

## ОТЗЫВ РЕЦЕНЗЕНТА

Баранова Дениса Евгеньевича

на магистерскую диссертацию Шевелева Александра Александровича

### «Применение методов топологического анализа данных при балансировке нагрузки в вычислительных сетях».

Рассматриваемая магистерская диссертация Шевелева А.А. посвящена исследованию применимости методов топологического анализа – *устойчивые гомологии* и *алгоритм Mapper*, для упрощения решения задач балансировки нагрузки на данных о задачах, решаемых вычислительными кластерами компании Google. И последующему сравнительному анализу качества определения типов задач с результатами, получаемыми при использовании более традиционных методов – *кластеризация к-средних* и *модель смеси гауссиан с использованием процесса Дирихле*, на идентичном наборе данных.

**Во введении** раскрыта проблематика непрерывно возрастающих вычислительных нагрузок, связанных с обработкой больших потоков данных. Так же представлено обоснование актуальности проводимого исследования.

**В первой главе** сделан довольно подробный обзор литературы в областях процессов балансировки нагрузки, топологических методов анализа данных и кратких обзор статей, описывающих методы, используемые для балансировки нагрузки. Здесь отражены передовые подходы к распределению вычислительных нагрузок и проблематика данной сферы. Рассмотрены типы задач, для решения которых необходимо использовать методы балансировки. Представлена классификация алгоритмов распределения нагрузки по различным критериям.

В данной главе особенно подчеркивается, что топологический анализ – это относительно новое направление в статистике и потенциальная сфера его использования до конца не определена, но задачи которые решаются с его помощью тесно связаны с задачами балансировки нагрузки — уменьшение размерности и выявление скрытых особенностей. Явно прослеживаются рассуждения автора относительно рассматриваемой информации и глубокая проработка используемого материала.

**Во второй главе** автор подробно описывает набор данных, который был использован для проведения исследований. Изложено грамотное обоснование необходимости предварительной обработки данных, а именно – нормализация и уменьшение количества учитываемых параметров; перед их непосредственным использованием для решения поставленных задач. Представлен статистически правильный подход, использованный для получения двух наборов репрезентативных выборок из генеральной совокупности данных, отличающихся размером выборок, общее количество которых равно 20.

Далее описываются особенности применения реализуемых алгоритмов в рамках рассматриваемой работы и реализаций используемых для анализа методов: *к-средних* и *модель смеси гауссиан с использованием процесса Дирихле* из пакетов библиотеки *sklearn* языка программирования *python*, *Mapper* из пакета *KeplerMapper* для *python* и *устойчивые гомологии* из пакета *TDA* языка *R*.

Одним из наиболее важных аспектов данной главы являются четко сформулированные критерии, по которым должно производиться сравнение результатов работы алгоритмов, основанные на особенностях поставленных задач.

**Третья глава** посвящена непосредственно практической реализации и анализу полученных результатов. Представлены данные попарной визуализации с целью выявления линейно зависимых характеристик. Описан финальный набор учитываемых параметров состоящий из 9 особенностей задач. Произведен подробный анализ и описание, с графическим

подкреплением, результатов работы алгоритмов. Для подбора параметров и верификации работы методов кластеризации *k-средних* и *модель смеси гауссиан с использованием процесса Дирихле* использован *метод силуэтов*. Для выявления устойчивых гомологий в наборе данных применяется *фильтр ВьеторисаРипса*, на основе которого был построен штрихкод для визуализации «времени жизни» гомологий.

С помощью алгоритма *Mapper* произведено выделение компонент сильной связанности путём проекции данных о задачах на двумерные плоскости, на основе пар учитываемых параметров. В данном случае наборы сходных по характеристикам объектов определялись на основе степени связанности между ними.

В целом магистерская диссертация Шевелева А.А. носит законченный характер и производит хорошее впечатление. Она содержит результаты обширной исследовательской и практической работ. Диссертация обладает актуальностью, научной новизной и прикладной составляющей. Материал изложен логично и связно. В работе присутствуют аналитические рассуждения автора и продуманность подходов к решению поставленной задачи.

В тоже время, в качестве **замечания** следует отметить:

1. В работе в явном виде не указаны временные затраты и ресурсоемкость работы реализованных методов, хотя данные замеры заявлены как критерий для сравнения эффективности работы алгоритмов.

Несмотря на выделенный недостаток, считаю что диссертационная работа Шевелева Александра Александровича «Применение методов топологического анализа данных при балансировке нагрузки в вычислительных сетях» заслуживает оценки «**отлично**», а автор присвоения требуемой квалификации.

Рецензент,  
ведущий разработчик,  
ООО «АйТи Либертас»



Баранов Д.Е.  
30.06.2016