

О Т З Ы В

на диссертационную работу Ю.Н. Синявского «Непараметрические методы и программно-алгоритмический инструментарий для сегментации мультиспектральных спутниковых изображений», представленную на соискание ученой степени кандидата технических наук по специальности 05.13.18 — «Математическое моделирование, численные методы и комплексы программ».

Актуальность темы

Технологии дистанционного зондирования Земли являются незаменимым инструментом, позволяющим проводить как глобальный мониторинг нашей планеты, так и обследование труднодоступных и опасных регионов. Существующие аэрокосмические системы позволяют получать цифровые данные об интенсивности излучения, исходящего от земной поверхности и атмосферы, на различных длинах волн. В зависимости от числа измерительных каналов используемой аппаратуры, полученные данные представляются в форме мультиспектральных или гиперспектральных изображений. В последние годы значительные усилия были направлены на разработку и усовершенствование методов автоматизированной тематической обработки дистанционных данных. Одним из наиболее важных этапов тематической обработки является сегментация — разбиение изображения на непересекающиеся области, которые ставятся в соответствие различным классам природных и техногенных объектов, составляющих ландшафт выбранного региона. Полученные результаты используются в совокупности с объектами энергетики, промышленности, сельского хозяйства, населенными пунктами, дорожно-транспортной сетью для определения контуров лесных массивов, лугов, болот, водоемов и последующего восстановления количественных и качественных параметров.

Основной целью диссертационной работы Синявского Ю.Н. является разработка новых эффективных непараметрических алгоритмов кластерного анализа в применении к задаче сегментации спутниковых изображений и создание современной платформы для стандартизованного доступа к ним. В настоящее время для решения данной задачи при обработке аэрокосмической информации используется ряд стандартных методов, реализованных в виде программного обеспечения: универсальное коммерческое программное обеспечение, предназначенное для работы с широким спектром данных дистанционного зондирования Земли; специализированные коммерческие программы, ориентированные на работу с конкретным оборудованием; свободно распространяемые программы с открытым кодом; программы в рамках универсальных научных математических пакетов. В указанных программных продуктах реализованы лишь некоторые из стандартных подходов к решению задачи классификации без учителя в то время, как более эффективные подходы остаются невостребованными.

Необходимость сегментации аэрокосмических изображений на основе кластерного анализа возникает в условиях недостатка априорных сведений о классах объектов, представленных на рассматриваемой сцене. Как правило, речь идет о форме дискриминантных поверхностей, вероятностных характеристиках и количестве классов. В этих условиях применение стандартных параметрических методов становится либо затруднительным, либо вовсе невозможным. Применение

непараметрических вероятностно-статистических методов ограничивается проблемами, связанными с неустойчивостью соответствующих оценок и высокой вычислительной сложностью, в особенности при большой размерности признакового пространства. Существующие экспериментальные исследования показывают, что перспективы развития методов кластерного анализа многоспектральных спутниковых изображений лежат в создании скользящих локально-параметрических моделей, что подтверждает актуальность исследований, представленных в диссертационной работе Синявского Ю.Н.

Содержание работы

Диссертация состоит из введения, пяти глав, заключения, списка литературы, списка публикаций автора и приложений. Общий объем диссертации составляет 151 страницу, включая 49 рисунков, 8 таблиц и 3 приложения. Библиография содержит 154 наименования.

Введение содержит обоснование актуальности темы исследования, формулировку целей и задач диссертационной работы, положения, составляющие научную новизну, обоснование теоретической и практической ценности, описание личного вклада автора.

В первой главе представлена постановка задачи кластеризации данных и приведен подробный обзор существующих методов ее решения. На основе анализа рассмотренных методов, выделены наиболее эффективные с учетом особенностей задачи сегментации спутниковых изображений. Обоснован выбор направления исследований, представленных в диссертационной работе.

Во второй главе представлена формальная постановка задачи кластеризации в рамках вероятностно-статистического подхода и приведено полное описание нового непараметрического алгоритма кластеризации MeanSC. Представлен анализ основных стратегий выбора параметров алгоритма, представлены результаты расчетов, демонстрирующие эффективность предлагаемого подхода на модельных данных, и перечислены основные идеи, позволившие снизить вычислительные затраты при сохранении качества получаемого результата.

В третьей главе выполнена формальная постановка задачи формирования коллективного решения и представлен новый метод кластеризации EMeanSC, представляющий собой ансамблевую модификацию алгоритма MeanSC. На основе метода статистического моделирования демонстрируется, что использование предлагаемого в диссертации ансамблевого подхода позволяет значительно упростить процедуру поиска оптимальных значений параметров. Также представлены результаты сравнительного анализа с сеточным алгоритмом GCOD на основе реальных данных.

Четвертая глава посвящена сравнению эффективности предложенных автором алгоритмов со стандартными алгоритмами кластеризации, реализованных в программных продуктах ENVI, ELKI и Smile. Приведенные результаты численных экспериментов демонстрируют явное преимущество методов MeanSC и EMeanSC как по показателям точности, так и по времени вычислений.

В пятой главе представлена структура разработанного программного обеспечения, в частности пакет программ «Image Processing Toolkit», и сформулированы предложения по его возможному использованию при решении практических задач. Представлены механизмы внедрения предложенных методов и алгоритмов в геоинформационную систему GRASS GIS и в виде веб-сервисов.

Степень достоверности и обоснованности научных положений, выводов и рекомендаций

Достоверность результатов работы обусловлена корректной постановкой решаемой задачи и использованием известных методов теории вероятности и математической статистики, непротиворечивостью теоретических и практических результатов и успешным решением практических задач. Все представленные в работе положения и выводы обоснованы и аргументированы. Выполнена программная реализация и исследование предложенных алгоритмов, продемонстрирована их эффективность (в сравнении с известными алгоритмами кластеризации).

Все результаты, представленные в диссертационной работе, многократно обсуждались на конференциях разного уровня и опубликованы в рецензируемых научных изданиях (в том числе, из списка ВАК), имеется 5 свидетельств о государственной регистрации программ и акт об использовании результатов исследования в Институте почвоведения и агрохимии СО РАН.

Научная новизна и практическая значимость результатов

Автору данной диссертационной работы Синявскому Ю.Н. удалось разработать новые статистические методы кластерного анализа данных. Предложенные методы позволяют выделять кластеры сложной структуры в условиях малой априорной информации о распределении обрабатываемых данных и неизвестного числа кластеров. Высокая вычислительная эффективность алгоритмов позволяет существенно повысить скорость сегментации многоспектральных аэрокосмических изображений по сравнению с рядом стандартных программных продуктов. Представленные результаты расчетов сопровождаются анализом их достоверности.

Предложенные алгоритмы сформулированы четко и могут быть использованы другими исследователями. Практическая ценность работы не вызывает сомнений. Результаты могут быть использованы различными службами, деятельность которых требует проведения автоматизированной тематической обработки аэрокосмических изображений.

Реализация предлагаемых соискателем методов в виде программно-алгоритмического инструментария и публикация их в виде веб-сервисов обеспечивает возможности практического использования полученных результатов рядом государственных организаций, в том числе и для выполнения государственных программ, связанных с использованием аэрокосмической информации.

Замечания

По содержанию диссертации имеются следующие замечания.

1. На стр. 5 предложение «Эффективность такого рода алгоритмов подтверждена многочисленными экспериментальными исследованиями» следовало бы дополнить литературными ссылками.
2. На стр. 15, 1 абзац указано, что разрабатываемые алгоритмы кластеризации должны позволить пользователю обрабатывать спутниковые изображения в диалоговом режиме. Здесь важно было бы указать допустимые размеры изображения и число каналов.

3. По результатам статистического моделирования, представленным на рис. 4.6, методы DBSCAN и SLINK справляются с поиском кластеров в форме спирали, а на рис. 4.11 и 4.12 - нет. С чем это связано?
4. Данные таблицы 4.1 демонстрируют явные преимущества предлагаемых методов по сравнению со стандартными. Однако, сравнение показателей MeanSC и EMeanSC показывает, что ансамблевый метод EMeanSC не дает увеличения точности, однако проигрывает по скорости методу MeanSC. В каких случаях предпочтительно использовать EMeanSC?
5. В работе присутствует ряд опечаток и неточностей:
 - В предложении (стр. 12) «Задача кластеризации заключается в разбиении выборки на сравнительно небольшое, заранее известное или нет, число $M \geq 2$ групп объектов (кластеров) так, чтобы элементы одного кластера были как можно более схожи, а элементы из разных кластеров существенно различались по заданному критерию сходства/различия» следовало бы избегать таких фраз, как «сравнительно небольшое» и «заранее известное или нет».
 - Рис. 1.1 (стр. 16) содержит опечатку: вместо «Непараметрические (Parametric, Model-based)» следует писать «Параметрические (Parametric, Model-based)». Также в данной схеме или в комментариях к ней следовало бы отразить, к какой группе следует отнести графовые алгоритмы кластерного анализа.
 - В первой формуле подпункта 1.4.2. не расшифрованы параметры N и M . Интуитивно понятно, что это объем выборки и число кластеров, но это можно было указать явно, например «Для нахождения глобального экстремума необходимо рассмотреть $S(N, M)$ возможных разбиений исходной выборки объема N на M кластеров».
 - В подпункте 1.4.2 при описании метода k -средних не указывается, что речь идет о стандартном (или наивном) алгоритме k -средних, для которого ниже приведена оценка сложности. Аналогичный вопрос возникает при описании метода k -представителей.
 - Стр. 26, 3 абзац. В предложении «Для запуска DBSCAN необходимо задать два параметра, оптимальные значения которых определить достаточно сложно.» следовало бы четко указать, какие это параметры. Можно предположить, что речь идет о параметрах ε и MinPts, однако лучше было бы это указать явно и расшифровать аббревиатуру MinPts (minimum points).
 - На стр. 43 (1 абзац), фразу «следующим удовлетворяющая условиям сходимости» следует заменить на «удовлетворяющая следующим условиям», причем слово «сходимости» нужно либо опустить, либо расшифровать о какой именно сходимости идет речь. Фразу «взбирание по градиенту» следует заменить на «градиентный подъем».
 - В формуле 2.2 (стр. 43) R^k следует совместить со знаком интеграла.

Указанные замечания несколько не изменяют общую положительную оценку данной диссертационной работы. Замечания либо носят технический характер, либо относятся к некоторым частным аспектам решения поставленных задач. Часть из них может рассматриваться автором в качестве пожеланий для дальнейшей работы.

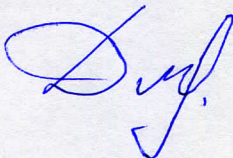
Заключение

Оценивая диссертацию Ю.Н. Синявского в целом, считаю, что она является законченной научно-исследовательской работой, имеющей важное научное и практическое значение в области автоматизации тематической обработки аэрокосмических изображений.

Результаты диссертации являются новыми и достоверными. Основные результаты опубликованы автором в научных журналах, доложены на всероссийских и международных конференциях. Автореферат правильно отражает основные положения диссертации.

Диссертационная работа соответствует требованиям, предъявляемым ВАК России к кандидатским диссертациям, а ее автор Синявский Юрий Николаевич заслуживает присуждения ему ученой степени кандидата технических наук по специальности 05.13.18 – математическое моделирование, численные методы и комплексы программ.

Официальный оппонент,
кандидат физико-математических наук,
старший научный сотрудник
ФГБУН Института вычислительной
математики им. Г.И. Марчука РАН



Дмитриев Егор Владимирович

“22” мая 2021 г.

Адрес места работы:

119333 Москва, ул. Губкина, дом 8,

<https://www.inm.ras.ru/>

Тел. 8 (495) 984-81-20, 8 (495) 989-80-24, факс: 8 (495) 989-80-23

E-mail: director@mail.inm.ras.ru

Подпись Е.В. Дмитриева заверяю
Ученый секретарь ИВМ РАН, в.н.с.,
доктор физико-математических наук



В.П. Шутяев

