

На правах рукописи

Е.И.К.

Коняева Елена Ивановна

**МЕТОДЫ КЛАСТЕРИЗАЦИИ
В ЗАДАЧАХ ОЦЕНКИ
ТЕХНИЧЕСКОГО СОСТОЯНИЯ
ЗДАНИЙ И СООРУЖЕНИЙ
В УСЛОВИЯХ НЕОПРЕДЕЛЁННОСТИ**

Специальность 05.13.01 –
«Системный анализ, управление и обработка информации
(технические системы)»

Автореферат
диссертации на соискание ученой степени
кандидата технических наук

Рязань 2010

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

Научный руководитель: доктор технических наук, доцент
ДЕМИДОВА Лилия Анатольевна

Официальные оппоненты: доктор технических наук, доцент
АНДРИАНОВ Дмитрий Евгеньевич

кандидат технических наук, доцент
ТАГАНОВ Александр Иванович

Ведущая организация: Липецкий государственный
технический университет

Защита диссертации состоится 28 июня 2010 г. в 12 часов на заседании диссертационного совета Д 212.211.01 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 « 21 » _____ мая _____ 2010 г.

Ученый секретарь
диссертационного совета
канд. техн. наук, доцент



В.Н. Пржегорлинский

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования. На сегодняшний день одной из наиболее актуальных градостроительных проблем является качество строительства, определяющее не только срок службы зданий, но и физическую безопасность жителей. Основной задачей при проведении реконструкции и капитального ремонта жилых и общественных зданий являются обеспечение сохранности основных фондов непродуцированной сферы, предотвращение их преждевременного выбытия, восстановление и улучшение их потребительских качеств. На капитальный ремонт зданий жилищно-гражданского назначения направляются большие финансовые, трудовые и материально-технические ресурсы. При этом ставится задача повышения эффективности использования этих ресурсов.

Значительный вклад в решение задач, связанных с оценкой, анализом и управлением качества строительства, внесли такие ученые, как Ю.В. Бейлезон, А.В. Гличев, О.П. Глудкин, В.В. Костюченко, В.В. Окрепилов, Ю. П. Панибратов, И.С. Степанов и др. Большое внимание анализу и решению градостроительных проблем уделяют МНИИ-ТЭП (Московский научно-исследовательский и проектный институт типологии, экспериментального проектирования), НИАЦ (Научно-исследовательский аналитический центр) Москомархитектуры, Центральный научно-исследовательский и проектный институт жилых и общественных зданий, Моспроект-1, Моспроект-2 и др.

Однако в последние годы в связи со значительным ростом объемов жилищного и промышленного строительства с применением новейших технологий и материалов проблема оценки качества строительства ощущается особенно остро. В настоящее время практически отсутствуют какие-либо действительно хорошие нормативы и методики по комплексному обследованию и мониторингу технического состояния зданий и сооружений современного города и их классификации, а прежние – безнадежно устарели. На данный момент не существует обоснованных однозначных рекомендаций по выбору конкретных значимых элементов строительных объектов (фундамент, крыша, стены и т.п.) для выполнения мониторинга, определению их степени важности, выбору количества классов принадлежности объектов мониторинга и т.п. Кроме того, значительной проблемой являются наличие типов строительных объектов с разным количеством элементов мониторинга (для малоэтажных и многоэтажных зданий, складских помещений и т.п.), а также существенно различающиеся количества строительных объектов разных типов. В последние несколько лет, по существу, заложены только основы для создания современной нормативной базы по комплексному обследованию и мониторингу технического состояния зданий и сооружений, в которой впервые будут системно объединены нормативные, методические и стоимостные аспекты проблемы.

Существующие методы оценивания технического состояния зданий и сооружений базируются в основном на инструментальных исследованиях, рассчитаны на проведение больших организационных мероприятий и требуют привлечения значительных трудовых и денежных ресурсов. Кроме того, современные здания и сооружения характеризуются наличием элементов и конструкций, точную информацию о которых невозможно получить в реальный отрезок времени из-за необходимости проведения дорогостоящих инструментальных исследований. Однако при решении задач обеспечения эффективной оценки качества строительства требуется учет и такой информации, которой присуща некоторая неопределенность. Ввиду практической сложности и высокой стоимости инструментального контроля существует необходимость в альтернативном подходе к оценке технического состояния зданий и сооружений. В качестве такого подхода можно использовать подход, основанный на экспертном оценивании технического состояния

зданий и сооружений. Этот подход может оказаться особенно полезным, если инструментальный мониторинг технического состояния зданий и сооружений затруднен в связи с большой трудоемкостью и длительностью его проведения во времени, в связи с существенными финансовыми затратами на его проведение, а также при неполноте и неточности необходимых для анализа данных. Еще более сложной является задача классификации технического состояния зданий и сооружений, так как недостаточно просто дать оценки некоторому объекту мониторинга по выбранному набору элементов мониторинга, необходимо принять обоснованное и адекватное решение, позволяющее определить класс принадлежности объекта. В случае мониторинга технического состояния большого количества зданий и сооружений, например с целью определения аварийных объектов при составлении плана штатных ремонтных работ, классификация может быть выполнена посредством кластеризации множества объектов мониторинга для выбранного набора элементов мониторинга, например, на заданное количество кластеров (классов). Таким образом, можно говорить о наличии задачи комплексной оценки и классификации технического состояния зданий и сооружений в условиях неопределенности.

Один из современных подходов, используемых в различных задачах принятия решений в условиях неопределенности, основан на применении инструментария теории нечетких множеств (ТНМ), основоположником которой является Л.А. Заде (1965 г.). Применение ТНМ и её приложений позволяет строить формальные схемы решения задач, характеризующихся той или иной степенью неопределенности, которая может быть обусловлена неполнотой, внутренней противоречивостью, неоднозначностью и размытостью исходных данных, представляющих собой приближенные количественные или качественные оценки параметров объектов. Эта неопределенность является систематической, так как обусловлена сложностью задач, дефицитом информации, лимитом времени на принятие решений, особенностями восприятия и т.п.

Неполнота и неточность информации могут заключаться: в принципиальной невозможности полного сбора и учета информации об анализируемом объекте; в некоторой недостоверности и недостаточности исходной информации об анализируемом объекте и др. Кроме того, неточность, неполнота и неопределенность исходных данных могут быть вызваны недостаточным знанием экспертов специфики конкретной прикладной задачи. Следовательно, можно говорить и о наличии «субъективного» человеческого фактора в задачах поддержки принятия решений в условиях неопределенности.

При разработке алгоритмов и методов ТНМ охватывается широкий круг математических и прикладных проблем, в решение которых значительный вклад внесли российские и зарубежные ученые: А.Н. Аверкин, А.В. Алексеев, Р. Беллман, В.В. Борисов, Л.А. Заде, А. Кофман, А.Н. Мелехов, Д.А. Поспелов, Т.Л. Саати, Н. Larsen, E. Mamdani, M. Sugeno. Алгоритмы нечеткой кластеризации (алгоритм нечетких c -средних и его модификации) предложены в работах таких ученых, как J. Bezdek, J. Dunn, R. Dave, J. Keller, R. Krishnapuram, Y. Ohashi. Значительное количество работ (G. Beni, H. Galda, I. Gath, A. Geva, D. Gustafson, W. Kessel, M. Halkidi, X. Xei) посвящено разработке и исследованию показателей качества кластеризации для соответствующих алгоритмов кластеризации.

Анализ известных алгоритмов кластеризации, основанных на применении ТНМ, показывает, что зачастую они не обеспечивают получение адекватных решений ввиду недостаточно обоснованного выбора их параметров, а поиск эффективных решений приводит к значительным временным затратам из-за необходимости выполнения многократных реализаций классических алгоритмов с целью выбора оптимальных параметров.

Одним из современных бионических принципов решения широкого класса прикладных задач, которые трудноразрешимы классическими методами, особенно в области *NP*-полных задач оптимизации, является применение генетических алгоритмов (ГА) – адаптивных методов поиска, реализующих эволюционные вычисления, основанные на генетических процессах биологических организмов. Общие принципы ГА были сформулированы Д.Х. Холландом (1975 г.) и описаны в работах: Д.И. Батищева, Л.А. Гладкова, Д.И. Голдберга, В.В. Емельянова, В.В. Курейчика, В.М. Курейчика и др.

Актуальность настоящей работы определяется необходимостью разработки эффективных методов кластеризации технического состояния зданий и сооружений, устраняющих недостатки существующих аналогов. Эти недостатки связаны: с неопределенностью выбора элементов мониторинга и оптимального количества кластеров, с особенностями инструментального и экспертного оценивания элементов мониторинга, а также с проблемой наличия строительных объектов с различным количеством элементов мониторинга. Использование новых методов кластеризации технического состояния зданий и сооружений позволит создать качественно новые программные средства, существенно расширяющие перечень задач поддержки принятия решений в условиях неопределенности и обеспечивающие повышение адекватности и объективности принятия решений при низких временных затратах.

Цель диссертационной работы состоит в разработке эффективных методов кластеризации технического состояния зданий и сооружений в условиях неопределенности, ориентированных на устранение недостатков существующих аналогов и обеспечивающих высокую обоснованность и адекватность принимаемых решений.

Для достижения поставленной цели необходимо решить следующие задачи.

1. Провести анализ существующих методов кластеризации технического состояния зданий и сооружений, выявить перспективные направления их развития.

2. Исследовать возможность комплексного использования инструментария ТНМ при разработке методов кластеризации технического состояния зданий и сооружений.

3. Исследовать возможность представления принадлежности объектов (технического состояния зданий и сооружений) к кластерам с помощью нечетких множеств первого типа и интервальных нечетких множеств второго типа.

4. Разработать методы кластеризации технического состояния зданий и сооружений с использованием алгоритмов кластеризации на основе нечетких множеств первого типа и генетических алгоритмов.

5. Разработать методы кластеризации технического состояния зданий и сооружений с использованием алгоритмов кластеризации на основе интервальных нечетких множеств второго типа и генетических алгоритмов.

6. Разработать пакет прикладных программ (ППП) для кластеризации технического состояния зданий и сооружений на основе разработанных методов кластеризации.

Методы исследования. Теоретические исследования выполнены с использованием методов системного анализа, теории вероятностей, математической статистики, теории нечетких множеств, генетических алгоритмов, математического моделирования и объектно-ориентированного программирования.

Научная новизна. В рамках диссертационной работы были получены следующие результаты.

1. Разработана методика кластеризации технического состояния зданий и сооружений с произвольным количеством элементов мониторинга.

2. Разработаны и исследованы методы кластеризации технического состояния зданий и сооружений с использованием алгоритма нечетких c -средних и алгоритма возможных c -средних на основе нечетких множеств первого типа и генетических алгоритмов с хромосомой постоянной и переменной длины.

3. Разработаны и исследованы методы кластеризации технического состояния зданий и сооружений с использованием алгоритма нечетких c -средних и алгоритма возможных c -средних на основе интервальных нечетких множеств второго типа и генетических алгоритмов с хромосомой постоянной длины.

4. Разработана методика выбора метода кластеризации технического состояния зданий и сооружений на основе нечетких множеств первого типа и интервальных нечетких множеств второго типа.

Практическая ценность работы. Практическая ценность работы состоит в том, что разработанные методы кластеризации позволяют реализовать новый подход к задаче кластеризации технического состояния зданий и сооружений в условиях неопределенности и обеспечивают:

- высокую обоснованность и адекватность принятия решения в условиях неопределенности и неточности исходной информации;
- минимизацию временных и финансовых затрат, связанных как с необходимостью сбора и учета точных и полных исходных данных (что может быть принципиально невозможным), так и с необходимостью многократной реализации классических алгоритмов кластеризации с целью выбора соответствующих оптимальных параметров, обеспечивающих принятие адекватных решений.

В конечном итоге предлагаемый подход обеспечивает эффективное решение задачи кластеризации технического состояния зданий и сооружений в условиях неопределенности как комплексной оценки с применением инструментальных исследований и с привлечением субъективного «человеческого фактора».

Достоверность полученных в диссертационной работе результатов подтверждается:

- использованием понятий и выводов теории нечетких множеств и теории генетических алгоритмов;
- результатами математического моделирования предложенных методов на ПЭВМ;
- разработкой действующих программных средств, подтвержденных свидетельствами об официальной регистрации;
- апробацией предложенных методик расчета для конкретных случаев;
- наличием актов внедрения результатов диссертационной работы.

На защиту выносятся:

1. Методика кластеризации технического состояния зданий и сооружений с произвольным количеством элементов мониторинга.

2. Методы кластеризации технического состояния зданий и сооружений с использованием алгоритма нечетких c -средних и алгоритма возможных c -средних на основе нечетких множеств первого типа и генетических алгоритмов с хромосомой постоянной и переменной длины, позволяющие учесть свойства кластерной относительности и типичности соответственно и обеспечивающие получение адекватных результатов кластеризации множества объектов, содержащего кластеры подобной плотности и подобного объема, с минимальными временными затратами.

3. Методы кластеризации технического состояния зданий и сооружений с использованием алгоритма нечетких c -средних и алгоритма возможных c -средних на основе

интервальных нечетких множеств второго типа и генетических алгоритмов с хромосомой постоянной длины, позволяющие учесть свойства кластерной относительности и кластерной типичности соответственно и обеспечивающие получение адекватных результатов кластеризации множества объектов, содержащего кластеры существенно разной плотности или существенно разного объема, с минимальными временными затратами.

4. Методика выбора метода кластеризации технического состояния зданий и сооружений на основе нечетких множеств первого типа и интервальных нечетких множеств второго типа, обеспечивающая получение адекватных результатов кластеризации.

5. ППП для кластеризации технического состояния зданий и сооружений на основе разработанных методов кластеризации.

Внедрение результатов. Результаты работы внедрены и используются в деятельности ООО «Независимый центр оценки и экспертиз», а также в работе Федерального бюджетного управления «Отдел капитального строительства и ремонта Управления Федеральной службы исполнения наказания (ОКСР УФСИН)» при решении задачи оценки технического состояния зданий и сооружений как задачи кластеризации при нечётком определении состояний многомерных объектов. Опытная эксплуатация ППП «Кластеризация технического состояния зданий и сооружений в условиях неопределенности» показала высокие характеристики надежности и эффективности разработанного программного обеспечения при решении задач кластеризации совокупностей объектов, содержащих как кластеры подобной плотности и подобного объема, так и кластеры существенно разной плотности или существенно разного объема.

Результаты полученных в диссертации исследований используются в учебном процессе Рязанского государственного радиотехнического университета при обучении студентов специальности 230105 «Программное обеспечение вычислительной техники и автоматизированных систем» в курсе «Проектирование искусственного интеллекта», специальности 080801 «Прикладная информатика в экономике» в курсах «Информационные технологии» и «Элементы теории нечетких множеств», а также в учебном процессе Рязанского института (филиала) Московского государственного открытого университета при обучении студентов специальностей 270102 «Промышленное и гражданское строительство» в курсе «Обследование и испытание зданий и сооружений», специальности 080502 «Экономика и управление на предприятии строительства» при изучении дисциплины «Анализ и диагностика финансово-хозяйственной деятельности на предприятии», специальности 270114 «Проектирование зданий» при изучении дисциплины «Основы реконструкции и реставрации», специальности 080507 «Производственный менеджмент в строительстве» при изучении дисциплины «Экономика реконструкций зданий и сооружений».

Использование результатов диссертационной работы на практике подтверждено соответствующими актами о внедрении.

Апробация работы. Основные научные положения диссертационной работы докладывались и обсуждались на научных конференциях: VIII всероссийской научно-технической конференции «Проблемы информатики в образовании, управлении, экономике и технике» (Пенза, 2008); VI международной научно-технической конференции «Искусственный интеллект в XXI веке. Решения в условиях неопределенности» (Пенза, 2008); 34-й всероссийской научно-технической конференции «Сети, системы связи и телекоммуникации. Деятельность вуза при переходе на Федеральный государственный

образовательный стандарт 3-го поколения» (Рязань, 2009); 17-й международной научной конференции «Методы и алгоритмы принятия эффективных решений» (Таганрог, 2009); XIV всероссийской научно-технической конференции студентов, молодых ученых и специалистов «Новые информационные технологии в научных исследованиях и в образовании» (Рязань, 2009); XIV и XV международных открытых научных конференциях «Современные проблемы информатизации в анализе и синтезе технологических и программно-телекоммуникационных систем» (г. Воронеж, 2009, 2010); международной научной конференции «Информационные системы и технологии» (Нижний Новгород, 2010 г.).

Публикации. По теме диссертации опубликовано 15 печатных работ. В их числе 1 статья в рецензируемой печати, 3 статьи в межвузовских сборниках, 1 статья в научно-техническом журнале, 4 доклада на международных конференциях, 4 доклада на всероссийских конференциях, 2 свидетельства об официальной регистрации подкомплексов программ в Отраслевом фонде алгоритмов и программ.

Структура и объем диссертации. Диссертационная работа состоит из введения, четырех глав, списка литературы и двух приложений. Содержит 291 страниц (из них 250 страниц – основная часть, 41 страниц – приложения), 16 таблиц, 65 рисунков. Список литературы состоит из 194 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 «Обзор и анализ подходов к решению задачи оценки и классификации технического состояния зданий и сооружений» сформулирована задача классификации технического состояния зданий и сооружений в условиях современного города, предполагающая использование современной нормативной базы по комплексному обследованию и мониторингу технического состояния зданий и сооружений.

Систематизированы требования, рекомендации и мероприятия по обследованию и мониторингу технического состояния зданий и сооружений. Показано, что наиболее актуальной задачей при мониторинге технического состояния зданий и сооружений является задача выявления проблемных ситуаций на основе принципиально новых подходов, позволяющих по своему характеру и финансовому исполнению поддерживать здания и сооружения современного города в рабочем состоянии и обеспечивающих принятие своевременных обоснованных и адекватных решений в штатных и аварийных ситуациях.

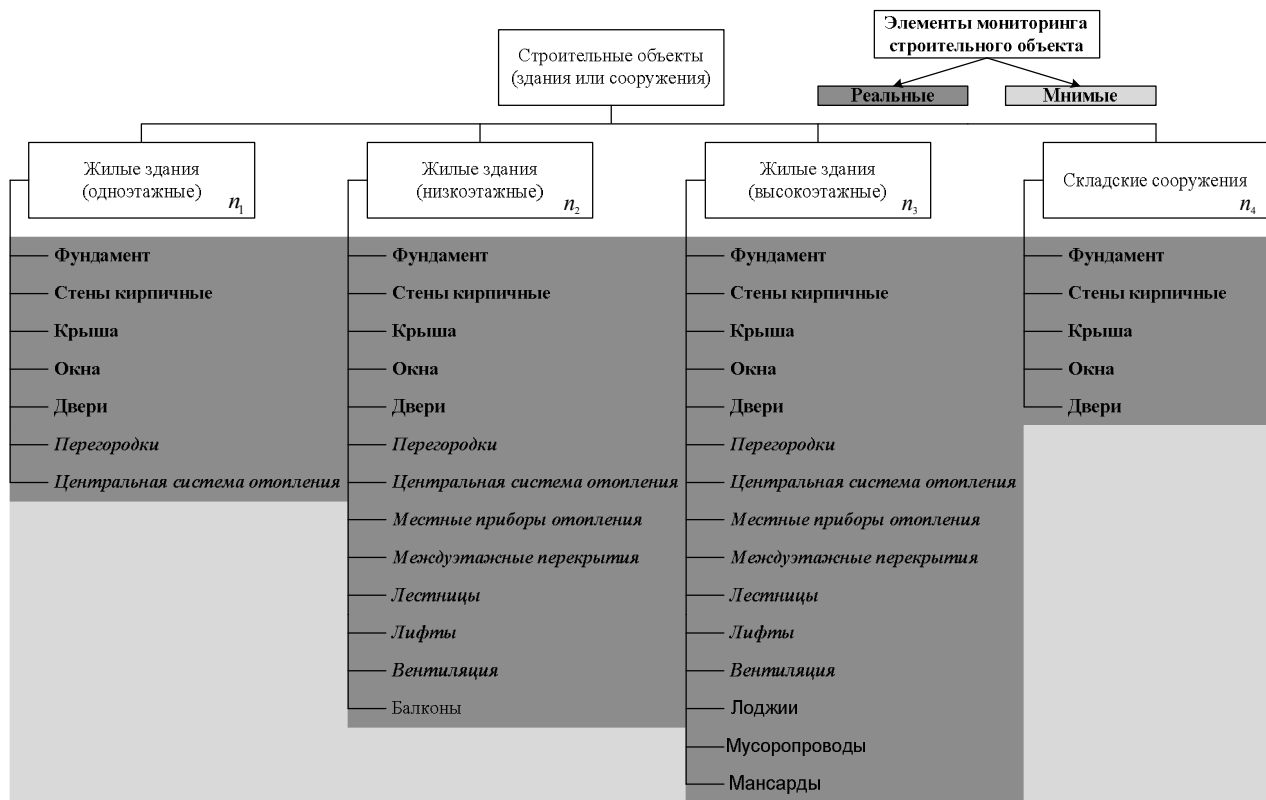
Сформулированы принципы классификации и кластеризации объектов в задачах многокритериального анализа. Выполнен анализ базового алгоритма кластеризации – алгоритма четких *c*-средних, показавший, что данный алгоритм обеспечивает получение адекватных результатов кластеризации, если кластеры компактны и хорошо отделимы. Выявлена проблема кластеризации множеств объектов, содержащих кластеры существенно разной плотности или существенно разного объема и т.п. Определены основные подходы к решению задачи классификации технического состояния зданий и сооружений, основанные на применении аппарата ТНМ и предполагающие использование алгоритмов кластеризации на основе НМТ1 и ИНМТ2. Принято решение о необходимости разработки новых методов кластеризации, позволяющих учитывать свойства кластерной

относительности и/или типичности объектов кластеризации и обеспечивающих получение обоснованных и адекватных результатов кластеризации с минимальными временными затратами.

Предложено для решения задачи поиска оптимальных параметров алгоритмов кластеризации с минимальными временными затратами использовать генетические алгоритмы, хорошо зарекомендовавшие себя при решении широкого спектра прикладных задач. Показано, что существующие в настоящее время подходы к оценке и кластеризации технического состояния зданий и сооружений имеют ряд существенных недостатков. В первую очередь, в большинстве известных подходов количество классов (кластеров) и элементов мониторинга невелико и жестко фиксировано. В то же время в последние годы количество возможных (потенциальных) элементов мониторинга неуклонно растет, что связано как с использованием новых строительных материалов, так и с применением новых методов и средств инструментальных исследований. Также показано, что недостатки существующих подходов к оценке и кластеризации технического состояния зданий и сооружений связаны с неопределенностью выбора элементов мониторинга и оптимального количества кластеров (что важно при планировании обычно ограниченных по своему размеру финансовых вложений на проведение капитального и текущего ремонта), а также с особенностями инструментального и экспертного оценивания элементов мониторинга: оценивание может выполняться либо посредством оценивания «процента износа» элемента мониторинга (обычно с помощью инструментальных средств контроля), либо с помощью балльного (экспертного) оценивания. Показано, что значительной проблемой являются наличие типов строительных объектов с разным количеством элементов мониторинга для малоэтажных и многоэтажных зданий, складских помещений и т.п. – $l_1, l_2, \dots, l_k$ (l_i – количество элементов мониторинга строительных объектов i -го типа, $i = 1, k$, k – количество типов строительных объектов), а также существенно различные количества строительных объектов разных типов – $n_1, n_2, \dots, n_k$ (n_i – количество строительных объектов i -го типа).

В связи с этим предложена методика кластеризации технического состояния зданий и сооружений с произвольным количеством элементов мониторинга. Для множества зданий и сооружений, подлежащих оценке их технического состояния и классификации, определяется полный перечень всех элементов мониторинга. Далее для каждого i -го типа строительных объектов все элементы мониторинга из общего перечня элементов мониторинга разбиваются на две группы: реальные (действительно имеющиеся у данного типа строительных объектов) элементы мониторинга и мнимые (отсутствующие у данного типа строительных объектов) элементы мониторинга (рис. 1; цветом показаны области реальных и мнимых элементов мониторинга). Оценивание реальных элементов мониторинга выполняется обычным образом (либо посредством оценивания «процента износа» элемента мониторинга, либо с помощью балльного оценивания).

При наличии у строительного объекта мнимого элемента мониторинга «процент износа» (или экспертный балл оценивания) по нему принимается равным нулю. В таком случае, независимо от того, сколько реальных элементов мониторинга имеется у строительного объекта каждого типа, удастся выполнить комплексную оценку всего множества оцениваемых зданий и сооружений в целом, а следовательно, и их адекватную кластеризацию, предполагающую (по своей сути) группирование объектов с близкими или наиболее различающимися оценками (в зависимости от применяемого алгоритма кластеризации) в один кластер.



n_i ($i = \overline{1,4}$) – количество строительных объектов данного типа

При наличии у строительного объекта «мнимого» элемента мониторинга «процент износа» по нему равен 0

Рис. 1. Кластеризация технического состояния строительных объектов с произвольным количеством элементов мониторинга

Показано, что использование общих принципов кластеризации объектов позволит разбить множество строительных объектов либо на заранее заданное количество кластеров, либо на некоторое предварительно неизвестное количество кластеров, являющееся оптимальным для данного множества строительных объектов с точки зрения выбранного показателя качества кластеризации.

Во второй главе «Разработка методов кластеризации на основе нечетких множеств первого типа и генетических алгоритмов» рассматривается проблема разработки методов кластеризации множества строительных объектов (на основе их технического состояния), содержащего кластеры подобного объема и подобной плотности, с использованием алгоритмов кластеризации на основе НМТ1 и ГА, с минимальными временными затратами.

Пусть n – количество строительных объектов, q – количество элементов мониторинга, c – количество кластеров. Для кластеризации строительных объектов предлагается использовать классический алгоритм нечеткой кластеризации – алгоритм нечетких c -средних (FCM-алгоритм на основе НМТ1) и его модификацию – алгоритм возможных c -средних (PCM-алгоритм на основе НМТ1).

FCM-алгоритм на основе НМТ1 реализует минимизацию целевой функции:

$$J(U, V) = \sum_{j=1}^c \sum_{i=1}^n (u_j(x_i))^m \cdot d_{ji}^2 \quad (d_{ji}^2 = \sum_{l=1}^q (x_i^l - v_j^l)^2), \quad (1)$$

где $U = [u_j(x_i)]$ – нечеткое c -разбиение множества объектов $\{x_i\}$ на основе функций принадлежности (ФП) $u_j(x_i)$; $V = (v_1, \dots, v_c)$ – центры кластеров; d_{ji} – евклидово расстояние между центром кластера v_j и объектом x_i : $d_{ji} = \|x_i - v_j\|$; m – фаззификатор ($m \in R, m > 1$); c – количество кластеров; n – количество объектов; q – количество элементов мониторинга.

FCM-алгоритм основан на учете свойства кластерной относительности, предполагает инициализацию начального разбиения $U = [u_j(x_i)]$ и выполнение следующих шагов.

1. Вычисление центров кластеров:

$$v_j = \frac{\sum_{i=1}^n (u_j(x_i))^m \cdot x_i}{\sum_{i=1}^n (u_j(x_i))^m}. \quad (2)$$

2. Вычисление ФП:

$$u_j(x_i) = 1 / \sum_{k=1}^c (d_{ji} / d_{ki})^{2/(m-1)} \quad (\sum_{j=1}^c u_j(x_i) = 1, \forall i = \overline{1, n}). \quad (3)$$

3. Шаги 1 и 2 повторяются до тех пор, пока не будет выполнено заданное число итераций s или не будет достигнута заданная точность: $|J(U, V) - J'(U, V)| \leq \varepsilon$, где $J(U, V)$, $J'(U, V)$ – значения целевой функции на двух последовательных итерациях.

Показано, что в качестве показателей качества кластеризации, которые должны быть минимизированы, с использованием FCM-алгоритма на основе НМТ1 целесообразно использовать индекс Се – Бени, характеризующий компактность и отделимость кластеров гиперсферической формы:

$$XB = \frac{\sum_{j=1}^c \sum_{i=1}^n (u_j(x_i))^2 \cdot \|x_i - v_j\|^2}{n \cdot \min_{i \neq j} \|v_i - v_j\|^2}, \quad (4)$$

и нечеткий общий гиперобъем для кластеров гиперэллиптической формы:

$$FH = \sum_{j=1}^c (det(R_j))^{1/2} \quad (R_j = \sum_{i=1}^n ((u_j(x_i))^m \cdot (x_i - v_j) \cdot (x_i - v_j)^T) / \sum_{i=1}^n u_j(x_i)^m), \quad (5)$$

где $u_j(x_i)$ – ФП объекта x_i кластеру X_j ; v_j – вектор координат центра j -го кластера; x_i – вектор координат i -го объекта; $j = \overline{1, c}$; $t = \overline{1, c}$; R_j – нечеткая ковариационная матрица j -го кластера; $det(R_j)$ – определитель матрицы R_j ; c – количество кластеров; n – количество объектов.

PCM-алгоритм на основе НМТ1 используется, если множество объектов содержит атипичные объекты для возможностной интерпретации неопределенности, что позволяет улучшить результаты кластеризации за счет ослабления свойства кластерной относительности и учета свойства кластерной типичности.

PCM-алгоритм на основе НМТ1 реализует минимизацию целевой функции:

$$J(W, V) = \sum_{j=1}^c \sum_{i=1}^n (w_j(x_i))^m \cdot d_{ji}^2 + \sum_{j=1}^c \eta_j^2 \cdot \sum_{i=1}^n (1 - w_j(x_i))^m, \quad (6)$$

где $W = [w_j(x_i)]$ – возможностное c -разбиение множества объектов $\{x_i\}$ на основе функций типичности (ФТ) $w_j(x_i)$; $V = (v_1, \dots, v_c)$ – центры кластеров; d_{ji} – расстояние между центром кластера v_j и объектом x_i ; m – фаззификатор; η_j ($j = \overline{1, c}$) – «ширина зоны», определяющая расстояние, на котором значение ФТ объекта j -му кластеру равно 0,5; c – количество кластеров; n – количество объектов кластеризации.

Возможностные ФТ, определяемые через абсолютное расстояние между центром кластера v_j и объектом x_i , могут быть вычислены как

$$w_j(x_i) = 1 / \left(1 + (d_{ji} / \eta_j)^{2/(m-1)} \right) \quad \left(\sum_{i=1}^n w_j(x_i) = 1, \forall j = \overline{1, c} \right). \quad (7)$$

Координаты центров кластеров находятся в соответствии с формулой (2).

«Ширина зоны» приблизительно может быть оценена как

$$\eta_j = \left(\frac{K}{n_j} \cdot \sum_{i=1}^n (u_j(x_i))^m \cdot d_{ij}^2 \right)^{1/2}, \quad (8)$$

где $u_j(x_i)$ – ФП, определяющая степень принадлежности объекта x_i кластеру X_j по формуле (3) для FCM-алгоритма; d_{ji} – расстояние между центром кластера v_j и объектом x_i ; n_j – количество объектов, отнесенных к j -му кластеру; c – количество кластеров; n – количество объектов; K – действительное число (обычно $K = 1$).

Инициализация начального разбиения в PCM-алгоритме на основе НМТ1 выполняется с помощью нескольких итераций FCM-алгоритма. Результаты кластеризации существенным образом зависят от правильности выбора значений «ширины зоны».

Для получения адекватных результатов кластеризации с использованием некоторого алгоритма кластеризации на основе НМТ1, определяющего локально-оптимальное разбиение, необходимо многократное выполнение этого алгоритма при заданном количестве кластеров c для различных исходных разбиений с целью определения разбиения, которому соответствует минимальное значение показателя качества кластеризации (целевой функции), для принятия окончательного решения об искомом разбиении, что приводит к значительным временным затратам. Для минимизации временных затрат при получении адекватных результатов кластеризации предлагается использовать методы кластеризации, реализующие поочередное выполнение соответствующего алгоритма кластеризации (нечетких c -средних или возможностных c -средних) и ГА с хромосомой постоянной и переменной длины.

Хромосомы разной длины при кластеризации n объектов на c кластеров по q элементам мониторинга



Рис. 2. Кодирование хромосомы переменной длины координатами центров кластеров

Для заданного количества кластеров c хромосома может быть закодирована координатами центров кластеров или степенями принадлежности (числами из интервала $[0, 1]$) объектов центрам кластеров. При кодировании хромосомы центрами кластеров длина хромосомы равна $c \cdot q$, где c – числа кластеров, q – количество элементов мониторинга: первые q координат соответствуют центру первого кластера, вторые q координат – центру второго кластера и т.п. При кодировании хромосомы степенями принадлежности объектов центрам нечетких кластеров длина хромосомы равна $c \cdot n$, где c – количество кластеров, n – количество объектов: первые c элементов хромосомы соответствуют степеням принадлежности первого объекта центрам кластеров, вторые c элементов хромосомы – степеням принадлежности второго объекта центрам кластеров и т.п.

При реализации ГА рассмотрены два случая: количество кластеров c заранее задано (фиксировано); количество кластеров c заранее не задано (неизвестно). В первом случае предлагается использовать ГА с хромосомой постоянной длины, а во втором – ГА с хромосомой переменной длины (рис. 2). ГА с хромосомой постоянной длины можно рассматривать как частный случай ГА с хромосомой переменной длины. Реализация ГА с хромосомой переменной длины аналогична реализации ГА с хромосомой постоянной длины, за исключением особенностей выполнения операции скрещивания. В качестве функции соответствия используется показатель качества кластеризации по формуле (4) или (5).

Ниже описан предлагаемый метод кластеризации с использованием алгоритма кластеризации на основе НМТ1 и ГА при кодировании хромосомы постоянной длины координатами центров кластеров, предполагающий выполнение следующих шагов.

1. Выполняется один шаг FCM-алгоритма (PCM-алгоритма) на основе НМТ1 при формировании хромосом начальной популяции размером P .

2. При $g < G$ (G и g – максимальное и текущее количество поколений ГА) выполняется один шаг ГА с реализацией операций скрещивания и мутации и вычислением значений функции соответствия для новой популяции хромосом размером $(P + R_c \cdot P)$.

3. Для новой популяции размером $(P + R_c \cdot P)$ выполняется один шаг FCM-алгоритма (PCM-алгоритма) на основе НМТ1 с вычислением новых значений ФП (ФТ) объектов центрам кластеров в соответствии с формулой (3) [формулой (7)], новых координат центров кластеров в соответствии с формулой (2) и значений функции соответствия.

4. Из расширенной популяции размером $(2 \cdot P + R_c \cdot P)$, полученной путем объединения популяции размером P предыдущего шага и популяции размером $(P + R_c \cdot P)$ текущего шага, удаляются «нежизнеспособные» $(P + R_c \cdot P)$ хромосом с максимальными значениями функции соответствия. Если $g < G$, осуществляется переход к шагу 2. Если $g \geq G$, то работа ГА завершается и осуществляется переход к шагу 5.

5. Выбирается лучшая хромосома, которая минимизирует функцию соответствия. Искомые координаты центров кластеров определяются на основе лучшей хромосомы. В качестве результирующих степеней принадлежности объектов центрам кластеров полагаются степени принадлежности объектов центрам кластеров, соответствующие лучшей хромосоме, вычисленные в ходе реализации комбинированного метода нечеткой (возможностной) кластеризации.

Методы кластеризации с использованием алгоритмов кластеризации на основе НМТ1 и ГА с хромосомой переменной длины реализуются аналогичным образом.

В процессе реализации предлагаемых методов кластеризации (при формировании начальной популяции и выполнении операций скрещивания и мутации) проверяется требование о разбиении множества объектов на заданное количество кластеров c (фактическое, а не формальное). При невыполнении этого требования хромосома признается «нежизнеспособной»: в соответствие ей ставится максимально возможное значение функции соответствия с целью дальнейшего исключения из популяции. В результате популяция состоит только из «жизнеспособных» хромосом. При реализации ГА выполняется односточечное скрещивание. Анализ результатов кластеризации при различных способах кодирования хромосом показал, что кодирование хромосомы степенями принадлежности объектов центрам кластеров следует использовать, если количество объектов n невелико и не превосходит количество элементов мониторинга q в 2-5 раз. Кодирование хромосомы координатами центров кластеров при любом количестве объектов n и

любом количестве элементов мониторинга q обеспечивает получение более стабильных результатов кластеризации. Оптимальный размер популяции, обеспечивающий получение адекватных результатов кластеризации множества объектов в 95-98 случаях из 100, составляет 40-50 хромосом.

Предлагаемые методы кластеризации позволяют существенно уменьшить время реализации, необходимое для получения адекватных результатов кластеризации, по сравнению со временем реализации классических FCM-алгоритма и РСМ-алгоритма на основе НМТ1 в 3-15 раз. Реализация ГА без использования FCM-алгоритма и РСМ-алгоритма на основе НМТ1 как самостоятельного шага вычислений оказывается менее эффективной, так как результаты кластеризации будут существенно зависеть от того, насколько хорошо выполнена инициализация центров кластеров.

Предложены методы кластеризации с использованием алгоритма нечетких c -средних и алгоритма возможностных c -средних и ГА с хромосомой постоянной и переменной длины при кодировании хромосомы координатами центров кластеров. Предложен метод кластеризации с использованием алгоритма нечетких c -средних и ГА с хромосомой постоянной длины при кодировании хромосомы степенями принадлежности объектов центрам кластеров. Реализация методов кластеризации с хромосомой переменной длины при кодировании хромосом степенями принадлежности объектов центрам кластеров не представляется возможной ввиду особенностей способа кодирования хромосомы степенями принадлежности объектов центрам кластеров и особенностей реализации операции скрещивания при данном способе кодирования.

Предложена методика выбора метода кластеризации (с использованием алгоритма нечетких c -средних или алгоритма возможностных c -средних) на основе НМТ1, обеспечивающего получение адекватных результатов кластеризации множества объектов, содержащего кластеры подобного объема и подобной плотности, являющихся лучшими по достижению минимального значения показателя качества кластеризации.

Эффективность применения метода кластеризации с использованием алгоритма нечетких c -средних на основе НМТ1 и ГА с хромосомой постоянной длины (при кодировании хромосом координатами центров кластеров) может быть продемонстрирована на примере кластеризации множества строительных объектов, экспертные оценки (по 10-балльной шкале) по пяти элементам мониторинга которых приведены в табл. 1.

Результаты кластеризации и параметры моделирования строительных объектов на за-

ранее заданное количество кластеров $c = 4$ с применением классического алгоритма нечетких c -средних на основе НМТ1 (модель «М 1»; I и II – худший и лучший варианты соответственно); с применением ГА (модель «М 2»); с применением метода кластеризации с использованием алгоритма нечетких c -средних на основе НМТ1 и ГА с хромосомой постоянной длины («М 3») приведены в табл. 2 и 3 соответственно. Жирным шрифтом в табл. 2 выделен 6-й объект, ошибочно кластеризуемый моделью «М 1».

Табл. 1. Оценки строительных объектов

Объект	Элементы мониторинга				
	Фундамент	Стена	Перекрытия	Крыша	Окна
1	1	1	1	1	1
2	1	2	3	4	2
3	2	2	2	2	2
4	5	5	5	5	5
5	10	10	10	10	10
6	9	8	8	7	10
7	5	7	6	8	7
8	4	5	7	2	4
9	7	6	8	7	9
10	6	6	6	3	7

Табл. 2. Результаты кластеризации при $c = 4$

Кластер	Модель			
	«М 1»		«М 2»	«М 3»
	«I»	«II»		
Ветхие здания	5, 6	5	5	5
Здания, требующие капитального ремонта	7, 9	6, 7, 9	6, 7, 9	6, 7, 9
Здания, требующие текущего ремонта	4, 8, 10	4, 8, 10	4, 8, 10	4, 8, 10
Здания, находящиеся в идеальном состоянии	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3

Несмотря на то, что результаты кластеризации (разбиение на нечеткие кластеры) на основе моделей «М 2» и «М 3» совпали, реализация модели «М 2» потребовала значительно большего количества поколений ГА, чем реализация модели «М 3», что подтверждает высокую эффективность метода кластеризации с использованием алгоритма нечетких c -средних на основе НМТ1 и ГА с хромосомой постоянной длины.

Табл. 3. Характеристики моделей при $c = 4$

Параметры моделирования	Модель			
	«М 1»		«М 2»	«М 3»
	«I»	«II»		
Размер популяции	-	-	20	20
Индекс Се – Бени	0,0225	0,0151	0,0151	0,0151
Количество итераций (поколений)	31	90	20000	144
Время кластеризации (с)	0,003	0,009	137,206	1,357

Кластеризация этого же множества строительных объектов на произвольное (заранее неизвестное) количество кластеров c с применением метода кластеризации с использованием алгоритма нечетких c -средних на основе НМТ1 и ГА с хромосомой переменной длины при варьировании количества кластеров от 2 до 8 показала,

что оптимальное количество кластеров равно 6, что подтверждается минимальным значением индекса Се – Бени, равным 0,0047 (рис. 3).

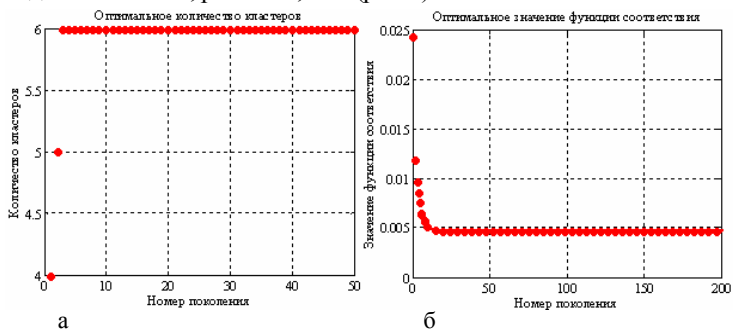


Рис. 3. Графические зависимости для генетического алгоритма с хромосомой переменной длины:
а – для оптимального количества кластеров;
б – для значений индекса Се – Бени

На рис. 4 приведен пример, демонстрирующий эффективность применения метода кластеризации с использованием алгоритма возможных c -средних на основе НМТ1 и ГА с хромосомой постоянной длины, реализующего кластеризацию с учетом свойства кластерной типичности. При этом можно предположить, что кластеризация выполняется по двум элементам мониторинга - «фундаменты» и «стены» - по величине «процента износа». Кластеризация множества объектов (рис. 4, а), содержащего три «вертикальных» кластера, близко расположенных по отношению друг к другу так, что расстояние между двумя любыми соседними объектами по горизонтали больше расстояния между

двумя любыми соседними объектами по вертикали (объекты разных кластеров помечены маркерами разной формы), с применением метода кластеризации с использованием алгоритма нечетких c -средних на основе НМТ1 и ГА с хромосомой постоянной длины дает неадекватные результаты (рис. 4, б).

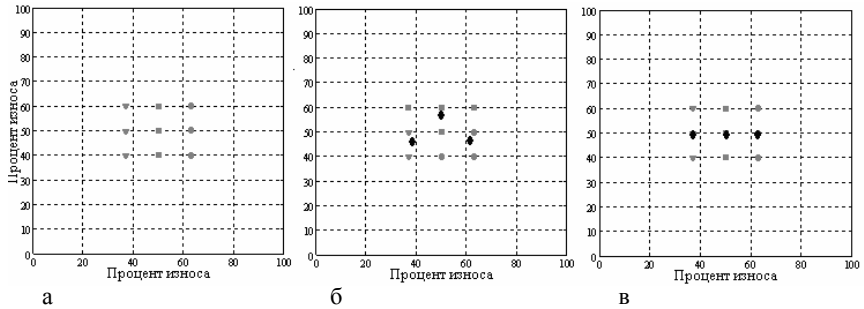


Рис. 4. Кластеризация объектов на основе НМТ1:

а – множество объектов кластеризации;

б – FCM-алгоритм на основе НМТ1 при $m = 2$;

в – PCM-алгоритм на основе НМТ1 при $m = 2$ с автоматическим выбором «ширины зоны»

В то же время применение метода кластеризации с использованием алгоритма возмозможностных c -средних на основе НМТ1 и ГА с хромосомой постоянной длины позволяет получить искомые адекватные результаты кластеризации (рис. 4, в). Центры кластеров на рисунках помечены черными маркерами ромбовидной формы.

В третьей главе «Разработка методов кластеризации на основе интервальных нечетких множеств второго типа и генетических алгоритмов» рассматривается проблема разработки методов кластеризации множества строительных объектов (на основе их технического состояния), содержащего кластеры существенно разной плотности или существенно разного объема, с использованием алгоритмов кластеризации на основе ИНМТ2 и ГА, с минимальными временными затратами.

Если кластеры в множестве объектов имеют существенно разную плотность или существенно разный объем, то работа алгоритма кластеризации на основе НМТ1 существенно зависит от выбора параметров алгоритма кластеризации. Так, например, FCM-алгоритм на основе НМТ1 может дать плохие результаты кластеризации при выборе несоответствующего фаззификатора m из-за различия в плотности или объеме между кластерами. В связи с этим предлагается использовать FCM-алгоритм на основе ИНМТ2, предполагающий расширение множества объектов кластеризации на ИНМТ2, обеспечивающее управление неопределенностью, связанной с заданием максимальной нечеткой области за счет введения двух фаззификаторов m_1 и m_2 , что позволяет существенно улучшить результаты кластеризации. При определении «отпечатка неопределенности» ИНМТ2 для объекта x_i рассматриваются «нижняя» и «верхняя» ФП с использованием двух различных значений фаззификатора m :

$$\bar{u}_j(x_i) = \begin{cases} u_j^1(x_i), & \text{если } u_j^1(x_i) > u_j^2(x_i) \\ u_j^2(x_i), & \text{если } u_j^1(x_i) \leq u_j^2(x_i) \end{cases}, \quad u_j(x_i) = \begin{cases} u_j^1(x_i), & \text{если } u_j^1(x_i) \leq u_j^2(x_i) \\ u_j^2(x_i), & \text{если } u_j^1(x_i) > u_j^2(x_i) \end{cases}, \quad (9)$$

где $u_j^p(x_i) = 1 / \sum_{k=1}^c (d_{ji} / d_{ki})^{2/(m_p-1)}$, $p = 1, 2$.

Использование фаззификаторов m_1 и m_2 , определяющих различные степени нечеткости, приводит к различным целевым функциям для FCM-алгоритма на основе ИНМТ2, которые необходимо минимизировать при $m = m_1$ и $m = m_2$:

$$J_{m_1}(U, V) = \sum_{j=1}^c \sum_{i=1}^n (u_j(x_i))^{m_1} \cdot d_{ji}^2, \quad (10)$$

$$J_{m_2}(U, V) = \sum_{j=1}^c \sum_{i=1}^n (u_j(x_i))^{m_2} \cdot d_{ji}^2. \quad (11)$$

Аналогичным образом определяется РСМ-алгоритм на основе ИНМТ2. При этом могут быть рассмотрены два случая: неопределенность, связанная с выбором значений фаззификаторов (как в FCM-алгоритме на основе ИНМТ2), и неопределенность, связанная с выбором значений «ширины зоны». Так, в первом случае для РСМ-алгоритма на основе ИНМТ2 рассматриваются две целевые функции при $m = m_1$ и $m = m_2$:

$$J_{m_1}(W, V) = \sum_{j=1}^c \sum_{i=1}^n (w_j(x_i))^{m_1} \cdot d_{ji}^2 + \sum_{j=1}^c \eta_j^2 \cdot \sum_{i=1}^n (1 - w_j(x_i))^{m_1}, \quad (12)$$

$$J_{m_2}(W, V) = \sum_{j=1}^c \sum_{i=1}^n (w_j(x_i))^{m_2} \cdot d_{ji}^2 + \sum_{j=1}^c \eta_j^2 \cdot \sum_{i=1}^n (1 - w_j(x_i))^{m_2}, \quad (13)$$

при этом в качестве $u_j^P(x_i)$ в формулах (9) используются ФТ $w_j^P(x_i)$ по формуле (7).

В качестве показателей качества кластеризации, которые должны быть минимизированы, для множества объектов, содержащего кластеры гиперсферической формы, целесообразно использовать индекс *Sph*:

$$Sph = \sum_{j=1}^c r_j / \min_{\substack{j=1,c; t=1,c \\ t \neq j}} d(v_j, v_t), (r_j = \max_{x_i \in X_j} d(x_i, v_j)), \quad (14)$$

а для кластеров гиперэллиптической формы – общий гиперобъем:

$$H = \sum_{j=1}^c (det(R_j))^{1/2}, (R_j = \frac{1}{n_j} \cdot \sum_{i=1}^n (x_i - v_j) \cdot (x_i - v_j)^T), \quad (15)$$

где v_j – вектор координат центра j -го кластера; x_i – вектор координат i -го объекта; $d(x_i, v_j)$ – расстояние между объектом x_i , принадлежащим кластеру X_j , и центром кластера v_j ; $d(v_j, v_t)$ – расстояние между центрами кластеров v_j и v_t ; n – количество объектов; c – количество кластеров; R_j – ковариационная матрица j -го кластера; n_j – количество объектов, отнесенных к j -му кластеру; $det(R_j)$ – определитель ковариационной матрицы R_j .

Управление неопределенностью в алгоритмах кластеризации на основе ИНМТ2 осуществляется с помощью вычисления центров кластеров и получения четкого разбиения для принятия конечного решения о результатах кластеризации. При оценке центра кластера используется итерационный алгоритм Карника – Менделя. При этом сначала находится интервал:

$$v_{\tilde{X}} = \sum_{u(x_1) \in J_{x_1}} \dots \sum_{u(x_n) \in J_{x_n}} 1 / \left(\sum_{i=1}^n x_i \cdot u(x_i)^m / \sum_{i=1}^n u(x_i)^m \right) = [v_{left}, v_{right}], \quad (16)$$

а искомый центр кластера (центроид ИНМТ2) находится как $v = (v_{left} + v_{right})/2$.

При неопределенности выбора фаззификатора для поиска оптимальной комбинации значений фаззификаторов m_1 и m_2 для FCM-алгоритма на основе ИНМТ2 и оптимальной

комбинации фаззификаторов m_1, m_2 и значений «ширины зоны» η_j ($j = \overline{1, c}$) для РСМ-алгоритма на основе ИНТМ2 предлагается использовать соответствующие ГА с хромосомой постоянной длины.

При этом, например, для FCM-алгоритма на основе ИНТМ2 хромосома задается в виде $s = (m_1, m_2)$, а для РСМ-алгоритма на основе ИНТМ2 – в виде $s = (m_1, m_2, \eta_1, \dots, \eta_c)$, где $m_1, m_2 \in (1, m_{\max}]$; $m_1 < m_2$; $\eta_j \in [\eta_j^{\min}, \eta_j^{\max}]$ ($j = \overline{1, c}$); $m_{\max}$, $\eta_j^{\min}$ и $\eta_j^{\max}$ – некоторые действительные числа.

При инициализации РСМ-алгоритма на основе ИНТМ2 с использованием результатов кластеризации, полученных с помощью FCM-алгоритма на основе ИНТМ2, хромосома целесообразно представлять в виде: $s = (\eta_1, \dots, \eta_c)$ (m_1, m_2 – фиксированы).

Предложены метод кластеризации с использованием FCM-алгоритма на основе ИНТМ2 и ГА с хромосомой постоянной длины: управление неопределенностью осуществляется за счет выбора двух фаззификаторов m_1 и m_2 ; два метода кластеризации с использованием РСМ-алгоритма на основе ИНТМ2 и ГА с хромосомой постоянной длины: в одном случае управление неопределенностью осуществляется за счет выбора двух фаззификаторов m_1 и m_2 (как описано выше), во втором случае – за счет выбора комбинации значений «ширины зоны» η_j : η_{j_1} и η_{j_2} , где $\eta_{j_1} < \eta_{j_2}$ ($j = \overline{1, c}$). Для каждого метода кластеризации реализуется вычисление «нижней» и «верхней» ФП по формулам (9) в соответствии с реализуемым способом представления неопределенности. Главной особенностью метода кластеризации с использованием FCM-алгоритма на основе ИНТМ2 и ГА с хромосомой постоянной длины и метода кластеризации с использованием РСМ-алгоритма на основе ИНТМ2 и ГА с хромосомой постоянной длины при неопределенности выбора значений фаззификаторов является необходимость проверки условия $m_1 < m_2$ при формировании начальной популяции и выполнении операций скрещивания и мутации; главной особенностью метода кластеризации с использованием РСМ-алгоритма на основе ИНТМ2 и ГА с хромосомой постоянной длины при неопределенности значений «ширины зоны» является необходимость проверки условий $\eta_{j_1} < \eta_{j_2}$ ($j = \overline{1, c}$) при формировании начальной популяции и выполнении операций скрещивания и мутации.

При этом если сложность реализации FCM-алгоритма на основе НМТ1 оценивается как $O(n^2)$, то сложность реализации FCM-алгоритма на основе ИНТМ2 – как $O(n^3)$ (за счет использования алгоритма Карника – Менделя). Тогда сложность реализации комбинированного метода кластеризации на основе НМТ1 оценивается как $(P+P \cdot Rc \cdot G) \cdot O(n^2)$, а сложность реализации метода кластеризации на основе ИНТМ2 – как $(P+P \cdot Rc \cdot G) \cdot O(n^3)$.

Показано, что расширение множества объектов на ИНТМ2 позволяет представлять и управлять неопределенностью, которая возникает, когда множество объектов содержит кластеры существенно разной плотности или существенно разного объема. Выполненный анализ показал, что алгоритм кластеризации на основе ИНТМ2 является расширением алгоритма кластеризации на основе НМТ1.

Предложена методика выбора метода кластеризации (нечеткой или возможностной) на основе ИНТМ2, обеспечивающего получение адекватных результатов кластеризации множества объектов, содержащего кластеры существенно разного объема или существенно разной плотности, – лучших по достижению минимального значения показателя качества кластеризации. Предложена методика выбора метода кластеризации на основе

НМТ1 или ИНМТ2, обеспечивающего получение адекватных результатов кластеризации произвольного множества объектов кластеризации.

На рис. 5, а показано множество объектов, содержащее 2 кластера подобной плотности и подобного объема (объекты разных кластеров помечены маркерами разной формы). Кластеры представляют собой множества, содержащие по 25 объектов, координаты которых были сгенерированы с использованием нормального закона распределения с математическим ожиданием по каждой координате для левого кластера – 44 и 44, а для правого кластера – 78 и 78. При этом дисперсия по обеим координатам одинакова и равна 16, а правый кластер получен «параллельным переносом» левого кластера. Использование FCM-алгоритма на основе НМТ1 при $m=2$ позволило получить адекватные результаты кластеризации (рис. 5, б). Центры кластеров на рисунках помечены черными маркерами ромбовидной формы. Далее к левому кластеру было добавлено еще 25 объектов, координаты которых так же были получены на основе нормального закона распределения с математическим ожиданием по каждой координате 44 и 44, но с существенной большей дисперсией - больше 841. В результате были получены 2 кластера существенно разной плотности и существенно разного объема (рисунок 5, в), содержащие по 50 и 25 объектов.

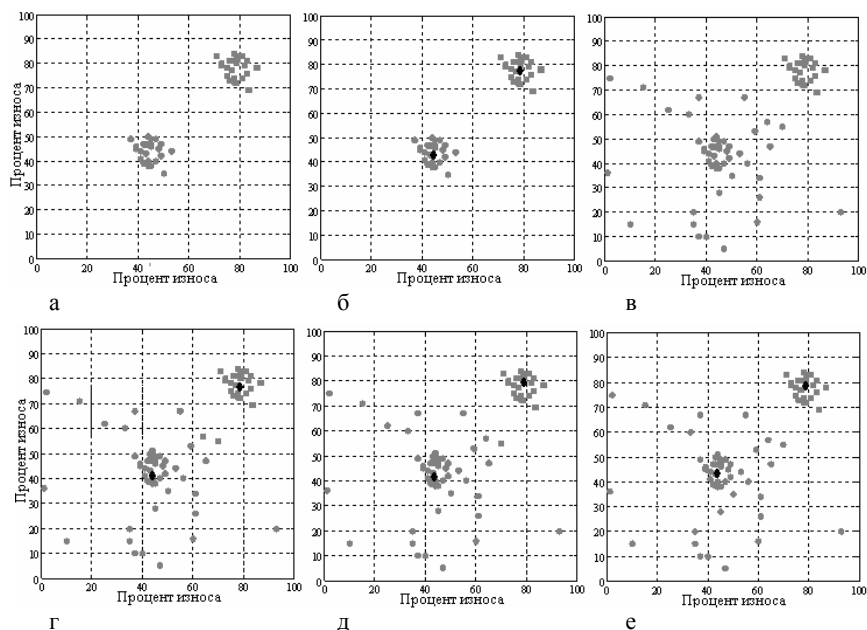


Рис. 5. Кластеризация объектов на основе НМТ1 и ИНМТ2:

а – множество объектов, содержащее кластеры идентичной структуры (подобной плотности и подобного объема);

б – FCM-алгоритм на основе НМТ1 при $m = 2$;

в – множество объектов, содержащее кластеры различной структуры (существенно разной плотности и существенно разного объема);

г – FCM-алгоритм на основе НМТ1 при $m = 2$;

д – FCM-алгоритм на основе ИНМТ2 при $m_1 = 123,566$ и $m_2 = 128,971$;

е – FCM-алгоритм на основе ИНМТ2 при $m_1 = 123,566$ и $m_2 = 128,971$;

$\eta_1 = 142,573$ и $\eta_2 = 74,273$

Табл. 4. Сравнительный анализ алгоритмов кластеризации

Алгоритм кластеризации	Общий гиперобъем	Ошибочно классифицир. объекты
FCM на основе НМТ1 при $m=2$	(нечетк. общ. гиперобъем) 256,803	3
Алгоритм четких c -средних	261,801	3
FCM-алгоритм на основе ИНМТ2	258,809	1
PCM-алгоритм на основе ИНМТ2	255,809	0

На рис. 5, б-е приведены результаты кластеризации с использованием FCM-алгоритма на основе НМТ1 и алгоритмов кластеризации на основе ИНМТ2, оптимальные параметры которых были определены с помощью разработанных методов кластеризации на основе ИНМТ2, демонстрирующие эффективность применения ИНМТ2. В табл. 4 приведены некоторые результаты анализа качества кластеризации, демонстрирующие уменьшение количества ошибочно классифицированных (распознанных) объектов при применении ИНМТ2.

Предлагаемые методы кластеризации на основе ИНМТ2 позволяют управлять неопределенностью, возникающей при анализе множества объектов, содержащего кластеры существенно разной плотности или существенно разного объема, а применение ГА с хромосомой постоянной длины позволяет найти оптимальную комбинацию параметров алгоритма кластеризации на основе ИНМТ2, обеспечивающую адекватные результаты кластеризации, что подтверждается минимальным значением функции соответствия.

В четвертой главе «Программная реализация методов кластеризации на основе нечетких множеств» разработаны программные средства, предназначенные для решения задачи кластеризации технического состояния зданий и сооружений в условиях неопределенности. При этом множество объектов кластеризации 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 жилищно-коммунального хозяйства, для грамотного распределения финансовых ресурсов на проведение капитального и текущего ремонта зданий и сооружений, а также для выявления строительных сооружений, не подлежащих восстановлению.

Подкомплексы программ, входящие в ППП, защищены авторским правом: получены 2 свидетельства об официальной регистрации программ для ЭВМ в Отраслевом фонде алгоритмов и программ.

ППП состоит из двух комплексов программ: «ClusteringT1FS» и «ClusteringT2FS», реализующих кластеризацию объектов с использованием разработанных методов кластеризации на основе НМТ1 и ИНМТ2 соответственно. При этом кластеризация объектов выполняется с использованием предложенной в диссертационной работе методики кластеризации технического состояния зданий и сооружений с произвольным количеством элементов мониторинга.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

Проведенные исследования позволяют сформулировать основные результаты.

1. Выполнено исследование проблемы кластеризации технического состояния зданий и сооружений в условиях неопределенности, выявлены достоинства и недостатки известных алгоритмов кластеризации объектов (как с использованием инструментальных

исследований, так и с применением экспертного оценивания). Проведенный анализ показал, что существующие алгоритмы кластеризации, основанные на применении инструментария ТНМ, зачастую не обеспечивают принятие объективных и адекватных решений ввиду недостаточно обоснованного выбора параметров моделирования, а поиск эффективных решений сопровождается значительными временными затратами из-за необходимости выполнения многократных реализаций классических алгоритмов кластеризации с целью выбора оптимальных параметров моделирования. Сделан вывод о целесообразности использования генетических алгоритмов для решения задачи поиска оптимальных параметров алгоритмов кластеризации объектов в условиях неопределенности.

2. Разработана методика кластеризации технического состояния зданий и сооружений с произвольным количеством элементов мониторинга.

3. Разработаны методы кластеризации технического состояния зданий и сооружений с использованием алгоритма нечетких c -средних и алгоритма возможностных c -средних на основе нечетких множеств первого типа и генетических алгоритмов с хромосомой постоянной и переменной длины, позволяющие учесть свойства кластерной относительности и типичности соответственно и обеспечивающие получение адекватных результатов кластеризации множества объектов, содержащего кластеры подобной плотности и подобного объема, с минимальными временными затратами.

4. Разработаны методы кластеризации технического состояния зданий и сооружений с использованием алгоритма нечетких c -средних и алгоритма возможностных c -средних на основе интервальных нечетких множеств второго типа и генетических алгоритмов с хромосомой постоянной длины, позволяющие учесть свойства кластерной относительности и кластерной типичности соответственно и обеспечивающие получение адекватных результатов кластеризации множества объектов, содержащего кластеры существенно разной плотности или существенно разного объема, с минимальными временными затратами.

5. Разработана методика выбора метода кластеризации технического состояния зданий и сооружений на основе нечетких множеств первого типа и интервальных нечетких множеств второго типа, обеспечивающая получение адекватных результатов кластеризации.

6. Разработан пакет прикладных программ для кластеризации технического состояния зданий и сооружений на основе разработанных методов кластеризации.

7. Исследования разработанных методов кластеризации 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 с необходимостью многократной реализации классических алгоритмов кластеризации с целью выбора соответствующих оптимальных параметров, обеспечивающих принятие адекватных решений.

ОСНОВНЫЕ ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

1. Демидова Л.А., Коняева Е.И. Кластеризация объектов с использованием FCM-алгоритма на основе нечетких множеств второго типа и генетического алгоритма // Вестник РГРТУ. – Рязань, 2008. – № 4 (выпуск 26). – С. 46-54 (в издании, рекомендованном ВАК РФ).
2. Демидова Л.А., Кираковский В.В., Коняева Е.И. Классификация объектов жилого фонда на основе FCM-алгоритма и генетического алгоритма // Математическое и программное обеспечение

вычислительных систем: межвуз. сб. науч. тр. / под ред. А.Н. Пылькина. – М.: Горячая линия – Телеком, 2008. – С. 21-32.

3. Демидова Л.А., Коняева Е.И. Генетический алгоритм поиска оптимальной комбинации значений фазификаторов для FCM-алгоритма на основе 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. 28-30.

4. Демидова Л.А., Коняева Е.И. Комплект программ для кластеризации строительных сооружений с помощью комбинирования генетического и FCM-алгоритмов / Свидетельство об отраслевой регистрации разработки в Отраслевом фонде алгоритмов и программ № 11502 от 08.09.2008.

5. Демидова Л.А., Коняева Е.И. Комплекс программ кластеризации объектов с использованием FCM-алгоритма на основе нечетких множеств второго типа и генетического алгоритма / Свидетельство об отраслевой регистрации разработки в Отраслевом фонде алгоритмов и программ № 11549 от 24.09.2008.

6. Демидова Л.А., Коняева Е.И., Коротаев А.Н. Сравнительный анализ методов кластеризации на основе нечетких множеств первого и второго типа // Современные проблемы информатизации в анализе и синтезе технологических и программно-телекоммуникационных систем: сборник трудов. – Вып. 14 / под ред. д-ра техн. наук, проф. О.Я. Кравца. – Воронеж: Научная книга, 2009. – С. 296-302.

7. Демидова Л.А., Коняева Е.И., Коротаев А.Н. Генетический алгоритм поиска оптимальной комбинации фазификаторов для FCM-алгоритма на основе нечетких множеств второго типа // Информационные технологии моделирования и управления. – Воронеж: Научная книга, 2008. – №6(49). – С. 657-665.

8. Демидова Л.А., Коняева Е.И., Коротаев А.Н. Использование модификаций генетических алгоритмов с переменной длиной хромосомы для решения задачи кластеризации объектов // Методы и алгоритмы принятия эффективных решений (МАПР-09): материалы 17-й международной научной конференции. – 2-я часть. – Таганрог: Изд-во ТТИ ЮФУ, 2009. – С. 22-27.

9. Демидова Л.А., Коняева Е.И., Коротаев 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.

10. Коняева Е.И. Подход к задаче классификации технического состояния зданий и сооружений на основе комбинации FCM-алгоритма и генетического алгоритма // Проблемы информатики в образовании, управлении, экономике и технике: сборник статей VIII Всероссийской научно-технической конференции. – Пенза: Приволжский Дом знаний, 2008. – С. 268-270.

11. Коняева Е.И. Сравнительный анализ способов кодирования хромосом при реализации комбинированного метода нечеткой кластеризации // Информационные и телекоммуникационные технологии. Подготовка специалистов для инфокоммуникационной среды: материалы 34-й Всероссийской научно-техн. конф. – В 2 ч. Ч.1. / РВВКУС, Рязань, 2009. – С. 189-192.

12. Коняева Е.И. Кластеризация объектов жилого фонда на основе FCM-алгоритма и генетического алгоритма с хромосомой переменной длины // Информационные технологии в процессе подготовки современного специалиста: межвуз. сб. статей. – Липецк: ЛГПУ. – 2009. – Выпуск 12. – Том 2. – С. 12-22.

13. Коняева Е.И. Исследование проблемы экспертной оценки технического состояния зданий и сооружений // Математическое и программное обеспечение вычислительных систем: межвуз. сб. науч. тр. / под ред. А.Н. Пылькина. – М.: Горячая линия – Телеком, 2009. – С. 12-16.

14. Коняева Е.И. Подход к проблеме кластеризации объектов с использованием алгоритма нечетких s -средних и генетического алгоритма с хромосомой переменной длины // Новые информационные технологии в научных исследованиях и образовании (НИТ – 2009): материалы XIV Всероссийской научно-технической конференции студентов, молодых ученых и специалистов. – Рязанский государственный радиотехнический университет. 2009. – С. 76-78.

15. Коняева Е.И., Коротаев А.Н. Использование генетических алгоритмов с постоянной и переменной длиной хромосомы для решения задачи кластеризации объектов // Современные проблемы информатизации в анализе и синтезе технологических и программно-телекоммуникационных систем: сборник трудов. – Вып. 15 / под ред. д-ра техн. наук, проф. О.Я. Кравца. – Воронеж: Научная книга, 2010. – С. 296-302.

Коняева Елена Ивановна

**МЕТОДЫ КЛАСТЕРИЗАЦИИ
В ЗАДАЧАХ ОЦЕНКИ
ТЕХНИЧЕСКОГО СОСТОЯНИЯ
ЗДАНИЙ И СООРУЖЕНИЙ
В УСЛОВИЯХ НЕОПРЕДЕЛЁННОСТИ**

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

Подписано в печать . Формат бумаги 60×84 1/16.

Бумага офсетная. Печать офсетная. Усл. печ. л. 2,0.

Тираж 100 экз. Заказ

Рязанский государственный радиотехнический университет.

390005, г. Рязань, ул. Гагарина, д. 59/1.

Редакционно-издательский центр РГРТУ.