

Утверждаю

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

Дроздов Игорь Геннадьевич

« 2017 г.

ОТЗЫВ

ведущей организации на диссертационную работу
Сташкова Дмитрия Викторовича на тему «Системы автоматической
группировки объектов на основе разделения смеси распределений»,
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

В работе Сташкова Д.В. решается актуальная задача создания относительно быстрых и достаточно точных алгоритмов нечеткой кластеризации, основанных на разделении смеси вероятностных распределений. Дефицит таких алгоритмов обусловлен тем, что алгоритмы, предназначенные для быстрого решения задач кластеризации, обычно основываются на более простых моделях k-средних или p-медиан, а алгоритмы, основанные на модели разделения смеси распределений, как правило, применяются к задачам с относительно небольшим объемом входных данных в пространстве небольшой размерности. С повышением размерности пространства данных усугубляется проблема неустойчивости результата к начальному решению алгоритма. Сама по себе модель разделения смеси распределений более удобна по сравнению с моделями k-средних. Она позволяет получить вероятностную оценку принадлежности объектов в той или иной группе (кластеру), позволяет получать кластеры различной формы в соответствии с видом разделяемых распределений. Работа Сташкова Д.В.

посвящена решению проблемы неустойчивости результата алгоритмов разделения смеси вероятностных распределений в условиях достаточно большого объема многомерных входных данных и ограниченного времени счета.

Основные научные результаты

Результаты, полученные в диссертации, а именно – семейство алгоритмов поиска с чередующимися окрестностями для задач разделения смеси распределений, семейство эволюционных алгоритмов для этих же задач и новая модель сборной партии электрорадиоизделий, являются новыми и получены автором лично.

В работе Сташкова Д.В. метод жадных эвристик для систем автоматической группировки объектов распространен на решение новых задач автоматической группировки, а именно - задачи нечеткой кластеризации на основе разделения смеси вероятностных распределений. При этом сам метод жадных эвристик дополняется новым подходом: для организации глобального поиска, кроме применявшихся ранее генетического алгоритма с вещественным алфавитом, мултистарта и метода изменяющихся вероятностей, в работе применяется оригинальный алгоритм поиска с чередующимися окрестностями. Изменяются и сами жадные агрегативные эвристические процедуры: процедуры, предусмотренные для задач четкой кластеризации методом жадных эвристик, оказались неудачными и неэффективными при решении задач разделения смеси распределений. Сташков Д.В. предлагает в своей работе новые, более эффективные процедуры.

Наиболее значимым из полученных результатов представляется семейство из трех алгоритмов поиска с чередующимися окрестностями для задач разделения смеси распределений, в которых окрестности известного образуются путем применения оригинальных агрегативных эвристических процедур, задействующих второе известное решение в качестве параметра окрестности. Два из трех алгоритмов этого семейства демонстрируют высокую сравнительную эффективность для большинства задач, за исключением задач с малым объемом входных данных, для которых применение известных алгоритмов является вполне достаточным.

Эти алгоритмы вписаны соискателем в общую схему метода жадных эвристик, что позволяет применять для других задач новые принципы организации локального поиска на основе идей поиска с чередующимися окрестностями.

Предложенные в работе алгоритмы с чередующимися окрестностями (VNS) обладают некоторыми преимуществами перед генетическими алгоритмами с жадной агломеративной эвристикой: они, очевидно, требуют меньшего объема памяти для своей работы, оперируя всего двумя известными решениями задачи, последовательно улучшая одно из них. Кроме того, как показали приведенные в работе Сташкова Д.В. вычислительные эксперименты, такие алгоритмы требуют меньшего времени счета для достижения приемлемых результатов решения задач, которые было бы трудно улучшить другими известными методами без увеличения времени счета. При этом генетические алгоритмы при увеличении времени счета все же позволяют получать результаты с несколько лучшим значением функции правдоподобия, что оправдывает их применение в системах автоматической группировки.

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

Генетические алгоритмы и алгоритмы поиска с чередующимися окрестностями, предложенные в работе Сташкова Д.В., позволяют улучшить результат решения практических задач разделения смеси вероятностных распределений и получать решения с лучшим значением целевой функции правдоподобия по сравнению с известными алгоритмами решения этих задач. Это свойство новых алгоритмов может быть использовано при построении систем автоматической группировки объектов на основе разделения смеси вероятностных распределений.

Кроме того, в третьей главе работы Сташкова Д.В. предлагается новая модель для решения практической задачи разделения сборной партии электрорадиоизделий на однородные производственные партии. В новой модели параметры электрорадиоизделий каждой из однородных производственных партий рассматривается как многомерная случайная величина, распределение которой близко к многомерному нормальному

распределению. Параметры, распределение которых отличается от нормального, и это отличие подтверждается статистическим критерием эксцесса, в модель не включаются. В работе на примерах реальных сборных партий электрорадиоизделий космического применения показано, что новая модель дает меньшее количество ошибок при разбиении сборной партии на однородные по сравнению с применявшимися ранее моделями. Данная модель внедрена в опытную эксплуатацию на ОАО "ИТЦ НПО-ПМ" при формировании специальных партий электрорадиоизделий космического применения.

Практические результаты диссертационной работы рекомендуются к применению на производствах, осуществляющих контроль однородности продукции либо сортировку продукции по однородным партиям на этапе контроля.

Замечания по диссертационной работе

1. Автор отмечает, что SEM-алгоритм обладает свойством большей устойчивости получаемого результата по сравнению с ЕМ-алгоритмом. В то же время, новые алгоритмы построены на основе ЕМ-алгоритма. Интересными представляются результаты включения SEM-алгоритма в аналогичные схемы.
2. Не понятно, что автор подразумевает под «ограниченным временем, приемлемым для интерактивной работы алгоритма». Если под интерактивностью понимать способность алгоритма обрабатывать данные по мере их поступления, то новые алгоритмы как раз этим свойством не обладают (!?).
3. Автор использует генетический алгоритм с вещественным кодированием генотипа. Для большинства оптимизационных задач такой способ представления хромосом показывает низкую сравнительную эффективность. Автор никак не поясняет, благодаря каким особенностям решаемых задач при таком способе кодирования достигается высокая сравнительная эффективность алгоритмов.
4. Не ясен принцип выбора тестовых наборов данных, для которых приведены результаты вычислительных экспериментов. Почему не выбраны особо тестовые наборы, соответствующие наиболее трудным задачам, например, с ограниченными распределениями, что характерно для большинства реальных задач.

5. В заключении на стр.128 автором делается вывод о том, что предложенные алгоритмы способны успешно решать широкий круг задач с повышенной точностью. Работа не содержит строгого математического доказательства данного утверждения. Статистическое доказательство преимущества новых алгоритмов приводится, но для определенных случаев (определенных типов и объемов данных, размерности данных).
6. Не ясно, почему не использован аппарат нелинейных дискриминантных функций (НЛДФ), особенно для случая многомерного нормального распределения (см. Айвазян С.А. (1974).

Заключение по диссертационной работе в целом

Приведенные замечания не снижают общей положительной оценки работы. Диссертация Сташкова Д.В. является научно-квалификационной работой, изложенные в которой положения о способах комбинации жадных агломеративных эвристических процедур, эволюционных алгоритмов, алгоритмов поиска с чередующимися окрестностями и известных алгоритмов разделения смеси распределений можно охарактеризовать как решение задачи повышения эффективности систем автоматической группировки объектов на основе модели разделения смеси распределений, что имеет существенное значение для науки и практики в области построения систем автоматической кластеризации.

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 172 страницах, состоит из введения, трех глав, заключения, библиографического списка и приложений. Изложена ясным стилем, иллюстрации хорошо оформлены. Основные результаты диссертации опубликованы в 14 печатных работах соискателя, включая работу в зарубежном рецензируемом издании, индексируемом в Scopus, и 5 работ в рецензируемых журналах, рекомендованных ВАК РФ.

Автореферат соответствует основному содержанию диссертации. Оформление автореферата и диссертации соответствует требованиям ВАК РФ.

Работа отвечает требованиям п.9«Положения о порядке присуждения ученых степеней» постановления Правительства Российской Федерации от

24.09.2013 №842 и от 21.04.2016 №335, предъявляемым к кандидатским диссертациям, а ее автор – Сташков Дмитрий Викторович – 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 rector@vorstu.ru

Подпись зав. кафедрой АВС
Подвального Семена Леонидовича
Зав
Дека
*(Подпись)*