

ОТЗЫВ

официального оппонента

Еремеева Антона Валентиновича

на диссертационную работу Шкабериной Гузели Шарипжановны
«Модели и алгоритмы автоматической классификации продукции»,
представленной на соискание ученой степени кандидата технических наук по
специальности 05.13.01 – Системный анализ, управление и обработка
информации (космические и информационные технологии)

Актуальность темы диссертационной работы

Предметом исследования диссертационной работы Шкабериной Гузели Шарипжановны являются модели и алгоритмы автоматической группировки продукции на основе модели k -средних с расстоянием Махаланобиса. Указанные модели и алгоритмы позволяют повысить точность автоматической классификации промышленной продукции и получать более стабильные значения целевой функции по сравнению с известными ранее алгоритмами.

Проведенный соискателем анализ известных моделей и алгоритмов решения задач автоматической группировки объектов с повышенными требованиями к точности разделения, показал, что существующие решения не позволяют одновременно получать высокую точность разделения объектов, получать стабильные результаты при многократных запусках алгоритмов, а также обеспечивать высокую скорость работы алгоритмов. Это дает основание утверждать, что задачи, которым посвящена работа Шкабериной Г.Ш., являются актуальными.

Структура диссертационной работы

Диссертация состоит из введения, четырех глав, заключения и приложений. Объем работы составляет 222 листа машинописного текста. Список литературы содержит 312 наименований.

Во введении автор обосновывает актуальность темы исследования, описывает степень разработанности темы, указывает цель и поставленные задачи, формулирует новые научные результаты и положения, выносимые на защиту.

Первая глава работы Шкабериной Г.Ш. посвящена анализу текущего состояния проблем, связанных с автоматической группировкой объектов, и обзору методов их решения. Проанализированы и представлены существующие проблемы в области автоматической группировки объектов при обязательных требованиях к высокой точности и стабильности получаемого результата при многократных запусках алгоритма.

Вторая глава посвящена проблеме понижения размерности данных с применением методов факторного анализа, а также разработке оптимизационной модели автоматической группировки объектов, основанной на модели k -средних, позволяющей повысить качество автоматической группировки (по индексу Рэнда). По результатам вычислительных экспериментов показано, методы факторного анализа не

позволяют существенно сократить размерность пространства без потери точности. Тем не менее, в некоторых случаях точность разбиения на однородные партии может быть существенно повышена. Принимая во внимание более высокое среднее значение индекса Рэнда, можно сделать вывод, что предложенная модель и алгоритм оптимизации, основанные на модели k-средних с расстоянием Махаланобиса и усредненной оценкой ковариационной матрицы, имеют преимущество перед традиционно используемыми моделями с евклидовыми или манхэттенскими расстояниями при кластеризации промышленной продукции на однородные производственные партии.

В третьей главе предложен генетический алгоритм для задачи k-средних с кодированием хромосомы действительными числами, где одна и та же процедура используется как в качестве оператора скрещивания, так и в качестве оператора мутации. Эксперименты показывают, что генетические алгоритмы с жадным агломеративным оператором скрещивания, построенные в соответствии с этой идеей, превосходят генетические алгоритмы без какой-либо процедуры мутации и алгоритмы с равномерной случайной мутацией по полученному значению целевой функции. При параллельной реализации (с использованием архитектуры CUDA) вычислительные эксперименты показывают, что способы поддержания популяционного разнообразия, такие как генетический оператор мутации, улучшают характеристики генетических алгоритмов с жадным эвристическим оператором скрещивания для крупномасштабной задачи k-средних. Более того, лучшие результаты могут показать алгоритмы с оператором мутации на основе жадного эвристического оператора скрещивания со случайно сгенерированной хромосомой (новая жадная эвристическая мутация).

Четвертая глава посвящена разработке алгоритма обучения двухслойной сигмоидальной нейронной сети с регуляризацией, а также новому подходу к нахождению начального приближения для искусственной нейронной сети и сохранению его положительных свойств при обучении с избыточностью описания в виде чрезмерно большого числа нейронов и одновременной недостаточной аппроксимацией в определенных частях области данных.

В заключении сформулированы основные выводы и результаты работы, показано, что цель диссертационной работы достигнута путем решения поставленных задач.

В приложении представлены дополнительные данные о результатах вычислительных экспериментов, акт об использовании результатов исследования и свидетельство о государственной регистрации программы для ЭВМ.

Степень достоверности и обоснованности научных положений, выводов и рекомендаций

В диссертационной работе Шкабериной Гузели Шарипжановны подробно проанализированы отечественная и зарубежная литература по

известным методам решения задач автоматической классификации объектов, эволюционным алгоритмам и жадным эвристическим процедурам, а также мерам сходства и критериям оценивания классификации объектов. Автор работы достаточно корректно использует известные научные методы обоснования полученных результатов, выводов и рекомендаций.

Достоверность результатов подтверждается применением современных методов исследования, которые были использованы в большом наборе экспериментов. Основные положения и результаты диссертационной работы апробированы на конференциях и опубликованы в 16 работах (из них 4 работы в ведущих рецензируемых журналах, рекомендуемых действующим перечнем ВАК, 11 – в международных изданиях, индексируемых в системах цитирования Web of Science и Scopus). Имеется свидетельство о государственной регистрации программы для ЭВМ. Предложенные автором модель и алгоритмы используются в деятельности АО «Испытательный технический центр – НПО ПМ».

Оценка новизны и достоверности

В качестве новых научных результатов автором работы предложены:

- 1) модель для решения задач автоматической группировки промышленной продукции на основе модели k -средних с расстоянием Махаланобиса с применением метода главных компонент. Применение новой модели позволяет повысить точность решения для задачи выделения однородных производственных партий изделий по данным тестовых испытаний;
- 2) алгоритм автоматической группировки объектов, основанный на оптимизационной модели k -средних с мерой расстояния Махаланобиса и средневзвешенной ковариационной матрицей, рассчитанной по обучающей выборке. Алгоритм позволяет снизить долю ошибок при выявлении однородных производственных партий продукции по результатам тестовых испытаний;
- 3) генетический алгоритм для задачи k -средних с применением единой жадной агломеративной эвристической процедуры в качестве оператора скрещивания и оператора мутации. Применение данного алгоритма позволяет статистически значимо повысить точность результата (улучшить достигаемое значение целевой функции в рамках выбранной математической модели решения задачи автоматической группировки), а также его стабильность, за фиксированное время, по сравнению с известными алгоритмами автоматической группировки;
- 4) алгоритм обучения двухслойной сигмоидальной искусственной нейронной сети с регуляризацией, демонстрирующий более высокую точность классификации промышленной продукции по данным тестовых испытаний в сравнении с методами обучения таких нейронных сетей при известных методах регуляризации.

Шкабериной Г.Ш. проведен значительный объем научной работы. Результаты получены лично автором и полностью опубликованы в

рецензируемых научных изданиях, прошли апробацию на конференциях различного уровня и внедрены в производственный процесс.

Научная значимость диссертационных исследований состоит в том, что предложенный комплекс алгоритмов и моделей дополняет список эффективных методов решения задач автоматической группировки объектов. Принцип использования единой процедуры в качестве оператора скрещивания и мутации создает основу для синтеза новых эвристических алгоритмов для более широкого круга NP-трудных задач.

Содержание диссертационной работы и полученные результаты соответствуют п. 4 паспорта специальности 05.13.01 – Системный анализ, управление и обработка информации.

Выявленные недостатки и замечания

1. В конце главы 3 производится проверка значимости по критериям Манна-Уитни и Стьюдента. При этом упоминается уровень значимости 0.99. Думаю, что в действительности имеется в виду уровень значимости 0.01. Поскольку в работе нормальность распределений не проверена статистически и не оправдана другими соображениями, то критерий Стьюдента нет смысла упоминать. Достаточно было бы привести только результаты критерия Манна-Уитни.

2. Перекрестная процедура в алгоритме 3.2 описана недостаточно ясно.

3. Поскольку в методе k-средних не определена окрестность, его не следует относить к алгоритмам локального поиска.

4. В третьей главе под точностью алгоритма понимается достигнутая величина целевой функции. С практической точки зрения, интересно было бы сравнить реальное разделение, если таковое известно, с полученным результатом работы алгоритма.

5. В третьей главе для наглядности можно было представить графики сходимости целевой функции для новых и известных алгоритмов, и таким образом продемонстрировать преимущество новых алгоритмов.

Заключение о соответствии диссертационной работы требованиям и критериям, установленным Положением о порядке присуждения ученых степеней

Работа базируется на достаточном объеме исходных данных, примеров и расчетов. По каждой главе и работе в целом сделаны четкие выводы. Выше указанные замечания не снижают положительной оценки работы. Диссертационная работа «Модели и алгоритмы автоматической классификации продукции» является завершенным научно-исследовательским трудом, выполненным Шкабериной Г.Ш. самостоятельно и на высоком научном уровне.

Автореферат достаточно полно и правильно отражает основное содержание диссертации. Оформление автореферата и диссертации соответствует требованиям ВАК РФ.

Представленная диссертационная работа отвечает требованиям п. 9 Положения о присуждении ученых степеней, постановление Правительства РФ от 24 сентября 2013, № 842 «О порядке присуждения ученых степеней», предъявляемым к диссертациям на соискание ученой степени кандидата наук, а ее автор Шкаберина Гузель Шарипжановна заслуживает присуждения ученой степени кандидата технических наук по специальности 05.13.01 – Системный анализ, управление и обработка информации (космические и информационные технологии).

Официальный оппонент,
зам. директора по научной работе –
директор ОФ ИМ СО РАН, главный научный сотрудник,
зав. лабораторией дискретной оптимизации
Омского филиала Федерального государственного
бюджетного учреждения науки Института математики
им. С.Л. Соболева СО РАН,
доктор физ.-мат. наук, доцент

Антон Валентинович Еремеев

Адрес организации:
644043 г. Омск, ул. Певцова 13
e-mail: eremeev@ofim.oscsbras.ru

Подпись
Еремеева Антона Валентиновича
Заверяю
Ученый секретарь

В.А. Планкова

