

Отзыв официального оппонента

кандидата технических наук Игнатова Дмитрия Игоревича на диссертацию Солнцевой-Чалей Марии Олеговны "Методы анализа многомерных данных посадочных траекторий самолета", представленную на соискание ученой степени кандидата технических наук по специальности 05.13.17 — Теоретические основы информатики.

1. Актуальность темы диссертационной работы

Диссертационная работа Солнцевой-Чалей Марии Олеговны посвящена анализу многомерных данных посадочных траекторий самолетов с целью выделения в выборке данных в трёхмерном пространстве установившихся потоков движения, моделированию как репрезентативной (усредненной) траектории потока, так и определению "выбросов" (outliers, т.е. нетипичных траекторий выделенного потока). Указанная задача не является тривиальной в силу того, что посадочные траектории представляют собой сложные многомерные данные, так, например, эти кривые обладают кривизной, кручением, а также содержат множественные пространственные пересечения. Существующие решения данной задачи основаны на кластеризации и сокращении размерности данных (в случае решения задачи на плоскости). В результате возникают следующие основные проблемы. Сокращение размерности ведет к существенной потере информации и невозможности использования полученного результата на практике, в частности, результат разбиения пространства на области в духе диаграмм Вороного не связан с установившимися потоками движения. При использовании для выделения групп траекторий со сходными характеристиками (в контексте диссертации называемых пучками траекторий) методов кластеризации возможно получение нестабильного результата, зависящего от параметров инициализации и требующих предварительного знания о числе содержащихся в выборке пучков. Перечисленные проблемы влекут необходимость разработки новых методов, учитывающих как высокую размерность анализируемых данных, так и их характерные геометрические особенности. Предлагаемый в рассматриваемой диссертации новый метод является методом без обучения, не требует предварительного знания о числе содержащихся в выборке пучков траекторий и иллюстрируется примерами обработки реальных данных. Диссертационное исследование актуально, в первую очередь, для оптимизации работы диспетчерских служб при разделении воздушного пространства на зоны ответственности в условиях постоянно увеличивающегося воздушного трафика. Следует отметить, что представленная диссертационная работа выполнена на должном уровне для столь важной

практической задачи и, в частности, содержит широкий спектр исследованных в ней методов.

2. Научная новизна

Новыми являются следующие результаты, полученные диссертантом:

1. Оригинальный способ нахождения устойчивого разбиения в трёхмерном пространстве множества многомерных временных рядов, представленных траекториями движения объектов к различным целям, на пучки (кластеры) таких траекторий, асимптотически сходящиеся с заданным значением порога, на основе косинусной меры близости многомерных временных рядов.

Предложенный способ описан в 3-й главе диссертации. Он позволяет выделять в трёхмерном пространстве группы траекторий движения, имеющих общие цели и разбивать множество многомерных временных рядов, представленных посадочными траекториями самолётов, на пучки, соответствующие посадкам с одного направления на заданные взлетно-посадочные полосы. Устойчивый результат достигается за счёт использования в качестве меры близости траекторий меры косинуса и сокращения размерности данных, т.е. перехода в пространство меньшей размерности, в котором определяется характерная особенность анализируемых данных – линия (геометрическая асимптота), касательная выделяемому пучку траекторий в окрестности сходимости. Затем следует возвращение в исходное пространство данных.

2. Способ определения (моделирования) в трёхмерном пространстве центроида для выделенного множества (пучка) многомерных временных рядов, представленных траекториями движения объектов, посредством решения оптимизационной задачи с ограничениями и использовании меры косинуса в качестве меры близости многомерных временных рядов.

В рамках данного способа при решении оптимизационной задачи с ограничениями моделируется траектория, представляющая собой пространственное среднее выделенного пучка траекторий таким образом, что сумма квадратов расстояния на основе косинуса от ее точек до соответствующих точек всех траекторий в рассматриваемом пучке минимальна,

что обеспечивает максимальное сходство формы траекторий. Достоверность результата подтверждается примерами обработки данных радара TRACON над заливом Сан-Франциско.

3. Способ определения выброса или нетипичной траектории (outlier) в выделенном пучке многомерных временных рядов, представленных траекториями движения объектов, основанный на оценке отклонения траекторий пучка от его центроида по мере косинуса.

Данный способ в результате решения оптимизационной задачи при ограничениях позволяет определять в выделенном пучке траекторий выброс – траекторию наиболее отклоняющуюся от смоделированного центроида, при этом сумма квадратов расстояния на основе косинуса от точек данной траектории до соответствующих точек траектории-центроида максимальна. Результат достигается за счёт того, что мера косинуса, выбранная в качестве меры близости многомерных пространственных посадочных траекторий, позволяет учесть характерную геометрическую форму траекторий, т.к. важным становится уже не близость точек в евклидовом пространстве, а угол между соответствующими векторами. Результаты убедительно демонстрируются на примере обработки данных радара TRACON для решения задачи деления пространства на сектора в плоском случае и описаны в главе 2.4.

Отдельно стоит отметить уверенное использование автором математического инструментария современного машинного обучения и многомерной статистики, а именно, непараметрическое оценивание плотности, использование ЕМ-алгоритма для обучения по данным, метода максимального правдоподобия, спектральной кластеризации и других.

3. Обоснованность и достоверность научных положений, выводов и рекомендаций

Рассматриваемая в диссертации задача анализа многомерных данных посадочных траекторий самолётов решается на базе достаточно широкого комплекса подходов и методов анализа данных. Использование математического аппарата корректно и обосновано. Достоверность научных результатов диссертации подтверждается примерами обработки выборки реальных данных посадочных траекторий, зарегистрированных радиолокационной системой TRACON, которые находятся в свободном доступе.

4. Значимость полученных результатов диссертационной работы

Теоретическая ценность результатов работы заключается во всестороннем исследовании различных стандартных методов анализа данных и разработке оригинальных, по-видимому, наиболее адекватных на данный момент, способов анализа многомерных данных посадочных траекторий самолётов, имеющих большую практическую значимость.

5. Соответствие специальности

Диссертационная работа, посвященная методам анализа многомерных данных посадочных траекторий самолетов, соответствует специальности 05.13.17 “Теоретические основы информатики”. Работа органично и убедительно демонстрирует содержание, указанное в паспорте специальности, а именно, разработку и исследование моделей и алгоритмов анализа данных, обнаружение закономерностей в данных.

6. Стиль изложения и оформления

Диссертация достаточно полно отражает значительный объем проделанной теоретической и практической работы, выполненной на высоком научно-техническом уровне. Стиль изложения свидетельствует о понимании автором применяемых методов и умении ясно описывать новые модели и сложные технические процессы. Использование автором рисунков, таблиц, псевдокода и схем делает изложение работы ещё более наглядным. Перечисленные выше научные результаты отличаются новизной, а работа в целом – практической значимостью.

Автореферат полностью соответствует содержанию диссертации. Оформление диссертации удовлетворяет соответствующим требованиям ВАК РФ.

7. Замечания по работе

Несмотря на высокое качество результатов и должное их представление, можно сделать некоторые замечания, которые не влияют на высокую оценку представленной диссертации:

1. Автор диссертации несколько раз использует слово “аутлаер” (см. стр. 8 и 12), являющееся буквальной транслитерацией англоязычного термина *outlier* там, где следует использовать установившийся термин выброс для нетипичной траектории в пучке.

2. Вольное обращение с терминами расстояние и близость может привести к некоторой неточности понимания: например, стр. 7 содержит фразу “использование евклидовой меры расстояния в качестве меры близости траекторий движения”. В работе предлагается использовать меру косинуса в качестве меры близости, подразумевая её, с одной стороны, как меру сходства пространственной формы траекторий, и с другой стороны как меру удалённости от центроида. Поэтому при моделировании центроида в пучке траекторий выполняется минимизация целевого критерия (формула 1.3.8), где используется нормированное скалярное произведение двух векторов (формула 1.3.7). Минимизация выполняется, потому что центроид удалён от всех траекторий в пучке. Тогда, при определении выброса (outlier), следует выделить траекторию наиболее удалённую от центроида, т.е. минимально близкую по мере косинуса (величина меры косинуса между такой траекторией и центроидом минимальна), что не ясно описано на шаге 3 предлагаемого алгоритма.
3. Имеются следующие опечатки и неточности. Стр. 13. International conference следует заменить на International conference. Используется английский термин data mining без перевода на стр. 8, 9 и 98. На стр. 54, 55 и 98 следует заменить “регрсии” на “регрессии”. В формуле 1.3.1 на стр. 27 допущена неточность в выражении $\forall (X_n, X_m) \in N_k$, т.к. согласно ранее данному определению множество пучков не содержит упорядоченные пары многомерных временных рядов, поэтому круглые скобки необходимо опустить.

8. Подтверждение опубликования основных результатов диссертации

По теме диссертационной работы опубликовано 20 научных работ (15 статей, сделано 5 докладов на международных и российских научно-технических конференциях), 12 из которых – в изданиях из списка ВАК. Автором также получено 4 патента на изобретение.

Диссертация Солнцевой-Чалей Марии Олеговны является самостоятельной законченной научно-исследовательской работой, в которой решена актуальная научно-техническая задача выделения на основании имеющихся данных установившихся потоков движения объектов в 3D-случае, благодаря использованию в качестве меры близости многомерных посадочных траекторий меры косинуса.

объектов в 3D-случае, благодаря использованию в качестве меры близости многомерных посадочных траекторий меры косинуса.

Отмеченные недостатки не снижают высокого качества работы, и позволяют оценить ее положительно.

Уровень диссертации соответствует требованиям п.7 Положения ВАК Минобрнауки России, а ее автор, Солнцева-Чалей Мария Олеговна, заслуживает присуждения ему ученой степени кандидата технических наук по специальности 05.13.17 "Теоретические основы информатики".

Официальный оппонент,

кандидат технических наук

доцент департамента анализа

данных и искусственного интеллекта

Национального исследовательского университета

Высшей школы экономики

Рабочий телефон: +7(495) 772-9590 добавочный 22930

Мобильный телефон: +7 926 3818033

Email: dignatov@hse.ru

9 мая 2018 года



Игнатов Дмитрий Игоревич

Подпись заверяю



СПЕЦИАЛИСТ ПО КАДРАМ

Исхакова Л.К.

