

ОТЗЫВ

официального оппонента Истомина Андрея Леонидовича на диссертацию Голованова Сергея Михайловича «Алгоритмы автоматической группировки электронных компонентов с учетом заданной эффективности разделения на группы», представленную на соискание ученой степени кандидата технических наук по специальности 2.3.1 - Системный анализ, управление и обработка информации, статистика

Актуальность темы работы

Работа Голованова Сергея Михайловича посвящена развитию специализированных алгоритмов кластерного анализа для обработки данных тестовых испытаний электронной компонентной базы (ЭКБ) космического применения. Основная идея работы базируется на том, что электронные компоненты, принадлежащие разным производственным партиям, в пространстве признаков образуют обособленные группы. Разделение элементов на производственные партии требуется для проведения выборочных разрушающих испытаний ЭКБ и выявления потенциально ненадежных элементов – элементов, обладающих скрытыми дефектами, которые могут привести к отказу при длительной эксплуатации. И если выборочный разрушающий контроль элементов испытываемой партии выявил дефекты, результаты тестирования распространяются на всю партию ЭКБ. Поэтому задача точного разделения ЭКБ на производственные партии и определения потенциально-ненадежных ЭКБ по данным неразрушающего контроля, является несомненно актуальной.

Для классификации объектов и разделения их на группы разработано множество методов и алгоритмов. Для решения поставленной цели в диссертационном исследовании автором выбран кластерный анализ. Кластерный анализ уже много лет входит в число востребованных исследований. Однако, универсальных алгоритмов кластерного анализа, обеспечивающих высокую эффективность разделения объектов при незначительных различиях в данных практически нет. Известные алгоритмы группировки электронных компонентов не обеспечивают высокую точность разделения, неэффективны при больших объемах данных. В работе разработаны алгоритмы машинного обучения с частичным привлечением учителя, осуществляющие поиск условно оптимального варианта кластеризации с ограничениями и обеспечивающие разделение объектов с заданной эффективностью разделения на группы. Полученные автором результаты имеют существенное значение для решения актуальной задачи повышения эффективности классификации и автоматической группировки объектов.

Структура диссертации

Диссертация Голованова С.М. с приложениями представлена на 195 страницах, основной текст состоит из введения, четырех разделов и заключения.

Во введении обосновывается актуальность диссертационного исследования, формулируется цель и основные задачи, характеризуется степень новизны полученных результатов и их апробация.

В разделе 1 дан обзор методов кластерного анализа, описаны сферы их применения, приведен актуальный практический пример – обработка результатов тестовых испытаний электронной компонентной базы космического применения. Даны основные определения, используемые в диссертации. Проведен обзор критериев эффективности, применяемых для оценки качества результата решения задачи кластерного анализа.

Раздел 2 посвящен разработке метода автоматической группировки множеств однотипных объектов с учетом заданной эффективности деления на группы, а также алгоритмов автоматической группировки электронных компонентов в соответствии с принадлежностью к производственным партиям. Введены характеристики, обеспечивающие заданную эффективность автоматической группировки множеств однотипных объектов (вектор показателей эффективности разделения на группы и область эффективного разделения множества на группы), разработаны алгоритмы кластерного анализа, выполняющие поиск варианта автоматической группировки, обеспечивающего заданную эффективность кластеризации. Предложен алгоритм определения границ области эффективного разделения на группы с использованием размеченных обучающих выборок. Предложен вариант вектора показателей эффективности разделения на группы и области эффективного деления для ЭКБ по результатам тестирования в испытательном техническом центре. Приведены примеры решения задач автоматической группировки элементов партий различных видов электронных компонентов.

В разделе 3 рассмотрены алгоритмы определения элементов-выбросов и приведены три алгоритма определения таких элементов, зарекомендовавшие себя при обработке результатов тестовых испытаний партий электронной компонентной базы. Для определения параметров алгоритмов, а также для определения наиболее эффективного алгоритма определения элементов-выбросов (ЭВ) предложен метод с использованием обучающих выборок – специально подобранного набора обучающих множеств, каждое из которых, представляющих из себя партию ЭКБ заданного типа.

Раздел 4 посвящен применению разработанных в диссертации алгоритмов в технологическом процессе дополнительных испытаний ЭКБ космического применения. Разработанные автором алгоритмы определения ЭВ применяются при проведении дополнительных испытаний ЭКБ космического применения в качестве дополнительных алгоритмов определения потенциально ненадежных элементов, что приводит к повышению эффективности выявления этих элементов. Приведенные примеры подтверждают работоспособность предложенных алгоритмов.

В заключении сформулированы основные выводы и результаты, показано, что решение поставленных задач привело к достижению цели диссертации.

Достоверность и обоснованность научных положений, выводов и рекомендаций

В диссертационной работе Голованова С.М. подробно проанализированы отечественная и зарубежная литература по методам и алгоритмам кластерного анализа, мерам сходства и критериям оценки результатов работы алгоритмов кластерного анализа. Автор корректно использует известные научные методы для обоснования полученных результатов, выводов и рекомендаций. Достоверность результатов подтверждается применением современных методов исследования, которые были использованы в большом наборе экспериментов. Предложенные автором алгоритмы внедрены в процесс дополнительных испытаний электронной компонентной базы космического применения в АО «Испытательный технический центр – НПО ПМ».

Оценка новизны результатов и выводов

Новыми научными результатами являются:

- метод поиска условно оптимального варианта автоматической группировки множеств однотипных объектов, обеспечивающий заданную эффективность разделения на группы;
- алгоритм автоматической группировки и разделения электронных компонентов на производственные партии;
- алгоритмы, выявляющие в составе партий ЭКБ потенциально-ненадежные элементы, являющихся носителями накопительного эффекта суммарных отклонений характеристик элементов от их средних по партии значений с применением размеченных обучающих выборок для определения параметров, обеспечивающих максимальную эффективность работы алгоритмов.

Основные положения и результаты работы прошли апробацию на международных научных конференциях и опубликованы в печати, в том числе в рецензируемых журналах.

По работе имеются следующие замечания и вопросы:

1. В литературном обзоре и в целом в работе под понятием «автоматическая группировка» автор подразумевает исключительно методы кластерного анализа. Но под автоматической группировкой понимается не только кластерный анализ, но и широкий набор подходов к классификации объектов, начиная со статистических методов (проверка статистических гипотез, многомерный статистический анализ, дискриминантный и факторный анализы), многочисленных задач классификации, сформулированных в виде задач математического программирования и исследования операций, и заканчивая, например, современными методами интеллектуального анализа данных и т.п. Поэтому верным было бы указать, что объектом исследования в работе является именно кластерный анализ, а предметом исследования – методы и алгоритмы кластерного анализа.

2. Автор изначально утверждает, что имеются системные различия между ЭКБ разных производственных партий, проявляющиеся в их характеристиках. Но с научной точки зрения, правильно было бы подтвердить эту гипотезу,

например, с помощью методов дисперсионного анализа, когда отличия зависят от качественных факторов (сырье, технология изготовления, оборудование и т.п.). Для этого все имеется: большой объем данных; данные имеют нормальный закон распределения; теоретически обоснованные, параметрические критерии. Тогда бы дальнейшее исследование было бы хорошо обоснованным.

3. Автор рассматривает несколько способов нормирования величин, в то же время остается неясным, почему проигнорирован наиболее известный в статистических исследованиях способ нормирования: $(x_j - \bar{X})/\sigma$, где $\bar{X}$ и σ – среднее значение и стандартное отклонение выборки из значений x_j соответственно, при котором все нормированные переменные полностью соизмеримы (средние значения равны нулю, а дисперсии и стандартные отклонения равны единице). Выбранные же автором способы нормирования приводят к существенным отличиям средних значений и дисперсий нормированных переменных, а значит и к несоизмеримому влиянию на полученное решение.

Кроме того, некорректным является термин «наиболее достоверный способ нормирования». Отличие решения при разных способах нормирования говорит не о достоверности или недостоверности нормирования, а о различии исходных данных при решении задачи автоматической группировки.

4. Не ясно, при каких условиях решается задача разделения элементов на партии (только с надежными элементами или с элементами-выбросами в составе партий) и как ведет себя алгоритм в этих случаях. Ведь одно дело – увидеть системные различия между однотипными элементами разных партий с малоотличающимися характеристиками (для этого требуется очень «тонкая» настройка алгоритма). Но тот же алгоритм, но с элементами-выбросами может разделить все множество объектов на надежные и ненадежные элементы (задача разделения на партии решена не будет).

5. Введены обозначения V_{qc} и S_{qd} , но не даны их описания, снова обозначение V_{qc} встречается в описании алгоритма разделения как вектор показателей эффективности разделения на группы, но ранее вектор показателей эффективности разделения на группы обозначен как V_{ed} .

6. Имеются стилистические ошибки, пропущенные слова, буквы и опечатки, например, вместо раздела 2 на стр. 17 автореферата указана глава 2, Alg_{EV}^2 вместо Alg_{EV2} и т.п.

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

Вышеуказанные замечания не снижают общей положительной оценки работы.

Диссертация Голованова С.М. является завершенным научно-исследовательским трудом на актуальную тему, выполненным самостоятельно и на высоком научном уровне. Полученные автором результаты достоверны, выводы и заключения являются обоснованными.

Представленная диссертационная работа отвечает требованиям п.9 «Положения о порядке присуждения ученых степеней» Постановления Правительства Российской Федерации от 24.09.2013 г. № 842, предъявляемым к кандидатским диссертациям, а ее автор Голованов Сергей Михайлович заслуживает присуждения ученой степени кандидата технических наук по специальности 2.3.1 - Системный анализ, управление и обработка информации, статистика.

Официальный оппонент,
профессор кафедры вычислительных
машин и комплексов, декан факультета
управления и бизнеса
ФГБОУ ВО «Ангарский государственный
технический университет»,
доктор технических наук, доцент

Подпись Истомина А.Л. удостоверяю

Проректор по научной работе
ФГБОУ ВО «Ангарский государственный
технический университет»,
доктор технических наук, доцент



А.В. Бальчугов

20 сентября 2023 г.

Адрес организации: 665830, г. Ангарск, ул. Чайковского, 60.
Тел. 8(3955) 67-89-15, postmast@angtu.ru