

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

Том 21
2015
№ 2

ТЕОРЕТИЧЕСКИЙ И ПРИКЛАДНОЙ НАУЧНО-ТЕХНИЧЕСКИЙ ЖУРНАЛ

Издается с ноября 1995 г.

УЧРЕДИТЕЛЬ
Издательство "Новые технологии"

СОДЕРЖАНИЕ

ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ

- Кулагин В. П. Тензорные методы исследования структур сетей Петри 83
Сирота А. А., Цуриков А. В. Модели и алгоритмы классификации фрагментов текста и их применение для создания контентно-зависимых цифровых водяных знаков. 95

МОДЕЛИРОВАНИЕ И ОПТИМИЗАЦИЯ

- Кухаренко Б. Г., Солнцева М. О. Анализ результатов кластеризации многомерных траекторий посредством моделей линейных динамических систем 104
Казakov П. В. Использование дифференциальной эволюции при определении множества Парето генетическими алгоритмами многокритериальной оптимизации 109
Вдовин А. Н., Четырбоцкий А. Н., Четырбоцкий В. А. Компьютерное моделирование динамики роста рыб (на примере южного одноперого терпуга *Pleurogrammus azonus*). Часть I 116

ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ И СЕТИ

- Коваленко В. Б., Дордопуло А. И., Гудков В. А., Гуленок А. А., Сластен Л. М. Использование софт-архитектур при решении задач цифровой обработки сигналов на реконфигурируемых вычислительных системах 121

ПРОГРАММНАЯ ИНЖЕНЕРИЯ

- Посконин А. В. Интеграция SQL-ориентированных СУБД и NoSQL-систем на уровне объектного отображения 128
Димов Э. М., Маслов О. Н., Трошин Ю. В. Выбор средств программного обеспечения процесса статистического имитационного моделирования 132

ЦИФРОВАЯ ОБРАБОТКА СИГНАЛОВ И ИЗОБРАЖЕНИЙ

- Дворников С. В., Манаенко С. С., Дворников С. С., Погорелов А. А. Синтез фазоманипулированных вейвлет-сигналов 140

Журнал в журнале НЕЙРОСЕТЕВЫЕ ТЕХНОЛОГИИ

- Галушкин А. И. Мемристоры в развитии высокопроизводительной вычислительной техники 146
Федосов В. В., Федосова А. В. Использование нейросетей для графической оценки загрязнения территории выбросами группы источников 156

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.
РЯБОВ Г. 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.
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.
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.

МОДЕЛИРОВАНИЕ И ОПТИМИЗАЦИЯ MODELING AND OPTIMIZATION

УДК 519.233

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

Анализ результатов кластеризации многомерных траекторий посредством моделей линейных динамических систем

Модели линейных динамических систем используются для анализа результатов кластеризации траекторий объектов по методу полиномиальных регрессий. Преимуществом этих моделей является уменьшение размерности анализируемого пространства. Для проекций траекторий кластера по каждому измерению выделяется наиболее информативная составляющая (полиномиальная регрессия) и проявляется тонкая структура кластера. Эффективность моделей линейных динамических систем демонстрируется на примере анализа результатов кластеризации для траекторий движения самолетов в воздушном пространстве аэропорта.

Ключевые слова: анализ данных, многомерные траектории, кластеризация, полиномиальная регрессия, фильтр Калмана, сглаживатель Рауха, линейные динамические системы, алгоритм ожидания и максимизации правдоподобия

Введение

Задача кластеризации траекторий движения обусловлена необходимостью предсказания движения для организации взаимодействия управляемых объектов на основе выделяемых паттернов в траекториях движения. Одновременная кластеризация и выравнивание траекторий, т. е. векторов переменной длины, представляющих их координатные временные зависимости, по методу полиномиальной регрессии представлена в работах [1, 2].

В работе [1] каждый вектор $\mathbf{z}_i \in \mathbb{R}^{N_i \times 1}$, $i = \overline{1, q}$, (одномерный временной ряд) состоит из последовательности измерений координатной зависимости $z_i = z_i(t)$ в моменты времени $t_i \in \mathbb{R}^{N_i \times 1}$. Вектор \mathbf{z}_i моделируется полиномиальной регрессией:

$$\mathbf{z}_i = \mathbf{T}_i \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i, \quad (1)$$

где $\boldsymbol{\beta}$ — вектор коэффициентов регрессии размерности $(m + 1) \times 1$, $\boldsymbol{\varepsilon}_i$ — Гауссов шум с нулевым средним, а \mathbf{T}_i — регрессионная матрица, которая зависит от типа используемой регрессионной модели. Для полиномиальной регрессии \mathbf{T}_i имеет вид стандартной матрицы Вандермонта:

$$\mathbf{T}_i = \begin{bmatrix} 1 & t_i[1] & (t_i[1])^2 & \dots & (t_i[1])^m \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ 1 & t_i[N_i] & (t_i[N_i])^2 & \dots & (t_i[N_i])^m \end{bmatrix}. \quad (2)$$

В основе модели одновременной кластеризации и выравнивания лежит модель смеси регрессий, в которую вводятся четыре независимых параметра преобразований выравнивания и масштабирования во времени и пространстве $\{\Phi_i\} = \{a_i, b_i, c_i, d_i\}$ (параметры a_i и b_i описывают масштабирование и сдвиг во времени, а параметры c_i и d_i — масштабирование и смещение в пространстве измерений) [3]. Полиномиальная регрессия для одномерного случая имеет вид

$$\mathbf{z}_i = c_i \Upsilon_i \boldsymbol{\beta}_k + d_i + \boldsymbol{\varepsilon}_i, \quad (3)$$

где матрица Υ_i получается из \mathbf{T}_i (2) подстановкой $t_i \rightarrow a_i t_i - b_i$; параметры $\boldsymbol{\beta}_k$ определяют полиномиальную регрессию для траекторий из k -го кластера ($k = \overline{1, K}$); $\boldsymbol{\varepsilon}_i$ — Гауссов шум с нулевым средним и дисперсией $\sigma_k^2 \mathbf{I}$. Поэтому распределение плотности условной вероятности имеет вид

$$p_k(\mathbf{z}_i | a_i, b_i, c_i, d_i) = \mathcal{N}(\mathbf{z}_i | c_i \Upsilon_i \boldsymbol{\beta}_k + d_i, \sigma_k^2 \mathbf{I}).$$

Плотность вероятности для кривой \mathbf{z}_i однозначно задается соответствующим множеством параметров $\{\Phi_i\}$, которые подлежат определению. Задача кластеризации кривых решается как стандартная задача оценки значений скрытых переменных. Каждый из параметров преобразования в выражении (3) рассматривается как характерная для \mathbf{z}_i случайная переменная с заранее известным распределением вероятности для кластера. Параметры преобразования в (3) и параметры $\boldsymbol{\beta}_k$, $k = \overline{1, K}$ полиномиальной регрессии оцениваются одновременно посредством

алгоритма ожидания и максимизации правдоподобия (*Expectation Maximization*) (EM-алгоритма) [4]. Однако число определяемых кластеров K не может быть большим, и на практике $K \leq 10$. Поэтому для набора достаточно неоднородных траекторий, определяемые кластеры траекторий также будут неоднородными. Для неоднородного кластера полиномиальная регрессия является весьма общим представлением. Чтобы продемонстрировать неоднородность кластеров (а следовательно, недоклассификацию траекторий) требуются методы сокращения размерности. Поэтому для анализа проекций траекторий кластера по каждой из пространственных координат могут быть использованы модели линейных динамических систем ограниченной размерности. При этом для удобства описания вместо матрицы траекторий (измерений)

$$\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_q]$$

с векторами-столбцами $\mathbf{z}_i, i = \overline{1, q}$, (1) и (3) используется транспонированная матрица, т. е.

$$\mathbf{Z} = \mathbf{Z}^T \in \mathbb{R}^{q \times N}.$$

1. Модель линейной динамической системы

Пусть $\mathbf{X} = \mathbf{X}[\overline{1, p}; \overline{1, N}]$ — многомерный временной ряд для вектора состояния, имеющего размерность $p \times N$, а $\mathbf{Z} = \mathbf{Z}[\overline{1, q}; \overline{1, N}]$ — многомерный наблюдаемый временной ряд размерности $q \times N$, причем в общем случае $p < q$ (размерность линейной динамической системы меньше размерности измерений (траекторий)). Как шум состояния $\mathbf{U} = \mathbf{U}[\overline{1, p}; \overline{1, N}]$, так и шум измерений $\mathbf{W} = \mathbf{W}[\overline{1, q}; \overline{1, N}]$ — многомерные временные ряды для Гауссовых случайных переменных с нулевым средним и ковариационными матрицами \mathbf{Q} и \mathbf{R} , соответственно. Линейные инвариантные относительно времени динамические системы, называемые также линейными Гауссовыми моделями в пространстве состояний, описываются двумя уравнениями для векторов-столбцов $\mathbf{x}[k] = \mathbf{X}[\overline{1, p}; k]$, $\mathbf{z}[k] = \mathbf{Z}[\overline{1, q}; k]$, $\mathbf{u}[k] = \mathbf{U}[\overline{1, p}; k]$ и $\mathbf{w}[k] = \mathbf{W}[\overline{1, q}; k]$ (индекс k представляет дискретное время):

$$\mathbf{x}[k+1] = \mathbf{F}\mathbf{x}[k] + \mathbf{u}[k]; \quad (4)$$

$$\mathbf{z}[k] = \mathbf{G}\mathbf{x}[k] + \mathbf{w}[k], \quad (5)$$

где \mathbf{F} — матрица переходов; \mathbf{G} — матрица наблюдений. В теории линейной фильтрации временной ряд \mathbf{Z} для вектора наблюдений рассматривается, как зашумленный детерминированный временной ряд \mathbf{X} для вектора состояния. В теории Байесовской фильтрации отличающиеся одним временным шагом векторы-столбцы (значения переменных состояния в последовательные дискретные моменты времени), на основе уравнения (4) объединяются в Гауссову случайную переменную со статистиче-

скими характеристиками шума состояния \mathbf{u} [5]. Для Гауссовой случайной переменной \mathbf{w} аналогичная комбинация одновременных векторов-столбцов переменной состояния системы и переменной наблюдения формируется на основе уравнения (5) [5]. В результате условные распределения вероятности для векторов наблюдения и состояния линейной динамической системы имеют вид

$$P(\mathbf{x}[k]|\mathbf{x}[k-1]) = (2\pi)^{-p/2}|\mathbf{Q}|^{-1/2}\exp\left(-\frac{1}{2}(\mathbf{x}[k] - \mathbf{F}\mathbf{x}[k-1])^T\mathbf{Q}^{-1}(\mathbf{x}[k] - \mathbf{F}\mathbf{x}[k-1])\right); \quad (6)$$

$$P(\mathbf{z}[k]|\mathbf{x}[k]) = (2\pi)^{-q/2}|\mathbf{R}|^{-1/2}\exp\left(-\frac{1}{2}(\mathbf{z}[k] - \mathbf{G}\mathbf{x}[k])^T\mathbf{R}^{-1}(\mathbf{z}[k] - \mathbf{G}\mathbf{x}[k])\right), \quad (7)$$

где $|\dots|$ обозначает определитель матрицы. Как и распределения (6) и (7), исходное распределение вероятности состояний в момент времени $k = 1$ также предполагается Гауссовым со средним $\pi[1]$ и вариацией $\mathbf{V}[1]$:

$$P(\mathbf{x}[1]) = (2\pi)^{-p/2}|\mathbf{V}[1]|^{-1/2}\exp\left(-\frac{1}{2}(\mathbf{x}[1] - \pi[1])^T\mathbf{V}[1]^{-1}(\mathbf{x}[1] - \pi[1])\right). \quad (8)$$

В уравнении (4) состояние системы (линейно) зависит только от предыдущего состояния, отстоящего на один временной шаг (марковский процесс). Поэтому с учетом свойства марковости условных вероятностей формула для совместной вероятности $P(\mathbf{X}, \mathbf{Z})$ имеет вид

$$P(\mathbf{X}, \mathbf{Z}) = P(\mathbf{x}[1])\left(\prod_{k=2}^N P(\mathbf{x}[k]|\mathbf{x}[k-1])\right)\left(\prod_{k=1}^N P(\mathbf{z}[k]|\mathbf{x}[k])\right). \quad (9)$$

Из выражений (6)–(8) следует, что логарифм $F(\mathbf{X}, \mathbf{Z})$ (9) является квадратичной формой (постоянный член опущен):

$$\begin{aligned} \log(P(\mathbf{X}, \mathbf{Z})) = & -\frac{1}{2} \sum_{k=2}^N (\mathbf{x}[k] - \mathbf{F}\mathbf{x}[k-1])^T\mathbf{Q}^{-1}(\mathbf{x}[k] - \mathbf{F}\mathbf{x}[k-1]) - \frac{N-1}{2} \log(|\mathbf{Q}|) - \frac{1}{2} \sum_{k=1}^N (\mathbf{z}[k] - \mathbf{G}\mathbf{x}[k])^T\mathbf{R}^{-1}(\mathbf{z}[k] - \mathbf{G}\mathbf{x}[k]) - \frac{N}{2} \log(|\mathbf{R}|) - \\ & - \frac{1}{2}(\mathbf{x}[1] - \pi[1])^T\mathbf{V}[1]^{-1}(\mathbf{x}[1] - \pi[1]) - \frac{1}{2} \log(|\mathbf{V}[1]|). \end{aligned} \quad (10)$$

EM-алгоритм. В работах [7–10] описано использование алгоритма ожидания и максимизации правдоподобия (EM-алгоритма) для оценки параметров линейной динамической системы (4)–(5) на основе многомерного наблюдаемого временного ряда \mathbf{Z} . E -шаг описываемого EM-алгоритма состоит в вычислении условного среднего логарифма правдоподобия (10):

$$LL = E[\log(P(\mathbf{X}, \mathbf{Z}))|\mathbf{Z}]. \quad (11)$$

Из выражения (10) следует, что LL (11) зависит от трех типов условных средних (ожиданий), для которых используются следующие обозначения:

$$\langle \mathbf{x}[k] \rangle = E[\mathbf{x}[k]|\mathbf{z}]; \quad (12)$$

$$\mathbf{P}[k] = E[\mathbf{x}[k]\mathbf{x}[k]^T|\mathbf{z}]; \quad (13)$$

$$\mathbf{P}[k; k-1] = E[\mathbf{x}[k]\mathbf{x}[k-1]^T|\mathbf{z}]. \quad (14)$$

В момент времени k оценка состояния $\langle \mathbf{x}[k] \rangle$ (12) зависит от прошлых $\mathbf{z}[1, k-1]$ и будущих $\mathbf{z}[k+1, N]$ наблюдений [6]. Следовательно, она отличается от оценки, вычисленной посредством фильтра Калмана, который оценивает состояние только на основе прошлых наблюдений [5]. Вычисление условных средних (ожиданий) (12)—(14) на E -шаге EM -алгоритма приводится после описания M -шага, т. е. оценки параметров линейной динамической системы (4)—(5).

M -шаг. Параметры линейной динамической системы (4) и (5) — это матрицы \mathbf{F} и \mathbf{G} , а также статистические характеристики \mathbf{Q} и \mathbf{R} распределений (6) и (7), соответственно, и характеристики $\pi[1]$ и $\mathbf{V}[1]$ распределения (8). На основе оценок (12)—(14) и наблюдения $\mathbf{z}[1, N]$ каждый из параметров линейной динамической системы оценивается из условия равенства нулю соответствующей частной производной условного среднего логарифма правдоподобия LL (11). В результате

$$\mathbf{F} = \left(\sum_{k=1}^N \mathbf{P}[k; k-1] \right) \left(\sum_{k=2}^N \mathbf{P}[k-1] \right)^{-1}; \quad (15)$$

$$\mathbf{Q} = \frac{1}{N-1} \left(\sum_{k=2}^N \mathbf{P}[k] - \mathbf{F} \sum_{k=2}^N \mathbf{P}[k; k-1] \right); \quad (16)$$

$$\mathbf{G} = \left(\sum_{k=1}^N \mathbf{z}[k]\langle \mathbf{x}[k] \rangle^T \right) \left(\sum_{k=1}^N \mathbf{P}[k] \right)^{-1}; \quad (17)$$

$$\mathbf{R} = \frac{1}{N-1} \sum_{k=1}^N (\mathbf{z}[k]\mathbf{z}[k]^T - \mathbf{G}\langle \mathbf{x}[k] \rangle\mathbf{z}[k]^T); \quad (18)$$

$$\pi[1] = \langle \mathbf{x}[1] \rangle; \quad (19)$$

$$\mathbf{V}[1] = E[(\mathbf{x}[1] - \langle \mathbf{x}[1] \rangle)(\mathbf{x}[1] - \langle \mathbf{x}[1] \rangle)^T|\mathbf{z}] = \mathbf{P}[1] - \langle \mathbf{x}[1] \rangle\langle \mathbf{x}[1] \rangle^T. \quad (20)$$

E -шаг. Используются следующие обозначения: $\langle \mathbf{x}[k; l] \rangle = E[\mathbf{x}[k]|\mathbf{z}[1, l]]$ (таким образом, для условного среднего (12) имеем $\langle \mathbf{x}[k] \rangle \equiv \langle \mathbf{x}[k; N] \rangle$) и $\mathbf{V}[k; l] = E[(\mathbf{x}[k] - \langle \mathbf{x}[k] \rangle)(\mathbf{x}[k] - \langle \mathbf{x}[k] \rangle)^T|\mathbf{z}[1, l]]$. Используя оценки (15)—(20), сначала приводятся рекурсии вперед для линейного фильтра Калмана

$$\langle \mathbf{x}[k; k-1] \rangle = \mathbf{F}\langle \mathbf{x}[k-1; k-1] \rangle; \quad (21)$$

$$\mathbf{V}[k; k-1] = \mathbf{F}\mathbf{V}[k-1; k-1]\mathbf{F}^T + \mathbf{Q} \quad (22)$$

и вычисляются матрицы усиления Калмана

$$\mathbf{K}[k] = \mathbf{V}[k; k-1]\mathbf{G}^T(\mathbf{G}\mathbf{V}[k; k-1]\mathbf{G}^T + \mathbf{R})^{-1}; \quad (23)$$

$$\langle \mathbf{x}[k; k] \rangle = \langle \mathbf{x}[k; k-1] \rangle + \mathbf{K}[k](\langle \mathbf{y}[k] \rangle - \mathbf{G}\langle \mathbf{x}[k; k-1] \rangle), \quad (24)$$

$$\mathbf{V}[k; k] = \mathbf{V}[k; k-1] - \mathbf{K}[k]\mathbf{G}\mathbf{V}[k; k-1], \quad (25)$$

где $\langle \mathbf{x}[1] \rangle = \pi[1]$ из (19) и $\mathbf{V}[1]$ (20). Вывод рекурсии (21)—(25) основан на Байесовском подходе [10].

На основании работ [7—10] для вычисления оценки $\langle \mathbf{x}[k] \rangle \equiv \langle \mathbf{x}[k; N] \rangle$ (12) и условного среднего $\mathbf{P}[k] \equiv \mathbf{V}[k; N] + \langle \mathbf{x}[k; N] \rangle\langle \mathbf{x}[k; N] \rangle^T$ (13) выполняется рекурсия назад (здесь $\mathbf{J}[k]$ — вспомогательная матрица):

$$\mathbf{J}[k-1] = \mathbf{V}[k-1; k-1]\mathbf{F}^T(\mathbf{V}[k; k-1])^{-1}; \quad (26)$$

$$\langle \mathbf{x}[k-1; N] \rangle = \langle \mathbf{x}[k-1; k-1] \rangle + \mathbf{J}[k-1](\langle \mathbf{x}[k; N] \rangle - \mathbf{F}\langle \mathbf{x}[k-1; k-1] \rangle); \quad (27)$$

$$\mathbf{V}[k-1; N] = \mathbf{V}[k-1; k-1] + \mathbf{J}[k-1](\mathbf{V}[k; N] - \mathbf{V}[k; k-1])\mathbf{J}[k-1]^T. \quad (28)$$

Условное среднее (14) имеет вид

$$\mathbf{P}[k; k-1] \equiv \mathbf{V}[k; k-1; N] + \langle \mathbf{x}[k; N] \rangle\langle \mathbf{x}[k-1; N] \rangle^T,$$

где

$$\mathbf{V}[k; k-1; N] = E[(\mathbf{x}[k] - \langle \mathbf{x}[k] \rangle)(\mathbf{x}[k-1] - \langle \mathbf{x}[k-1] \rangle)^T|\mathbf{z}[1, N]].$$

Значения $\mathbf{V}[k; k-1; N]$ также вычисляются рекурсией назад

$$\mathbf{V}[k-1; k-2; N] = \mathbf{V}[k-1; k-1; N]\mathbf{J}[k-2]^T + \mathbf{J}[k-1](\mathbf{V}[k; k-1; N] - \mathbf{F}\mathbf{V}[k-1; k-1; N])\mathbf{J}[k-2]^T, \quad (29)$$

которая инициализируется как

$$\mathbf{V}[N; N-1; N] = (\mathbf{I} - \mathbf{K}[N]\mathbf{G})\mathbf{F}\mathbf{V}[N-1; N-1].$$

Итерации в виде последовательного чередования M - и E -шагов с оценкой изменения ожидаемого условного логарифма правдоподобия LL (11) обеспечивают оценку $\langle \mathbf{x}[1, N] \rangle$ (12) временного ряда состояния размерности $p \times N$.

Эффективность описанного EM -алгоритма демонстрируется на примере анализа результатов кластеризации траекторий по данным радаров международного аэропорта г. Сан-Франциско, находящихся в свободном доступе на сайте <https://c3.nasa.gov/daslilink/resources/132/>.

2. Численный эксперимент

В настоящей работе выполняется анализ неоднородности кластеров траекторий движения самолетов, полученных с помощью метода полиномиальных регрессий [1, 2]. Этот метод, несмотря на то, что является одним из надежных, так как учитывает реальную форму траекторий в пространстве, тем не менее, может приводить к довольно абстрактной линии тренда в неоднородных кластерах. Последующее применение моделей линейных динамических систем в случае неоднородных кластеров проявляет их тонкую структуру. Модели линейных динамических систем позволяют значительно сократить размерность многомерного временного ряда. Неоднородная структура кластера выявляется при рассмотрении проекций траекторий, соответствующих этому кластеру, на координатные оси

трехмерного пространства. Эффективность такого подхода демонстрируется на примере выборки траекторий первых 110 самолетов, идущих на посадку в международном аэропорту и зарегистрированных радаром TRACON 1 января 2006 года ([https:// c3.nasa.gov/das-MinK/resources/132/](https://c3.nasa.gov/das-MinK/resources/132/)). Начало координат (0,0,0) совпадает с положением радара, интервал времени между точками регистрации составляет около 5 с. В работе рассматриваются только 160 последних точек каждой траектории, чтобы исключить случайные маневры самолетов до начала захода на посадку.

На рис. 1 (см. вторую сторону обложки) показано трехмерное представление всех анализируемых траекторий самолетов.

Методом полиномиальной регрессии [1, 2] анализируемые траектории самолетов (см. рис. 1) разделяют на пять кластеров, как показано на рис. 2 (см. вторую сторону обложки): розовый — с 16 траекториями; зеленый — с 13 траекториями; синий — с 3 траекториями; черный — с 37 траекториями; красный — с 38 траекториями. Каждый кластер соответствует "посадочному" паттерну, который задается существующим маршрутом посадки.

Кроме того, в анализируемой выборке траекторий (см. рис. 1) выделяются три посторонние траектории (рис. 3), которые не могут быть отнесены ни к одному из выявленных кластеров.

На рис. 4 (см. вторую сторону обложки) показано изменение координат самолетов отдельно по каждой оси x , y , z в соответствии с последовательностью моментов времени k регистрации радаром. Линия тренда (полиномиальная регрессия), выделенная жирной линией, представляет обобщенную форму траекторий в кластере. Несмотря на то что снижение самолетов происходит неодновременно, сходство их траекторий обусловлено существованием типичных маршрутов посадки, например, когда самолеты выстраиваются в "караван", ожидая своей очереди приземления на посадочную полосу.

Для траекторий пяти выявленных кластеров, выделенных на рис. 2 и 4 (см. вторую сторону обложки) различными цветами, на рис. 5—9 показаны их соответствующие проекции на оси x и y вместе с компонентами вектора состояний линейной динамической системы размерности $p = 2$. Линия 1 на рис. 5—9 соответствует единственной компоненте вектора состояний при $p = 1$ и представляет полиномиальную регрессию на рис. 4. Линия 2 на рис. 5—9 соответствует второй компоненте вектора состояний при $p = 2$ и описывает тонкую структуру кластера. Проекция на оси x и y траекторий в кластерах

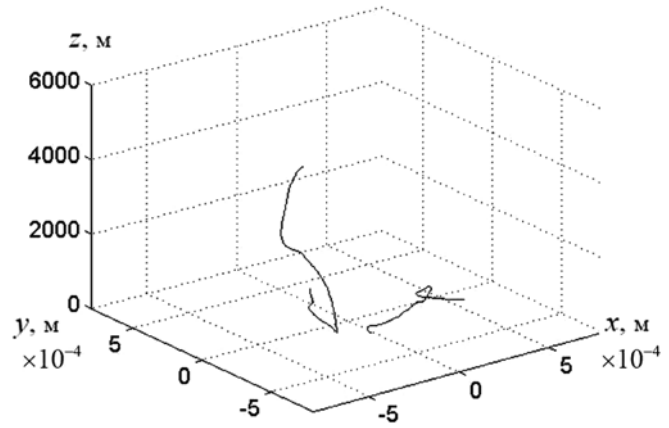


Рис. 3. Три посторонние траектории

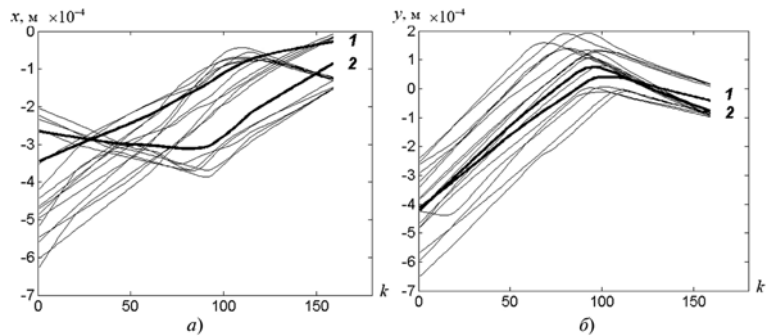


Рис. 5. Проекция на оси x и y траекторий розового кластера (см. рис. 2 и 4) показаны совместно с компонентами вектора состояний (линии 1 и 2): а — линейная модель проекций на ось x ; б — линейная модель проекций на ось y

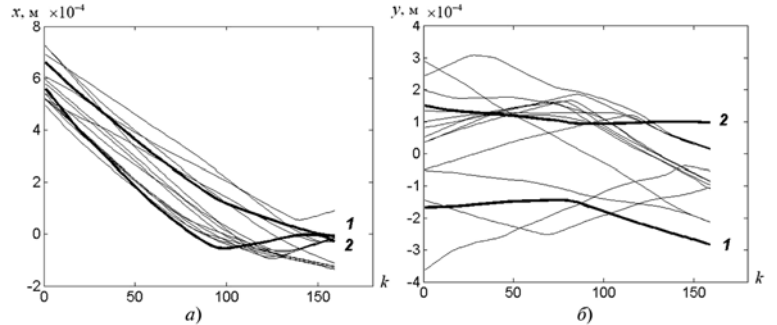


Рис. 6. Зеленый кластер на рис. 2 и 4: а — линейная модель проекций на ось x ; б — линейная модель проекций на ось y

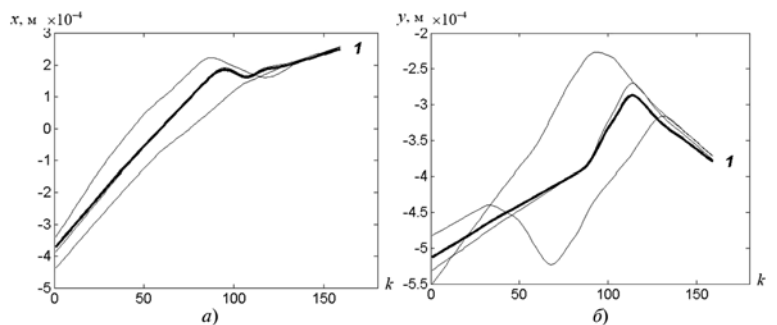


Рис. 7. Синий кластер на рис. 2 и 4: а — линейная модель проекций на ось x ; б — линейная модель проекций на ось y (определена одна компонента)

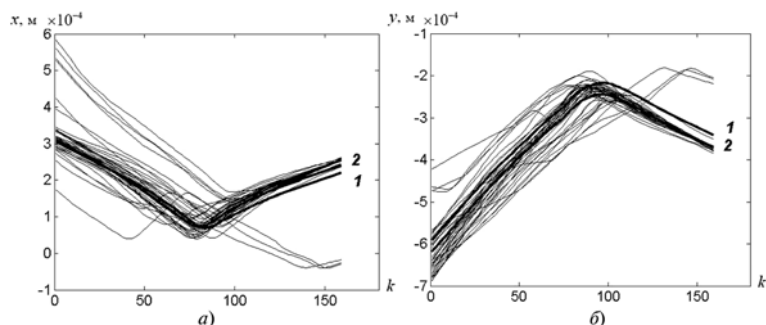


Рис. 8. Черный кластер на рис. 2 и 4:
 а — линейная модель проекций на ось x ; б — линейная модель проекций на ось y

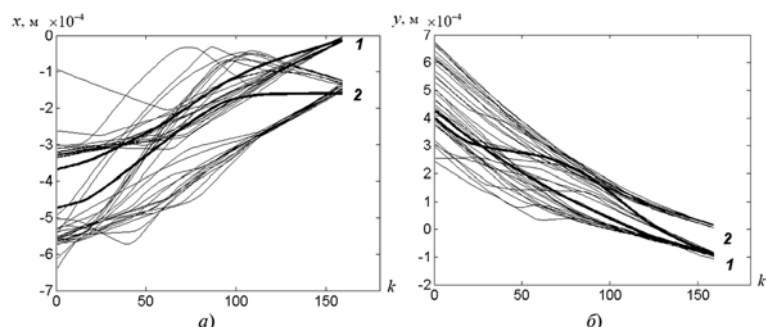


Рис. 9. Красный кластер на рис. 2 и 4:
 а — линейная модель проекций на ось x ; б — линейная модель проекций на ось y

на рис. 4 представляют переходные зависимости от времени и поэтому описываются линейной моделью. Проекция на ось z траекторий в кластерах (см. рис. 2 и 4) не позволяют дифференцировать эти кластеры, поэтому далее они не анализируются.

На рис. 5—9 видно, как в случае неоднородных кластеров модели линейных динамических систем выявляют тонкую структуру кластера, разделяя исходные кластеры на два подкластера.

Заключение

В настоящей работе метод линейных динамических систем демонстрируется на примере снижения размерности для результатов кластеризации траекторий самолетов, идущих на посадку в зоне крупного международного аэропорта. Эффективность метода линейных динамических систем достигается благодаря обучению параметров посредством EM-алгоритма.

Список литературы

1. Gaffney S., Smyth P. Joint probabilistic curve clustering and alignment / Saul L., Weiss Y., Bottou L., eds // Proc. of Neural Information Processing Systems (NIPS 2004). December 13–18, 2004. Vancouver, British Columbia, Canada. Advances in Neural Information Processing Systems. V. 17. Cambridge, MA: MIT Press, 2005. P. 473–480.
2. Кухаренко Б. Г., Солнцева М. О. Кластеризация управляемых объектов на основе сходства их многомерных траекторий // Информационные технологии. 2014. № 5. С. 3–7.
3. Gaffney S., Smyth P. Trajectory clustering with mixtures of regression models / Chaudhuri S., Madigan D., eds. // Proceedings of Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. August 15–18 1999. New York, NY: ACM Press, 1999. P. 63–72.
4. Dempster A. P., Laird N. M., Rubin D. B. Maximum likelihood from incomplete data via the EM algorithm // Proc. of the Royal Statistical Society. 1976. P. 1–38.
5. Simon D. Optimal State Estimation: Kalman, H_∞ and Nonlinear Approaches. Hoboken, New Jersey: John Wiley & Sons, Inc., 2006.
6. Rauch H. E. Solutions to the linear smoothing problem // IEEE Transactions on Automatic Control. 1963. V. 8. P. 371–372.
7. Shumway R. H., Stoffer D. S. Time Series Analysis and Its Applications. New York: Springer, 2011.
8. Shumway R. H., Stoffer D. S. An approach to time series smoothing and forecasting using the EM algorithm // Journal of Time Series Analysis. 1982. V. 3, N. 4. P. 253–264.
9. Shumway R., Stoffer D. Dynamic linear models with switching // Journal of the American Statistical Association. 1992. V. 86. P. 763–769.
10. Roweis S., Ghaliramani Z. A unifying review of linear Gaussian models // Neural Computation. 1999. V. 11, N. 2. P. 305–345.

B. G. Kukharenko, Leading research scientist, Blagonravov Institute of Engineering Science of RAS,
M. O. Solntseva, Post-graduate student, Moscow Institute of Physics and Technology (SU)

Analysis of Multi Dimensional Trajectory Clustering by Models of Linear Dynamical System

For clustering multidimensional trajectories polynomial regression method with parameter leaning by the Expectation-Maximization algorithm is in use. The method based on polynomial regression is characterized by the joint clustering and continuous alignment of curve sets in time and space. Nevertheless, a number of defined clusters can't be large enough. Thus, for a set of sufficiently heterogeneous trajectories, the defined clusters are heterogeneous also. For the inhomogeneous cluster, its polynomial regression is a too strong abstraction. To demonstrate the cluster heterogeneity (and, thus, non full clustering trajectories) a method of dimension reducing is in need. So, linear dynamical system models are applied to object multi dimensional trajectory clustering by polynomial regression method. An advantage of linear dynamical systems is reducing clustering result dimension. For trajectory coordinate projections in a cluster most informative component (polynomial regression) is extracted and the cluster fine structure is appeared. An efficiency of linear dynamical systems is demonstrated by example of clustering results of airplane flight tracks in an airport space.

Keywords: data mining, multi-dimensional trajectories, clustering, polynomial regression, Kalman filter, Rauch smoother, linear dynamical systems, Expectation-Maximization algorithm

References

1. **Gaffney S., Smyth P.** Joint probabilistic curve clustering and alignment / Saul L., Weiss Y., Bottou L. (eds.). *Proc. of Neural Information Processing Systems (NIPS 2004)*. December 13–18, 2004, Vancouver, British Columbia, Canada. Advances in Neural Information Processing Systems. V. 17. Cambridge, MA: MIT Press, 2005. P. 473–480.
2. **Kukhareno B. G., Solntseva M. O.** Klasterizacia upravlyayemykh objektov na osnove shodstva ih mnogomernykh trajektoriy. *Informacionnye tehnologii*. 2014, N. 5. P. 3–7.
3. **Gaffney S., Smyth P.** Trajectory clustering with mixtures of regression models / Chaudhuri S., Madigan D., eds. *Proceedings of Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. August 15–18, 1999. New York, NY: ACM Press, 1999. P. 63–72.
4. **Dempster A. P., Laird N. M., Rubin D. B.** Maximum likelihood from incomplete data via the EM algorithm. *Proceedings of the Royal Statistical Society*. 1976. P. 1–38.
5. **Simon D.** Optimal State Estimation: Kalman, H_∞ and Nonlinear Approaches. Hoboken, New Jersey: John Wiley & Sons, Inc., 2006.
6. **Rauch H. E.** Solutions to the linear smoothing problem. *IEEE Transactions on Automatic Control*. 1963. V. 8. P. 371–372.
7. **Shumway R. H., Stoffer D. S.** Time Series Analysis and Its Applications. New York: Springer, 2011.
8. **Shumway R. H., Stoffer D. S.** An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis*. 1982. V. 3, N. 4. P. 253–264.
9. **Shumway R., Stoffer D.** Dynamic linear models with switching. *Journal of the American Statistical Association*. 1992. V. 86. P. 763–769.
10. **Roweis S., Ghahramani Z.** A unifying review of linear Gaussian models. *Neural Computation*. 1999. V. 11, N. 2. P. 305–345.

УДК 004.89 + 004.021

П. В. Казаков, канд. техн. наук, доц., e-mail: pvk_mail@list.ru
Брянский государственный технический университет

Использование дифференциальной эволюции при определении множества Парето генетическими алгоритмами многокритериальной оптимизации

Рассматривается новый способ повышения эффективности работы генетических алгоритмов при определении множества Парето в задачах многокритериальной оптимизации. Он основан на использовании принципов дифференциальной эволюции при развитии популяции и формировании ее новых индивидов. Приводится сравнительный анализ эффективности использования дифференциальной эволюции при решении задач многокритериальной оптимизации разной сложности.

Ключевые слова: многокритериальная оптимизация, множество и граница Парето, многокритериальные генетические алгоритмы, дифференциальная эволюция

Введение

Необходимость решения задач многокритериальной оптимизации (МО) возникает в самых разных теоретических и прикладных областях. Качество решения задач МО зависит от точности найденного множества Парето [1, 2]. В то же время, его определение остается сложным и плохо автоматизированным процессом ввиду теоретически неограниченного числа недоминируемых решений, образующих множество Парето, а также вычислительно трудоемкой процедуры его построения. Применяемые для этих целей традиционные численные методы МО во многом ориентированы на преобразование задачи со множеством критериев к однокритериальной ее постановке. Подобные упрощения не только снижают точность найденных решений, но и существенно ограничивают возможность нахождения всего множества Парето. Поэтому разработка и совершенствование алгоритмов МО остаются актуальным направлением исследований.

В настоящее время одним из перспективных подходов к определению множества Парето в задачах МО является использование эволюционных методов, а именно специальных версий генетических

алгоритмов для многокритериальной оптимизации (МГА) [3–5]. Они относятся к многоточечным методам оптимизации, где каждое решение ассоциируется с так называемыми индивидами [6]. Вместе они образуют популяцию, итерационное изменение которой в процессе оптимизации позволяет эффективно исследовать пространства поиска (переменных и критериев). В то же время возможное многообразие постановок задачи МО, сложность анализа многомерных пространств переменных и критериев, вычислительная трудоемкость определения множества Парето не позволяют выбрать универсальный МГА. Так, исследование возможностей наиболее известных МГА показало снижение качества их результата при увеличении числа критериев в решаемых задачах МО [7]. Вместе с тем потенциал применения МГА для определения множества Парето остается достаточно большим, поэтому актуальным является разработка новых и совершенствование существующих МГА. Достигается это прежде всего интеграцией в МГА новых способов исследования пространства поиска, например, с использованием методов параллельных вычислений [8], коэволюционных технологий [9, 10] и др. В настоящей работе предлагается подход к развитию