

УДК 004.654

ОБНАРУЖЕНИЕ СТАТИСТИЧЕСКИХ АНОМАЛИЙ В ДАННЫХ СИСТЕМ АВТОМАТИЧЕСКОГО ЗАВИСИМОГО НАБЛЮДЕНИЯ

Гриняк В.М., Сапунов В.А., Гусев Е.Г.

*Владивостокский государственный университет экономики и сервиса,
Владивосток, e-mail: auscol@mail.ru*

Статья посвящена задаче сбора данных о движении воздушных объектов в зонах ответственности систем управления воздушным движением. Разработка технологий обеспечения безопасности воздушного движения, построения и верификации моделей управления коллективным воздушным движением и моделей управления безопасностью полетов, как правило, связана с постановкой вычислительных экспериментов на реальных данных о движении воздушных судов. Это требует предварительного сбора таких данных. В работе показан практически реализованный способ сбора необходимых данных с различных открытых интернет-ресурсов. При этом собираемые данные нередко содержат в себе некорректную информацию, что требует их предварительной обработки, идентификации и удаления некорректных записей. Для решения этой задачи разработан подход, связанный с разбиением траектории движения на отдельные участки полёта, идентификации их кластерами в фазовом пространстве данных о движении и поиском аномалий. В статье продемонстрирован результат работы созданных алгоритмов на примере конкретных данных о траектории движения воздушного судна, сделан вывод об эффективности и практической пригодности разработанного подхода.

Ключевые слова: управление воздушным движением, моделирование движения, системы управления безопасностью полетов, траектория воздушного судна, обработка данных, кластеризация

DETECTION OF STATISTICAL ANOMALIES IN DATA OF ADS-B SYSTEMS

Grinyak V.M., Sapunov V.A., Gusev E.G.

Vladivostok State University of Economics and Service, Vladivostok, e-mail: auscol@mail.ru

Current paper is devoted to the problem of collecting data on the movement of air targets in the areas of responsibility of air traffic control systems. Development of technologies for ensuring air traffic safety, building and verification of collective air traffic control models and safety management models is usually associated with the establishment of computational experiments on real data on the movement of aircraft. This requires preliminary collection of such data. The paper shows a practically realized way of collecting necessary data from various open Internet resources. In this case, the collected data often contains incorrect information, which requires their preliminary processing, identification and removal of incorrect records. To solve this problem, the approach associated with splitting the trajectory of motion into individual sections of the flight, identifying them with clusters in the phase space of motion data, and searching for anomalies turned out to be productive. The paper demonstrates the result of the work of the created algorithms on the example of concrete data on the trajectory of the aircraft's motion, a conclusion is made about the effectiveness and practical suitability of the developed approach.

Keywords: air traffic control, motion modeling, flight safety control systems, airplane trace, data processing, clustering

Постоянное повышение уровня безопасности полетов лежит в основе развития воздушного транспорта. Вопрос безопасности является важнейшей составляющей систем организации воздушного движения [1, 2, 5].

Первичные данные о движении воздушных судов традиционно предоставляются наземными радарными системами; как правило, такие системы позволяют с достаточной точностью определить координаты и скорость движения воздушного судна, однако дистанция их действия ограничена. Системы автоматического зависимого наблюдения в режиме радиовещания (ADS-B) [8, 11–13] и система мультilaterации (MLAT) представляют собой альтернативные системы наблюдения за воздушным трафиком. Их привлекательность обусловлена в том числе сравнительно низкими затратами на установку и обслуживание оборудования. Ожидается, что системы ADS-B и MLAT будут все шире использоваться

в тех районах, где использование радарных систем экономически нецелесообразно.

Организация воздушного движения связана с решением целого ряда специфических навигационных задач, в том числе – с целью предотвращения опасных ситуаций (чрезмерное сближение, уход с эшелона и т.д.). Эти навигационные задачи описываются математическими моделями и алгоритмами, требующими верификации на предмет их пригодности для использования в реальных системах управления движением воздушных судов. Подготовка данных для экспериментов, связанных такой верификацией, предполагает формирование модельных траекторий движения воздушного судна, максимально приближенных к реальным. Один из путей, позволяющих получить доступ к таким данным, – использование открытых интернет-источников данных о движении воздушных судов, которые, в свою очередь, используют системы

автоматического зависимого наблюдения за движением воздушных судов. Уже сейчас в широком доступе в глобальной сети Интернет доступен целый ряд сервисов, предоставляющих данные как морского, так и воздушного трафика [3], например *flightradar24.com*, *marinetraffic.com* и др.

Некоторые вопросы исследования массивов данных о движении морских и воздушных судов уже рассматривались авторами ранее [3, 4, 6, 7]. Так, были предложены разные модели обработки и хранения данных о движении, изучались вопросы структуры и частоты обновления данных о движении морских и воздушных судов, предоставляемых системами с открытым доступом, анализировались возможности по интерполяции данных для решения различных прикладных задач обеспечения безопасности движения.

Настоящая статья посвящена задаче детектирования аномалий (выбросов) в данных о движении воздушных судов посредством применения статистических методов Data Mining. Работа ориентирована на расширение функций существующих систем управления воздушным движением.

Сбор, структура и характеристики данных о движении воздушных судов

Упомянутые интернет-сервисы имеют API-интерфейс, позволяющий обрабатывать данные протокола HTTP. В систему подаётся GET запрос в соответствующем API формате поставщика данных и принимается ответ в формате JSON. Получаемые таким образом данные представляют собой запись следующего вида:

```
{...
  «90103962»:»co1»:»KAL»,»co2»:»KE»,»
  dst1»:»RJGG»,»dst2»:»NGO»,»fal»:34500,»f
  ar»:»HL7783»,»farnd»:»,»fat»:»B77W»,»fch
  »:1479465000637,»fgs»:450,»fhd»:28,»fid»:9
  0103962,»fn1»:»KAL751»,»fn2»:»KE751»,»
  falm»:0,»fgskmh»:17.594,»fgsnah»:8.255272,
  »fsr»:»ADSB»,»fvr»:»,»lat»:37.460003,»lng
  »:126.451515,»org1»:»RKSI»,»org2»:»ICN»,
  »std»:»PGANRB300711»
  ...}
```

В данном случае запись представляет собой строку символов, содержащих информацию о наблюдаемом воздушном судне. Сопоставляя информацию, отображаемую на сайте, определяется формат полученных данных: *dst1* – код аэропорта вылета, *dst2* – код аэропорта назначения, *fal* – высота полета ВС, *far* – регистрационный номер, *fat* – модель ВС, *fch* – временная метка (UNIX-время),

fgs – скорость движения ВС, *fhd* – истинный курс ВС, *lng* – долгота, *lat* – широта.

В рассматриваемых данных о движении воздушных судов можно выделить статическую и динамическую часть. Статические данные содержат информацию о регистрационных параметрах и маршруте движения воздушного судна, а динамические данные отражают координаты и физические параметры движения. Запрос полетных данных осуществляется периодически. Объем получаемых и обрабатываемых при этом данных о движении зависит от наблюдаемого региона, периода опроса и длительности сбора полетных данных; он может достигать нескольких сотен записей в минуту. Обработка и хранение данных в этом случае с использованием традиционной реляционной модели бывает затруднена вследствие повышенных требований к производительности соответствующего оборудования, что является побудительным мотивом обращения к СУБД не реляционного типа [14].

Как показали проведенные наблюдения, данные о движении самолётов, поступающие из доступных интернет-источников, нередко содержат некорректную информацию. Это требует их предварительной обработки, идентификации и удаления некорректных записей. Для решения этой задачи оказался продуктивным подход, связанный с разбиением траектории движения на отдельные участки полёта, идентификации их кластерами в фазовом пространстве данных о движении и поиском аномалий.

Постановка задачи

Опишем рассматриваемое множество полетных данных о движении воздушного судна в пространстве и введём обозначения. Будем рассматривать следующие параметры траектории: высота полета над уровнем моря h , горизонтальная скорость полета v_{xy} , вертикальная скорость полёта v_h . Таким образом, будем иметь множество полётных данных $X = \{h(t_i), v_{xy}(t_i), v_h(t_i)\}$.

Исходя из специфики проблемной области и исходных данных введём следующие предположения:

1. Каждый полет состоит из набора последовательных участков – фаз полета, в которых параметры движения близки друг к другу. Это соответствует понятию «кластер» в фазовом пространстве переменных h, v_{xy}, v_h .

2. Каждый выделенный кластер может быть разбит на более мелкие участки – сегменты фазы полёта.

3. Могут иметь место выбросы – аномальные значения данных, которые в значительной степени отличаются от значений параметров выбранной фазы полёта.

В работе ставится задача синтеза подходящего алгоритма идентификации характерных фаз полёта (кластеров) и детектирования выбросов. Выбросы считаются некорректной информацией о движении воздушного судна.

Метод решения задачи выявления некорректной информации

Описываемый далее метод обнаружения выбросов данных о движении основан на анализе траектории движения воздушного судна. Алгоритм состоит из следующей последовательности этапов:

1. Выделение фаз полета.
2. Кластеризация фаз.
3. Анализ выбросов.

Выделение и кластеризация фаз полёта. На этом этапе траектория движения воздушного судна разбивается на участки, каждый из них несет определенный физический смысл. Как правило, выделяют следующие фазы полета [5, 10]:

- стоянка;
- буксировка;
- руление до места старта;
- взлет;
- набор круизной высоты;
- круиз;
- снижение;
- маневрирование;
- приближение;
- посадка;
- руление до места стоянки.

Фазы можно выделить, используя кластеризацию траектории движения воздушного судна методами общего назначения, в которых расстояние между объектами при-

нимается евклидовым, а расстояние между кластерами – расстоянием Уорда [9]. Основное ограничение, накладываемое на процесс кластеризации в рассматриваемой задаче, состоит в том, что объединяться в кластеры могут только соседние по времени данные.

Исследовались различные варианты кластеризации данных траектории движения воздушного судна.

- По трём параметрам (h, v_{xy}, v_h) .
- По двум параметрам $(h, v_{xy}), (h, v_h)$ или (v_{xy}, v_h) .
- По одному из параметров h, v_{xy} или v_h .
- По одному параметру – суперпозиции (взвешенная сумма высоты и горизонтальной скорости, высоты и вертикальной скорости, горизонтальной и вертикальной скорости, произведение высоты и горизонтальной скорости и др.)

Наиболее устойчивые результаты дала кластеризация по одному параметру – суперпозиции высоты и горизонтальной скорости. При этом рассматривались разные варианты суперпозиции: взвешенная сумма $a_h h + a_v v_{xy}$ и произведение параметров $h \times v_{xy}$. Окончательный выбор был остановлен на произведении в силу отсутствия необходимости обоснования выбора весовых коэффициентов.

Анализируя результаты эксперимента на реальных данных, можно отметить, что проведенная кластеризация полного маршрута движения воздушного судна достаточно точно соотносится с фактической градацией фаз движения воздушного судна. На рис. 1 показаны относительные значения параметра кластеризации на протяжении траектории полёта (по оси абсцисс отложено приведённое время движения) и центры найденных кластеров. В данном случае было выделено 13 кластеров, соответствующих основным фазам движения воздушного судна (таблица).

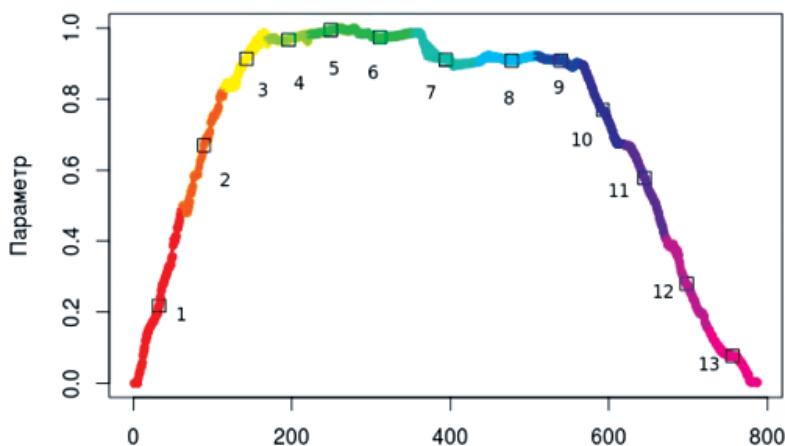


Рис. 1. Выделенные кластеры параметров траектории движения

Выделенные кластеры и фазы движения воздушного судна

Номер кластера	Фаза движения
1	Буксировка, руление до места старта, взлёт
2, 3	Набор круизной высоты
4, 6, 9	Круиз
10, 11, 12	Снижение
13	Посадка

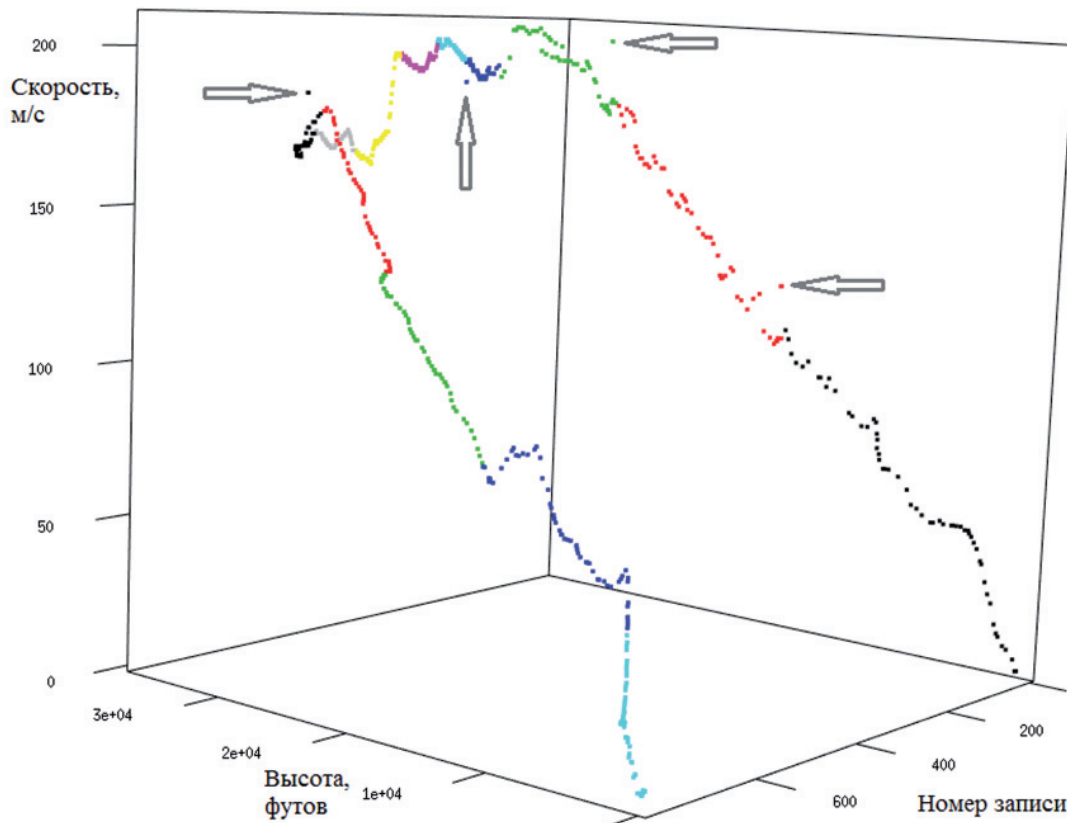


Рис. 2. Пример траектории движения воздушного судна с выделенными некорректными данными

Анализ выбросов. Используя полученные данные на предыдущем этапе о кластерах и фазах движения воздушного судна, примем известную кусочно-линейную регрессионную модель [9] данных о движении для каждого кластера. Будем считать выбросами такие значения параметров, которые не входят в границы доверительного интервала с вероятностью 95%, имея в виду функционал метода наименьших квадратов.

На рис. 2 показан пример решения задачи детектирования выбросов, которые считаются некорректными данными. В данном случае по координатным осям отложены данные, по суперпозиции (произведению $h \times v_{xy}$) которых выполняется кластеризация: высота полёта и скорость движения,

а также номер записи (приведённое время полёта). Различным цветом показаны данные различных фаз полёта. Стрелками показаны точки, оцененные как некорректные данные – случайные выбросы (видно, что они лежат далеко от фазовой траектории), видно, что они явно выделяются на фоне валидных данных.

Рис. 3 представляет собой увеличенную копию части рис. 2, на которой показан только участок траектории взлёта и набора высоты воздушного судна.

Как видно из рис. 2 и 3, описанный подход вполне позволяет идентифицировать некорректные данные о траектории движения самолёта, обусловленные сбоями работы навигационного оборудования и специализированных веб-ресурсов.

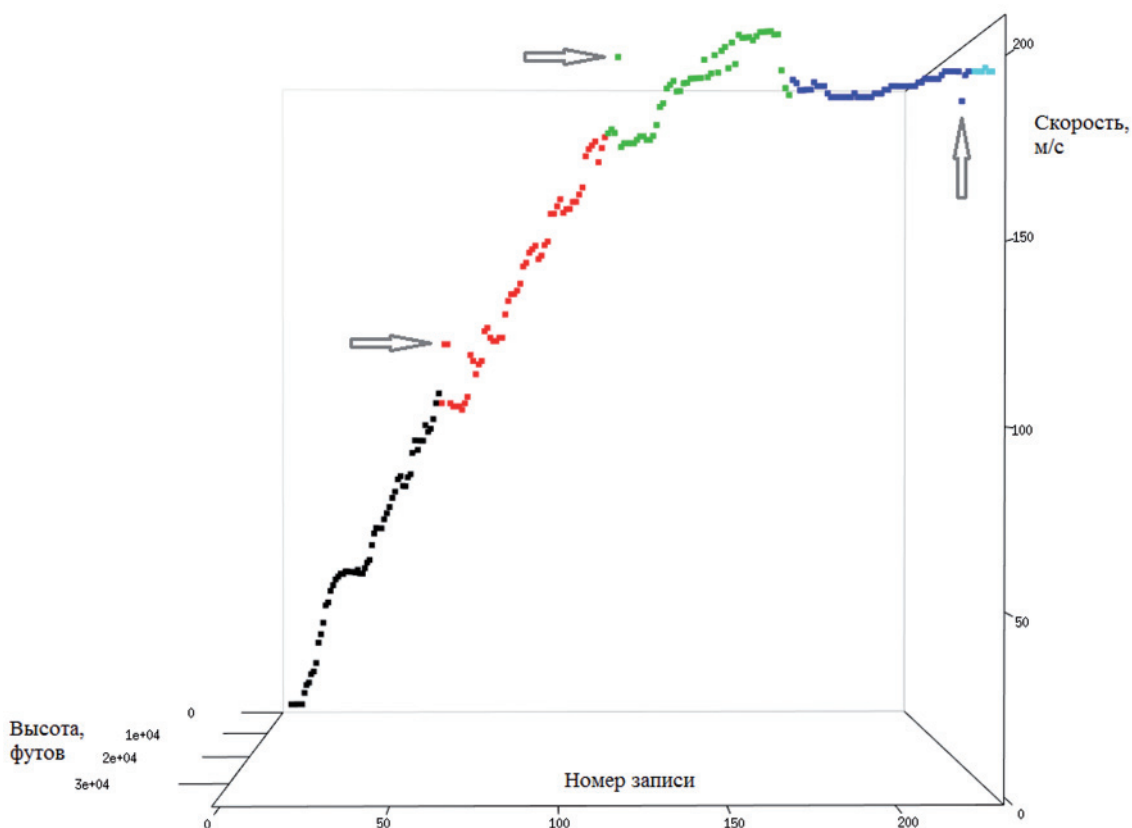


Рис. 3. Пример траектории движения воздушного судна – участок взлёта и набора высоты

Заключение

Реальные данные о трафике воздушных судов представляют собой исключительную ценность для моделирования и исследования задач в области управления воздушным движением и обеспечения безопасности полётов. Они требуются для оценки работоспособности, свойств и ограничений создаваемых моделей и алгоритмов в условиях реального воздушного трафика. Для сбора и анализа таких данных авторами была разработана информационная система, позволяющая собирать и хранить информацию о движении воздушных судов по всему земному шару в зонах доступа базовых станций ADS-B. При этом была выявлена проблема достоверности предоставляемых системой данных: они нередко содержат некорректную информацию. В настоящей работе предлагается подход к выделению (идентификации) таких некорректных данных, который характеризуется следующими аспектами:

- Данные о траектории движения воздушного судна запрашиваются на открытых интернет-сайтах, затем загружаются,

парсируются и хранятся в базе нереляционного типа.

- Полученные данные обрабатываются алгоритмом кластеризации общего назначения, в котором параметром кластеризации выступает произведение высоты и горизонтальной скорости движения воздушного судна; при этом объединяться в кластеры могут только соседние по времени данные.

- Внутри каждого кластера – фазы движения воздушного судна – реализуется кусочно-линейная регрессионная модель данных, в которой доверительный интервал определяется пороговым уровнем вероятности 95%. Данные, лежащие вне доверительного интервала, считаются ошибочными.

Указанные положения определяют научную новизну настоящей работы.

Предложенный способ идентификации некорректных данных в массивах информации о движении воздушных судов, загружаемых из доступных интернет-источников, был реализован и апробирован в реально действующей информационной системе и подтвердил свою реализуемость и эффективность. Получаемые указанной системой

данные используются для верификации разрабатываемых новых методов, моделей и алгоритмов обеспечения безопасности воздушного движения.

Список литературы

1. Баженов С.Г., Кулида Е.Л., Лебедев В.Г. Формирование бесконфликтных траекторий предпосадочного маневрирования с учетом ограничений на маневренные возможности самолёта // Проблемы управления. – 2012. – № 2. – С. 70–75.
2. Варфоломеев К.С., Щепилов Ю.Н. О подходе к количественной оценке качества процедур маневрирования в районе аэродрома // Транспорт: наука, техника, управление. – 2012. – № 6. – С. 27–29.
3. Головченко Б.С., Гриняк В.М. Информационная система сбора данных трафика морской акватории // Научно-техническая информация. Сер. 2: Информационные процессы и системы. – 2014. – № 8. – С. 24–28.
4. Гриняк В.М., Будников А.И. Информационная система сбора данных о движении воздушных объектов гражданской авиации // Транспорт: наука, техника, управление. – 2015. – № 8. – С. 44–47.
5. Дашков И.Д., Зубков Б.В. Определение и оценка состояний функциональных систем воздушных судов в системе управления безопасностью полетов // Научный вестник Московского государственного технического университета гражданской авиации. – 2014. – № 205. – С. 32–36.
6. Девятисильный А.С., Дорожко В.М., Гриняк В.М. Нейроподобные алгоритмы высотной классификации движущихся объектов // Информационные технологии. – 2001. – № 12. – С. 45–51.
7. Девятисильный А.С., Дорожко В.М., Гриняк В.М. Определение гидродинамического сопротивления по траекторным данным инерционного движения объекта // Журнал технической физики. – 2003. – № 2. – С. 38–42.
8. Игнатюк В.А., Сметанин С.И., Марус В.С. Способ организации расширенной системы спутникового GNSS мониторинга // Территория новых возможностей. Вестник Владивостокского государственного университета экономики и сервиса. – 2015. – № 1. – С. 72–79.
9. Пронина О.Ю., Баженов Р.И. Исследование методов регрессионного анализа программной среды EVIEWS // Nauka-Rastudent.ru. – 2015. – № 1. – С. 45.
10. Смирнова Ю.В. О совершенствовании системы управления безопасностью полетов в России // Безопасность в техносфере. – 2012. – № 3. – С. 43–45.
11. Сметанин С.И., Игнатюк В.А., Евстифеев А.А. Способ реализации программной веб-части системы спутникового мониторинга // Информационные технологии. – 2015. – № 6. – С. 448–455.
12. ADS-B Технология (TIS-B, FIS-B) [Электронный ресурс]. – Режим доступа: http://adsbradar.ru/ads-b_fis-b_tis-b_technology (дата обращения 01.04.17).
13. ADS-B Technologies Website [Электронный ресурс]. – Режим доступа: <http://www.ads-b.com> (дата обращения 01.04.17).
14. The JavaScript Object Notation (JSON) Data Interchange Format [Электронный ресурс]. – Режим доступа: <http://www.rfc-editor.org/rfc/rfc7159.txt> (дата обращения 01.04.17).