение – ситуация, когда некоторые полученные аргументы опровергают выводы ранее полученных рассуждений. Подрыв – несимметричная форма атаки, когда один аргумент отрицает связь между посылками и заключением другого аргумента. Для логики предикатов первого порядка поиск конфликтов осложняется необходимостью обрабатывать кванторы общности и существования. Ключевая идея заключается в использовании механизма унификации для поиска конфликтов обоих типов.

**Степени обоснования в пересматриваемых рассуждениях**

Прежде всего, рассмотрим, как степени обоснования могут задаваться. В данной статье для задания степеней обоснования используется числовая шкала [0,1], где 0 соответствует пораженному аргументу, 1 -наиболее обоснованному аргументу. Степени обоснования присваиваются каждому исходному аргументу, и представляют собой некую оценку достоверности источника, из которого получен данный аргумент. Например, по телевизору сказали, что вероятность осадков 70%. Соответственно мы можем построить аргумент *A1:Завтра(дождь)* со степенью обоснования 0.7. Степени обоснования будем записывать функцией *Jus(A)*, то есть для приведенного примера *Jus(Завтра(дождь))=0.7*. Конкретные механизмы получения степеней обоснования зависят прежде всего от предметной области. Например, это могут быть статистические данные (в 90% этот источник дает верные данные) или экспертные оценки, например вероятность роста акций 60%.

Итак, нам необходимо задать функцию *Jus(A)* для вычисления степени обоснования любого из аргументов в графе вывода. Будем считать, что для начальных аргументов эта величина является определенной. На значение этой функции будут оказывать влияние два фактора – дерево вывода аргумента (т.е. степень обоснования аргументов, которые использовались в выводе данного аргумента) и конфликты с другими аргументами. Для удобства рассмотрим эти два фактора отдельно: *Jus<sub>anc</sub>(A)* – унаследованная степень обоснования и *Jus<sub>con</sub>(A)* – в какой степени конфликт подрывает обоснование аргумента.

$$Jus_{anc}(A)=\min(\{Jus(A1), Jus(A2) \dots Jus(An)\}), \tag{1}$$

где *A1, A2 ... An* – аргументы, участвовавшие в выводе аргумента *A*.

Формулу (1) называют принципом слабой связи [3]. Отметим, что из формулы (1) следует, что, если производить вычисление степеней обоснования рекурсивно, начиная от исходно заданных аргументов, то можно искать минимум, не из всех аргументов в базисе, а только на предыдущем шаге. Таким образом, если у аргумента один предок, то его унаследованная степень обоснования будет равна степени обоснования его предка.

Если при вычислении *Jus<sub>anc</sub>* ищутся наиболее слабые аргументы, то при определении того, насколько конфликт уменьшает обоснования, используются наиболее сильные аргументы. Пусть *Aconfl* – множество аргументов, вступающих в конфликт с *A*, тогда

$$Jus_{con}(A) = \begin{cases} \text{Max}(\{Jus_{anc}(Aconfl_1), Jus_{anc}(Aconfl_2), \dots, Jus_{anc}(Aconfl_n)\}), & |Aconfl| > 0; \\ 0, & \text{в противном случае.} \end{cases} \tag{2}$$

В формуле (2) используется *Jus<sub>anc</sub>* для того, чтобы верно обрабатывать случаи, когда между аргументами есть конфликт типа опровержение.

Итак, окончательно:

$$Jus(A) = \begin{cases} Jus_{anc}(A) - Jus_{con}(A), & Jus_{anc}(A) > Jus_{con}(A); \\ 0, & \text{в противном случае.} \end{cases} \tag{3}$$

**Методы обучения при наличии противоречий.**

Важным компонентом ИСППР является подсистема обучения и индуктивного формирования понятий. С помощью методов обобщения в интеллектуальных системах имеется возможность выделять наиболее важные закономерности, характерные для различных классов ситуаций, возникающих на сложном объекте. Это делается посредством анализа уже классифицированных объектов и формирования некоторого набора правил

(обобщенной модели). Затем эта обобщенная модель может быть использована для распознавания ситуаций, не известных системе ранее.

Обнаружение знаний в базах данных тесно связано с решением задачи индуктивного формирования понятий, или задачи обобщения. Формулировка задачи обобщения понятий по признакам дана в [4]. Пусть  $O = \{o_1, o_2, \dots, o_n\}$  – множество объектов, которые могут быть представлены в интеллектуальной системе  $S$ . Каждый объект в  $S$  характеризуется  $r$  признаками. Такое описание объекта называется признаковым описанием. В качестве признаков объектов могут использоваться количественные, качественные либо шкалированные признаки.

Пусть среди всех объектов, представленных в некоторой системе знаний, имеется  $V$  - множество положительных объектов – примеров формируемого понятия, и  $W$  - множество отрицательных объектов (контрпримеров понятия). Справедливо  $O = V \cup W$ ,  $V \cap W = \emptyset$ . Пусть  $K$  - непустое множество объектов такое, что  $K = K^+ \cup K^-$ , где  $K^+ \subset V$  и  $K^- \subset W$ . Будем называть  $K$  обучающей выборкой. На основании обучающей выборки надо построить правило, разделяющее положительные и отрицательные объекты обучающей выборки. Алгоритмы обобщения формируют решение в виде правил типа «ЕСЛИ <условия> ТО <искомое понятие>». Решающее правило является корректным, если оно в дальнейшем успешно распознаёт объекты, не вошедшие первоначально в обучающую выборку.

Для решения задачи обобщения в работе предлагается использовать алгоритм, построенный на основе теории приближенных множеств [5]. В соответствии с данной теорией приближенное множество определяется посредством задания верхней и нижней границ некоторого множества, называемых приближениями этого множества. Подобно нечетким множествам, приближенные множества являются математической концепцией для работы с нечеткостями в данных. Алгоритм GIRS (Generalized Iterative algorithm based on the Rough Set approach) был разработан авторами и подробно описан в [6]. Основные этапы алгоритма: поиск классов эквивалентности по отношению неразличимости, построение верхнего и нижнего приближений, поиск среза за счёт выделения наиболее существенных атрибутов, формирование системы продукционных правил. В алгоритме GIRS предлагается совместить выполнение этапа дискретизации с поиском среза для того, чтобы производить дискретизацию только тех количественных атрибутов, которые окажутся существенными для среза. Помимо этого, в качестве существенных признаков предлагается рассматривать признаки, входящие в приближительные срезы достаточно высокого качества. Понятие приближительного среза представляет собой обобщение понятия среза, и рассмотрено в [6]. Применение идеи приближительных срезов весьма полезно при обработке противоречивых и зашумленных данных.

Понятие «шум» включает в себя различные искажения, которые могут присутствовать в исходных данных. Шум – это несистематические ошибки в описании объектов, возникающие при преобразовании истинных входных величин в процессе измерения и пересылки данных. Основной характеристикой шума является величина  $p_0$  – вероятность искажения значения признака. Последняя величина называется «уровень шума» и выражается в процентах. Мы исследуем следующие модели шума: «отсутствие значений» и «перемешивание значений»; эти модели подробно описаны в [7].

Была рассмотрена работа алгоритма обобщения GIRS при наличии в исходных данных шума вышеназванных моделей. Наша цель – оценить точность классификации примеров контрольной выборки при возрастании уровня шума в исходной выборке, когда воздействию шума подвергался решающий атрибут.

В таблице 1 приведены результаты, полученные для модели шума «перемешивание значений». Шум уровня 5, 10, 15, 20 и 25 процентов вносился в решающий атрибут. Внесение шума в решающий атрибут оказывает существенное влияние на точность классификации объектов поскольку при внесении шума в такой атрибут в обучающее множество появляются не просто ошибочные, но даже противоречивые примеры.

Экспериментальные результаты представлены на известной коллекции данных Калифорнийского университета «UCI Machine Learning Repository»[8]. Были выбраны «задачи монахов», поскольку такие наборы данных содержат только качественные атрибуты.

Проведенные опыты показали, что при увеличении шума до 25% в решающем атрибуте наблюдается снижение точности классификации тестовых примеров, причем этот процесс не всегда является монотонным. Это свидетельствует об ухудшении качества сформированных алгоритмом продукционных правил. Следует заметить, что при уровне шума свыше 10 – 15 % точность классификации тестовых примеров значительно снижается. Это связано как с появлением большого числа противоречивых примеров в выборках при перемешивании, так и с затруднениями при восстановлении отсутствующих значений методом «ближайших соседей».

Таблица 1

Точность классификации примеров при наличии шума «перемешивание значений» в обучающем множестве

Набор данных	Точность классификации, %					
	No noise	Шум 5%	Шум 10%	Шум 15%	Шум 20%	Шум 25%
Monks-1	100	96,60	93,06	83,3	79,75	68,46
Monks-2	74,31	71,64	65,86	65,62	61,92	63,31
Monks-3	94,44	92,48	88,89	86,92	87,5	77,2

Таблица 2

Точность классификации примеров при наличии шума «отсутствие значений» в обучающем множестве

Набор данных	Точность классификации, %					
	No noise	Шум 5%	Шум 10%	Шум 15%	Шум 20%	Шум 25%
Monks-1	100,00	99,64	95,83	97,38	96,43	87,04
Monks-2	74,31	72,53	70,22	69,56	71,76	68,06
Monks-3	94,44	94,44	91,59	92,51	90,01	92,55

Предложенный алгоритм GIRS совместно с алгоритмом дискретизации позволяет снизить время поиска существенных атрибутов. Для рассмотренных наборов данных при наличии шума точность классификации тестовых примеров данным алгоритмом не хуже, чем у таких алгоритмов обобщения, как C 4.5, CN2 [4], а в некоторых случаях превосходит их.

### **Заключение**

Были рассмотрены проблемы, возникающие при обработке противоречивой информации. Предложены подходы для решения этих проблем, а именно использовать теорию аргументации и методы теории приближенных множеств.

Система аргументации, основанная на теории пересматриваемых рассуждений, может быть применена в различных ИСППР, использующих ЛППП в качестве языка представления знаний. Однако, проблема вычисления степеней обоснования остается открытой, и реализация различных методов их вычисления является предметом дальнейших исследований. Также были рассмотрены модели шума в таблицах баз данных, следствием которых является отсутствие значений признаков, либо искажение значения признаков в обучающей выборке. Предложен алгоритм GIRS, позволяющий обрабатывать обучающие выборки, содержащие примеры с неизвестными или искажёнными значениями, в сочетании с алгоритмами восстановления. Разработана и программно реализована система построения обобщенных понятий, которая использует полученные теоретические результаты и создана на основе предложенных алгоритмов.

Полученные результаты моделирования показали, что алгоритм GIRS в сочетании с алгоритмами восстановления позволяет повысить точность классификации примеров при наличии шума в данных.

### Литература

1. *Vagin V.N., Ereemeev A.P.* Methods and Tools for Modelling Reasoning in Diagnostic Systems. ISEIS 2009, Proceedings of the 11th International Conference on Enterprise Information Systems, vol. AIDSS, Milan, Italy, 2009. pp 271-276.
2. *Вагин В.Н., Головина Е.Ю., Загорянская А.А., Фомина М.В.* Достоверный и правдоподобный вывод в интеллектуальных системах / Под ред. В.Н. Вагина, Д.А. Поспелова. – 2-е издание дополненное и исправленное. – М.: ФИЗМАТЛИТ, 2008. – 712 с.
3. *Pawlak Z.* Rough sets and intelligent data analysis / Information Sciences, Elsevier Science. 2002. Vol. 147. iss. 1. pp. 1-12.
4. *Vagin V.N., Kulikov A.V., Fomina M.V.* The Development of the Generalization Algorithm based on the Rough Set Theory. International Journal INFORMATION Theories & Applications (ITHEA), 2006. Vol.13 № 3, pp. 255-262.
5. *Vagin V., Fomina M.* Problem of Knowledge Discovery in Noisy Databases. In Int. J. Mach. Learn. & Cyber. 2011. Vol.2, num. 3, Springer Ferlag, Berlin, pp. 135-145.
6. *Merz C, Murphy P.* 1998. UCI Repository of Machine Learning Datasets. Information and Computer Science University of California, Irvine, CA 92697-3425 <http://archive.ics.uci.edu/ml/>

### Argumentation in inductive notion formation

*Vadim Nikolaevich Vagin, Doctor of Engineering, Professor, Chair of applied mathematics*

*Marina Vladimirovna Fomina, Associate professor, Candidate of Technical Sciences, Chair of computer facilities*

*National Research University Moscow Power Institute*

*The methods of processing incomplete and inconsistent information in such subsystems of intelligent decision support systems as a decision searching subsystem of acquiring and storing knowledge are considered. In the decision searching subsystem, the main stress is made on the application of argumentation with justification degrees. In the acquiring and completion subsystem, the generalization problem in conditions of inconsistent and incomplete data is solved. For data generalization, the algorithm based on the rough set theory is developed. Research of noise influence on the work of the suggested generalization algorithm is carried out. The results of program modelling are presented.*

*Keywords: argumentation, defeasible reasoning, degrees of justification, rough sets, notion formation, generalization, noisy data.*

УДК 519.673

### СХЕМА ДЕЙСТВУЮЩЕГО ПРОТОТИПА МУЛЬТИАГЕНТНОЙ МОДЕЛИ РАСПРОСТРАНЕНИЯ ЭПИДЕМИЙ

*Светлана Юрьевна Лапшина, научный сотрудник*

*Тел.: 8 916 518 65 80, e-mail: lapshina@jscc.ru*

*Федеральное государственное бюджетное учреждение науки Межведомственный суперкомпьютерный центр Российской академии наук (МСЦ РАН)*

*<http://www.jscc.ru>*