

Утверждаю
Проректор по науке и инновациям
Воронежского государственного
технического университета
д-р техн. наук, профессор

Прозоров Игорь Геннадьевич

« 2017 г.

ОТЗЫВ

ведущей организации на диссертационную работу

Гудымы Михаила Николаевича на тему «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 диссертации для науки и практики

Диссертация Гудымы М.Н. посвящена исследованию и разработке алгоритмов автоматической группировки (кластеризации) многомерных данных. Тема достаточно хорошо исследована, однако большинство алгоритмов, разработанных для решения подобных задач, требует априорного указания числа групп (кластеров). В работе Гудымы М.Н. задача автоматической группировки с неизвестным числом кластеров решается как серия задач (множество задач, различающихся только числом групп, на которые разбиваются объекты) в ходе одного запуска алгоритма. Результаты решения этой серии задач в дальнейшем могут быть проанализированы с применением любого критерия, приемлемого для определения числа групп в конкретной производственной задаче, такой как, например, задача разделения сборной партии электрорадиоизделий космического применения на однородные производственные партии. Внедрение разработанного алгоритма позволяет решать актуальную задачу повышения качества и эффективности разрушающих отбраковочных испытаний электрорадиоизделий, используемых в космическом производстве, за счет эффективной обработки данных уже имеющихся данных неразрушающих тестовых испытаний.

Научная новизна

Результаты, полученные в диссертации, а именно – генетические алгоритмы для одновременного решения серии задач автоматической группировки объектов, генетический алгоритм для той же задачи на основе

разделения смеси распределений и новый алгоритм решения задачи Вебера с допустимыми зонами, ограниченными окружностями, являются новыми и получены автором лично.

Генетические алгоритмы метода жадных эвристик для систем автоматической группировки объектов являются компромиссными по времени счета алгоритмами, которые позволяют получать за результаты, которые трудно улучшить другими методами без значительного увеличения времени счета. При этом время счета для решения этими алгоритмами задач с тысячами векторов данных размерностью в десятки-сотни измерений должно быть значительным, измеряемым минутами и десятками минут при работе на современной компьютерной технике. Учитывая, что фактическое количество групп (кластеров) в большинстве таких задач заранее неизвестно, возможны два подхода к решению таких задач:

А) последовательный многократный запуск алгоритма для решения задачи с числом кластеров от 2 до некоторого заранее установленного максимального числа кластеров,

Б) применение агломеративно-диссоциативных алгоритмов наподобие X-means, позволяющих одновременно с решением задачи кластеризации решать также задачу числа кластеров в наборе данных.

Недостаток второго подхода заключается в том, что такие методы отличаются более низкой точностью по сравнению с алгоритмами метода жадных эвристик, а также в том, что для определения числа кластеров используется единственный заранее установленный достаточно простой критерий - как правило, байесов информационный критерий или критерий Акаике. Для таких задач, как разделение сборной партии электрорадиоизделий на однородные производственные партии, данные критерии неадекватны. Оценка решения применяемым на практике критерием силуэта - сама по себе вычислительно сложная задача, и ее включение в итеративный процесс оценки решений вряд ли возможно. Недостаток первого подхода очевиден - низкая производительность вычислений.

Гудыма М.Н. предлагает третий подход, в котором задействована важная особенность агломеративных эвристических процедур: они последовательно уменьшают число кластеров в решении. Если фиксировать эти промежуточные результаты, можно получить решение сразу серии задач, отличающихся только числом кластеров. Гудыма М.Н. предлагает в одной популяции генетического алгоритма с вещественным алфавитом оперировать решениями разных задач, отличающихся только числом кластеров. Оператор скрещивания генетического алгоритма с жадной эвристикой позволяет скрещивать такие решения с различным числом кластеров. При этом повышается вариативность популяции, за счет чего повышается и точность получаемых решений, выраженная значением целевой функции.

Кроме того, в работе Гудымы М.Н. предлагается вспомогательный алгоритм для решения задачи Вебера с особой мерой расстояния, ограниченной снизу. Данный алгоритм используется при решении задач автоматической группировки, основанных на p -медианной модели с данной мерой расстояния.

Предложен новый алгоритм пчелиной колонии, позволяющий решать данные задачи с большей точностью. Важно, что для ряда задач доказана статистическая значимость превосходства новых алгоритмов над известными по результатам серии экспериментов.

Автором использован инструментарий ряда технических научных дисциплин, что позволило получить совокупность новых результатов, совместное применение которых позволяет усовершенствовать решение поставленной практической задачи о выделении однородных партий электрорадиоизделий по результатам тестовых испытаний. Это, в свою очередь, позволяет снизить процент ошибок при классификации электрорадиоизделий, и поднять качество продукции специализированных тестовых центров.

Практическое значение и рекомендации по использованию

Подход к решению задач автоматической группировки генетическим алгоритмом с неоднородной популяцией применен Гудымой М.Н. к решению классических задач автоматической группировки, основанных на моделях к-средних и к-медиан, а также к решению задач автоматической группировки на основе разделения смеси распределений. И в том, и в другом случае новые генетические алгоритмы для одновременного решения серии задач, разработанные с применением этого подхода, показали свою конкурентоспособность по сравнению предложенными ранее алгоритмами, что продемонстрировано на достаточном количестве практических примеров. Получаемый результат во многих случаях лучше, чем результат алгоритмов решения единственной задачи известными алгоритмами.

Таким образом, практическая значимость работы Гудымы М.Н. заключается в том, что в ней предложен эффективный подход к увеличению производительности генетических алгоритмов решения задач автоматической группировки за счет использования в генетическом алгоритме неоднородной популяции, содержащей в своем составе решения задач автоматической группировки, различающихся количеством кластеров.

Практические результаты диссертационной работы рекомендуются к применению на производствах, осуществляющих контроль однородности партий промышленной продукции либо сортировку продукции по однородным партиям на этапе контроля, а также при разработке других систем автоматической группировки объектов, к которым предъявляются повышенные требования точности и стабильности получаемых результатов.

Замечания по диссертационной работе

1. Не приведено теоретическое или иное обоснование выбора вещественного алфавита для генетического алгоритма. При этом известно, что для большинства задач генетические алгоритмы с вещественным кодированием решений, несмотря на свою сложность и громоздкость, демонстрируют низкую эффективность.

2. Формулировки поставленных оптимизационных задач отличаются от распространенных в литературе, что затрудняет восприятие текста. Скажем, целевая функция задачи k-средних обычно записывается как:

$$D = \sum_{i=1}^k \sum_{X_j \in S_i} \|X_j - c_i\|^2,$$

где k – число кластеров, S_i – полученные кластеры, c_i – центры кластеров. Задача формулируется как минимизация этой суммы квадратов расстояний. В тексте приведена более громоздкая запись (стр.21)

То же касается и p-медианной задачи. Строго говоря, эквивалентность приведенных формулировок не очевидна.

3. Результаты работы рандомизированных алгоритмов зависят от времени счета. Эту зависимость следовало в работе показать, например, графиками сходимости новых и известных алгоритмов. В противном случае, следовало бы обосновать выбор времени счета, которое приведено в результатах вычислительных экспериментов.

4. В обзоре литературы рассуждения об уровне пригодности математических алгоритмов для решения прикладных задач выглядят некорректно. Не следует смешивать понятия, оценивать качество решения такого алгоритма можно лишь применительно к строго сформулированной математической проблеме, но никак не к прикладной задаче. Соответственно, не вполне корректны сравнения этих алгоритмов с эвристическими.

5. Утверждение о разработке универсального генетического алгоритма (их уже разработано несколько сот!) является спорным. Известно, что универсального алгоритма, который бы с высокой эффективностью решал задачи различных классов не существует: повышение эффективности решения задачи одного класса является платой за понижение эффективности для другого класса (по free lunch [Wolpert D. (1995)], [Батищев Д.И. (2011)]).

6. Замечания по оформлению:

- завышенный объем диссертации (201 стр.);
- список цитируемой литературы (289 наименований) следовало разместить по алфавиту;
- вместо многочисленных таблиц (всего 49) следовало прибегнуть к более наглядной графической форме представления сравнительных результатов вычислительных экспериментов.

Заключение по диссертационной работе в целом

Приведенные замечания не снижают общей положительной оценки работы. Диссертация Гудымы М.Н. является научно-квалификационной работой, изложенные в которой положения о способах комбинации жадных агломеративных эвристических процедур, эволюционных алгоритмов, метаэвристических алгоритмов можно охарактеризовать как решение задачи повышения эффективности работы систем автоматической группировки объектов, что имеет существенное значение для науки и практики в области построения систем автоматической кластеризации. Работа выполнена соискателем самостоятельно и на высоком уровне.

Достоверность результатов диссертации подтверждается проведением достаточного количества вычислительных экспериментов при различных постановках задач как на классических наборах данных, так и на практических задачах разделения партий электрорадиоизделий.

Работа изложена на 201 странице, состоит из введения, трех глав, заключения, библиографического списка и приложений. Основные результаты диссертации опубликованы в 17 печатных работах соискателя, включая 2 работы в изданиях, индексируемых в Scopus, и 7 работ в рецензируемых журналах, рекомендованных ВАК РФ.

Автореферат соответствует основному содержанию диссертации.

Работа отвечает требованиям «Положения о порядке присуждения ученых степеней», предъявляемым к кандидатским диссертациям, а ее автор – Гудыма Михаил Николаевич – заслуживает присуждения ему ученой степени кандидата технических наук по специальности 05.13.01 – Системный анализ, управление и обработка информации (космические и информационные технологии).

Отзыв обсужден и одобрен на заседании кафедры автоматизированных и вычислительных систем ВГТУ (протокол № 4 от « 14 » ноября 2017 г.), присутствовали 16 человек.

Заведующий кафедрой
автоматизированных и вычислительных систем
Воронежского государственного
технического университета
Заслуженный деятель науки Российской Федерации,
доктор технических наук, профессор

 Подвальный Семен Леонидович

ФГБОУ ВО «Воронежский государственный технический университет»
394026, Воронежская область, г.Воронеж, Московский проспект, д.14.
Телефон +7(473)221-09-19, веб-сайт <http://www.vorstu.ru>,
адрес электронной почты: spodvalny2@mail.ru

*Подпись заведующего
кафедрой автоматизированных
и вычислительных систем
Подвального Семена Леонидовича*
*Заведующий кафедрой
автоматизированных
и вычислительных систем*
*Маслауров С.С.*