

9. Курейчик В.М., Писаренко В.И., Кравченко Ю.А. Инновационные образовательные технологии в построении систем поддержки принятия групповых решений // Известия ЮФУ. Технические науки. Тематический выпуск "Интеллектуальные САПР". – 2008. – № 4 (81). – С. 216-221.
10. Городецкий В.И. Многоагентные системы (обзор) / В.И. Городецкий, М.С. Грушинский, А.В. Хабалов // Новости искусственного интеллекта. – М.: Изд-во ЦНИЭИуголь, 1998. – № 2. – 196 с.

Кравченко Юрий Алексеевич

Технологический институт федерального государственного образовательного учреждения высшего профессионального образования «Южный федеральный университет» в г. Таганроге.

E-mail: krav-jura@yandex.ru.

347928, г. Таганрог, пер. Некрасовский, 44.

Тел.: 88634371651.

Кафедра систем автоматизированного проектирования, доцент.

Kravchenko Yuriy Alekseevich

Taganrog Institute of Technology – Federal State-Owned Educational Establishment of Higher Vocational Education “Southern Federal University”.

E-mail: krav-jura@yandex.ru.

44, Nekrasovskiy, Taganrog, 347928, Russia.

Phone: 88634371651.

The Department of Computer Aided Design; associate professor.

УДК 519.712.2

Л.А. Гладков, Н.В. Гладкова

ОСОБЕННОСТИ ПОСТРОЕНИЯ СИСТЕМ АНАЛИЗА И ИЗВЛЕЧЕНИЯ ЗНАНИЙ НА ОСНОВЕ НЕЧЕТКИХ ГИБРИДНЫХ МЕТОДОВ*

В работе приводятся основные отличия методов Data Mining от традиционных методов анализа. Также обсуждаются преимущества и недостатки методов Data Mining и приводятся предложения по их решению на основе использования гибридных интеллектуальных технологий и методов вычислительного интеллекта. Также рассмотрены основные аспекты применения нечетких генетических алгоритмов для решения задач извлечения знаний. Приведена обобщенная структура нечеткого логического контроллера и описана основная идея его применения. Описаны основные компоненты организации и процесса взаимодействия генетического алгоритма и нечеткого логического контроллера. Кратко представлены основные направления и проблемы создания нечетких генетических алгоритмов. В заключении приведено обоснование актуальности разработки новых гибридных методов анализа и извлечения данных.

Анализ и извлечение знаний; нечеткий генетический алгоритм; нечеткий логический контроллер; фаззификация; дефаззификация.

L.A. Gladkov, N.V. Gladkova

FEATURES OF CONSTRUCTION OF DATA MINING SYSTEMS ON THE BASIS OF FUZZY HYBRID METHODS

In article the basic differences of methods Data Mining from traditional methods of the analysis are resulted. Also advantages and lacks of methods Data Mining are discussed and offers under their decision on the basis of use of hybrid intellectual technologies and methods of compu-

* Работа выполнена при поддержке: РФФИ (грант № 08-01-00473), г/б № 2.1.2.1652.

ting intelligence are resulted. Also the basic aspects of application of fuzzy genetic algorithms for the decision of problems of extraction of knowledge are considered. The generalized structure of the fuzzy logic controller is resulted and the basic idea of its application is described. The basic components of the organization and process of interaction of genetic algorithm and the fuzzy logic controller are described. The basic directions and problems of creation of fuzzy genetic algorithms are short presented. In the conclusion the substantiation of an urgency of working out of new hybrid methods of the analysis and extraction of data is resulted.

Data mining; fuzzy genetic algorithm; fuzzy logic controller; fuzzification; defuzzification..

Проблема создания эффективных систем интеллектуального анализа и извлечения знаний (Data Mining) из имеющихся массивов данных сегодня чрезвычайно актуальна. Поэтому разработку новых теоретических и прикладных подходов к решению данной проблемы целесообразно объединить в отдельную междисциплинарную область. Междисциплинарный характер, отсутствие четких границ проблемной области, а также использование инструментария различных научных направлений, позволяет рассматривать данную научную задачу в качестве одного из направлений искусственного интеллекта.

Задача сводится к разработке новых эффективных технологий выявления в больших массивах данных неявной и неструктурированной информации, неочевидных, но полезных закономерностей. По сравнению с другими понятиями данной предметной области, такими как данные, информация, понятие «знание» определяется как совокупность фактов, закономерностей и эвристических правил, с помощью которых решается поставленная задача [1]. Согласно другому определению, «знания – это абсолютное использование информации и данных, совместно с потенциалом практического опыта людей, способностями, идеями, интуицией, убежденностью и мотивациями».

Отличительными особенностями знаний как особой понятийной категории являются: структурированность, компактность и внутренняя непротиворечивость. При этом к понятию компактности знаний можно отнести и такое свойство, как лаконичность, отсутствие посторонних, не относящихся к изучаемому предмету данных, и удобство доступа и усвоения новых знаний и т.д.

Для решения задачи анализа и извлечения знаний, очевидным представляется использование методов, относящихся к вычислительному интеллекту: нейросетевые алгоритмы, нечеткие модели и методы, биоинспирированные алгоритмы, экспертные системы. Эти технологии, с одной стороны, позволяют эффективно работать с нечеткой, плохо формализованной информацией, с другой стороны, они имеют серьезную математическую основу, обеспечивающую достаточный запас прочности. Важно, что технологии вычислительного интеллекта уже давно и эффективно используются при решении различных задач анализа и принятия решений в условиях нечеткой, плохо формализованной, а зачастую и противоречивой входной информации.

Так, например, искусственные нейронные сети эффективно используются для решения задач прогнозирования и оценки, эволюционные и генетические алгоритмы представляют собой мощное средство поиска оптимальных решений на больших объемах данных, математический аппарат нечеткой логики является основой для построения различных систем логического вывода и представления знаний и т.д.

Основными проблемами при построении современных систем анализа и извлечения знаний являются [1]:

- ◆ сложность разработки и эксплуатации;
- ◆ сложность подготовки данных;

- ◆ большой процент недостоверных или бессмысленных решений;
- ◆ высокая стоимость.

По мнению различных экспертов из-за существенных различий между инструментами разработчиков программного обеспечения технологии анализа и извлечения знаний перед применением необходимо тщательно изучить на предмет их совместимости и корректности будущих результатов. При этом считается, что результаты применения технологий Data Mining на восемьдесят процентов зависят от уровня подготовки исходных данных, который выполняется до начала работы собственно алгоритма.

Среди задач, решаемых в процессе выделения знаний, особое место занимают задачи классификации и кластеризации. Эти задачи имеют много общего, но в то же время между ними имеются принципиальные различия. Классификация – это распределение изучаемых предметов по каким-либо существенным признакам и расположение их в определенном порядке, отражающем степень этого сходства. Таким образом, процесс классификации заключается в разбиении начального множества объектов на классы по определенному критерию [1].

Основное отличие задачи кластеризации от задачи классификации в том, что при кластеризации классы на которые необходимо разбить исходное множество, заранее не определены. Таким образом, целью кластеризации является поиска существующих в изучаемом множестве однородных структур [1].

Для успешного решения вышеуказанных проблем необходимо привлечение из различных предметных областей, тщательный предварительный анализ и выбор модели данных, а также эффективная проверка адекватности полученных моделей.

Одним из перспективных направлений развития методов Data Mining является выделение типов предметных областей с соответствующими эвристиками, что позволяет решать различные задачи, а также создание эффективных методов Data Mining, способных не только извлекать закономерности из массивов данных, но и формулировать на основе полученных эмпирических данных определенные теоретические положения и выкладки.

Это определяет необходимость использования в Data Mining возможностей математических методов и последних разработок в области создания интеллектуальных информационных систем. Сюда можно отнести искусственные нейронные сети, кластерный анализ, байесовские сети, эволюционное программирование и генетические алгоритмы, методы визуализации данных и др.

Практика использования современных интеллектуальных систем показывает, что разработка четких математически обоснованных моделей и методов, либо экономически неприемлема, либо практически нереализуема. В то же время системы, функционирующие на основе использования интегрированных, нечетких гибридных механизмов и моделей прекрасно зарекомендовали себя при решении такого рода задач, и представляют собой наиболее разумный компромисс.

В настоящее время наибольшие успехи в интеграции систем и подходов нечеткой логики и генетических алгоритмов (ГА) достигнуты в следующих двух областях [2]:

- 1) применение механизмов генетических и эволюционных алгоритмов для решения проблем оптимизации и поиска в условиях нечеткой, неопределенной или недостаточной информации об объекте, параметрах и критериях решаемой задачи, совместно с использованием систем нечетких логики;
- 2) использование нечетких инструментов и методов, основанных на нечеткой логике для моделирования различных компонентов и операторов генетических

алгоритмов, а также для адаптации и управления основными параметрами генетического алгоритма для динамической настройки и улучшения работы ГА.

Также с точки зрения развития методов Data Mining интерес представляет применение нечетких и генетических алгоритмов для выявления и использования формализованных закономерностей и формирования и оценки шаблонов.

Математический аппарат теории нечетких систем используется для кодирования, подбора оптимальных параметров генетических алгоритмов, значений вероятности генетических операторов, выбора функции пригодности и критерия останова, создания нечетких генетических операторов. Основными направлениями разработки НГА можно считать следующие.

- ♦ **Контроль и динамическое изменение основных управляющих параметров генетического алгоритма.** Для контроля и динамического изменения соответствующих параметров генетического алгоритма в систему вводится нечеткий логический контроллер (НЛК), который на основе накопленного опыта либо доступных экспертных оценок, соответствующим образом изменяет параметры генетического поиска в ходе выполнения ГА для того, чтобы избежать преждевременной сходимости.

В генетических алгоритмах (ГА) управляющими параметрами, как правило, являются значения вероятности выполнения генетических операторов кроссинговера (P_c) и мутации (P_m), а также размер популяции. В нечетком генетическом алгоритме происходит динамическое изменение этих параметров при помощи нечеткого логического контроллера (НЛК).

НЛК преобразует заданные параметры к нечеткому виду, затем на основе имеющихся в системе знаний и правил определяет управляющее воздействие и возвращает скорректированные значения контрольных параметров.

В составе НЛК можно выделить следующие блоки [3]:

- 1) база знаний, включающая в себя базу правил и базу данных;
- 2) блок фаззификации;
- 3) блок дефаззификации;
- 4) система вывода решения;
- 5) система контроля.

В общем, схему работу НЛК можно описать следующим образом. НЛК использует знания экспертов в форме лингвистических правил контроля. Система выработки правил на основе знаний экспертов и, используя рассуждения, делает определенный вывод, который после дефаззификации, превращается из нечеткого правила в реальное воздействие на параметры алгоритма. Изменение параметров алгоритма влечет за собой изменение процесса поиска и текущих результатов, которые затем в блоке фаззификации из переменных состояния преобразуются в нечеткие множества [4].

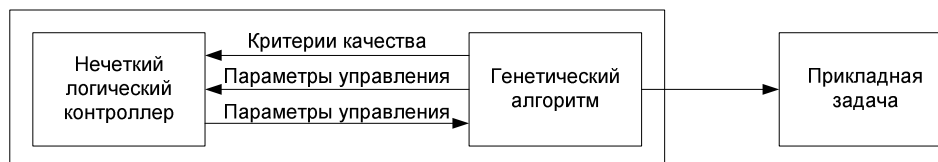


Рис. 1. Схема взаимодействия блоков нечеткого генетического алгоритма

На вход НЛК подается некая комбинация критериев качества работы ГА или текущих параметров контроля, а на выходе получают новые значения параметров контроля. При этом вероятности выполнения кроссинговера и мутации могут оп-

ределяться НЛК исходя из оценки не всей популяции, а по определенной выборке решений учитывающей значения функции пригодности и разнообразие популяции. Также для решения одной задачи могут использоваться несколько НЛК [5].

Известно, что работа генетического алгоритма напрямую зависит от выбора его параметров. Использование нечетких логических контроллеров, для изменения параметров генетического алгоритма позволяет более осторожно, взвешенно и целенаправленно контролировать работу генетического алгоритма.

- ◆ **Разработка и модификация операторов кроссинговера.** Нечеткие связи и треугольные функции распределения вероятности можно использовать для создания эффективных операторов кроссинговера, которые устанавливают адекватные уровни разнообразия популяции и таким образом позволяют решать проблему преждевременной сходимости.
- ◆ **Кодирование решений.** Классическое бинарное представление решений, когда гены принимают значения ноль или единица может быть преобразовано в нечеткое представление, где гены принимают значения в интервале между нулем и единицей. Это позволит выражать более сложные особенности, как генотипа, так и фенотипа различных решений популяции наподобие тех, которые встречаются в природе.
- ◆ **Критерии останова.** Такие характеристики нечеткой логики как степень истинности, оценка достоверности той или иной логической функции могут быть использованы для получения оптимальных решений с заданной пользователем точностью, прогнозирования и регулирования критериев останова генетического алгоритма по мере достижения заданных значений.

Еще одним перспективным подходом к организации структуры методов Data Mining является использование мультиагентных архитектур.

Понятие агента применительно к различным информационным системам может трактоваться по-разному. Многоагентная система может рассматриваться как популяция простых и независимых агентов, каждый агент которой самостоятельно реализуется в локальной среде и взаимодействует с другими агентами. Связи между различными агентами являются горизонтальными, а глобальное поведение агентов определяется на основе расплывчатых правил.

Интеграция различных направлений и методов вычислительного интеллекта и создание на этой основе новых гибридных технологий решения слабоформализованных задач одно из перспективных направлений исследований в области Data Mining. Основой для подобной интеграции является их терпимость к нечеткости и противоречивости используемых данных, гибкость и относительно низкая себестоимость [6]. Примерами таких гибридных технологий являются нечеткие нейронные сети, нечеткие генетические, эволюционные и адаптивные алгоритмы и т.д. [7,8]. Активная разработка новых форм и направлений подобной интеграции сейчас активно ведется как России, так и за рубежом.

Очевидно, что создание эффективных методов решения задач извлечения знаний на основе гибридных нечетких генетических методов находится только в самом начале пути. Однако уже имеется достаточно большое число примеров успешного использования таких методов для решения различных прикладных задач извлечения знаний.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Чубукова И.А. Data Mining. Учебное пособие. – М.: Интернет-Университет Информационных Технологий; БИНОМ. Лаборатория знаний, 2006.
2. Herrera F., Lozano M. Fuzzy Genetic Algorithms: Issues and Models. – Source unknown.
3. Herrera F., Lozano M. Fuzzy Adaptive Genetic Algorithms: design, taxonomy, and future directions // Soft Computing 7(2003), Springer-Verlag, 2003. – P. 545-562.
4. Hongbo Liu, Zhanguo Xu, Ajith Abraham. Hybrid Fuzzy-Genetic Algorithm Approach for Crew Grouping. – Source unknown.
5. Курейчик В.В., Сороколетов П.В., Щеглов С.Н. Анализ современного состояния автоматизированных систем приобретения и представления знаний // Известия ЮФУ. Технические науки. – 2008. – № 9 (86). – С. 120-125
6. Борисов В.В., Круглов В.В., Федюлов А.С. Нечеткие модели и сети. – М.: Горячая линия-Телеком, 2007.
7. Курейчик В.В., Курейчик В.М., Родзин С.И. Концепция эволюционных вычислений, инспирированных природными системами // Известия ЮФУ. Технические науки. Тематический выпуск "Интеллектуальные САПР". – 2009. – № 4 (93). – С. 16-24.
8. Курейчик В.В., Курейчик В.М., Ковалев С.М. Оптонечеткие системы // Известия ЮФУ. Технические науки. Тематический выпуск "Интеллектуальные САПР". – 2009. – № 4 (93). – С. 99-106.

Гладков Леонид Анатольевич

Технологический институт федерального государственного образовательного учреждения высшего профессионального образования «Южный федеральный университет» в г. Таганроге.

E-mail: leo@tsure.ru.

347928, г. Таганрог, пер. Некрасовский, 44.

Тел.: 88634371625.

Кафедра систем автоматизированного проектирования; доцент.

Gladkov Leonid Anatolievich

Taganrog Institute of Technology – Federal State-Owned Educational Establishment of Higher Vocational Education “Southern Federal University”.

E-mail: leo@tsure.ru.

44, Nekrasovskiy, Taganrog, 347928, Russia.

Phone: 88634371625.

The Department of Computer Aided Design; associated professor.

Гладкова Надежда Викторовна

Технологический институт федерального государственного образовательного учреждения высшего профессионального образования «Южный федеральный университет» в г. Таганроге.

E-mail: leo@tsure.ru.

347928, г. Таганрог, пер. Некрасовский, 44.

Тел.: 88634371625.

Кафедра систем автоматизированного проектирования; старший преподаватель.

Gladkova Nadegda Viktorovna

Taganrog Institute of Technology – Federal State-Owned Educational Establishment of Higher Vocational Education “Southern Federal University”.

E-mail: leo@tsure.ru.

44, Nekrasovskiy, Taganrog, 347928, Russia.

Phone: 88634371625.

The Department of Computer Aided Design; senior teacher.