УДК 519.246

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-Чалей, аспирант,
Московский физико-технический институт (ГУ), e-mail: solnceva.chalei@gmail.com

Применение моделей непрерывного профиля для анализа результатов кластеризации многомерных траекторий

Показано применение моделей непрерывного скрытого профиля для определения центральной траектории в достаточно однородных кластерах проекций многомерных траекторий. Модели непрерывного профиля являются скрытыми марковскими моделями с внутренними состояниями времени и масштаба. Определение центральной траектории кластера позволяет, используя оценку близости по мере косинуса, обнаружить потенциально посторонние траектории кластера, которые представляют движения, находящиеся в зоне риска.

Ключевые слова: *анализ данных, многомерные траектории, кластеризация, полиномиальная регрессия, модели непрерывного профиля, линейные динамические системы*

Введение

Настоящая работа посвящена определению центральных траекторий, представляющих "безопасные коридоры" для самолетов, идущих на посадку. Знание этих центральных траекторий способствует снижению рисков, возникающих из-за расположения взлетно-посадочной полосы в зоне сложного географического ландшафта в горных и прибрежных районах. В работах [1—2] анализируются (дискретные) многомерные траектории самолетов, совершивших посадку в международном аэропорту, находящемся рядом с заливом Сан-Франциско. С помощью метода полиномиальных регрессий, учитывающих сходство геометрической формы траекторий, анализируемые траектории разделяются на пять кластеров [1]. Однако вследствие того, что полиномиальная регрессия дает обобщенное представление о форме траекторий, полученные кластеры являются неоднородными.

Анализ этих кластеров на основе моделей линейных динамических систем определяет компоненты вектора состояний и обнаруживает тонкую структуру кластеров [2]. Компоненты вектора состояний позволяют провести расщепление неоднородного кластера на подкластеры, однако эти компоненты не являются центральными для новых подкластеров и не могут использоваться для обнаружения потенциально посторонних траекторий, которые представляют движения в зоне риска.

Для определения центральных траекторий для новых подкластеров может использоваться модель непрерывного (скрытого) профиля (*Continuous Profile Model* — CPM) [3—4], основанная на скрытой марковской модели (*Hidden Markov Model* — HMM) [5—6]. Модель непрерывного профиля предполагает, что каждый наблюдаемый временной ряд является зашумленной масштабированной выборкой из некоторого непрерывного скрытого профиля, или скрытой записи (*latent trace*). Наблюдаемые временные ряды генерируются при прохождении последовательности скрытых марковских состояний. CPM обучается с помощью алгоритма ожидания — максимизации правдоподобия (*Expectation-Maximization algorithm* — EM), который в контексте скрытых марковских моделей называется алгоритмом Баума — Уолша (*Baum — Welch algorithm*) [5—6]. В результате последовательных итераций обучения CPM определяется наиболее вероятная скрытая запись.

1. Модель непрерывного профиля

В CPM элементы каждого наблюдаемого временного ряда являются результатом эмиссии, т. е. генерируются непрерывным профилем, или скрытой записью при прохождении через последовательность скрытых состояний с конкретными индексами, как это имеет место в скрытой марковской модели [5—6]. Для учета изменения амплитуды в пределах конкретного временного ряда и между наблюдаемыми временными рядами к состояниям скрытого времени добавляются состояния скрытого масштаба относительно соответствующих элементов скрытой записи.

1.1. Скрытая марковская модель

Пусть $\mathbf{x}[i] = \{x[i;\, k],\ k = \overline{1, N}\}$, $i = \overline{1, K}$, — наблюдаемые временные ряды длиной N. Пусть набор возможных состояний НММ $\mathbf{S} = \{s[j],\ j = \overline{1, M}\}$ (M — число скрытых состояний НММ). Для наблюдаемого временного ряда $\mathbf{x}[i] = \{x[i;\, k],\ k = \overline{1, N}\}$ последовательность скрытых состояний времени $\boldsymbol{\varphi}[i] = \{\varphi[i;\, k] \in \mathbf{S},\ k = \overline{1, N}\}$. В марковской модели вероятности переходов между состояниями $s[m]$ и $s[j]$ — $\mathcal{T}_{s[m],\, s[j]} \equiv p(\varphi[i;\, k] = s[j] \| \varphi[i;\, k-1] = s[m])$ не зависят от номера i временного ряда. При условии, что в скрытой последовательности $\boldsymbol{\varphi}[i]$ в момент времени k имеется состояние $s[j]$, вероятность эмиссии $x[i;\, k]$ равна $\mathcal{A}^{i}_{k,\, s[j]} \equiv p(x[i;\, k] \| \varphi[i;\, k] = s[j])$. В логарифме правдоподобия $\mathcal{L}$ наблюдаемых временных рядов учитывается, что, поскольку состояния $\varphi[i;\, k]$ скрытые, вероятности переходов между состояниями $\mathcal{T}_{s[m],\, s[j]}$ и вероятности эмиссии $\mathcal{A}^{i}_{k,\, s[j]}$ следует суммировать по всем их возможным значениям:

$$\mathcal{L} = \log\left(\prod_{i=1}^{K} p(\mathbf{x}[i],\ |\{\mathcal{T}_{s[m],\, s[j]}\},\ \{\mathcal{A}^{i}_{k,\, s[j]}\})\right) = \sum_{i=1}^{K} \log\left(\sum_{\varphi[i]} \mathcal{T}_{0,\, \varphi[i;k]} \left(\prod_{k=1}^{N} \mathcal{A}^{i}_{k,\, \varphi[i;k]}\right) \times \left(\prod_{k=1}^{N} \mathcal{T}_{\varphi[i;k],\, \varphi[i;k-1]}\right)\right). \quad (1)$$

В формуле (1) правдоподобие для каждого наблюдаемого временного ряда факторизуется на три множителя, поскольку условные вероятности состояний в НММ зависят только от предыдущих состояний на одном временном шаге [5—6]. Первый член под логарифмом в формуле (1) — это вероятность начала ряда в конкретном скрытом состоянии, второй — вероятность эмиссии в каждом состоянии и третий — вероятность переходов между состояниями.

Поскольку рассматривается модель со скрытыми переменными, для оценки параметров $\{\mathcal{T}_{s[m],\, s[j]}\}$, $\{\mathcal{A}^{i}_{k,\, s[j]}\}$ в (1) используется ЕМ-алгоритм. Для этого необходимо вычислить предельные апостериорные вероятности каждого состояния $p(\varphi[i;\, k] \| \mathbf{x}[i])$ и определить вероятности переходов. Кроме этого, необходимо вычислить парные предельные апостериорные вероятности $p(\varphi[i;\, k],\ \varphi[i;\, k+1] \| \mathbf{x}[i])$. Их вычисление в НММ основано на алгоритме динамического программирования, в контексте этой модели называемого алгоритмом прямой и обратной рекурсии.

1.2. Алгоритм прямой и обратной рекурсии

Для K наблюдаемых временных рядов длиной N и числе скрытых состояний M алгоритм прямой и обратной рекурсии (*Forward-Backward algorithm*) выполняет вычисления с оценкой временной сложности $O(KMN)$ (если матрица переходов разреженная, временная оценка сложности будет меньше). Введем обозначение совместной вероятности $\alpha[i;\, k;\, j] \equiv p(x[i;\, 1],\ x[i;\, 2],\ \ldots,\ x[i;\, k],\ \varphi[i;\, k] = s[j])$. Для всех i и j после инициализации

$$\alpha[i;\, 1;\, j] \equiv \mathcal{A}^{i}_{1,\, s[j]}\, \mathcal{T}_{0,\, s[j]}$$

выполняется прямая рекурсия для $k = \overline{1, N}$

$$\alpha[i;\, k;\, j] \equiv \mathcal{A}^{i}_{k,\, s[j]} \sum_{m=1}^{M} \alpha[i;\, k-1;\, m]\, \mathcal{T}_{s[m],\, s[j]}. \quad (2)$$

Если рекурсия (2) завершена, то можно вычислить

$$p(\mathbf{x}[i]) = \sum_{j=1}^{M} p(x[i;\, 1],\ x[i;\, 2],\ \ldots,\ x[i;\, N],\ \varphi[i;\, N] = s[j]) = \sum_{j=1}^{M} \alpha[i;\, N;\, j]. \quad (3)$$

С учетом (3) логарифм правдоподобия $\mathcal{L}$ (1) определяется как

$$\mathcal{L} = \sum_{i=1}^{K} \log(p(\mathbf{x}[i])) = \sum_{i=1}^{K} \log\left(\sum_{j=1}^{M} \alpha[i;\, N;\, j]\right). \quad (4)$$

Для вычисления предельных апостериорных вероятностей (*marginals of the posterior*) $p(\varphi[i;\, k] \| \mathbf{x}[i])$ и $p(\varphi[i;\, k],\ \varphi[i;\, k+1] \| \mathbf{x}[i])$ используются формулы Байеса для условной вероятности:

$$p(\varphi[i;\, k] = s[j] \| \mathbf{x}[i]) = \frac{p(\mathbf{x}[i] | \varphi[i;k] = s[j])\, p(\varphi[i;k] = s[j])}{p(\mathbf{x}[i])} \equiv \frac{\alpha[i;k;j]\, p(x[i;k+1],\, \ldots,\, x[i] | \varphi[i;k] = s[j])}{p(\mathbf{x}[i])} \quad (5)$$

и

$$p(\varphi[i;\, k-1] = s[j],\ \varphi[i;\, k] = s[m] \| \mathbf{x}[i]) = \frac{\alpha[i;k-1;j]\, p(x[i;k],\, \ldots,\, x[i,\, N] | \varphi[i,\, k] = s[m])\, \mathcal{T}_{s[j],\, s[m]}}{p(\mathbf{x}[i])}. \quad (6)$$

Вычисление совместных вероятностей $\beta[i;\, k;\, j] \equiv p(x[i;\, k+1],\ x[i;\, k+2],\ \ldots,\ x[i;\, N],\ \varphi[i;\, k] = s[j])$

позволяет вычислить вероятности (5), (6). Для всех k и j выполняются инициализация

$$\beta[i;\ 1;\ j] = 1$$

и обратная рекурсия для $k = \overline{N-1;1}$

$$\beta[i;\ k;\ j] \equiv \sum_{m=1}^{M} \mathcal{T}_{s[j],\,s[m]}\, \mathcal{A}^{i}_{k+1,\,s[m]}\, \beta[i;\ k+1;\ m]. \quad (7)$$

Если $\alpha[i;\ k;\ j]$ и $\beta[i;\ k;\ j]$ определены в результате прямой (2) и обратной (7) рекурсий, то предельные апостериорные вероятности (5) и (6) вычисляются по формулам

$$p(\mathbf{x}[i]) = \sum_{j=1}^{M} \alpha[i;\ N;\ j] \equiv \sum_{j=1}^{M} \alpha[i;\ k;\ j]\beta[i;\ k;\ j]; \quad (8)$$

$$p(\varphi[i;\ k] = s[j]|\mathbf{x}[i]) = \frac{\alpha[i;k;j]\beta[i;k;j]}{p(\mathbf{x}[i])}; \quad (9)$$

$$p(\varphi[i;\ k-1] = s[j],\ \varphi[i;\ k] = s[m]|\mathbf{x}[i] = $$
$$= \frac{\alpha[i;k-1;j]\ \mathcal{T}_{s[j],\,s[m]}\mathcal{A}^{i}_{k,\,s[m]}\beta[i;k;m]}{p(\mathbf{x}[i])}. \quad (10)$$

Эффективное вычисление правдоподобия и предельных апостериорных вероятностей (9), (10) необходимо для оценки параметров в формуле (1) посредством EM-алгоритма. Апостериорная вероятность $p(\boldsymbol{\varphi}[i]|\mathbf{x}[i])$ оценивает, что именно эта последовательность скрытых состояний генерирует наблюдаемый временной ряд. Для оценки наиболее вероятной последовательности скрытых состояний на основе наблюдаемого временного ряда используется алгоритм Витерби [5, 6].

1.3. Модель непрерывного профиля

При обучении СРМ определяется непрерывный профиль, т. е. скрытая запись, вероятности переходов, управляющие марковской эволюцией состояний времени и масштаба, суммарный уровень шума наблюдаемого временного ряда и его глобальный фактор масштаба. После обучения скрытая запись

$$\mathbf{z} = \{z[k],\ k = \overline{1, M}\} \quad (11)$$

имеет более высокое разрешение по сравнению с ее зашумленными масштабированными копиями — наблюдаемыми временными рядами

$$\mathbf{x}[i] = \{x[i;\ k],\ k = \overline{1, N}\},\ i = \overline{1, K}.$$

В идеале $M \gg N$, тогда каждый элемент наблюдаемого временного ряда $\mathbf{x}[i]$, $i = \overline{1, K}$, в точности отображается в элемент скрытой записи $\mathbf{z}$ (11). Поскольку разрешение скрытой записи $\mathbf{z}$ (11) выше, чем у наблюдаемого временного ряда $\mathbf{x}[i]$, $i = \overline{1, K}$, его наблюдаемое время может эффективно ускоряться или замедляться при продвижении вдоль скрытой записи.

Неоднородность выборки элементов скрытой записи (11) и локальный масштаб, используемые при генерации элементов наблюдаемого временного ряда, определяются соответствующей ему последовательностью скрытых состояний СРМ. Каждое скрытое состояние представляет собой пару состояние масштаба/состояние времени

$$\varphi[i;\ k] = \{\tau[i;\ k],\ \chi[k]\},\ k = \overline{1, N}. \quad (12)$$

Состояния времени $\tau[i;\ k]$, $k = \overline{1, N}$, принадлежат последовательности натуральных чисел $\overline{1, M}$, представляющих скрытое время, которое индексирует скрытую запись (11). Состояния масштаба $\chi[k]$, $k = \overline{1, N}$, принадлежат упорядоченному набору $\overline{1, L}$. Распределение вероятности эмиссии элементов наблюдаемого временного ряда $\mathbf{x}[i] = x[i;\ k]$, $k = \overline{1, N}$, производимых последовательностью скрытых состояний $\varphi[i;\ k]$, $k = \overline{1, N}$, на основе скрытой записи $\mathbf{z} = z[k]$, $k = \overline{1, M}$, имеет вид

$$p(x[i;\ k]|\mathbf{z},\ \varphi[i;\ k],\ u[i],\ \sigma[i]) \equiv$$
$$\equiv \mathcal{N}(x[i;\ k]|u[i]z[\tau[i;\ k]]\chi[k],\ (\sigma[i])^2), \quad (13)$$

где $\mathcal{N}(x|z,\ \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-z)^2}{2\sigma^2}\right)$ — Гауссово нормальное распределение; $(\sigma[i])^2$ — это вариация шума в наблюдаемом временном ряду; $u[i]$ — глобальный параметр масштаба, уникальный для наблюдаемого временного ряда, который корректирует (глобальную) разницу масштаба временного ряда $\mathbf{x}[i]$ и скрытой записи $\mathbf{z}$ (11).

Для полного описания СРМ необходимо определить вероятности переходов между скрытыми состояниями. Поскольку вероятности переходов между состояниями времени и вероятности переходов между состояниями масштаба определяются отдельно, совместная вероятность переходов между скрытыми состояниями факторизуется следующим образом:

$$\mathcal{T}^{i}_{\varphi[i;k],\,\varphi[i;l]} \equiv p(\varphi[i;\ k]|\varphi[i;\ l]) =$$
$$= p(\tau[i;\ k]|\tau[i;\ l])p(\chi[k]|\chi[l]). \quad (14)$$

1.4. Логарифм правдоподобия набора наблюдаемых временных рядов

Логарифм правдоподобия $\mathcal{L}^p$ наблюдаемых временных рядов $\mathbf{x}[i] = \{x[i;\ k],\ k = \overline{1, N}\}$, $i = \overline{1, K}$, имеет вид $\mathcal{L}^p = \mathcal{L} + \mathcal{P}$, где $\mathcal{L}$ — это член правдоподобия, происходящий из HMM и состоящий из вероятностей эмиссии (13) и вероятностей переходов между скрытыми состояниями (14)

$$\mathcal{L} = \sum_{i=1}^{K} \log\left(\sum_{\varphi[i]} p(\varphi[i;\ 1]) \times\right.$$

$$\times \left(\prod_{i=1}^{N} \mathcal{N}(x[i;\ k] | u[i] z[\tau[i;\ k]]\chi[k],\ (\sigma[i])^2)\right) \times$$

$$\left.\times \left(\prod_{i=2}^{N} \mathcal{T}^{i}_{\varphi[i;k-1],\ \varphi[i;k]}\right)\right), \qquad (15)$$

где $p(\varphi[i;\ 1])$ — априорные вероятности начальных скрытых состояний, а $\mathcal{P}$ — это логарифм априорной вероятности или штрафной член, обеспечивающий регуляризацию СРМ [4, 5]. Максимум $\mathcal{L}^p$ определяется с помощью ЕМ-алгоритма.

1.5. Оценка предельных апостериорных вероятностей (Е-шаг)

На Е-шаге (ЕМ-алгоритма) предельные апостериорные вероятности $p(\varphi[i;\ k] = s|\mathbf{x}[i])$ и $p(\varphi[i;\ k-1] = s,\ \varphi[i;\ k] = s'|\mathbf{x}[i])$ СРМ вычисляются посредством алгоритма прямой и обратной рекурсии, как в НММ (подраздел 1.2). Ввиду разреженности матрицы переходов между скрытыми состояниями (13) вычислительная сложность этого алгоритма линейная по числу скрытых состояний (12) [4, 5].

1.6. Оценка параметров СРМ (М-шаг)

Оценки параметров на М-шаге, которые максимизируют ожидаемый логарифм правдоподобия

$$\langle \mathcal{L}^{p}_{\mathrm{comp}} \rangle \equiv \langle \mathcal{P} \rangle + \sum_{i=1}^{K} \langle \log(p(\boldsymbol{\varphi}[i],\ \mathbf{x}[i])) \rangle, \qquad (16)$$

представляются аналитическими формулами [4, 5]. Для оценки элементов скрытой записи (11) используется численная процедура [4, 5]. Производная $\langle \mathcal{L}^{p}_{\mathrm{comp}} \rangle$ (16) по $z[k]$ (элементу скрытой записи) имеет вид

$$\frac{\partial \langle \mathcal{L}^{p}_{\mathrm{comp}} \rangle}{\partial (z[k'])} = \frac{\partial}{\partial (z[k'])} \sum_{s=1}^{S} \sum_{k=1}^{N} p(\varphi[i;\ k] = s|\mathbf{x}[i]) \times$$

$$\times \log(\mathcal{N}(x[i;\ k] | u[i] z[\tau[s]]\chi[s],\ (\sigma[i])^2)) - \lambda \bar{u} \frac{\partial}{\partial (z[k'])} \times$$

$$\times \sum_{k=1}^{M-1} (z[k+1] - z[k])^2 =$$

$$= -\sum_{\{s|\tau[s] = k'\}} \sum_{k=1}^{N} p(\varphi[i;\ k] = s|\mathbf{x}[i]) \frac{\partial}{\partial (z[k'])} \times$$

$$\times \frac{(x[i;k] - u[i] z[\tau[s]]\chi[s])^2}{2(\sigma[i])^2} - \lambda \bar{u} (2(z[k'] - z[k'-1]) -$$

$$- 2(z[k'+1] - z[k'])) =$$

$$= -\sum_{\{s|\tau[s] = k'\}} \sum_{k=1}^{N} p(\varphi[i;\ k] = s|\mathbf{x}[i]) u[i]\chi[s] \times$$

$$\times \frac{(x[i;k] - u[i] z[\tau[s]]\chi[s])^2}{2(\sigma[i])^2} -$$

$$- \lambda \bar{u} (4z[k'] - 2z[k'-1] - 2z[k'+1]), \qquad (17)$$

где λ — коэффициент при штрафном члене, обеспечивающем гладкость скрытой записи (11).

Таким образом, при $\lambda \neq 0$ условие $\frac{\partial \langle \mathcal{L}^{p}_{\mathrm{comp}} \rangle}{\partial (z[k])} = 0$, $k = \overline{1, M}$, приводит к трехдиагональной системе уравнений (17).

2. Численный эксперимент

В работе [1] представлены результаты кластеризации 112 (трехмерных) посадочных траекторий самолетов на пять кластеров. Эта исходная кластеризация выполняется по методу полиномиальных регрессий и приводит к слишком обобщенному представлению формы траекторий, поэтому в полученных кластерах присутствует неоднородность. Анализ проекций траекторий этих кластеров на основе моделей линейных динамических систем определяет компоненты вектора состояний и обнаруживает тонкую структуру кластеров [2]. В работе [2] эта тонкая структура кластеров показана в проекциях траекторий на координатные оси (ниже рассматриваются только проекции на ось x). На рис. 1 (см. третью сторону обложки) показаны проекции на ось x 16 многомерных траекторий (длина траекторий $N = 156$), входящих в один из исходных кластеров, полученных по методу полиномиальных регрессий (розовый кластер в [1]). Жирные линии *1* и *2* на рис. 1 представляют две компоненты вектора состояний линейной динамической системы, выявляющие тонкую структуру кластера [2]. Приписывание проекций траекторий $[x_i[k],\ k = \overline{1, 156}]$, $i = \overline{1, 16}$, к одному из подкластеров *1* (красному) или *2* (синему) на рис. 1 выполняется в соответствии с их максимальной близостью к одной из компонент вектора состояний $[y_j[k],\ k = \overline{1, 156}]$, $j = \overline{1, 2}$, с использованием меры косинуса

$$R = \left(\sum_{k=1}^{156} x_i[k] y_j[k]\right) \Big/ \sqrt{\sum_{k=1}^{156} (x_i[k])^2} \sqrt{\sum_{k=1}^{156} (y_j[k])^2},$$

которая учитывает разницу в направлении векторов в пространстве состояний (безотносительно Евклидова расстояния между ними).

В настоящем разделе при анализе тонкой структуры кластера траекторий (рис. 1) в результате применения СРМ в каждом подкластере определяется (согласующаяся с большинством) обобщенная форма проекций траекторий.

На рис. 2 (см. третью сторону обложки) показаны проекции траекторий красного подкластера и жирной кривой *1* соответствующая компонента вектора состояний (компонента *1* на рис. 1). На рис. 3 (см. третью сторону обложки) показаны проекции траекторий синего подкластера и жирной кривой *1* соответствующая компонента вектора состояний (компонента *2* на рис. 1). Штриховые жирные линии *2* на рис. 2 и 3 показывают скрытые записи (непрерывные скрытые профили) (11) соответствующих CPM. Непрерывные профили на рис. 2 и 3 отображают проекции траекторий подкластеров по принципу голосования большинства [7]. Непрерывный профиль *2* на рис. 2 отражает ход проекций семи траекторий из 11. Проекция траектории, показанная на рис. 2 зеленым цветом, наиболее удалена (по мере косинуса) от непрерывного профиля *2*.

На рис. 3 непрерывный профиль *2* отражает ход проекций четырех траекторий, при этом одна траектория игнорируется. Эта наиболее удаленная (по мере косинуса) от непрерывного профиля *2* проекция траектории показана голубым цветом на рис. 3.

Заключение

В настоящей работе показано применение CPM для описания обобщенной формы проекций траекторий в однородных кластерах. Непрерывный скрытый профиль достаточно точно отражает форму большинства траекторий в кластере, представляя, таким образом, центральную траекторию этого кластера. В результате определения такой центральной траектории кластера можно улучшить однородность структуры кластера и обнаружить потенциально посторонние траектории кластера, которые представляют движения, находящиеся в зоне риска.

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. **Кухаренко Б. Г., Солнцева М. О.** Анализ результатов кластеризации многомерных траекторий посредством моделей линейных динамических систем // Информационные технологии. 2015. № 2. С. 104—109.
3. **Listgarten J., Neal R. M., Roweis S. T., Emili A.** Multiple alignment of continuous time series / Saul L. K., 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. 5—13.
4. **Listgarten J.** Analysis of sibling time series data: Alignment and difference detection. PhD thesis. University of Toronto: Graduate Department of Computer Science. 2007.
5. **Кухаренко Б. Г.** Анализ независимых компонент и скрытая Марковская модель для определения доминантных компонент многомерных временных рядов // Информационные технологии. 2010. № 11. Приложение. С. 1—32.
6. **Poritz A. B.** Hidden Markov models: A guided tour // Proceedings of the IEEE Conference on Acoustics, Speech and Signal Processing (ICASSP). Morgan Kaufmann. 1988. P. 7—13.
7. **Dasgupta P., Maskin E.** On the robustness of majority rule // Journal of the European Economic Association. 2008. Vol. 6, N 5. P. 949—973.

B. G. Kukharenko[1], Leading Research Scientist, e-mail: kukharenkobg@hotmail.com,
M. O. Solntseva-Chaley[2], Postgraduate Studtent, e-mail: solnceva.chalei@gmail.com
[1]Blagonravov Institute of Engineering Science of RAS
[2]Moscow Institute of Physics and Technology (SU)

Applying Continuous Profile Models for Analysis of Multidimensional Trajectory Clustering Results

Problem of clustering trajectories is pre-conditioned by a need to organize motion of objects under control. The polynomial regression method is the best approach to trajectory cluster selection, which estimates a form of general trajectory of each cluster. 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 the polynomial regression method. Extracted components of system state vector give an opportunity to elicit cluster fine structure in accordance with cosine measure. In present paper applying continuous latent profile models is shown for determining a central trajectory in sufficiently homogeneous clusters of multi-dimensional trajectory projections. Continuous profile models are the hidden Markov models with internal states of time and scale. Defining the cluster central trajectory with affinity estimation by cosine measure in use gives an opportunity to determine the cluster outlying trajectories in potential, which represent motions located in a risk zone

Keywords: *data mining, multi-dimensional trajectories, clustering, polynomial regression, continuous profile models, linear dynamical systems*

References

1. **Kukharenko B. G., Solntseva M. O.** Klasterizacia upravlyaemyh objektov na osnove shodstva ih mnogomernyh trajektoryi. *Informacionnye tehnologii.* 2014, no. 5, pp. 3—7.
2. **Kukharenko B. G., Solntseva M. O.** Analiz resultatov klasterizacii mnogomernyh trajektoryi posredstvom modelei lineinyh dinamicheskih sistem. *Informacionnye tehnologii.* 2015, no. 2, pp. 104—109.
3. **Listgarten J., Neal R. M., Roweis S. T., Emili A.** Mlultiple alignment of continuous time series / Saul L. K., 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, vol. 17, pp. 5—13.
4. **Listgarten J.** *Analysis of sibling time series data: Alignment and difference detection. PhD thesis.* University of Toronto: Graduate Department of Computer Science. 2007.
5. **Kukharenko B. G.** Analiz nezavisimyh komponent i skrytaja Markovskaja model dlja opredelenija dominantnyh komponent mnogomernyh vremennyh rjadov. *Informacionnye tehnologii.* 2010, no. 11. Prilogenie, pp. 1—32.
6. **Poritz A. B.** Hidden Markov models: A guided tour. *Proceedingsof the IEEE Conference on Acoustics, Speech and Signal Processing (ICASSP).* Morgan Kaufmann. 1988, pp. 7—13.
7. **Dasgupta P., Maskin E.** On the robustness of majority rule // *Journal of the European Economic Association*. 2008, vol. 6, no. 5, pp. 949—973.

УДК 621.391

А. В. Волынская, канд. техн. наук, доц., e-mail: anna-volinskaya@mail.ru,
Уральский государственный университет путей сообщения, г. Екатеринбург,
И. В. Сапожников, аспирант, fbjumper@mail.ru,
Уральский технический институт связи и информатики, г. Екатеринбург

Моделирование линейного мультипараметрического модема для каналов передачи информации с низким отношением сигнал/помеха

Приведены результаты компьютерного моделирования алгоритма мультипараметрической модуляции и демодуляции с оценкой его помехоустойчивости путем корреляционного анализа принятых сигналов.

Ключевые слова: *модуляция, помехоустойчивость, свертка, двоичные сигналы, сложные шумоподобные сигналы, коды Баркера, корреляция*

Введение

Существуют системы передачи информации, в которых отношение сигнал/помеха (с/п) на входе приемника меньше единицы. К ним относятся, например, системы беспроводной аварийной сигнализации в шахтах, PLC-R каналы (связь с электровозом по контактному проводу), военные системы связи, работающие в условиях радиопротиводействия. В этих системах применение классических видов модуляции затруднено ввиду так называемого порогового эффекта помехоустойчивости, когда отношение сигнал/помеха на входе приемника снижается до единицы, а на выходе — лавинообразно падает [1]. Причина этого явления — нелинейность модуляции, возникающая в результате перемножения сигналов. Кроме того, модуляции подвергается, как правило, один параметр простого гармонического колебания, максимум — два (в комбинированных видах модуляции). Теоретическая возможность применения более сложной шумоподобной несущей обсуждается с середины прошлого века, но только в варианте однопараметрической модуляции, например дисперсии или средней частоты [2]. В начале 80-х годов А. И. Самойловым предложены алгоритм и варианты технической реализации мультипараметрической линейной модуляции шумоподобной несущей путем ее свертки (*convolution*) с передаваемым сигналом и развертки (*deconvolution*) на приемном конце [3]. Назовем такой модем линейным мультипараметрическим модемом (ЛММ). Преимуществом ЛММ является его способность работать при наличии шумов с адаптацией спектра передаваемого сигнала к спектру помех. В упрощенном виде эта технология реализована в стандарте IEEE 802.22 (когнитивное радио) [4] и PLC-каналах передачи сигналов по проводам электроснабжения [5]. Идея общая — энергия сигнала перераспределяется на участки спектра, где минимальна энергия помехи.

Основные положения

На рис. 1 представлены схема и временные диаграммы, поясняющие суть ЛММ, который работает следующим образом. Пусть передаваемое сообщение $s(t)$, имеющее ширину спектра F_s и длительность T_s, разбивается на фрагменты по $2F_sT_s$ отсчетных значений, которые записываются в регистр; шумоподобное несущее колебание $y(t)$ имеет период $2F_sT_s$ отсчетных значений и продвигается по цифровой линии задержки, при этом каждое его