

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

ТЕОРЕТИЧЕСКИЙ И ПРИКЛАДНОЙ НАУЧНО-ТЕХНИЧЕСКИЙ ЖУРНАЛ

8 (216)
2014

САПР

КОМПЬЮТЕРНАЯ ГРАФИКА

МЕТОДЫ ПРОГРАММИРОВАНИЯ

ОПЕРАЦИОННЫЕ СИСТЕМЫ И СРЕДЫ

ТЕЛЕКОММУНИКАЦИИ
И ВЫЧИСЛИТЕЛЬНЫЕ СЕТИ

ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ

НЕЙРОСЕТИ И
НЕЙРОКОМПЬЮТЕРЫ

СТРУКТУРНЫЙ СИНТЕЗ

ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ

ПРИКЛАДНЫЕ ИНФОРМАЦИОННЫЕ
СИСТЕМЫ

ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ

ОПТИМИЗАЦИЯ И МОДЕЛИРОВАНИЕ

ИТ В ОБРАЗОВАНИИ

ГИС

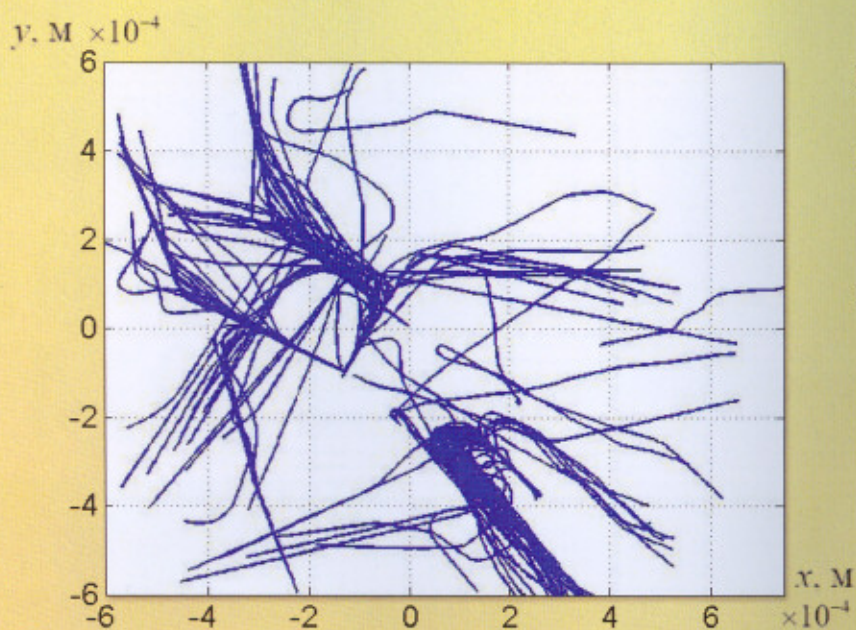


Рис. 1. Исходный набор из 143 траекторий самолетов

Рисунки к статье Б. Г. Кухаренко,
М. О. Солнцевой
**«Итеративная кластеризация
траекторий управляемых
объектов в многомерном
пространстве характеристик»**

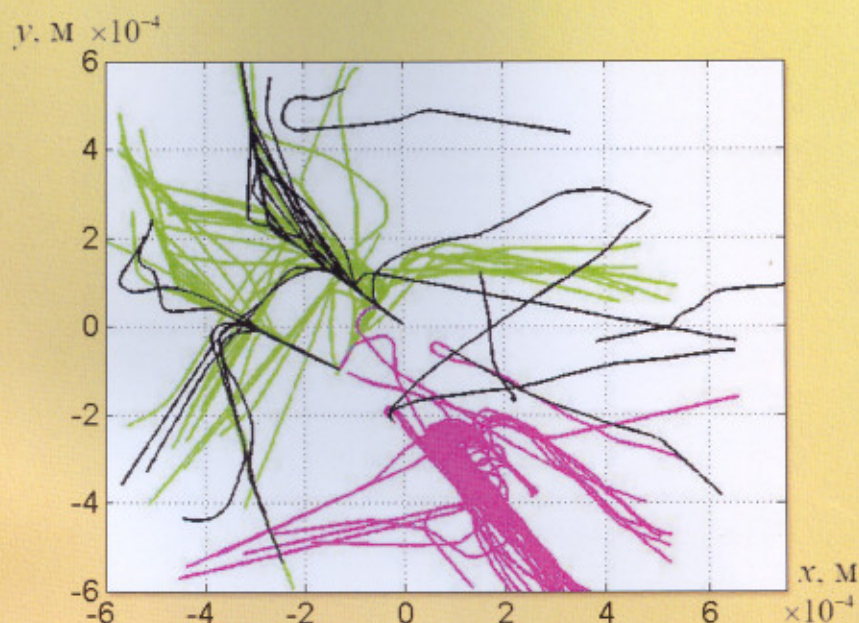


Рис. 2. Кластеры траекторий самолетов после первой итерации

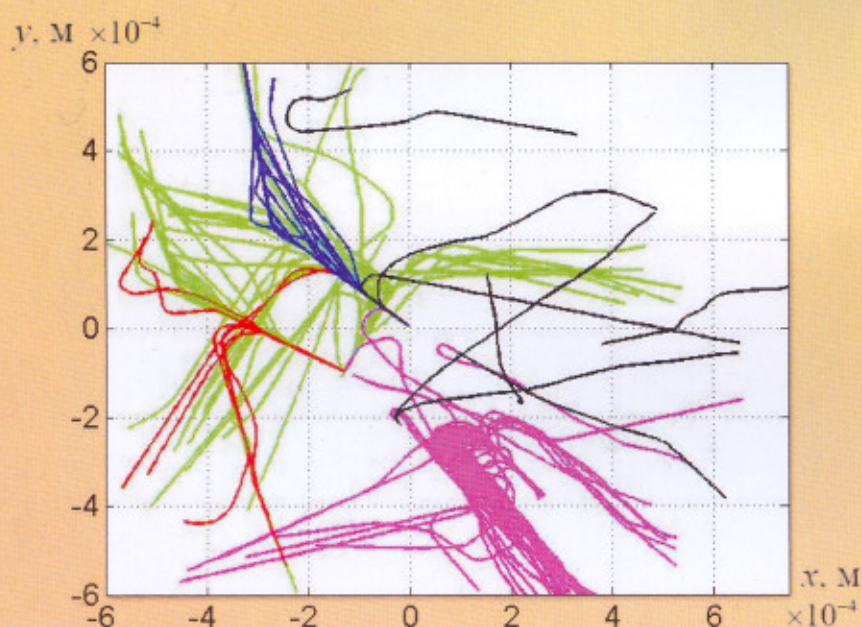


Рис. 3. Окончательный результат кластеризации траекторий на рис.1

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

8(216)
2014

ТЕОРЕТИЧЕСКИЙ И ПРИКЛАДНОЙ НАУЧНО-ТЕХНИЧЕСКИЙ ЖУРНАЛ

Издается с ноября 1995 г.

УЧРЕДИТЕЛЬ
Издательство "Новые технологии"

СОДЕРЖАНИЕ

ОБЩИЕ ВОПРОСЫ

Нечаев В. В. Современные бионические информационные технологии . . . 3

ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ

Кухаренко Б. Г., Солнцева М. О. Итеративная кластеризация траекторий управляемых объектов в многомерном пространстве характеристик . . . 11

Соколов И. В., Агеев А. М., Волобуев М. Ф., Замыслов М. А., Мальцев А. М., Михайленко С. Б. Оценка статистических характеристик сигнала на выходе резервированной системы, управляемой нечетким контроллером . . . 17

МОДЕЛИРОВАНИЕ И ОПТИМИЗАЦИЯ

Имамвердиев Я. Н. Модель ситуационного управления информационной безопасностью электронного правительства . . . 24

Куравский Л. С., Мармалюк П. А., Баранов С. Н., Алхимов В. И., Юрьев Г. А., Артюхина С. В. Марковские модели глазодвигательной активности и их применение для тестирования профессиональных навыков . . . 34

СИСТЕМЫ АВТОМАТИЗИРОВАННОГО ПРОЕКТИРОВАНИЯ

Гаврилов С. В., Иванова Г. А., Манукян А. А. Новые проблемы логико-топологического синтеза заказных сложно-функциональных блоков и методы их решения . . . 44

Ракитин В. В., Русаков С. Г. Электрические модели мемристорных элементов . . . 50

Анисимов В. И., Гридин В. Н., Шабани М. А. Методы построения высокопроизводительных распределенных систем автоматизации схемотехнического проектирования . . . 59

ПРОГРАММНАЯ ИНЖЕНЕРИЯ

Авдошин С. М., Песоцкая Е. Ю. Экосистемы программного обеспечения: инновации в ИТ . . . 64

ПРИКЛАДНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

Амирасланова З. Н. Автоматизация формирования чертежей арматуры железобетонных конструкций . . . 70

Новак Е. В. Информационно-коммуникационные технологии: оценка эффективности . . . 74

Главный редактор:

СТЕМПКОВСКИЙ А. Л.,
акад. РАН, д. т. н., проф.

Зам. главного редактора:

ИВАННИКОВ А. Д., д. т. н., проф.
ФИЛИМОНОВ Н. Б., д. т. н., с.н.с.

Редакционный совет:

БЫЧКОВ И. В., акад. РАН, д. т. н.
ЖУРАВЛЕВ Ю. И.,

акад. РАН, д. ф.-м. н., проф.

КУЛЕШОВ А. П.,

акад. РАН, д. т. н., проф.

ПОПКОВ Ю. С.,

чл.-корр. РАН, д. т. н., проф.

РУСАКОВ С. Г.,

чл.-корр. РАН, д. т. н., проф.

СОЙФЕР В. А.,

чл.-корр. РАН, д. т. н., проф.

СОКОЛОВ И. А., акад.

РАН, д. т. н., проф.

СУЕТИН Н. В., д. ф.-м. н., проф.

ЧАПЛЫГИН Ю. А.,

чл.-корр. РАН, д. т. н., проф.

ШАХНОВ В. А.,

чл.-корр. РАН, д. т. н., проф.

ШОКИН Ю. И.,

акад. РАН, д. т. н., проф.

ЮСУПОВ Р. М.,

чл.-корр. РАН, д. т. н., проф.

Редакционная коллегия:

АВДОШИН С. М., к. т. н., доц.

АНТОНОВ Б. И.

БАРСКИЙ А. Б., д. т. н., проф.

ВАСЕНИН В. А., д. ф.-м. н., проф.

ГАЛУШКИН А. И., д. т. н., проф.

ДИМИТРИЕНКО Ю. И., д. ф.-м. н., проф.

ДОМРАЧЕВ В. Г., д. т. н., проф.

ЗАГИДУЛЛИН Р. Ш., к. т. н., доц.

ЗАРУБИН В. С., д. т. н., проф.

ИСАЕНКО Р. О., к. т. н., с.н.с.

КАРПЕНКО А. П., д. ф.-м. н., проф.

КОЛИН К. К., д. т. н., проф.

КУЛАГИН В. П., д. т. н., проф.

КУРЕЙЧИК В. М., д. т. н., проф.

КУХАРЕНКО Б. Г., к. ф.-м. н., доц.

ЛЬВОВИЧ Я. Е., д. т. н., проф.

МИХАЙЛОВ Б. М., д. т. н., проф.

НЕЧАЕВ В. В., к. т. н., проф.

РЯБОВ Г. Г., чл.-корр. РАН, д. т. н., проф.

СОКОЛОВ Б. В., д. т. н., проф.

УСКОВ В. Л., к. т. н. (США)

ФОМИЧЕВ В. А., д. т. н., проф.

ЧЕРМОШЕНЦЕВ С. Ф., д. т. н., проф.

ШИЛОВ В. В., к. т. н., доц.

Редакция:

БЕЗМЕНОВА М. Ю.

ГРИГОРИН-РЯБОВА Е. В.

ЛЫСЕНКО А. В.

ЧУГУНОВА А. В.

Информация о журнале доступна по сети Internet по адресу <http://novtex.ru/IT>.

Журнал включен в систему Российского индекса научного цитирования.

Журнал входит в Перечень научных журналов, в которых по рекомендации ВАК РФ должны быть опубликованы научные результаты диссертаций на соискание ученой степени доктора и кандидата наук.

УДК 519.233

Б. Г. Кухаренко, канд. физ.-мат. наук, ст. науч. сотр., вед. науч. сотр.,
Институт машиноведения РАН, г. Москва, e-mail: kukharenko@imash.ru
М. О. Солнцева, аспирант, Московский физико-технический институт (ГУ),
e-mail: solnceva.chalei@gmail.com

Итеративная кластеризация траекторий управляемых объектов в многомерном пространстве характеристик

Итеративное применение алгоритма кластеризации многомерных траекторий управляемых объектов позволяет выделить характерные паттерны движения и изолирует посторонние траектории. Первичная кластеризация траекторий выполняется последовательно в различных пространствах характеристик с помощью алгоритма сдвига среднего. На основании результатов кластеризации по всем пространствам характеристик выделяются обобщенные кластеры. Итеративная кластеризация демонстрируется на примере анализа траекторий движения самолетов в зоне аэропорта.

Ключевые слова: анализ данных, многомерные траектории, кластеризация

Введение

Траектории объектов видеонаблюдения являются важными элементами для представления и анализа движения объектов. Кластеризация является ключевым методом анализа траекторий, когда вместо моделирования или исследования движения индивидуального объекта множество траекторий обрабатывается одновременно, раскрывая структуру потоков движущихся объектов. Кластеризация позволяет разделить траектории на два основных типа: нормальные траектории, которые относятся к общим моделям движения, и посторонние траектории (outliers), которые отклоняются от общих моделей. В результате кластеризации в один кластер попадают траектории, имеющие сходные характеристики, а характеристики траекторий разных кластеров значительно различаются. В трехмерном декартовом пространстве траектории движущихся объектов описываются последовательностью трех пространственных координат, абсолютной скоростью и временем. Для классификации траекторий объектов применяется как одновременная кластеризация траекторий в многомерном пространстве [1, 2], так и кластеризация моделей траекторий в различных абстрактных пространствах характеристик с последующим объединением результатов [3].

В настоящей работе предлагается итеративный подход к кластеризации траекторий, отображенных в абстрактные пространства характеристик. Все абстрактные пространства характеристик траекторий используются одновременно для получения более высокой степени информативности, чем при исполь-

зовании единственного пространства характеристик. В каждом пространстве характеристик используется эмпирическое распределение вероятности векторов, представляющих траектории, для которого с помощью алгоритма сдвига среднего (mean shift) [4, 5] определяются моды — локальные максимумы этого распределения. Представляющие траектории вектора, которые относятся к конкретным модам, образуют кластеры. Далее результаты кластеризации в пространствах характеристик анализируются совместно для определения окончательных кластеров. Полученные на первой итерации кластеры с малым числом элементов и векторы, соответствующие траекториям, которые находятся далеко от центра кластера, считаются посторонними. Затем для множества всех посторонних траекторий процедура кластеризации повторяется до достижения стационарного результата, когда последующие итерации не меняют число посторонних траекторий. Результаты итеративной кластеризации траекторий в абстрактных пространствах характеристик демонстрируются на примере анализа данных радаров международного аэропорта г. Сан-Франциско, находящихся в свободном доступе на сайте <http://c3.nasa.gov/dashlink/resources/132>.

Представление траекторий в пространствах характеристик

Выбор модели представления траектории является основой алгоритма кластеризации [3]. Модели представления траекторий делятся на модели с обу-

чением и модели без обучения. Модели с обучением чувствительны к начальному выбору параметров, что может привести к ненадежным моделям кластеров. В отличие от моделей с обучением, модели без обучения не используют обучающую выборку траекторий и предварительное задание числа кластеров.

Данные анализируемых траекторий представляются в различных абстрактных пространствах характеристик, в каждом из которых выполняется непараметрическая кластеризация. Анализ множества траекторий с помощью алгоритма сдвига среднего [4, 5] выявляет моды и соответствующие им кластеры. Окончательные кластеры формируются в результате объединения результатов кластеризации по всем используемым пространствам характеристик. Кластеры, соответствующие похожим траекториям, рассматриваются как нормальные, в то время как отдельные, не вошедшие в кластеры, траектории считаются посторонними.

В трехмерном пространстве траектории объектов представляются в виде

$$T_j = \{(x_j[i], y_j[i], z_j[i]), i = \overline{1, N_j}\}, j = \overline{1, J},$$

где декартовы координаты $(x_j[i], y_j[i], z_j[i])$ соответствуют оцениваемому положению j -го объекта; N_j — число точек траектории (траектории объектов могут иметь различные длины).

Пусть функции $\Phi_m(\cdot), m = \overline{1, M}$, переводят траекторию T_j в d_m -мерные пространства характеристик $\{\Psi_m^{d_m}, m = \overline{1, M}\}$, т. е.

$$\Phi_m(\cdot): T_j \rightarrow \Psi_m^{d_m}, m = \overline{1, M},$$

где M — число пространств характеристик.

Пространства характеристик $\{\Psi_m^{d_m}, m = \overline{1, M}\}$ считаются независимыми, во-первых, чтобы избежать нормирования, необходимого при совместном анализе характеристик, и, во-вторых, чтобы упростить совместный анализ неортогональных пространств характеристик с различными областями определения (пространственными и угловыми). Кроме того, этот подход допускает параллельную кластеризацию с использованием различных характеристик и дальнейшую интеграцию результатов, полученную в различных пространствах характеристик.

Выбор характеристик траектории определяется решаемой задачей, и каждое пространство характеристик способствует выделению кластеров из множества анализируемых траекторий. В качестве характеристик траектории рассматриваются пространственные и угловые представления траектории j -го объекта:

- средняя скорость объекта (average velocity) $\langle v_j \rangle$;

- вектор перемещения объекта (directional distance) $\mathbf{d}_j$;
- среднее траектории объекта (mean) $\mathbf{m}_j$;
- модель формы траектории объекта в виде полиномиальной регрессии (shape);
- представление траектории объекта с помощью анализа главных компонент (Principal Component Analysis — PCA) [6, 7];
- гистограмма изгибов траектории объекта (directional histogram) [8].

Средняя скорость $\langle v_j \rangle$ описывает изменение положения j -го объекта

$$\langle v_j \rangle = \frac{1}{N_j - 1} \sum_{i=1}^{N_j-1} (x_j[i+1] - x_j[i], y_j[i+1] - y_j[i], z_j[i+1] - z_j[i]). \quad (1)$$

Средняя скорость помогает разделению траекторий объектов, движущихся с переменным шагом. Вектор перемещения $\mathbf{d}_j$ j -го объекта представляет длину проекций траекторий на оси координат и описывает направление движения:

$$\mathbf{d}_j = (x_j[N_j] - x_j[1], y_j[N_j] - y_j[1], z_j[N_j] - z_j[1]). \quad (2)$$

Эта характеристика помогает разделить длинные и короткие траектории, а также противоположно направленные траектории. Пространственные координаты среднего положения $\mathbf{m}_j$ j -й траектории

$$\mathbf{m}_j = \frac{1}{N_j} \sum_{i=1}^{N_j} (x_j[i], y_j[i], z_j[i]). \quad (3)$$

Эта характеристика хорошо разделяет траектории, относящиеся к различным областям в трехмерном пространстве. Для моделирования формы траектории T_j независимо от ее длины и набора точек используется полиномиальная регрессия

$$\begin{cases} \mathbf{x}_j^* = (1 \mathbf{z}_j (\mathbf{z}_j)^2 \dots (\mathbf{z}_j)^q) \times (\beta_0^{(1)} \beta_1^{(1)} \dots \beta_q^{(1)})^T + \boldsymbol{\varepsilon}^{(1)}; \\ \mathbf{y}_j^* = (1 \mathbf{x}_j (\mathbf{x}_j)^2 \dots (\mathbf{x}_j)^q) \times (\beta_0^{(2)} \beta_1^{(2)} \dots \beta_q^{(2)})^T + \boldsymbol{\varepsilon}^{(2)}; \\ \mathbf{z}_j^* = (1 \mathbf{y}_j (\mathbf{y}_j)^2 \dots (\mathbf{y}_j)^q) \times (\beta_0^{(3)} \beta_1^{(3)} \dots \beta_q^{(3)})^T + \boldsymbol{\varepsilon}^{(3)}, \end{cases} \quad (4)$$

где первые множители в правой части уравнений — стандартные матрицы Вандермонда размерности $N_j \times (q+1)$ с компонентами $\mathbf{z}_j = \{z_j[i], i = \overline{1, N_j}\}$, $\mathbf{x}_j = \{x_j[i], i = \overline{1, N_j}\}$ и $\mathbf{y}_j = \{y_j[i], i = \overline{1, N_j}\}$ соответственно. Второй множитель в (4) — вектор размерности $(q+1) \times 1$, и последний компонент — это вектор размерности $N_j \times 1$. Наилучшая модель определяется нахождением оптимальных значений коэффициентов $(\beta_0^{(k)}, \beta_1^{(k)}, \beta_2^{(k)})$, $k = 1, 2, 3$, в (4), для которых $\boldsymbol{\varepsilon}^{(1)} = |\mathbf{x}^* - \mathbf{x}|$, $\boldsymbol{\varepsilon}^{(2)} = |\mathbf{y}^* - \mathbf{y}|$, $\boldsymbol{\varepsilon}^{(3)} = |\mathbf{z}^* - \mathbf{z}|$ минимальны. Этот поиск требует некоторого ком-

промисса между точностью и эффективностью. С увеличением степени полинома точность приближения растет, но только до некоторой величины. В настоящей работе фиксируется $q = 2$, так как рост значения q не влияет на общую точность. Коэффициенты $(\beta_0^{(k)}, \beta_1^{(k)}, \beta_2^{(k)})$, $k = 1, 2, 3$, определяют начальное положение, скорость и ускорение объекта.

Анализ главных компонент (РСА) используется для уменьшения размерности набора данных перед кластеризацией с минимальными потерями информативности [6, 7]. РСА-анализ хорошо работает в случае данных с единым гауссовым распределением. Этот метод требует точной оценки ковариационной матрицы шума на основе исходных данных. Пусть $\mathbf{r}_j[i] = (x_j[i], y_j[i], z_j[i])^T$, тогда траектория T_j может быть представлена в виде набора этих векторов $T_j = \{\mathbf{r}_j[i], i = \overline{1, N_j}\}$. Затем из каждой точки траектории T_j вычитаются координаты среднего положения $\mathbf{m}_j$ этой траектории

$$\tilde{T}_j = \{\mathbf{r}_j[i] - \mathbf{m}_j, i = \overline{1, N_j}\}. \quad (5)$$

Матрица ковариации имеет вид

$$\Xi_j = \frac{1}{N_j} \tilde{T}_j \tilde{T}_j^T. \quad (6)$$

Разложение собственных чисел Ξ_j (6) дает собственные значения $\alpha = \{\alpha_i, i = \overline{1, N_j}\}$ и соответствующие собственные векторы $\Phi = \{\Phi_i, i = \overline{1, N_j}\}$. Упорядочив $\alpha_i, i = \overline{1, N_j}$, по убыванию, рассматривают первые два собственных вектора $\Phi_k, \Phi_l \in \Phi$, соответствующие двум наибольшему собственным значениям $\alpha_k, \alpha_l \in \alpha$, поскольку наибольшее изменение приходится как раз на эти две собственные компоненты [6, 7].

Для оценки резкости поворотов траектории T_j вычисляется гистограмма направленности (trajectory directional histogram) в каждой координатной плоскости. Эта характеристика описывает статистически направленное распределение траекторий. Гистограмма направленности вычисляется по методу, предложенному в работе [8]:

$$\mathbf{h}_j = H(\theta_j[i]), \quad (7)$$

где $H(\theta_j[i])$ — гистограмма углов направлений $\theta_j[i] = \arctg((y_j[i+1] - y_j[i]) / (x_j[i+1] - x_j[i]))$.

Аналогичные гистограммы вводятся для плоскостей (x, z) и (z, y) . Однако гистограмма направленности сама по себе недостаточна, так как не содержит пространственную информацию. Поэтому две траектории, которые разнесены в пространстве, попадут в один кластер, если их гистограммы на-

правленности подобны. Индексы трех самых больших значений $\mathbf{h}_j$ соответствуют доминирующим углам в траектории.

Таким образом, данные анализируемых траекторий преобразуются в векторы пространства характеристик. Затем в этих пространствах выполняется поиск кластеров. Поскольку нет никаких предварительных знаний о типе анализируемых траекторий, все их характеристики считаются равнозначными и при кластеризации имеют равные веса. В следующем разделе описывается процедура кластеризации.

Кластеризация

В настоящей работе пространства характеристик считаются независимыми, что упрощает совместный анализ неортогональных пространств характеристик с различными областями определения. Каждому пространству характеристик ставится в соответствие эмпирическое распределение вероятности векторов, представляющих траектории в этом пространстве [4]. Метод сдвига среднего поднимается по градиенту распределения вероятности, достигая ближайший пик распределения, т. е. доминантную моду [4, 5, 9]. После этого каждый вектор, представляющий траекторию, приписывается к ближайшей моде, т. е. к соответствующему кластеру.

Пусть $\mathbf{x}_j \in \Psi_m^{d_m}, j = \overline{1, J}$, является набором векторов, представляющих траектории $T_j, j = \overline{1, J}$, в m -м пространстве характеристик. Эмпирическое распределение вероятности векторов имеет вид

$$p(\mathbf{x}, h_m^{d_m}) = \frac{1}{J h_m^{d_m}} \sum_{j=1}^J K\left(\frac{\mathbf{x} - \mathbf{x}_j}{h_m^{d_m}}\right), \quad (8)$$

где $h_m^{d_m}$ — ширина полосы ядра $K(\cdot)$. Правильный

выбор ширины полосы $h_m^{d_m}$ важен при кластеризации методом сдвига среднего. Для этого используется итеративная процедура. Первоначальная ширина полосы составляет 10 % от каждой размерности m -го пространства характеристик и итеративно возрастает до 80 %. Нижняя граница отсекает кластеры из одного вектора (для единственной траектории), а верхняя граница исключает кластер, объ-

единяющий векторы $\mathbf{x}_j \in \Psi_m^{d_m}, j = \overline{1, J}$, для всех траекторий. Малая величина $h_m^{d_m}$ снижает отклонение оценки распределения (8), но увеличивает дисперсию. В работе [10] показано, что оптимальным ядром в формуле (8), обеспечивающим минимум средней интегральной квадратичной ошибки

(mean integrated square error), является радиально-симметричное ядро Епанечникова

$$K(\mathbf{x}) = \begin{cases} \frac{1}{2V_m^{d_m}}(d_m + 2)(1 - \|\mathbf{x}\|^2), & \|\mathbf{x}\| < 1, \\ 0, & \|\mathbf{x}\| \geq 1, \end{cases} \quad (9)$$

где $V_m^{d_m}$ представляет объем d_m -мерной единичной сферы. Градиент эмпирического распределения (8) имеет вид

$$\nabla p(\mathbf{x}) = \frac{d+2}{h_m^{d_m} V_m^{d_m}} \left(\frac{1}{J_c} \sum_{\mathbf{x}_j \in S(\mathbf{x})} (\mathbf{x} - \mathbf{x}_j) \right), \quad (10)$$

где $S(\mathbf{x})$ — гиперсфера радиуса $h_m^{d_m}$ с объемом $h_m^{d_m} V_m^{d_m}$ и центром в точке $\mathbf{x}$, содержащая J_c векторов [5]. В результате вектор сдвига среднего

$$\mathbf{M}_h(\mathbf{x}) = \frac{1}{J_c} \sum_{\mathbf{x}_j \in S(\mathbf{x})} (\mathbf{x} - \mathbf{x}_j)$$

имеет вид

$$\mathbf{M}_h(\mathbf{x}) = \frac{h_m^{d_m} V_m^{d_m}}{(d_m + 2)} \frac{\nabla p(\mathbf{x})}{p(\mathbf{x})}. \quad (11)$$

Результатом выполнения процедуры сдвига среднего является множество векторов, связанных с конкретной модой [9]. После выбора начального вектора, представляющего некоторую траекторию, процедура сдвига среднего сходится к локальной моде, и все векторы, ограниченные шириной полосы $h_m^{d_m}$ ядра $K(\cdot)$ в формуле (8), приписываются к этой моде. Малая ширина полосы способствует росту числа мод и большой дисперсии, которая означает неустойчивость оценки локального распределения (8). Этот недостаток устраняется путем объединения близко расположенных мод [11]. В настоящей работе смежные кластеры объединяются, если соответствующие им моды распределения находятся на расстоянии, меньшем $h_m^{d_m} + 0,1\%(h_m^{d_m})$.

Окончательная кластеризация траекторий движения

После объединения смежных кластеров в каждом пространстве характеристик окончательная кластеризация исходных траекторий обеспечивается совместным анализом кластеров во всех пространствах характеристик. Эта кластеризация проходит в три этапа: сначала оценивается среднее число кластеров, затем устанавливается соответствие между кластерами в различных пространствах

характеристик и в результате каждая траектория приписывается к конкретному кластеру.

Пусть множество $\{N_m, m = \overline{1, M}\}$ содержит числа кластеров для каждого из пространства характеристик $\{\Psi_m^{d_m}, m = \overline{1, M}\}$. Окончательное число кластеров

$$N = \frac{1}{M} \sum_{m=1}^M N_m.$$

После определения числа кластеров N оптимизируется их структура. Оптимизация кластеризации начинается с выбора пространств характеристик $\Psi_l^{d_l} \in \{\Psi_m^{d_m}, m = \overline{1, M}\}$, для которого $N_l = N$. Исходными кластерами считаются

кластеры $\{\mathbb{C}_i^l, i = \overline{1, N}\} \subset \Psi_l^{d_l}$. Затем ищется соответствие этих кластеров кластерам во всех остальных пространствах $\Psi_n^{d_n} \in \{\Psi_m^{d_m}, m = \overline{1, M}\}, n \neq l$. Если кластер $\mathbb{C}_j^n \subset \Psi_n^{d_n}$ имеет максимальное число перекрывающихся элементов с кластером $\mathbb{C}_i^l \subset \Psi_l^{d_l}$, т. е.

$$j = \arg \max_k |\mathbb{C}_i^l \cap \mathbb{C}_k^n|, \quad (12)$$

то кластер $\mathbb{C}_i^l$ корректируется согласно $\mathbb{C}_i^l = \mathbb{C}_i^l \cap \mathbb{C}_j^n$. Такая корректировка выполняется для всех кластеров $\{\mathbb{C}_i^l, i = \overline{1, N}\} \subset \Psi_l^{d_l}$, и в результате формируется N согласованных кластеров траекторий $\{\mathbb{C}_i^{final}, i = \overline{1, N}\}$. После этого к кластерам $\{\mathbb{C}_i^{final}, i = \overline{1, N}\}$ присоединяется множество траекторий $\{T' \subseteq \{T_j, j = \overline{1, J}\}\}$, которые не вошли в кластеры. Каждый кластер моделируется собственным распределением Гаусса с шириной полосы, определяемой дисперсией кластера. Условная вероятность того, что траектория $T_0 \in \{T'\}$ генерируется в соответствии с окончательной моделью кластеров, имеет вид

$$p(T_0 | \mathbb{C}_k^{final}) = \frac{1}{\sqrt{2\pi}\sigma_k^{final}} \exp \left(- \left(\frac{\mathbf{m}_{T_0} - \boldsymbol{\mu}_k^{final}}{\sigma_k^{final}} \right)^2 \right), \quad (13)$$

где $\mathbf{m}_{T_0}$ — среднее траектории T_0 ; $\boldsymbol{\mu}_k^{final} = |\mathbb{C}_k^{final}|^{-1} \times \sum_{j=1}^{|\mathbb{C}_k^{final}|} \mathbf{m}_j$ (со средними траекторий (3)) — среднее

для кластера $\mathbb{C}_k^{final}$; $\sigma_k^{final} = |\mathbb{C}_k^{final}|^{-1} \sum_{j=1}^{|\mathbb{C}_k^{final}|} \Xi_j$ (с мат-

рицами ковариации траекторий (6)) — стандартная дисперсия для кластера $\mathbb{C}_k^{final}$ [5]. Траектория T_0 приписывается к кластеру $\mathbb{C}_k^{final}$, если вероятность

$$p(T_0 | \mathbb{C}_k^{final}) > p(T_0 | \mathbb{C}_l^{final}), l = \overline{1, N}, l \neq k. \quad (14)$$

Описанный подход к кластеризации траекторий, использующий выделение кластеров в пространствах характеристик, позволяет группировать даже траектории, удаленные в декартовом трехмерном пространстве, и выявляет кластеры траекторий, представляющих паттерны движения.

Выявление посторонних траекторий

Существует два типа посторонних траекторий. Посторонние траектории первого типа располагаются в областях с высокой плотностью траекторий, но демонстрируют поведение, отличное от общего паттерна движения. Если траектория $T_j \in \mathbb{C}_k^{final}$ со средним $\mathbf{m}_{T_j}$, лежит далеко от центра μ_k^{final} кластера, к которому она принадлежит, она считается посторонней при условии

$$\frac{\mathbf{m}_{T_j} - \mu_k^{final}}{\sigma_k^{final}} > \tau, \quad (15)$$

где $\tau = 0,95$.

Второй тип посторонних траекторий располагается в районах с низкой плотностью траекторий. Если кластер содержит небольшое число траекторий и не может быть объединен с ближайшим кластером, то считается, что он состоит из посторонних траекторий. Пороговое значение для выделения кластера посторонних траекторий составляет 10 % от среднего числа элементов по всем кластерам.

Численный эксперимент

Описываемый в настоящей работе подход к кластеризации кривых тестируется на данных радары TRACON (Terminal Radar Approach Control), регистрирующего траектории полетов воздушных судов над заливом Сан-Франциско (данные находятся в открытом доступе на сайте <http://c3.nasa.gov/dash-link/resources/132/>). Подробное описание этих данных приводится в работах [9, 12]. В настоящей работе анализируются траектории 143 самолетов, совершивших посадку в течение одного дня наблюдения. Анализируются только траектории, содержащие от 150 до 200 точек наблюдения, регистрируемых через равные интервалы времени. Начало координат совпадает с положением радара.

На рис. 1 (см. вторую сторону обложки) показан исходный набор траекторий самолетов, обрабаты-

ваемых с помощью итеративного подхода к кластеризации траекторий.

После кластеризации траекторий (рис. 1, см. вторую сторону обложки) на первом этапе выделяются два кластера (показаны зеленым и розовым цветом на рис. 2, см. вторую сторону обложки) и посторонние траектории (показаны черным цветом на рис. 2). Как видно на рис. 2, в каждый из двух основных кластеров траектории объединяются не только по мере близости, но и по сходству формы. В работе [11] кластеризация аналогичных данных траекторий самолетов объединяет в кластеры пространственно близкие траектории. В настоящей работе объединение траекторий похожей формы в общие кластеры оказывается возможным, поскольку в двух из рассматриваемых пространств характеристик моделируется форма траектории (это является очевидным преимуществом применяемого подхода).

Траектории объектов, демонстрирующие движение, не характерное для первых двух кластеров на рис. 2 (см. вторую сторону обложки), ранее были отнесены к посторонним. Однако следующая итерация, состоящая в кластеризации посторонних траекторий на рис. 2, определяет два дополнительных кластера (рис. 3, см. вторую сторону обложки). Оставшиеся траектории (показаны черным цветом на рис. 3), которые не классифицированы при этой итерации, являются настоящими посторонними траекториями. На рис. 3 видны посадочные паттерны. Таким образом, применение итеративной процедуры кластеризации в пространствах характеристик позволяет выявить посадочные паттерны воздушных судов, соответствующие реальным посадочным полосам.

Заключение

Применение итеративного подхода к кластеризации траекторий управляемых объектов на основе представления траекторий в многомерных пространствах характеристик обеспечивает более полную кластеризацию этих траекторий, сокращает число посторонних траекторий и выявляет основные паттерны движения объектов.

Список литературы

1. Кухаренко Б. Г., Солнцева М. О. Кластеризация управляемых объектов на основе сходства их многомерных траекторий // Информационные технологии. 2014. № 5. С. 3—7.
2. Gaffney S., Smyth P. Joint probabilistic curve clustering and alignment / Saul L., Weiss Y., Bottou L., eds. // Proceedings of Neural Information Processing Systems (NIPS 2004). December 13—18, 2004, Vancouver, British Columbia, Canada. Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press. 2005. V 17. P. 473—480.
3. Anjum N., Cavallaro A. Trajectory clustering for scene context learning and outlier detection / Schonfeld D., Shan C., Tao D., Wang L., eds. // Video Search and Mining. Berlin. Studies in Computational Intelligence. Heidelberg: Springer. 2010. V. 287. P. 33—51.

4. **Fukunaga K., Hostetler L. D.** The estimation of the gradient of a density function, with applications in pattern recognition // *Transactions on Information Theory*. 1975. V. 21, N. 1. P. 32–40.
5. **Comaniciu D., Meer P.** Mean shift: A robust approach toward feature space analysis // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2002. V. 24, N. 5. P. 603–619.
6. **Jolliffe I. T.** *Principal Component Analysis*, 2nd ed. Springer Series in Statistics. New York: Springer. 2002.
7. **Yang K., Shahabi C.** A PCA-based similarity measure for multivariate time series // *Proceedings of the 2nd ACM International Workshop on Multimedia Databases*. New York: ACM Press. 2004. P. 65–74.
8. **Li X., Hu W.** A coarse-to-fine strategy for vehicle motion trajectory clustering // *Proceedings of 18th International Conference on Pattern Recognition (ICPR 2006)*. Washington, DC: IEEE Computer Society. 2006. V. 1. P. 591–594.
9. **Кухаренко Б. Г., Солнцева М. О.** Использование методов сокращения фона при сегментировании телеметрических изображений для идентификации групп объектов // *Информационные технологии*. 2014. N. 2. P. 3–8.
10. **Comaniciu D., Meer P.** Distribution free decomposition of multivariate data // *Pattern Analysis and Applications*. 1999. V. 2, N. 1. P. 22–30.
11. **Wang H., Suter D.** False-peaks-avoiding mean shift method for unsupervised peak-valley sliding image segmentation // *Proceedings of 7th International Conference on Digital Image Computing: Techniques and Applications (DICTA'03)*. 10–12 December 2003. Sydney, Australia: Macquarie University. 2003. P. 581–590.
12. **Gariel M., Srivastava A. N., Feron E.** Trajectory clustering and an application to airspace monitoring // *IEEE Transactions on Intelligent Transportation Systems*. 2011. V. 12. P. 1511–1524.

B. G. Kukharensko, Leader Research Scientist,
M. O. Solntseva, Graduate Student, e-mail: solntseva.chalei@gmail.com,
A. A. Blagonravova, Institute of Engineering, Russian Academy of Sciences

Iterative Clustering Trajectories of Objects under Control in Multi-Dimensional Characteristic Space

In present paper an iterative application of multi-dimensional trajectory clustering algorithm in abstract characteristic spaces is proposed for objects under control, which gives an opportunity to extract typical patterns of motion and isolate outliers. As trajectory features, spatial and angle characteristics are under consideration, such as average velocity, directional distance, trajectory mean, trajectory shape, principal components, and directional histogram. First vectors related to object trajectories are clustering sequentially in different characteristic spaces by means of mean-shift algorithm. Local maxima of each empirical density distribution are determined, which define the distribution modes related to trajectory clusters. Based on clustering results in all characteristic spaces, general trajectory clusters are extracted. An outlier set obtained at first iteration is an input to second iteration and so on. The iterations continue until the outlier number can't be changed. So, the stationary trajectory clustering result is obtained. The iterative clustering trajectories in abstract characteristic spaces is demonstrated by analysis of airplane trajectories in an airport zone.

Keywords: data mining, multi-dimensional trajectories, clustering

References

1. **Kukharensko B. G., Solntseva M. O.** Klasterizatsia upravlyayemykh ob'ektov na osnove skhodstva ikh mnogomernykh trayektoriy. *Informatsionnye tekhnologii*. 2014. N. 5. P. 3–7.
2. **Gaffney S., Smyth P.** Joint probabilistic curve clustering and alignment / Saul L., Weiss Y., Bottou L., eds. *Proceedings of Neural Information Processing Systems (NIPS 2004)*. December 13–18, 2004, Vancouver, British Columbia, Canada. *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press. 2005. V. 17. P. 473–480.
3. **Anjum N., Cavallaro A.** Trajectory clustering for scene context learning and outlier detection / Schonfeld D., Shan C., Tao D., Wang L., eds. *Video Search and Mining*. Berlin. *Studies in Computational Intelligence*. Heidelberg: Springer. 2010. V. 287. P. 33–51.
4. **Fukunaga K., Hostetler L. D.** The estimation of the gradient of a density function, with applications in pattern recognition. *Transactions on Information Theory*. 1975. V. 21, N. 1. P. 32–40.
5. **Comaniciu D., Meer P.** Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2002. V. 24, N. 5. P. 603–619.
6. **Jolliffe I. T.** *Principal Component Analysis*, 2nd ed. Springer Series in Statistics. New York: Springer. 2002.
7. **Yang K., Shahabi C.** A PCA-based similarity measure for multivariate time series. *Proceedings of the 2nd ACM International Workshop on Multimedia Databases*. New York: ACM Press. 2004. P. 65–74.
8. **Li X., Hu W.** A coarse-to-fine strategy for vehicle motion trajectory clustering. *Proceedings of 18th International Conference on Pattern Recognition (ICPR 2006)*. Washington, DC: IEEE Computer Society. 2006. V. 1. P. 591–594.
9. **Kukharensko B. G., Solntseva M. O.** Ispolzovanie metodov sokrashchenia fona pri segmentirovanii telemektricheskikh izobrazheniy dlya identifikatsii grupp ob'ektov. *Informatsionnye tekhnologii*. 2014. N. 2. P. 3–8.
10. **Comaniciu D., Meer P.** Distribution free decomposition of multivariate data. *Pattern Analysis and Applications*. 1999. V. 2, N. 1. P. 22–30.
11. **Wang H., Suter D.** False-peaks-avoiding mean shift method for unsupervised peak-valley sliding image segmentation. *Proceedings of 7th International Conference on Digital Image Computing: Techniques and Applications (DICTA'03)*. 10–12 December 2003. Sydney, Australia: Macquarie University. 2003. P. 581–590.
12. **Gariel M., Srivastava A. N., Feron E.** Trajectory clustering and an application to airspace monitoring. *IEEE Transactions on Intelligent Transportation Systems*. 2011. V. 12. P. 1511–1524.