



На правах рукописи

Коротаев Александр Николаевич

**РАЗРАБОТКА АЛГОРИТМОВ
КЛАССИФИКАЦИИ ПРОЕКТОВ
ПЛАНИРОВКИ ТЕРРИТОРИИ
С ИСПОЛЬЗОВАНИЕМ ТЕОРИИ
НЕЧЕТКИХ МНОЖЕСТВ**

Специальность 05.13.10 –
«Управление в социальных и экономических системах»

Автореферат
диссертации на соискание ученой степени
кандидата технических наук

Рязань 2013

Работа выполнена в федеральном государственном бюджетном образовательном учреждении высшего профессионального образования «Рязанский государственный радиотехнический университет»

Научный руководитель:

Кираковский Валерий Владимирович
кандидат технических наук, доцент,
генеральный директор ООО «Проектный
институт «Промгражданпроект»

Официальные оппоненты:

Майков Константин Анатольевич
доктор технических наук, профессор
каф. Программного обеспечения ЭВМ
и информационных технологий
МГТУ имени Н.Э. Баумана

Рудомин Евгений Николаевич
кандидат технических наук, доцент
каф. Промышленного и гражданского
строительства
РИ(ф) МГОУ имени В.С. Черномырдина

Ведущая организация:

**ФГБОУ ВПО «Липецкий государ-
ственный технический университет»**

Защита диссертации состоится **18 декабря 2013 г. в 14 часов** на заседании диссертационного совета Д 212.211.02 в Рязанском государственном радиотехническом университете по адресу: **390005, г. Рязань, ул. Гагарина, 59/1.**

С диссертацией можно ознакомиться в библиотеке
Рязанского государственного радиотехнического университета.

Автореферат разослан 15 ноября 2013 г.

Ученый секретарь
диссертационного совета
канд. техн. наук, доцент



Д.А. Перепелкин

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования. На сегодняшний день проблема оценки проектов планировки территории входит в число часто встречаемых, поскольку, в связи с проведением большого количества конкурсов или тендеров на разработку документации, появилась проблема быстрой и эффективной оценки материалов, предоставляемых участниками. Также участились жалобы на необъективность этих оценок, и, в связи с этим, возросло число просьб обосновать принятое решение. Эти претензии не безосновательны, потому что, как показывает практика, оценка выставляется исходя из предпочтений членов конкурсной комиссии. Поскольку эксперты не обосновывают свое решение, то возникает возможность ошибки. В связи с увеличением объема строительства за последнее десятилетие, а, следовательно, и конкурсов по проектам планировки территории, вопрос о повышении объективности их оценок приобрел особую актуальность. Эта задача относится к задачам классификации.

Методы классификации можно разделить на несколько групп. По способу задания показателя качества классификации методы делятся на эвристические и оптимизационные. По способу объединения – на дивизимые, агломеративные и итеративные.

Наиболее распространенными на сегодняшний день являются оптимизационные методы классификации. Они обычно основываются на кластерном анализе.

Специфика рассматриваемой задачи заключается в выборе наиболее значимых для конкретных проектов данных. Помимо основных показателей по каждому проекту может предоставляться дополнительная информация, которую претендент считает важной. Она предоставляется в произвольной форме, поэтому число показателей по конкурирующим проектам может различаться. Проблема отбора наиболее значимых показателей является актуальной. Для ее решения существуют различные алгоритмы, такие как Forward Selection, Backward Elimination, Stepwise, Best Subsets и др.

В связи с тем, что число групп, на которые будут распределяться проекты изначально неизвестно, а также неизвестны четкие критерии отнесения проекта к той или иной группе, возникает проблема классификации проектов. В таких условиях очень хорошо себя зарекомендовал аппарат теории нечетких множеств (ТНМ), основы которой были заложены Л. Заде в 1965 г. Использование ТНМ предоставляет широкие возможности для моделирования зависимостей (взаимосвязей) между показателями проектов, что позволяет разрабатывать различные варианты решения задачи классификации проектов планировки территории.

При разработке алгоритмов кластеризации проектов планировки территории рассматривались математические и прикладные задачи, в решение которых значительный вклад внесли такие ученые, как: С.А. Айвазян, В.М. Бухштабер, В.Н. Вапник, И.С. Енюков, Ю.И. Журавлев, Н.Г. Загоруйко, В.В. Рязанов, М. Шлезингер, J. Friedman, T. Hastie, T. Mitchell, и др.

Существует достаточно большое число алгоритмов кластеризации, основанных на применении аппарата ТНМ. Однако их большим недостатком является большое время поиска субоптимального набора параметров в следствие необходимости многократного выполнения этих алгоритмов.

Для оптимизации времени поиска искомого набора параметров алгоритма кластеризации могут быть использованы генетические алгоритмы (ГА) – эвристические алгоритмы поиска, используемые для решения оптимизационных задач путём случайного подбора, комбинирования и вариации искомым параметров с использованием механизмов, аналогичных механизмам естественного отбора в природе.

Значительный вклад в решение математических и прикладных задач с использованием ГА внесли такие ученые, как: J.H. Holland, N.A. Bariccelli, Л.А. Гладков, В.В. Курейчик, В.М. Курейчик, Д. Рутковская и др.

Однако использование алгоритмов кластеризации и ГА позволяет найти только субоптимальное решение. Это означает, что при классификации проектов могут быть получены ошибочные результаты. В связи с этим целесообразно использовать алгоритмы уточнения результатов классификации. В решении этой проблемы очень хорошо себя зарекомендовали нейро-нечеткие сети и, в частности, адаптивная сеть нечеткого вывода ANFIS (Adaptive-Network-Based Fuzzy Inference System).

Значительный вклад в решение математических и прикладных задач с использованием нейро-нечетких сетей внесли такие ученые, как: J.S.R. Jang, A. Abraham, Y. Jin, B. Kosko, C. Quek, R.W. Zhou и др.

Еще один прогрессивный подход, обеспечивающий повышение качества классификационных решений, основан на разработке так называемых кластерных ансамблей, консолидирующих частные классификационные решения.

Цель диссертационной работы состоит в разработке алгоритмов классификации проектов планировки территории, обеспечивающих высокую точность и обоснованность принимаемых решений.

Для достижения поставленной цели были решены следующие задачи.

1. Проанализировать существующие методы классификации проектов планировки территории, выявить их недостатки и возможные пути их устранения.
2. Разработать алгоритмы отбора наиболее значимых показателей для оценки проектов планировки территории.
3. Разработать алгоритмы классификации проектов планировки территории с использованием алгоритмов кластеризации на основе нечетких множеств первого типа и генетических алгоритмов с хромосомой произвольной длины.
4. Разработать алгоритмы классификации проектов планировки территории с использованием алгоритмов кластеризации на основе интервальных нечетких множеств второго типа и генетических алгоритмов с хромосомой произвольной длины.
5. Разработать алгоритм классификации проектов планировки территории для кластеров произвольной формы.
6. Разработать алгоритм уточнения результатов классификации проектов планировки территории с использованием нейро-нечеткого логического вывода.
7. Разработать кластерные ансамбли уточнения результатов классификации проектов планировки территории.
8. Выполнить программную реализацию разработанных алгоритмов классификации проектов планировки территории.

Методы исследования. Для проведения исследований использовались инструментарий Data Mining, аппарат теории нечетких множеств, генетические алгоритмы и методы объектно-ориентированного программирования.

Научная новизна. При проведении диссертационных исследований были получены следующие результаты:

1. Разработаны алгоритмы классификации проектов планировки территории для проектов с разбиением на кластеры подобного объема с учетом свойств кластерной относительности и типичности с использованием алгоритмов кластеризации на основе нечетких множеств первого типа и генетических алгоритмов с хромосомой произвольной длины.

2. Разработаны алгоритмы классификации проектов планировки территории для проектов с разбиением на кластеры существенно разного объема с использованием алгоритмов кластеризации на основе интервальных нечетких множеств второго типа и генетических алгоритмов с хромосомой произвольной длины в условиях неопределенности выбора «ширины зоны» и фаззификаторов.

3. Разработан алгоритм классификации проектов планировки территории для кластеров произвольной формы с использованием алгоритмов кластеризации на основе нечетких множеств первого типа и генетических алгоритмов с хромосомой произвольной длины.

4. Разработаны алгоритмы уточнения результатов классификации проектов планировки территории с использованием нейро-нечеткого логического вывода.

5. Разработаны кластерные ансамбли уточнения результатов классификации проектов планировки территории, обеспечивающие консолидацию частных классификационных решений.

Практическая ценность работы. Практическая ценность работы заключается в том, что полученные алгоритмы классификации позволяют реализовать новый подход к задаче классификации проектов планировки территории и обеспечивают:

- высокую обоснованность принятия решения;
- минимизацию трудовых затрат, связанных с необходимостью экспертной оценки проектов по множеству показателей большого объема с целью выбора проектов-победителей.

В конечном итоге предлагаемые алгоритмы обеспечивают математически обоснованное решение задачи классификации проектов планировки территории.

Практическая ценность результатов диссертации подтверждается актами внедрения.

Достоверность полученных в диссертационной работе результатов подтверждается:

- обоснованным использованием инструментария Data Mining, аппарата ТНМ, генетических алгоритмов и нейро-нечетких сетей;
- результатами программной реализации разработанных алгоритмов;
- наличием актов внедрения результатов диссертационной работы.

На защиту выносятся:

1. Алгоритмы классификации проектов планировки территории для проектов с разбиением на кластеры подобного объема с учетом свойств кластерной относительности и типичности с использованием алгоритмов кластеризации на основе нечетких множеств первого типа и генетических алгоритмов с хромосомой произвольной длины.

2. Алгоритмы классификации проектов планировки территории для проектов с разбиением на кластеры существенно разного объема с использованием алгоритмов кластеризации на основе интервальных нечетких множеств второго типа и генетических алгоритмов с хромосомой произвольной длины в условиях неопределенности выбора «ширины зоны» и фазификаторов.

3. Алгоритм классификации проектов планировки территории для кластеров произвольной формы с использованием алгоритмов кластеризации на основе нечетких множеств первого типа и генетических алгоритмов с хромосомой произвольной длины.

4. Алгоритмы уточнения результатов классификации проектов планировки территории с использованием нейро-нечеткого логического вывода.

5. Кластерные ансамбли уточнения результатов классификации проектов планировки территории, обеспечивающие консолидацию частных классификационных решений.

Внедрение результатов. Исследования по тематике диссертационной работы велись:

– в рамках госбюджетной НИР 9-07Г «Разработка математических моделей, методов и алгоритмов обработки больших потоков информации в сложно организованных вычислительных структурах» (2007 г.);

– в рамках госбюджетной НИР 7-09Г «Разработка математических методов и алгоритмов передачи и обработки цифровой информации для поддержки интеллектуальных систем управления» (2009 г.);

– в рамках госбюджетной НИР 11-12Г «Разработка математических моделей, методов и алгоритмов обработки больших объемов информации в сложно организованных системах искусственного интеллекта» (2012 г.).

Результаты работы внедрены и используются:

– в Министерстве строительного комплекса Рязанской области (акт внедрения от 7.09.2012г.);

– в Главном управлении архитектуры и строительства Рязанской области (акт внедрения от 7.09.2012г.);

– в Управлении капитального строительства администрации г. Рязани (акт внедрения от 11.09.2012г.);

Апробация работы. Основные положения диссертации докладывались на следующих конференциях:

– VI Всероссийская научно-техническая конференция «Искусственный интеллект в XXI веке. Решения в условиях неопределенности» (Пенза, 2008);

– VIII Всероссийская научно-техническая конференция «Проблемы информатики в образовании, управлении, экономике и технике» (Пенза, 2008);

- XIV Международная открытая научная конференция «Современные проблемы информатизации в анализе и синтезе программных и телекоммуникационных систем» (Воронеж, 2009);
- «Методы и алгоритмы принятия эффективных решений» (Таганрог, 2009);
- XV Международная открытая научная конференция «Современные проблемы информатизации в анализе и синтезе программных и телекоммуникационных систем» (Воронеж, 2010);
- XVII Международная научная конференция «Методы и алгоритмы принятия решений» (Нижний Новгород, 2010);
- XVIII Всероссийская научно-техническая конференция студентов, молодых ученых и специалистов «Новые информационные технологии в научных исследованиях» (Рязань, 2013).

Публикации. По теме диссертации опубликовано 15 печатных работ. В их числе 3 статьи в изданиях, рекомендованных ВАК РФ, 1 монография, 3 статьи в межвузовских сборниках, 3 доклада на Международных конференциях, 4 доклада на Всероссийских конференциях, 1 свидетельство о регистрации программы для ЭВМ.

Структура и объем диссертации. Диссертационная работа состоит из введения, четырех глав, списка литературы и приложения. Содержит 155 страниц (из них 150 страниц – основная часть, 5 страниц – приложения), 5 таблиц, 59 рисунков. Список литературы состоит из 79 наименований.

СОДЕРЖАНИЕ РАБОТЫ

Во введении обосновывается актуальность выбора темы диссертации, формулируются цель и задачи исследований, научная новизна и практическая ценность основных результатов диссертационной работы.

В первой главе «Задача классификации проектов планировки территории и методы ее решения» сформулирована задача классификации проектов планировки территории, предполагающая использование нормативной документации и опыта экспертов в выставлении оценок проектам.

Сформулированы требования по решению задачи классификации проектов планировки территории. Показано, что наиболее актуальной задачей при оценке проектов планировки территории является задача обоснования выбора того или иного проекта на основе принципиально новых подходов, позволяющих математически обосновать этот выбор. Также такое решение позволит снизить недовольство проигравших конкурсантов и показать им недостатки их проектов.

Рассмотрены существующие методы классификации объектов, в том числе и кластерный анализ. Выявлены их достоинства и недостатки, исследована применимость к решению задачи классификации проектов планировки территории.

Показано, что существующие в настоящее время подходы к классификации проектов планировки территории имеют ряд существенных недостатков. Эти недостатки связаны с неопределенностью выбора показателей оценивания и оптимального количества кластеров, а так же с особенностями экспертного

оценивания, выполняемого посредством выставления баллов по показателям оценивания. Показано, что значительной проблемой является наличие проектов планировки территории с разным количеством показателей оценивания (в разных проектах могут присутствовать или отсутствовать некоторые объекты, например, стадионы, детские сады и т.д.).

Показано, что одной из важнейших проблем при классификации и отборе проектов планировки территории является выбор наиболее значимых показателей оценивания. Поскольку эти показатели не регламентируются, то их значимость и отбор осуществляется каждым экспертом в отдельности. Это может привести с одной стороны к поверхностной оценке проектов (при необоснованно сжатом наборе показателей), а с другой – к излишним затратам времени на детальное изучение простых проектов (при необоснованно расширенном наборе показателей).

В связи с этим предложен новый подход, заключающийся в анализе всех возможных комбинаций показателей в наборе с последующим отбором того набора показателей, который обладал бы наилучшей описательной способностью анализируемых проектов и при этом содержал бы минимум независимых показателей. При этом подробно исследованы такие алгоритмы отбора показателей оценивания, как: Forward Selection (прямой отбор), Backward Elimination (обратное исключение), Stepwise и даны рекомендации по их использованию при решении задачи отбора показателей оценивания проектов планировки территорий.

Во второй главе «Разработка алгоритмов классификации проектов планировки территории с использованием алгоритмов кластеризации на основе нечетких множеств первого типа, интервальных нечетких множеств второго типа и генетических алгоритмов с хромосомой произвольной длины» рассматривается разработка алгоритмов классификации проектов планировки территории с использованием алгоритмов ТНМ и ГА.

При разработке алгоритмов классификации проектов планировки территории предлагается использовать наиболее известные алгоритмы кластеризации: алгоритм нечетких c -средних (FCM-алгоритм; fuzzy c -means) и алгоритм возможностей c -средних (PCM-алгоритм; possibilistic c -means).

Данные алгоритмы используются для нахождения субоптимального разбиения проектов на кластеры, определяемого среди вариантов разбиений по критерию минимальности значения целевой функции.

Поскольку использование данных алгоритмов предполагает многократное их выполнение, то для сокращения времени поиска субоптимального разбиения предлагается использовать ГА. Для определения оптимального количества кластеров, на которые разбиваются проекты, предлагается использовать ГА с хромосомой произвольной длины.

В ГА были рассмотрены три варианта (алгоритма) скрещивания хромосом: одноточечное скрещивание с обменом составными частями хромосом (рисунок 1); одноточечное скрещивание с обменом одноименных генов хромосом (рисунок 2); многоточечное скрещивание с сохранением постоянного расстояния до краев правил (рисунок 3).

Алгоритм классификации проектов планировки территории с разбиением их на кластеры подобного объема с учетом свойства кластерной относительности с использованием FCM-алгоритма на основе нечетких множеств первого типа (FCM-алгоритм на основе НМТ1) и ГА с хромосомой произвольной длины может быть описан следующей последовательностью шагов.

1. Производится начальное разбиение проектов на произвольное количество кластеров c ($c_{min} \leq c \leq c_{max}$) с вычислением значений функций принадлежности $u_j(x_i)$ i -го проекта j -му кластеру.

2. Вычисляются координаты центров кластеров, которые используются для кодирования хромосомы.

$$v_j = \frac{\sum_{i=1}^n (u_j(x_i))^m \cdot x_i}{\sum_{i=1}^n (u_j(x_i))^m}, \quad (1)$$

где x_i – i -ый проект планировки территории; $u_j(x_i)$ – функция принадлежности i -го проекта j -му кластеру; n – количество проектов планировки территории; m – фаззификатор; v_j – вектор координат центра j -го кластера.

3. Шаги 1-2 повторяются до тех пор, пока не будет сформирована популяция хромосом размером P , закодированных координатами центров кластеров.

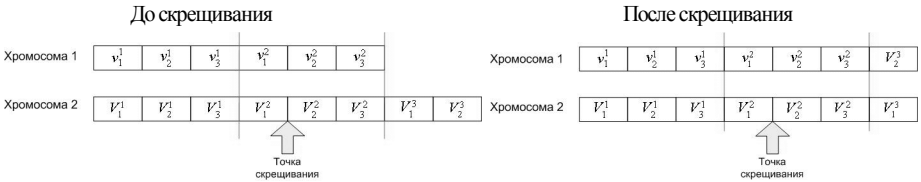


Рисунок 1 – Операция односточечного скрещивания хромосом с обменом составными частями

4. Вычисляются новые значения функции принадлежности хромосомы кластерам по формуле (2) и новые координаты центров кластеров по формуле (1).

$$u_j(x_i) = \frac{1}{\sum_{t=1}^c \left(\frac{d_{ji}}{d_{ti}} \right)^{\frac{2}{m-1}}}, \quad (2)$$

где d_{ji} – расстояние между проектом x_i и центром кластера v_j .

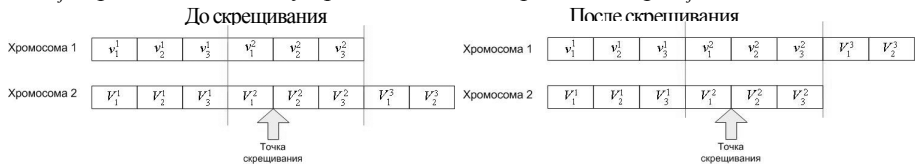


Рисунок 2 – Операция односточечного скрещивания хромосом с обменом одноименных генов

5. Если номер текущего поколения ГА не превышает максимальный, то для хромосом популяции вычисляются значения функции соответствия на основе индекса Se – Бени, который должен быть минимизирован:

$$XB = \frac{\sum_{j=1}^c \sum_{i=1}^n (u_j(x_i))^2 \cdot \|x_i - v_j\|^2}{n \cdot \min_{t \neq j} \|v_t - v_j\|^2}, \quad (3)$$

иначе производится переход к шагу 14.



Рисунок 3 – Операция многоточечного скрещивания хромосом с сохранением постоянного расстояния до краев правил

6. Производится выбор хромосом-родителей, для которых выполняется операция скрещивания в соответствии с одним из вариантов на рисунках 1-3.

7. Производится операция мутации хромосом-отпрысков, полученных на шаге 6.

8. Формируется промежуточная популяция хромосом, состоящая из популяции хромосом предыдущего поколения и хромосом-отпрысков.

9. Для хромосом расширенной популяции вычисляются новые значения функций принадлежности по формуле (2) и вычисляются новые координаты центров кластеров по формуле (1).

10. Для хромосом расширенной популяции вычисляются значения функции соответствия по формуле (2).

11. Формируется расширенная популяция хромосом, состоящая из хромосом предыдущего поколения и хромосом промежуточной популяции.

12. Из расширенной популяции хромосом удаляются хромосомы с максимальными значениями функции соответствия так, чтобы количество хромосом в популяции стало равным P .

13. Переход к шагу 4.

14. Производится выбор хромосомы, которая имеет минимальное значение функции соответствия по формуле (2). Координаты центров кластеров, закодированные в этой хромосоме, принимаются за искомые и осуществляется завершение алгоритма.

Если кластеры содержат атипичные объекты, то применение FCM-алгоритм на основе НМТ1 может давать не всегда адекватные результаты кластеризации.

В этом случае предлагается использовать алгоритм возможностных c -средних на основе НМТ1 (PCM-алгоритм на основе НМТ1).

В целом совместная реализация PCM-алгоритма на основе НМТ1 и ГА с хромосомой произвольной длины идентична реализации FCM-алгоритма на основе НМТ1 и ГА с хромосомой произвольной длины с соответствующей заменой функций принадлежности на функции типичности и заменой формул, по которым вычисляются координаты центров кластеров и функций соответствия.

В частности, функция типичности вычисляется как:

$$w_j(x_i) = \frac{1}{1 + \left(\frac{d_{ji}}{\eta_j}\right)^{m-1}}, \quad (4)$$

где η_j – «ширина зоны», которая определяет расстояние, на котором значение функции типичности проекта j -му кластеру равно 0,5.

Формула для вычисления координат центров кластеров идентична формуле (1) с заменой функций принадлежности на функции типичности, а в качестве функции соответствия может быть использована целевая функция PCM-алгоритма:

$$J(W, V) = \sum_{j=1}^c \sum_{i=1}^n (w_j(x_i))^m \cdot d_{ji}^2 + \sum_{j=1}^c \eta_j^2 \cdot \sum_{i=1}^n (1 - w_j(x_i))^m \quad (5)$$

При этом значение «ширины зоны» может быть вычислено с применением одной итерации FCM-алгоритма как:

$$\eta_j = \left(\frac{\sum_{i=1}^n (u_j(x_i))^m \cdot d_{ji}^2}{\sum_{i=1}^n (u_j(x_i))^m} \right)^{\frac{1}{2}} \quad (6)$$

Если множество проектов содержит кластеры существенно различные по объему, то алгоритмы кластеризации на основе НМТ1 могут давать неадекватные результаты кластеризации. Для решения этой проблемы предлагается использовать алгоритмы кластеризации на основе интервальных нечетких множеств второго типа (ИНМТ2).

В этом случае предлагается использовать совместную реализацию PCM-алгоритма на основе ИНМТ2 в условиях неопределенности выбора фаззификаторов и ГА.

1. Производится начальное разбиение проектов на произвольное количество кластеров c с проверкой условия $m_1 < m_2$.

2. Если номер текущего поколения ГА не превышает максимальный, то выполняется PCM-алгоритм на основе ИНМТ2, иначе производится переход к шагу 10.

3. Вычисляется функция соответствия для каждой хромосомы.

$$H = \sum_{j=1}^c (\det(R_j))_2^{\frac{1}{2}} \rightarrow \min, \quad (7)$$

где R_j – ковариационная матрица j -го кластера.

4. Производится выбор хромосом-родителей, для которых выполняется операция скрещивания в соответствии с одним из вариантов на рисунках 1-3 с проверкой условия $m_1 < m_2$.

5. Производится операция мутации хромосом-отпрысков, полученных на шаге 4 с проверкой условия $m_1 < m_2$.

6. Для хромосом-отпрысков выполняется РСМ-алгоритм на основе ИНМТ2 и вычисляются функции соответствия по формуле (7).

7. Формируется расширенная популяция хромосом, состоящая из хромосом предыдущего поколения и хромосом промежуточной популяции.

8. Из расширенной популяции хромосом удаляются хромосомы с максимальными значениями функции соответствия так, чтобы количество хромосом в популяции стало равным P .

9. Переход к шагу 2.

10. Производится выбор хромосомы, которая имеет минимальное значение функции соответствия по формуле (7). Для каждого объекта определяется его принадлежность к кластерам.

Кроме того, была выполнена совместная реализация РСМ-алгоритм на основе ИНМТ2 в условиях неопределенности выбора «ширины зоны» и ГА.

В целом совместная реализация РСМ-алгоритма на основе ИНМТ2 и ГА с хромосомой произвольной длины при управлении неопределенностью «ширины зоны» идентична реализации РСМ-алгоритм на основе ИНМТ2 и ГА с хромосомой произвольной длины при управлении неопределенностью фаззификаторов с соответствующей заменой условия $m_1 < m_2$ на $\eta_{j1} < \eta_{j2}$ ($j = \overline{1, c}$).

Если множество проектов содержит кластеры, форма которых отличается от гиперсферической (гиперэллипсоидной), то разработанные алгоритмы классификации показывают плохие результаты. Для решения этой проблемы предлагается использовать FCM-алгоритм на основе НМТ1 с применением кривой Безье для задания контура кластеров (VFCM-алгоритм на основе НМТ1).

В целом совместная реализация VFCM-алгоритма на основе НМТ1 и ГА с хромосомой произвольной длины идентична реализации FCM-алгоритма на основе НМТ1 и ГА с хромосомой произвольной длины с соответствующей заменой формул, по которым вычисляются координаты центров кластеров и функций соответствия.

В частности, координаты центров кластеров вычисляются по формуле:

$$v_j = \frac{\sum_{i=1}^n (u_j(x_i)^m \cdot (f_1, f_2, \dots, f_n))}{2 \cdot \sum_{i=1}^n u_j(x_i)^m}, \quad (8)$$

где $f_n = x_{in} - d_{ij} \cdot \frac{x'_{in} - v_{jn}}{d'_{ij}} + x'_{in} - d'_{ij} \cdot \frac{x_{in} - v_{jn}}{d_{ij}}$, x'_{in} – точка пересечения прямой li ,

проходящей через x_i и v_j , и кривой Безье, образованной точкой x_i и двумя ближайшими к ней точками, принадлежащими кластеру j , расположенных по разным сторонам от li .

Для нахождения x'_i все проекты переводятся в полярную систему координат.

В качестве функции соответствия может быть использована функция вида:

$$J_m(U, V) = \sum_{j=1}^c \sum_{i=1}^n (u_j(x_i))^m \cdot \left(\frac{\sum_{i=1}^n d_{ij} - d_{ij}}{z - 1} \right)^2, \quad (9)$$

где $z = u_j(x_i)^{m-1} / \sum_{i=1}^n u_j(x_i)^{m-1}$.

На практике была решена задача классификации проектов планировки территории с отбором наиболее значимых показателей оценивания одним из рассмотренных в главе 1 алгоритмов (Forward Selection, Backward Elimination, Stepwise).

При этом была выполнена классификация 50 проектов планировки территории на произвольное (заранее неизвестное) количество кластеров по набору из 20 показателей оценивания (мониторинга): расположение участков домов «ветхого фонда», инженерная инфраструктура, транспортная инфраструктура, объекты социально-бытового обслуживания, экология, имущественно-земельные отношения, экономическая эффективность использования территорий, модернизация производств, реорганизация производственной зоны, границы районов, подземные коммуникации, надземные коммуникации, уличное освещение, размещение светофоров, газоснабжение, водоснабжение, канализация, электричество, демография, озеленение.

Применение алгоритмов отбора показателей оценивания проектов позволило сократить список показателей оценивания, при этом алгоритм Forward Selection выявил 16 значимых показателей; алгоритм Backward Elimination – 17 значимых показателей; алгоритм Stepwise – 17 значимых показателей.

При этом все алгоритмы исключили из рассмотрения такие показатели, как: озеленение, размещение светофоров, реорганизация производственной зоны и уличное освещение. При выполнении дальнейших исследований использовался алгоритм Forward Selection, т.к. два других алгоритма отбора показателей оценивания существенно увеличивали время классификации, не уменьшая при этом количество ошибок классификации.

Результаты классификации проектов представлены в таблице 1. При этом используемые при классификации пары алгоритмов обозначены парой цифр, разделенных между собой дефисом. Первая цифра означает один из пяти вышеописанных алгоритмов кластеризации (1 – FCM-алгоритм на основе НМТ1; 2 – РСМ-алгоритм на основе НМТ1; 3 – РСМ-алгоритм на основе ИНМТ2 в условиях неопределенности выбора «ширины зоны»; 4 – РСМ-алгоритм на основе ИНМТ2 в условиях неопределенности выбора фаззификаторов; 5 – VFCM-алгоритм на основе НМТ1), а вторая цифра означает алгоритм скрещивания (1 – одноточечное скрещивание с обменом составными частями хромосом; 2 – одноточечное скрещивание с обменом одноименных генов хромосом; 3 – многоточечное скрещивание с сохранением постоянного расстояния до краев правил).

Таблица 1 – Результаты классификации

Алгоритм	Время классификации, сек	Ошибочно классифицированные проекты
1-1	8,33	2
1-2	8,28	2
1-3	8,19	2
2-1	9,24	1
2-2	9,26	1
2-3	9,31	1
3-1	23,98	0
3-2	24,09	0
3-3	24,15	0
4-1	25,42	0
4-2	27,40	0
4-3	36,72	0
5-1	28,31	0
5-2	31,71	0
5-3	33,49	0

Как показывает анализ результатов классификации, наименьшее время классификации у FCM-алгоритма на основе НМТ1, но, в то же время, он дает большее количество ошибок классификации. Если при классификации более 100 проектов образуются кластеры не круглой или эллиптической формы, то наиболее точным оказывается последний алгоритм. Однако время его выполнения не самое минимальное.

На практике все предлагаемые алгоритмы классификации применялись при выборе проектов планировки территории Чучковского, Елатомского и Ухоловского городских поселений. Как показали результаты исследований, FCM-алгоритма на основе НМТ1 хорошо справляется с небольшими наборами конкурсных проектов поселений.

В третьей главе «Разработка алгоритмов классификации проектов планировки территории на основе нейро-нечеткой сети ANFIS и кластерных ансамблей» предлагаются алгоритмы уточнения результатов классификации проектов планировки территории.

Классификация проектов планировки территории с использованием рассмотренных в главе 2 алгоритмов является субоптимальной, то есть полученные результаты классификации могут содержать ошибки относительно принадлежности проектов тем или иным классам. В связи с этим рассмотрены два подхода к уточнению результатов классификации проектов.

Первый подход основан на применении адаптивной сети нейро-нечеткого вывода ANFIS (Adaptive-Network-Based Fuzzy Inference System) для уточнения результатов классификации, полученных с применением того или иного алгоритма классификации, предложенного в главе 2.

Нейро-нечеткая сеть ANFIS основана на реализации алгоритма нечеткого вывода Сугено и представляет собой пятислойную нейронную сеть прямого

распространения сигнала (первый слой – термы входных лингвистических переменных; второй слой – antecedentes (посылки) нечетких правил; третий слой – нормализация степеней выполнения правил; четвертый слой – заключения правил; пятый слой – агрегирование результата, полученного по различным правилам).

Разработка нейро-нечеткой сети ANFIS осуществляется с применением обучающей выборки, сформированной по результатам кластеризации проектов на кластеры с применением некоторого алгоритма главы 2. При этом в обучающую выборку из каждого кластера заносится половина проектов, наиболее близких к центру вычисленного кластера, и номер этого кластера.

Разработка нейро-нечеткой сети ANFIS может быть выполнена в соответствии со следующей последовательностью шагов.

1. Генерируется *fis*-структура. Задаются: количество входных лингвистических переменных, количество функций принадлежности для каждой входной лингвистической переменной, начальные параметры всех функций принадлежности.

2. Задаются параметры обучения: количество эпох, минимальная ошибка обучения и скорость обучения сети и пр. На основе обучающей выборки генерируется база правил; определяется количество правил, предпосылки и заключения для каждого правила. Обучение сети по обучающей выборке осуществляется методом обратного распространения ошибки.

3. Если количество эпох меньше заданного или ошибка обучения, рассчитываемая как среднеквадратичное отклонение, больше некоторого порогового значения, реализуется новая эпоха обучения, иначе осуществляется переход на шаг 4.

4. Нейро-нечеткая сеть полагается обученной.

Обученная таким образом нейро-нечеткая сеть используется для классификации оставшихся проектов.

Результаты экспериментальных исследований свидетельствуют об эффективности использования нейро-нечеткой сети для уточнения результатов классификации проектов планировки территорий (на тестовых примерах это улучшение составляет 2-4%).

Разработанная таким образом нейро-нечеткая сеть ANFIS используется для уточнения кластерной принадлежности остальных проектов, не попавших в обучающую выборку. Целесообразность использования данного подхода может быть обоснована тем, что алгоритмы кластеризации из главы 2 реализуют разбиение только на кластеры гиперсферической (гиперэллипсоидной) формы, в то время как нейро-нечеткая сеть ANFIS не выдвигает таких жестких требований к форме кластеров, а, следовательно, ее применение позволяет уточнить форму кластеров и кластерную принадлежность некоторых проектов.

Второй подход основан на разработке кластерных ансамблей, реализующих консолидацию r частных результатов классификаций проектов, полученных с применением алгоритмов классификации, предложенных в главе 2.

В настоящее время наиболее известными и низкокзатратными с вычислительной точки зрения (имеющими линейную сложность) алгоритмами разра-

ботки кластерных ансамблей являются алгоритмы HGPA (HyperGraph Partitioning Algorithm) и MCLA (Meta-Clustering Algorithm), предложенные в работах A. Strehl и J.Ghosh.

Классический алгоритм HPGA заключается в построении многомерного графа (гиперграфа), в котором все проекты и кластеры полагаются равновесными. Предложено модифицировать алгоритм HPGA посредством назначения весов (баллов) кластерам за повторы в разбиениях: если кластер не повторяется ни в одном разбиении, ему присваивается 1 балл; если кластер повторяется в n разбиениях, ему присваивается n баллов.

Модифицированный алгоритм HPGA может быть описан следующей последовательностью шагов.

1. Из вариантов разбиения проектов планировки территории, полученных с помощью алгоритмов классификации, строится гиперграф.

2. Гиперграф делится на кластеры, полученные с помощью различных алгоритмов классификации.

3. Определяются веса кластеров и выявляются кластеры с максимальными весами.

4. Производится разделение гиперграфа на c несвязанных кластеров в соответствии с весами кластеров. Объединяются кластеры с одинаковыми весами, или с незначительно отличающимися весами.

Классический алгоритм MCLA заключается в представлении каждого кластера на гиперграфе гиперячейкой, которые затем группируются и свертываются. Предложено модифицировать алгоритм MCLA посредством адаптации под использование матрицы вместо гиперграфа.

Модифицированный алгоритм MCLA может быть описан следующей последовательностью шагов.

1. Каждому кластеру c в каждом разбиении r (где r – количество частных классификационных решений) назначаются n коэффициентов принадлежности проекта кластеру $h \in [0, 1]$. Это отображается в матрице размером $c \times n \times r$.

2. Производится перераспределение проектов между кластерами так, чтобы получилось c сбалансированных метакластеров. Для определения баланса вычисляется вес между кластерами $\omega_{a,b}$ с помощью измерения Жаккарда по формуле $\omega_{a,b} = (h_a * h_b) / (h_a^2 + h_b^2 - h_a * h_b)$, где h_a и h_b – коэффициенты принадлежности кластерам a и b соответственно.

3. Производится свертывание метакластеров. Размер матрицы становится равным $c \times n$. В ячейках матрицы размещаются сводные коэффициенты h , определяющиеся как среднее арифметическое свернутых ячеек.

4. По коэффициентам h производится окончательное распределение проектов по кластерам.

Предложенные модификации алгоритмов HGPA и MCLA обеспечивают повышение обоснованности принятия решения в задачах классификации проектов планировки территории. Алгоритм HPGA хорошо работает при разбиении на схожие по размеру кластеры, что на практике встречается довольно редко. В связи с этим рекомендуется использовать алгоритм HPGA только, если расхождении размеров вычисленных кластеров составляет не более, чем 5%.

Результаты экспериментальных исследований свидетельствуют об эффективности использования кластерного ансамбля для уточнения результатов классификации проектов планировки территорий (на тестовых примерах это улучшение составляет 5-6%).

В четвертой главе «Программная реализация алгоритмов классификации проектов планировки территории» разработан ППП₂ предназначенный для решения задачи классификации проектов планировки территории. При этом набор проектов, подлежащих классификации, может содержать как кластеры подобной плотности и подобного объема, так и кластеры существенно разной плотности или существенно разного объема.

ППП может быть рекомендован к использованию специалистами, работающими в сфере строительства, для экономии времени на оценку, классификацию и отбор проектов планировки территории для повышения обоснованности классификационных решений.

ППП состоит из 8-и подкомплексов программ, реализующих классификацию проектов с использованием разработанных алгоритмов, и включает в себя подкомплекс программ, реализующей отбор наиболее значимых показателей оценивания проектов. При этом классификация проектов выполняется с использованием предложенных в диссертационной работе алгоритмов классификации проектов планировки территории с произвольным количеством показателей оценивания.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

Проведенные исследования позволяют сформулировать основные результаты.

1. Проведен анализ существующих подходов к классификации проектов планировки территории, выявлены перспективные направления их развития.
2. Исследована возможность комплексного использования инструментария Data Mining, аппарата ТНМ, ГА и алгоритмов нейро-нечеткого вывода при разработке алгоритмов классификации проектов планировки территории.
3. Разработаны алгоритмы классификации проектов планировки территории для проектов с разбиением на кластеры подобного объема с учетом свойств кластерной относительности и типичности с использованием алгоритмов кластеризации на основе нечетких множеств первого типа и генетических алгоритмов с хромосомой произвольной длины.
4. Разработаны алгоритмы классификации проектов планировки территории для проектов с разбиением на кластеры существенно разного объема с использованием алгоритмов кластеризации на основе интервальных нечетких множеств второго типа и генетических алгоритмов с хромосомой произвольной длины в условиях неопределенности выбора «ширины зоны» и фазсификаторов.
5. Разработан алгоритм классификации проектов планировки территории для кластеров произвольной формы с использованием алгоритмов кластеризации на основе нечетких множеств первого типа и генетических алгоритмов с хромосомой произвольной длины.

6. Разработаны алгоритмы уточнения результатов классификации проектов планировки территории с использованием нейро-нечеткого логического вывода.

7. Разработаны кластерные ансамбли уточнения результатов классификации проектов планировки территории, обеспечивающие консолидацию частных классификационных решений.

8. Разработан пакет прикладных программ для классификации проектов планировки территории на основе предложенных алгоритмов классификации.

9. Исследования предложенных алгоритмов классификации проектов планировки территории показали:

– высокую обоснованность и адекватность принятия решения в условиях неопределенности и неточности исходной информации (в том числе, экспертной);

– минимизацию временных и финансовых затрат, связанных как с необходимостью сбора и учета точных и полных исходных данных (что может быть принципиально невозможным), так и необходимостью многократной реализации классических алгоритмов кластеризации с целью выбора соответствующих оптимальных параметров, обеспечивающих принятие адекватных решений.

ОСНОВНЫЕ ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи в изданиях по списку ВАК

1. Демидова Л.А., Кираковский В.В., **Корогаев А.Н.** Подход к проблеме классификации технического состояния зданий и сооружений с использованием алгоритмов возможностной кластеризации и генетических алгоритмов // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. Санкт-Петербург. Издательство Политехнического университета, №1. – 2010. – С. 151-157.
2. Демидова Л.А., Кираковский В.В., **Корогаев А.Н.** Кластеризация объектов с использованием РСМ-алгоритма на основе интервальных нечетких множеств второго типа и генетического алгоритма // Вестник ТОГУ, №3(18). – 2010. – С. 53-62.
3. Кираковский В.В., **Корогаев А.Н.** Подход к оценке инвестиционных проектов для застройки территории с использованием теории нечетких множеств // Промышленное и гражданское строительство, №11. – 2010. – С. 29-30.

Монография

4. Кираковский В.В., **Корогаев А.Н.**, Пылькин А.Н. и др. Теория, методология и практика экономического, социального и инновационного развития организационных систем // Приволжский дом знаний, Пенза, 2012. – 108 с.

Свидетельство о регистрации

5. Демидова Л.А., **Корогаев А.Н.** Реализация двухуровневого генетического алгоритма кластеризации объектов на основе РСМ-алгоритма с использованием нечетких множеств второго типа // Свидетельство об отраслевой регистрации разработки в Отраслевом фонде алгоритмов и программ №11813 от 20.11.2008 г.

Статьи в сборниках научных трудов

6. Демидова Л.А., **Корогаев А.Н.** Решение задачи выбора алгоритма кластеризации на основе интервальных нечетких множеств второго типа с использованием гене-

- тического алгоритма с переменной длиной хромосомы // Математическое и программное обеспечение вычислительных систем: Межвуз. сб. науч. тр. / Под ред. А.Н. Пылькина – М.: Горячая линия – Телеком, 2009. – С. 23-30.
7. **Корогаев А.Н.** Решение задачи кластеризации объектов с использованием генетических алгоритмов с изменяющейся длиной хромосомы // Программные информационные системы: Межвуз. сб. науч. тр. / Под ред. А.Н. Пылькина – Рязань: РГРТУ, 2010. – С. 57-61.
 8. Демидова Л.А., **Корогаев А.Н.** Разработка гибридного подхода к решению задачи классификации объектов в условиях неопределенности // Математическое и программное обеспечение вычислительных систем: Межвуз. сб. науч. тр. / Под. ред. А.Н. Пылькина – Рязань (РГРТУ), 2011. – С. 22-24.
- Тезисы докладов на международных и всероссийских конференциях*
9. Демидова Л.А., **Корогаев А.Н.** Двухуровневый генетический алгоритм кластеризации объектов на основе FCM-алгоритма с использованием нечетких множеств второго типа // Искусственный интеллект в XXI веке. Решения в условиях неопределенности: сборник статей VI Всероссийской научно-технической конференции. – Пенза: Приволжский Дом знаний, 2008. – С. 25-27.
 10. Демидова Л.А., **Корогаев А.Н.** Генетический алгоритм настройки параметров системы нечеткого вывода на основе нечетких множеств второго типа // Проблемы информатики в образовании, управлении, экономике и технике: сборник статей VIII Всероссийской научно-технической конференции. – Пенза: Приволжский Дом знаний, 2008. – С. 79-81.
 11. Демидова Л.А., Коняева Е.И., **Корогаев А.Н.** Сравнительный анализ методов кластеризации на основе нечетких множеств первого и второго типа // Современные проблемы информатизации в анализе и синтезе программных и телекоммуникационных систем: Сб. трудов. Вып. 14 / Под ред. д.т.н., проф. О.Я. Кравца. – Воронеж: «Научная книга», 2009. – С. 296-302.
 12. Демидова Л.А., Коняева Е.И., **Корогаев А.Н.** Использование модификаций генетических алгоритмов с переменной длиной хромосомы для решения задачи кластеризации объектов // Материалы международной научной конференции. «Методы и алгоритмы принятия эффективных решений» - 2 часть – Таганрог: Изд-во ТТИ ЮФУ, 2009. – С. 22-27.
 13. Коняева Е.И., **Корогаев А.Н.** Использование генетических алгоритмов с постоянной и переменной длиной хромосомы для решения задачи кластеризации объектов // Современные проблемы информатизации в анализе и синтезе программных и телекоммуникационных систем: Сб. трудов. Вып. 15/ Под ред. д.т.н., проф. О.Я. Кравца. – Воронеж: «Научная книга», 2010. – С. 251-254.
 14. Демидова Л.А., Коняева Е.И., **Корогаев А.Н.** Подход к проблеме классификации технического состояния зданий и сооружений с использованием алгоритмов нечеткой кластеризации и генетических алгоритмов // Методы и алгоритмы принятия решений: материалы 17-й международной научной конференции. – Нижний Новгород: Изд-во ТТИ ЮФУ, 2010. – С. 22-27.
 15. Белогубец А.И., **Корогаев А.Н.** Классификация проектов планировки территории с использованием кластерных ансамблей // Новые информационные технологии в научных исследованиях (НИТ-2013): материалы XVIII Всероссийской научно-технической конференции студентов, молодых ученых и специалистов. – Рязань: РГРТУ, 2013. – С. 121-122.

К о р о т а е в А л е к с а н д р Н и к о л а е в и ч

**РАЗРАБОТКА АЛГОРИТМОВ
КЛАССИФИКАЦИИ ПРОЕКТОВ
ПЛАНИРОВКИ ТЕРРИТОРИИ
С ИСПОЛЬЗОВАНИЕМ ТЕОРИИ
НЕЧЕТКИХ МНОЖЕСТВ**

А в т о р е ф е р а т
диссертации на соискание ученой степени
кандидата технических наук

Подписано в печать 31.10.13 г. Формат бумаги 60×84 1/16.
Бумага офсетная. Печать офсетная. Усл. печ. л. 2,5.
Уч.-изд. л. 2,5. Тираж 100 экз. Бесплатно.
Рязанский государственный радиотехнический университет.
390005, г. Рязань, ул. Гагарина, д. 59/1.
Редакционно-издательский центр РГРТУ.