

ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

на диссертационную работу Гудымы Михаила Николаевича
«Алгоритмы решения серии задач автоматической группировки»,
представленную на соискание ученой степени кандидата технических наук
по специальности 05.13.01 – Системный анализ, управление и обработка информации (космические и информационные технологии)

Актуальность темы

Задачи автоматической группировки (кластеризации) объектов и данных широко используются в системах интеллектуального анализа данных. Идея наиболее популярных методов кластеризации, т.е. разделения объектов на группы, основана на минимизации суммарных расстояний между объектами одного кластера или между объектами кластера и его центром. Такие задачи NP-трудны, а методам их решения посвящено много научных публикаций. Используемые при этом меры расстояния зависят от особенностей решаемых задач. Определение количества групп обычно выделяется в отдельную, довольно сложную задачу. Диссертационная работа Гудымы М.Н. посвящена разработке генетических алгоритмов для решения задач автоматической группировки с заранее неизвестным числом групп, способных одновременно получать приближенные решения нескольких задач, различающихся только числом групп. Таким образом, в диссертации ставится и решается актуальная задача повышения эффективности генетических алгоритмов с жадной агломеративной процедурой кроссинговера, имеющая важное значение для разработки алгоритмов и систем автоматической группировки (кластеризации) объектов и данных.

Структура работы

Текст диссертации представлен на 201 странице, состоит из введения, трех глав и заключения, содержит 289 библиографических источников и два приложения.

Во введении автор обосновывает актуальность выбранной темы, показывает научную новизну и формулирует положения, выносимые на защиту. Приводятся аргументы, подтверждающие практическую значимость работы.

В первой главе приводится обзор современных постановок задач автоматической группировки и методов их решения. Иллюстрируется связь с задачами размещения в непрерывной постановке.

Во второй главе представлены новые алгоритмы для решения серии задач автоматической группировки. Приведены результаты численных экспериментов для новых алгоритмов, показана их эффективность.

Третья глава посвящена расширению круга метрик и мер расстояния, которые могут быть задействованы в моделях и алгоритмах автоматической группировки. Показано, что метаэвристические подходы могут успешно применяться для решения задач с ограничениями на максимальные и минимальные расстояния.

В заключении диссертации подведены итоги, сформулированы основные результаты и сделаны выводы по работе.

Степень обоснованности научных положений, выводов и рекомендаций

Автор корректно использует известные научные методы обоснования полученных результатов и выводов. Разработанные алгоритмы внедрены в опытную эксплуатацию на промышленном предприятии и позволили повысить эффективность решения задач автоматической классификации электрорадиоизделий по производственным партиям.

Оценка новизны и достоверности результатов

– Разработан генетический алгоритм с вещественным алфавитом для одновременного решения серий непрерывных и дискретных задач автоматической группировки объектов с различными мерами расстояний.

– Разработан генетический алгоритм с вещественным алфавитом для одновременного решения серий непрерывных задач автоматической группировки объектов на основе моделей разделения смеси распределений.

– Построен алгоритм решения задачи Вебера с допустимыми зонами, ограниченными окружностями.

Новизна и достоверность подтверждаются теоретическими и экспериментальными данными, опубликованными в 17 работах, из которых 7 работ опубликованы в журналах, входящих в Перечень российских рецензируемых научных журналов, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученых степеней доктора и кандидата наук.

Результаты диссертационной работы неоднократно обсуждались на конференциях и семинарах.

Замечания по диссертационной работе

1) В разделе 1.4. (стр. 19–26) приводится обзор работ, посвященных методам решения непрерывной задачи размещения. Далее эти методы совершенствуются и применяются к задачам большой размерности. Но при большой размерности решение дискретной задачи будет хорошим приближенным для непрерывной задачи. Непонятно, почему игнорируются известные методы для дискретной задачи, например метод Лагранжевых релаксаций (см. работы: *Avella P., Boccia M., Salerno*

S., Vasilyev I. *An aggregation heuristic for large scale p-median problem* // *Comput. Oper. Res.* 2012. Vol. 39, no. 7. P. 1625–1632; Avella P., Sassano A., Vasil'ev I. *A heuristic for large-scale p-median instances* // *Electron. Notes in Discrete Math.* 2003. Vol. 13, P. 14–17.). Сравнение с таким подходом только украсило бы работу.

2) В разделе 2.1 «Генетический алгоритм решения задачи автоматической группировки с динамической популяцией» приводятся схемы генетических алгоритмов без мутаций. Непонятно, почему такой важный оператор, гарантирующий в асимптотике получение точного решения, не используется ни в одной из версий генетического алгоритма.

3) Стр.58. При описании результатов вычислительных экспериментов не описаны временные характеристики работы алгоритмов: из таблиц не ясно, сопоставимо ли между собой время счета для известных и новых алгоритмов. При этом в тексте необходимые разъяснения даны, но такой способ представления результатов затрудняет чтение. Отсутствуют графики сходимости алгоритмов.

4) На стр. 126–128 приводятся результаты расчетов для алгоритма КО-АПК и его сравнение с другими алгоритмами. Утверждается, что средние значения целевой функции и стандартных отклонений дают возможность судить о надежности предлагаемого подхода. На самом деле, оптимум в этих тестовых примерах неизвестен, и судить можно только о превосходстве нового подхода над другими. Насколько сильно ошибаются все эти алгоритмы, нам неизвестно. То же самое замечание касается утверждения о высокой точности получаемых результатов на стр. 101.

5) В разделе 2.5. «Стабильность получаемых решений» (стр. 80–84) обсуждается вопрос о стабильности и воспроизводимости получаемых решений при многократных испытаниях алгоритмов кластеризации. Но задачи кластеризации могут иметь много (экспоненциально много) оптимальных решений. Даже точные алгоритмы на таких тестовых примерах могут не давать один и тот же результат. Вероятностные алгоритмы тем более не будут отличаться стабильностью. Результаты расчетов на стр. 80–83 свидетельствуют скорее о том, что тестовые примеры были простыми и не содержали «плато» с одинаковыми значениями целевой функции. Интересно было бы проверить эффективность алгоритмов на действительно трудных тестовых примерах, в частности, на примерах из работы Kochetov Y., Ivanenko D. *Computationally difficult instances for the uncapacitated facility location problem* // *Metaheuristics: progress as real problem solvers* / Ed. by T. Ibaraki, K. Nonobe, M. Yagiura. New York: Springer, 2005. P. 351–367.

Указанные недостатки не влияют на высокую оценку выполненных исследований.

Заключение по диссертационной работе в целом

Результаты диссертации являются новыми и актуальными, имеют большое прикладное значение. Их достоверность обусловлена использованием современных научных методов, корректностью математических выкладок, подтверждается вычислительными экспериментами и не вызывает сомнений. Результаты диссертационной работы опубликованы в 17 печатных работах, в том числе в рецензируемых журналах из списка ВАК (7 статей), докладывались на различных российских и международных научных конференциях и семинарах. Диссертация написана ясным математическим языком, сопровождается иллюстративным материалом. Имеются 3 свидетельства о государственной регистрации программ для ЭВМ, акт о внедрении на промышленном предприятии. Автореферат достаточно полно и правильно отражает содержание диссертации.

Диссертационная работа М.Н. Гудымы представляет собой завершенную научно-квалификационную работу, полностью соответствующую паспорту специальности 05.13.01 – Системный анализ, управление и обработка информации (космические и информационные технологии). Считаю, что диссертация соответствует всем требованиям ВАК РФ, предъявляемым к кандидатским диссертациям, а ее автор М.Н. Гудыма заслуживает присуждения ученой степени кандидата технических наук по указанной специальности.

Официальный оппонент
доктор физико-математических наук,
профессор, главный научный сотрудник
лаборатории «Математические модели
принятия решений» ФГБУН Института
математики им. С.Л.Соболева СО РАН



Кочетов Юрий Андреевич

Адрес организации:
просп. Академика Коптюга, 4,
Новосибирск, 630090, Россия
E-mail: im@math.nsc.ru
Телефон: +7 (383) 329 75 83

