

Описание структуры фрейма. Сразу после добавления свойств терма в слоты соответствующих фреймов происходит вызов демона *IF-ADDED*, запуск которого осуществляется при каждой очередной вставке значения в тот или иной слот. Демон *IF-ADDED*, в зависимости от того, каким фреймом он был вызван, вызывает соответствующую процедуру приведения значения свойства терма, хранящегося в слоте к форме на целевом языке.

Так, например, при вставке в слот «число» значения «множественное» демон *IF-ADDED* вызовет процедуру преобразования формы терма из множественного числа исходного языка к множественному числу целевого языка. Для этого преобразующая процедура обращается к имеющемуся в системе словарю окончаний для пары «ИЯ–ЦЯ». Таким образом, формируется база знаний на основе фреймового представления (рис. 2), хранящая информацию о структуре сегментов текста и термов, образующих эти сегменты.

Данная модель пригодна для осуществления сопоставления исходного и целевых текстов на этапе синтеза ЦЯ-текста в IP-системе машинного перевода,

а также для выполнения посегментного перевода фраз ИЯ-текста в фразы ЦЯ-текста.

Рассмотренная фреймовая модель является эффективным средством представления знаний в IP-системе машинного перевода на этапе синтеза текста, так как позволяет управлять формой термов при переходе от исходного языка к целевому, является менее громоздкой, чем представление через нейронную сеть, и более гибкой, чем продукционное представление. Приведенный алгоритм сегментации текста через связь позволяет быстро и эффективно производить разбиение текстового массива на фрагменты, что ускоряет процесс их анализа.

Библиографические ссылки

1. Мультилингвистическая модель распределенной системы на основе тезауруса / П. В. Зеленков, И. В. Ковалев, М. В. Карасева, С. В. Рогов // Вестник СибГАУ. Вып. 1(18). 2008. С. 26.
2. Заболеева-Зотова А. В., Камаев В. А. Лингвистическое обеспечение автоматизированных систем. М. : Высш. шк., 2008. С. 174–177.

K. V. Polyansky

TRANSLATING FRAME MODEL CONSTRUCTION WITH USE OF TERMS CLUSTERING

The knowledge representation frame model in IP-systems of machine translation is considered. The segmentation algorithm of the source and target text through communication is offered. Various terms clustering methods are analysed, the most suitable are offered to clusters text splitting.

Keywords: machine translation, text segmentation, terms clustering, frame model.

© Полянский К. В., 2011

УДК 519.8

Е. С. Семенкин, А. А. Шабалов, С. Н. Ефимов

АВТОМАТИЗИРОВАННОЕ ПРОЕКТИРОВАНИЕ КОЛЛЕКТИВОВ ИНТЕЛЛЕКТУАЛЬНЫХ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ МЕТОДОМ ГЕНЕТИЧЕСКОГО ПРОГРАММИРОВАНИЯ*

Рассмотрены методы автоматизированного проектирования интеллектуальных информационных технологий (ИИТ) для решения сложных задач анализа данных и принятия решений. При генерации нейросетевых моделей, систем на нечеткой логике и нейро-нечетких систем применяются эволюционные алгоритмы. В проектировании коллектива ИИТ с целью повышения эффективности и надежности системы предложено применять метод генетического программирования.

Ключевые слова: нейронные сети, системы на нечеткой логике, нейро-нечеткие системы, эволюционные алгоритмы, генетическое программирование, коллективное принятие решений.

На сегодняшний день интеллектуальные системы получили широкое распространение при решении сложных задач анализа данных в различных областях человеческой деятельности. Искусственные нейронные сети [1], нечеткая логика [2], нейро-нечеткие сис-

темы [3], эволюционные алгоритмы [4] и другие методики и технологии являются популярным объектом исследования в силу их способности решать сложные интеллектуальные задачи, которые трудно решить с помощью классических методов [5].

* Работа выполнена при финансовой поддержке ФЦП «Научные и научно-педагогические кадры инновационной России» (НИР НК-136П/3, гос. контракт П11007) и ФЦП «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007–2013 годы» (НИР 2011-1.9-519-005-042, гос. контракт 11.519.11.4002).

Гибридные подходы привели к значительному росту интереса к сфере проектирования ИИТ. Наиболее популярными являются гибридизация нейронных сетей и генетических алгоритмов (ГА), нечеткой логики и ГА, нейронных сетей и аппарата нечеткой логики.

Автоматическое проектирование ИИТ – сложная оптимизационная задача, структура которой не позволяет эффективно решить ее с помощью классических методов. Генетические алгоритмы представляют собой параллельную, робастную поисковую оптимизационную процедуру, основанную на принципах естественной эволюции. Популяция индивидов, представляющих решение, приспособляется в процессе эволюции посредством генетических операторов, таких как селекция, рекомбинация и мутация, максимизируя таким образом пригодность (минимизируя функцию стоимости) [6]. ГА продемонстрировали высокую эффективность при решении сложных практических задач с многоэкстремальными целевыми функциями. Кроме их возможности находить субоптимальные решения в сложных пространствах, структура кодирования и независимость от свойств меры качества делает их подходящим инструментом, позволяющим включить в себя априорные знания об исследуемом объекте или процессе [7].

Алгоритм генетического программирования оперирует компьютерными программами, представленными деревьями. Желаемая цель достигается путем выращивания популяции деревьев с использованием принципа естественного отбора и генетических операторов [8].

Для решения поставленной задачи необходимо определить функциональное множество (множество используемых функций) и терминальное множество (набор переменных функций, используемые типы констант), которые должны обладать свойствами замкнутости и достаточности.

Автоматизированное проектирование ИИТ. Рассмотрим применение генетических алгоритмов при генерации нейросетевых моделей, систем на нечеткой логике и нейро-нечетких систем.

Нейросетевые модели. В качестве структуры нейронной сети был взят многослойный перцептрон как наиболее распространенный вид нейросетевых моделей. Обучение данной структуры, как правило, осуществляется на основе алгоритма обратного распространения ошибки. Сложность проектирования заключается в исходном выборе количества скрытых слоев и числа нейронов на каждом скрытом слое, структура которых для каждой задачи заранее неизвестна, а также в выборе функций активации нейронов. Недостатками алгоритма обратного распространения ошибки являются низкая скорость сходимости, чувствительность к шуму, зависимость качества функционирования от эвристически выбираемого шага обучения, и, как правило, недостижение глобально экстремума ошибки моделирования ввиду ее сложности [9].

Для преодоления упомянутых проблем предлагается применить генетические алгоритмы для генерирования структуры перцептрона и настройки его весовых коэффициентов.

Хромосома структуры нейронной сети кодируется следующим образом. Изначально исследователем задается максимальное количество скрытых слоев и максимальное количество нейронов на каждом слое. При кодировании проход по сети проводится сверху вниз, слева направо по каждому нейрону. Длина хромосомы каждого нейрона будет состоять из 4 или 5 бит при использовании набора из 8 или 16 активационных функций соответственно. Первый бит указывает на наличие или отсутствие нейрона в сети, остальные биты кодируют информацию о номере функции активации. Таким образом, длина хромосомы, кодирующей структуру нейронной сети, будет равна максимальному количеству нейронов, умноженному на длину хромосомы одного нейрона.

Весовые коэффициенты кодируются аналогично. Исследователем задается интервал изменения весовых коэффициентов и точность (дискретизация числа), на основании чего определяется количество бит n для кодирования вещественного числа. Длина хромосомы, кодирующей весовые коэффициенты, равно числу n , умноженному на число всех коэффициентов текущей сети.

Таким образом, необходимо сгенерировать популяцию индивидов, представляющих структуру нейронной сети, и для каждой структуры – отдельную популяцию весовых коэффициентов. Условием останова является либо число поколений структуры, либо ошибка моделирования (обычно среднеквадратичная).

Системы на нечеткой логике. При проектировании нечетких систем эксперт сталкивается с проблемой выбора исходных нечетких правил, набор которых может оказаться неполным или противоречивым. Выбор параметров функций принадлежности, описывающих входные и выходные переменные объекта, происходит субъективно и может некорректно отображать реальную действительность [10].

В силу этого для повышения обоснованности принятия решений при проектировании нечетких систем были применены ГА. При генерации нечеткой системы использовался питтсбургский подход. Изначально необходимо задать максимальное количество правил N_{\max} . В процессе обучения настраиваются следующие параметры базы знаний: параметры функций принадлежности, определяющие центр и ширину термов, количество правил, общая длина всей базы правил благодаря включению дополнительного термина – термина «игнорирование». Схемы кодирования хромосомы по Мамдани и по Такаги–Сугено (нулевого порядка) представлены на рис. 1 и 2 соответственно.

Нейро-нечеткие системы. Процесс генерирования нейро-нечетких моделей происходит в два этапа [11; 12]. Первый этап («обучение без учителя») заключается в кластеризации исходных числовых данных методами конкурентного обучения с механизмом штрафования соперника, алгоритмом адаптивной резонансной теории и др., после чего получаем «грубые» нечеткие правила. Второй этап («обучение с учителем») состоит в «тонкой» настройке базы правил (семантики лингвистических термов) на основе градиентных алгоритмов, недостатки которых широко известны и пре-

пятствуют эффективному использованию нейронечетких систем. Поэтому в работе применяется ГА, эффективность использования которого на практических задачах показана в [13; 14] и превосходит метод наискорейшего спуска.

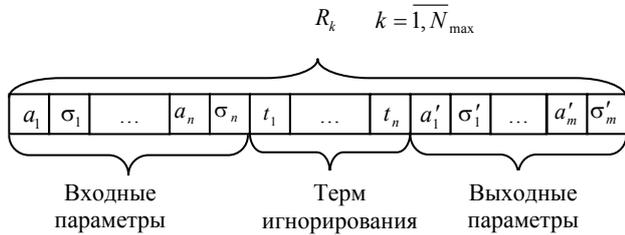


Рис. 1. Представление кодирования хромосомы по Мамдани

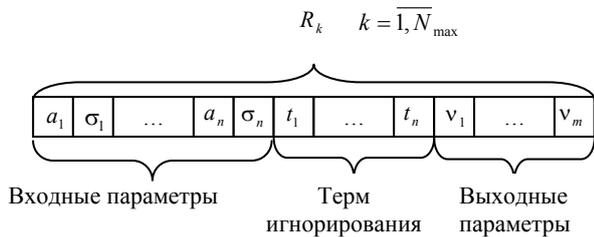


Рис. 2. Представление кодирования хромосомы по Такаги–Сугено нулевого порядка

Формирование коллектива на основе метода генетического проектирования. С целью повышения эффективности и надежности системы предлагается применять метод генетического проектирования для формирования состава коллектива ИИТ при решении сложных задач и формы взаимодействия участников коллектива при получении результирующего решения, основанного на частных решениях индивидуальных технологий.

При таком подходе существует два варианта гибридизации ИИТ.

Первый вариант заключается в построении математического выражения из результатов отдельных членов коллектива. В данном случае элементами терминального множества алгоритма генетического программирования являются частные решения отдельных систем. Функциональное множество при данном подходе включает в себя математические функции и арифметические операции.

На предварительном этапе необходимо сгенерировать и обучить заданное (установленное пользователем) число элементов терминального множества, которые будут в дальнейшем использоваться в работе алгоритма. В данной постановке задачи существует два способа реализации мутации: либо случайным образом выбирать обученный элемент из терминального множества, либо генерировать полностью новую интеллектуальную систему.

Второй вариант подхода заключается в формировании гибридной многослойной системы, состоящей из отдельных элементов этого коллектива. Терминальное множество представляется отдельной структурой ИИТ. Элементами функционального множества в этом случае являются правила соединения отдельных технологий между собой.

При реализации данного подхода заранее необходимо задать число циклов обучения каждой технологии. При генерации новой популяции индивидов получают новые структуры, вследствие чего каждая отдельная технология заново проектируется.

Апробация предложенных подходов на практических задачах. Для апробации предложенных схем автоматизированного проектирования интеллектуальных информационных технологий была разработана программная система для решения прикладных задач моделирования, классификации, кластеризации, прогнозирования и управления (рис. 3).

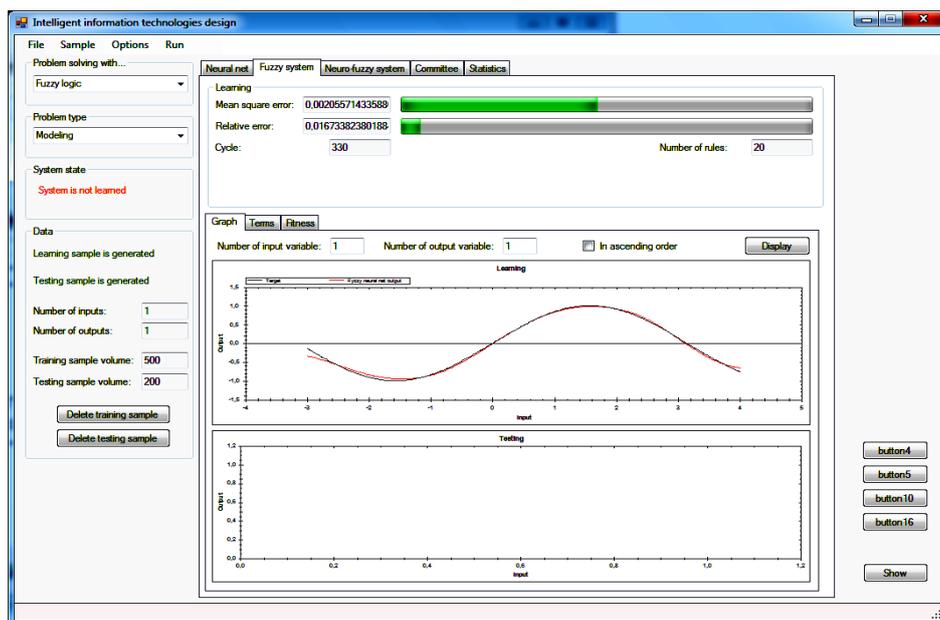


Рис. 3. Главное окно программы

Апробация методов автоматизации ИИТ

Задача	Размерность входа	Размерность выхода	Объем выборки	
			Обучение	Экзамен
Репозиторий машинного обучения UCI				
Классификация ирисов	4	3	135	15
Классификация вин	13	3	163	15
Предсказание лесных пожаров	12	1	477	40
Распознавание типов земель по спутниковым изображениям	36	6	4 435	2 000
Практические задачи				
Прогнозирование состояния турбины по вибрационным характеристикам	11	12	1 000	400
Моделирование процесса рудно-термической плавки	9	1	47	10
Прогнозирование деградации электрических характеристик солнечных батарей космического аппарата	7	4	177	20

При реализации генетического алгоритма за основу был принят ГА с модифицированным оператором множественной равномерной рекомбинации ввиду его более высокой эффективности по сравнению со стандартным ГА [15; 16].

Для апробации подходов использовались практические задачи (табл. 1). Часть из них была взята из репозитория машинного обучения UCI.

Для каждого типа ИИТ было выполнено 20 запусков программы. При каждом запуске были получены работоспособные системы. Лучшие результаты приведены в табл. 2.

Таблица 2

Результаты решения практических задач

№	Нейронная сеть		Система на нечеткой логике		Нейро-нечеткая система	
	О	К	О	К	О	К
	Ошибка					
1	5	1	2	0	2	0
2	1	1	0	0	0	0
3	–	–	16,87 %	19,61 %	15,67 %	17,5 %
4	1,78 %	1,79 %	1,11 %	1,11 %	1,45 %	1,46 %
5	9,11 %	9,14 %	8,07 %	8,09 %	7,99 %	7,97 %
6	4,86 %	4,97 %	2,99 %	3,01 %	2,81 %	2,92 %
7	–	–	5,66 %	7,66 %	5,05 %	5,87 %

Примечания. Приняты следующие обозначения: О – обучающая выборка, К – контрольная выборка. В задачах классификации в качестве ошибки приводится количество неправильно классифицируемых признаков, в остальных задачах – средняя относительная ошибка.

По данным табл. 2 видно, что во многих случаях наилучшее качество моделирования имеет нейро-нечеткая система. При этом все типы ИИТ демонстрируют эффективность, сопоставимую с известными результатами.

Ниже приведены примеры построения коллектива на основе математического выражения из частных решений отдельных технологий. На начальной стадии заранее были сгенерированы и обучены по 10 ИИТ каждого типа.

На примере задачи классификации вин была получена следующая формула:

$$C = \sin \left(NFN_4 \cdot \sqrt{e^{NFN_{10}}} \right),$$

где C – номер класса. При этом ошибка распознавания составила 0 % как на обучающей выборке, так и на контрольной, что лучше, чем у нейросетевых и нечетких моделей по отдельности и сравнимо с нейро-нечеткими. При этом отдельные ИИТ, превосходящие остальные по качеству моделирования, в коллектив не вошли.

В задаче моделирования процесса рудно-термической плавки получено выражение

$$Ni(\%) = NFN_{10} \cdot e^{\frac{FIS_6}{FIS_6 \cdot e^{NFN_9}} \cdot \frac{1}{FIS_{10}}},$$

определяющее процентное содержание никеля в отработанном шлаке. Относительная ошибка составила 2,21 % на обучающей выборке и 2,33 % на тестовой, что лучше, чем у каждой индивидуальной ИИТ в отдельности.

В ходе многочисленных экспериментов было выявлено, что отдельные технологии, превосходящие по качеству моделирования остальные, не всегда присутствуют в конечной формуле. В композиции могут принимать участие технологии с разным качеством моделирования, коллектив которых позволяет тем самым увеличить эффективность и надежность системы в целом.

Таким образом, программная система, реализующая разработанный подход, дает возможность автоматически генерировать нейросетевые модели, нечеткие и нейро-нечеткие системы, т. е. позволяет решать сложные задачи анализа данных и принятия решений (задачи моделирования, классификации, прогнозирования, управления и др.). Автоматическое формирование коллектива ИИТ позволяет увеличить надежность и эффективность системы. Полученные результаты подтверждены решением ряда практических задач.

Дальнейшее развитие системы направлено на расширение ее функциональности за счет включения других типов ИИТ (динамических нейросетей, сетей Кохонена и Хопфилда–Танка, деревьев решений, многокритериального выбора и т. д.), повышения адаптивности за счет автоматизации настройки эволюционных алгоритмов, а также доработки методов проектирования коллективов при сетевом подходе.

Библиографические ссылки

1. Rojas R. Neural networks: a systematic introduction. Berlin : Springer, 1996.
2. Yager R. R., Filev D. P. Essentials of fuzzy modeling and control. N. Y. : Wiley, 1994.
3. Tommi O. Neuro-fuzzy in control // Tampere Univ. of Technology. Tampere, 1995.
4. Eiben A. E., Smith J. E. Introduction to evolutionary computation. Berlin : Springer, 2003.
5. Konar A. Computational Intelligence: Principles, techniques and applications. Berlin : Springer, 2005.
6. Haupt R. L., Haupt S. E. Practical Genetic Algorithms. N. Y. : Wiley-Interscience, 2004.
7. Goldberg D. E. Genetic algorithms in search, optimization, and machine learning. Reading : Addison-Wesley, 1989.
8. Koza J. R. Genetic programming. London : MIT Press, 1998.
9. Уоссермен Ф. Нейрокомпьютерная техника: теория и практика : пер. с англ. М. : Мир, 1992.
10. Круглов В. В., Дли М. И., Голунов Р. Ю. Нечеткая логика и искусственные нейронные сети. М. : Физматлит, 2001.
11. Castellano G., Fanelli A. M. A self-organizing neural fuzzy inference network // Proc. of IEEE Int. Joint Conf. on Neural Networks (IJCNN-2000). Vol. 5. Como, 2000. P. 14–19.
12. Castellano G., Fanelli A. M. Information granulation via neural network based learning // Proc. of Joint 9th IFSA World Congress and 20th NAFIPS Intern. Conf. (IFSA-NAFIPS 2001). Vancouver, 2001. P. 3059–3064.
13. Шабалов А. А. Автоматическая настройка параметров нейро-нечетких систем генетическим алгоритмом // Информ. технологии и мат. моделирование : сб. науч. тр. VIII Всерос. науч.-практ. конф. с междунар. участием. Ч. 2. Томск, 2009. С. 296.
14. Шабалов А. А. Автоматизация проектирования нейро-нечетких систем с помощью эволюционного алгоритма // Студент и науч.-техн. прогресс // Материалы XLVIII междунар. науч. студ. конф. Новосибирск, 2010. С. 290.
15. Шабалов А. А. Исследование эффективности модификации оператора равномерного скрещивания эволюционного алгоритма в задаче управления реальными инвестициями // Информ. технологии и мат. моделирование : сб. науч. тр. VII Всерос. науч.-практ. конф. с междунар. участием. Ч. 1. Томск, 2008. С. 175–177.
16. Шабалов А. А. Применение эволюционных алгоритмов с модифицированным оператором равномерного скрещивания в задаче поддержки принятия решений при управлении реальными инвестициями // Решетневские чтения : материалы XII междунар. науч. конф. Красноярск, 2008. С. 284–285.

E. S. Semenkin, A. A. Shabalov, S. N. Efimov

AUTOMATED DESIGN OF INTELLIGENT INFORMATION TECHNOLOGY KOLLEKTIVES BY MEANS OF GENETIC PROGRAMMING TECHNIQUE

Intelligent information technology automated design for solving complex problem of data analysis and decision support is considered in this paper. Evolutionary algorithms are applied for generating connectionist models, fuzzy systems and neuro-fuzzy systems. Genetic programming technique application for intelligent information technologies kollektives design is proposed for system effectiveness and reliability improvement.

Keywords: neural networks, fuzzy rules based systems, neuro-fuzzy systems, evolutionary algorithms, genetic programming, group decision making.

© Семенкин Е. С., Шабалов А. А., Ефимов С. Н., 2011