

МЕТОД КЛАССИФИКАЦИИ СИГНАЛОВ НА ОСНОВЕ СПЕКТРАЛЬНОГО АНАЛИЗА ГРАФА КЛАСТЕРНОГО РАЗБИЕНИЯ

С. С. Юлин^а, инженер-программист

И. Н. Паламарь^б, канд. техн. наук, профессор

^аОАО «КБ «ЛУЧ», Рыбинск, РФ

^бРыбинский государственный авиационный технический университет им. П. А. Соловьева, Рыбинск, РФ

Постановка проблемы: большинство методов классификации сигналов учитывают внутреннюю структуру объектов, основываясь только на марковских предположениях, что приводит к потере значительной части дискриминантной информации, содержащейся в сигнале. Целью данной работы является улучшение качества работы методов классификации сигналов за счет анализа информации о структуре их кластерного разбиения. **Результаты:** предложен метод классификации на основе спектрального анализа графов, формируемых по обучаемой с учителем модели, и разработан способ классификации сигналов на основе анализа структуры кластеров, получаемых в результате отображения сигнала на нелинейное главное многообразие малой размерности с помощью алгоритма формирования самоорганизующейся карты Кохонена. Полученное кластерное разбиение представляется в виде графа. Определены граф данных и граф модели, задаваемые на топологической, упорядоченной структуре карты Кохонена. Целочисленная мера сходства графа данных и графа модели вычисляется с использованием методов спектральной теории графов. Экспериментально доказано улучшение качества классификации при объединении предложенного в работе метода с широко известными методами, такими как HMM и HCRF, а также с ранее разработанным нами методом NPM-PGM. **Практическая значимость:** разработанный метод может быть использован для решения задач распознавания сигналов (последовательностей наблюдений), таких как, например, распознавание рукописных символов, написанных стилусом на планшете, и распознавание действий человека по данным акселерометра наручного браслета.

Ключевые слова — классификация сигналов, самоорганизующаяся карта Кохонена, спектральная теория графов, сравнение графов.

Введение

Большинство методов классификации временных последовательностей (сигналов), например скрытые марковские модели (HMM — *Hidden Markov Model*) [1] или условные случайные поля со скрытыми состояниями (HCRF — *Hidden Conditional Random Fields*) [2], учитывают внутреннюю структуру объектов, основываясь только на марковских предположениях [3]. Это приводит к потере значительной части дискриминантной информации, содержащейся в сигнале, и, как следствие, к снижению качества классификации. В связи с этим целью данной работы является повышение качества классификации сигналов путем извлечения информации о структуре сигнала на основе анализа графа кластерного разбиения данного сигнала.

Многие проблемы, представляющие интерес в машинном обучении, могут быть сформулированы как задачи нахождения соответствия между двумя наборами точек [4]. Поскольку набор точек может иметь внутреннюю структуру, то он часто рассматривается не просто как множество точек, а как граф. В этих условиях вершины графа представляют собой характерные точки, извлеченные из каждого экземпляра образа (например, тестового и эталонного), а ребра пред-

ставляют связи между характерными точками. Следовательно, задача сопоставления наборов точек сводится к проблеме сравнения графов, состоящей в нахождении соответствия между двумя наборами вершин с идентичной смежностью. Из-за своей комбинаторной природы эта задача решается либо точно с введением ограничений, либо приблизительно. Большая часть современной литературы о сравнении графов посвящена разработке приближенных схем решения задачи сравнения графов.

Как правило, в задаче распознавания образов можно выделить два графа: G_M — граф модели и G_D — граф данных. Процедура сравнения этих графов выполняется в целях определения того, какой модели (классу образов) соответствуют те или иные данные. Формально задача сравнения графов формулируется так: дано два графа $G_M = (V_M, E_M, A_M)$ и $G_D = (V_D, E_D, A_D)$, где V — множество вершин; E — множество ребер; A — множество атрибутов ребер, и выполняется условие $V_M \neq V_D$. Найти отображение $f: V_M \rightarrow V_D$ такое, что любые две вершины u и v графа G_D смежны тогда и только тогда, когда смежны вершины $f(u)$ и $f(v)$ графа G_M [5].

Широко используются два подхода к сравнению графов: на основе спектральной теории графов [6, 7] и полуопределенного программиро-

вания. В данной работе используется метод спектрального сравнения графов [8]. Спектр графа представляет собой множество всех собственных чисел матрицы смежности. Вычисление спектра позволяет получить информацию о вершине (охарактеризовать вершину) по инцидентным ей ребрам. Несмотря на то, что спектр инвариантен не только при преобразовании посредством групп перестановок, но и под действием групп всех ортогональных преобразований, спектральные методы сравнения графов получили широкое распространение, особенно при сравнении неориентированных графов, так как найти пару изоспектральных неизоморфных неориентированных графов является довольно трудной задачей [9].

Очевидно, что не в любых данных легко отыскать характерные точки, способные являться вершинами графа. В настоящей работе предлагается универсальный способ формирования характерных точек путем формирования кластерного разбиения классифицируемого образа. Группа кластеров и связей между ними формируется путем отображения обучающих данных на нелинейное главное многообразие малой размерности, формируемое методом самоорганизующейся карты Кохонена [10, 11]. Определим граф кластерного разбиения (граф самоорганизующейся карты Кохонена) как граф $G_M = (V_M, E_M, A_M)$, где V_M — множество узлов карты Кохонена; E_M — множество связей между узлами, задаваемое топологией карты; $A_M = \{A_{ij}\}$ — множество атрибутов связей, значение которых равно евклидову расстоянию между весами двух смежных узлов карты i и j .

Решение задачи сравнения графов подразумевает формирование матрицы соответствия этих графов, т. е. определение того, какие пары вершин двух графов $i^{(M)}, j^{(M)}$ и $i^{(D)}, j^{(D)}$ нужно проверять на существование смежных ребер (и эквивалентность атрибутов ребер) между ними, другими словами, значения каких пар вершин учитывать при формировании меры сходства графов. В данной работе мы предлагаем метод формирования матрицы соответствия на основе особенностей карты Кохонена, состоящих в ее фиксированной топологии, а также в ее упорядоченности. Упорядоченность карты говорит о том, что близкие значения узлов карты будут расположены на карте топологически близко друг к другу. Соответственно, схожие области двух различных карт топологически будут расположены в одном и том же месте на карте. Таким образом, упорядоченность карты Кохонена, доказанная в работе [12], позволяет формировать матрицу соответствия по принципу «один к одному», т. е. однозначно сопоставлять вершины с одинаковыми номерами. Следует обратить внимание на то, что

задача сравнения графов в общем случае отличается от задачи определения изоморфизма гораздо меньшей строгостью в том плане, что соответствие между вершинами может формироваться по принципу «один ко многим».

В работе рассматривается задача классификации сигналов на примере практических задач распознавания рукописных символов по траектории пера и распознавания действия человека (*Human Motion Primitives*) [13] по показаниям акселерометра наручного браслета. Представленный в работе метод классификации основывается на анализе отношений между узлами карты Кохонена. Ранее нами был разработан метод, основанный на анализе непосредственно значений самих узлов карты [14, 15]. Объединение этих методов позволило получить значительное увеличение качества классификации, что подтверждается сравнительной оценкой объединения классификаторов различными методами: методом голосования, а также вероятностным методом по правилу Байеса и его упрощениями.

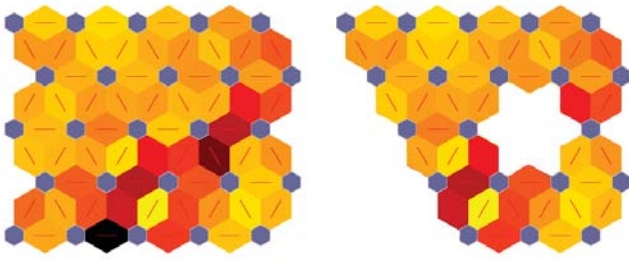
Метод и алгоритм классификации на основе спектрального анализа графа кластерного разбиения

Пусть дано обучающее множество $TS_y = \{X_1, X_2, \dots, X_m\}$ — неупорядоченный набор последовательностей наблюдений длины m , соответствующих классу $y, y = 1..C$, где C — количество классов; $X = \{x_1, x_2, \dots, x_T\}$ — последовательность наблюдений длины T , представляющая собой упорядоченный набор векторов-признаков $x_t \in R^d$, где $t = 1..T$; d — размерность пространства признаков. Отображение обучающих данных класса y на карту Кохонена выполняется как решение следующей оптимизационной задачи (для каждого класса обучается своя карта):

$$\sum_{\forall x_t \in TS_y} \sum_{j=1}^n \|x_t - w_j\| \rightarrow \min, \quad (1)$$

где n — количество узлов карты; w_j — координата j -го узла карты y в $R^d, w_j \in R^d$.

Тогда граф модели определим как $G_M = (V_M, E_M, A_M)$, где V_M — множество узлов карты Кохонена, имеющих вес $w_j, j \in V_M, |V_M| = n$; E_M — множество связей $e \in E_M$ (ребра e будем обозначать по номерам инцидентным им вершин — $e = (i, j)$) карты Кохонена; $A_M = \{A_{ij}\}$ — множество атрибутов ребер, $A_{ij} = \|w_i - w_j\|$. Построение графа данных $G_D = (V_D, E_D, A_D)$ будем производить путем выполнения операции пересечения множества узлов карты, соответствующих наблюдаемым данным, и множества вершин графа модели G_M (рис. 1). Фактически выполняем отображение по-



■ **Рис. 1.** Пример графа модели (слева) и графа данных (справа): серо-синими фигурами обозначены вершины, красными линиями — ребра, цвет объекта вокруг красной линии характеризует атрибут ребра

следовательности наблюдений $X = \{x_1, x_2, \dots, x_T\}$ на карту, тем самым формируя множество узлов $B = \{b_t\}, t = 1..T$, соответствующих последовательности X :

$$b_t = \operatorname{argmin} \sum_{i=1}^T \sum_{j=1}^n \|x_t - w_j\|, t = 1..T, \quad (2)$$

и определяем множество вершин V_D как

$$V_D = V_M \cap B. \quad (3)$$

Тогда множество ребер E_D определим как подмножество множества ребер E_M , инцидентных вершинам из множества V_D . Соответственно, множество атрибутов A_M также однозначно отображается на множество ребер E_D , формируя множество A_D . Для каждой последовательности X будет построено y графов данных G_D , по одному на модель каждого класса. Чем больше граф G_D последовательности X «похож» на граф G_M модели y , тем больше вероятность соотнесения последовательности X к классу y .

Сама задача сравнения графов в общем виде формулируется как задача целочисленного квадратичного программирования

$$S = z^T \cdot M \cdot z \rightarrow \max, \quad (4)$$

где S — целочисленная мера сходства графов; M — матрицы соответствия $M_{i^{(M)}i^{(D)}, j^{(M)}j^{(D)}} = f(A_{i^{(M)}j^{(M)}}, A_{i^{(D)}j^{(D)}})$, f — функция сравнения атрибутов ребер, $M : n^{(M)} \cdot n^{(D)} \times n^{(M)} \cdot n^{(D)}$; z — бинарный вектор неизвестных параметров, $z \in \{0, 1\}^{n^{(M)} \cdot n^{(D)}}$.

В качестве функции f сравнения атрибутов ребер будем использовать отношение равенства

$$M_{i^{(M)}i^{(D)}, j^{(M)}j^{(D)}} = \begin{cases} 1, & \text{если } A_{i^{(M)}j^{(M)}} = A_{i^{(D)}j^{(D)}} \\ 0 & \text{в противном случае} \end{cases}. \quad (5)$$

Решение задачи (4) выполним методом, описанным в работе [3], в результате чего получим целочисленную меру S сходства двух графов.

Суть метода [3] состоит в следующем: максимизация $z^T \cdot M \cdot z \rightarrow \max$ сводится к максимизации отношения $\frac{z^T \cdot M \cdot z}{z^T \cdot z} \rightarrow \max$, которое в свою очередь достигает максимума при условии, что вектор z является собственным вектором матрицы M . Следовательно, вычисление собственного вектора z матрицы M приводит непосредственно к нахождению меры S . Вычисление собственного вектора z производится с помощью ряда методик, описанных в работе [3]. В отличие от NP-трудной задачи (4) такая задача имеет линейную вычислительную сложность.

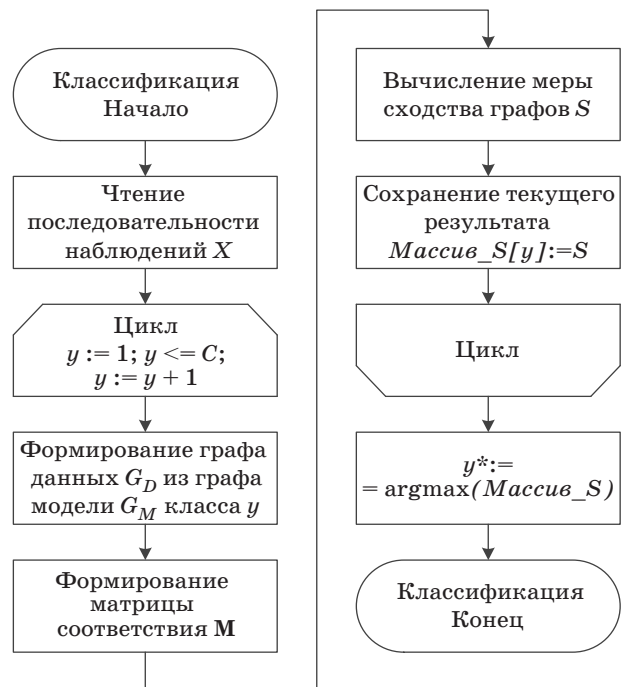
Тогда итоговая задача соотнесения последовательности X одному из классов y решается как

$$y^* = \operatorname{argmax}_{y=1..C} (z^T \cdot M \cdot z), \quad (6)$$

где M — матрица соответствия, построенная для модели каждого класса y от 1 до C для последовательности X .

Предложенный метод классификации на основе спектрального анализа графа кластерного разбиения состоит из двух этапов — обучения и классификации. Рассмотрим алгоритмы реализации каждого из этапов.

Обучение: обучение модели состоит только из обучения нескольких самоорганизующихся карт Кохонена на обучающем наборе данных TS_y , что соответствует построению графов G_M , по одному для каждого класса $y, y = 1..C$. Обучение происходит стандартным алгоритмом формирования самоорганизующейся карты Кохонена.



■ **Рис. 2.** Блок-схема алгоритма классификации

Классификация: алгоритм классификации последовательности $X = \{x_1, x_2, \dots, x_T\}$ состоит из следующих шагов (рис. 2): последовательное отображение данной последовательности на каждую из карт Кохонена, представляющую собой модель класса $y, y = 1..C$; формирование графа модели и графа данных по текущей карте и вычисление меры их сходства. Тот класс, чья модель имеет наибольшую меру сходства, является искомым классом.

Экспериментальное исследование и анализ результатов

Проведем сравнительную оценку следующих методов в решении задачи классификации на наборах данных из репозитория машинного обучения UCI [16, 17]: классических методов HMM и HCRF; ранее разработанного нами метода NPM-PGM [14]; предложенного в данной работе метода SAUDM (*Spectral Analysis of Unified Distance Matrix*); их объединений (NPM-PGM & SAUDM, HMM & SAUDM, HCRF & SAUDM, HCRF & HMM). Описание данных приведено в табл. 1. Результаты оценки качества классификации на наборе данных Dataset for ADL Recognition with Wrist-worn Accelerometer [16] приведены в табл. 2. Результаты оценки качества классификации на

наборе данных Character Trajectories Data Set [17] приведены в табл. 3.

Характеристики используемых наборов данных следующие:

1) набор данных Dataset for ADL Recognition with Wrist-worn Accelerometer — 5212 последовательностей показаний трехосного акселерометра, закрепленного на правой руке 16 людей различного пола и возраста. Человек с закрепленным на руке акселерометром совершает семь различных действий: *climb_stairs* (поднимается по лестнице), *drink_glass* (пьет из стакана: берет стакан, пьет, ставит на стол), *getup_bed* (встает с кровати из положения лежа), *pour_water* (берет бутылку со стола, наливает из бутылки в стакан, стоящий на столе, ставит бутылку обратно на стол), *sitdown_chair* (садится на стул), *standup_chair* (встает со стула), *walk* (шагает). Частота считывания координат — 32 Гц;

2) набор данных Character Trajectories Data Set — траектории движения пера, полученные при написании букв английского алфавита на планшете Wacom. Данные состоят из трех параметров: координаты точки по оси абсцисс, оси ординат и силы нажима. Данные сглажены и нормированы. Частота считывания координат — 200 Гц.

Разделение на обучающую и тестовую выборки произведено самими авторами наборов дан-

■ **Таблица 1.** Описание данных для оценки качества классификаторов

Наименование данных	Dataset for ADL Recognition with Wrist-worn Accelerometer	Character Trajectories Data Set
Количество классов	7	20
Размерность пространства признаков	3	3
Количество экземпляров каждого класса для обучения	20	50
Количество экземпляров каждого класса для тестирования	100	50
Источник	Репозиторий UCI [16]	Репозиторий UCI [17]

■ **Таблица 2.** Точность классификации на наборе Dataset for ADL Recognition with Wrist-worn Accelerometer

Метод классификации	Метод объединения классификаторов						
	Vote	Bayes MAX	Bayes SUM	Bayes MIN	Bayes AVG	Bayes PRO	Bayes
HMM	0,6728						
HCRF	0,6928						
SAUDM	0,6871						
NPM-PGM	0,6929						
HMM & SAUDM	0,6971	0,6943	0,6729	0,6729	0,6729	0,6729	0,6300
HCRF & SAUDM	0,7043	0,6986	0,7214	0,7186	0,7214	0,7214	0,6200
NPM-PGM & SAUDM	0,7143	0,7086	0,7343	0,7243	0,7343	0,7329	0,6457
HCRF & HMM	0,6963	0,6857	0,6729	0,6729	0,6729	0,6729	0,6314

■ Таблица 3. Точность классификации на наборе Character Trajectories Data Set

Метод классификации	Метод объединения классификаторов						
	Vote	Bayes MAX	Bayes SUM	Bayes MIN	Bayes AVG	Bayes PRO	Bayes
HMM	0,7570						
HCRF	0,7650						
SAUDM	0,7450						
NPM-PGM	0,7480						
HMM & SAUDM	0,7990	0,7980	0,7570	0,7570	0,7570	0,7570	0,7050
HCRF & SAUDM	0,7770	0,7660	0,8310	0,8230	0,8310	0,8330	0,6740
NPM-PGM & SAUDM	0,7550	0,7450	0,7990	0,8000	0,7990	0,6630	0,6520
HCRF & HMM	0,8110	0,7960	0,7570	0,7570	0,7570	0,7570	0,6770

ных и в ходе проводимых нами экспериментов не менялось, кроме того, исходное пространство признаков также не подвергалось никаким изменениям.

Параметры HMM: первоначальная инициализация центров распределения вероятности в каждом состоянии производится алгоритмом k -средних; для обучения используется алгоритм Баума — Велша; нормальное распределение вероятности; количество скрытых состояний для Dataset for ADL Recognition with Wrist-worn Accelerometer — 9 (24 гауссовы компоненты в смеси), для Character Trajectories Data Set — 7 (одна гауссова компонента в смеси).

Параметры HCRF: квазиньютоновский алгоритм оптимизации — BFGS с L_2 -регуляризацией; количество скрытых состояний для Dataset for ADL Recognition with Wrist-worn Accelerometer — 9, для Character Trajectories Data Set — 7.

Параметры NPM-PGM и SAUDM: гексагональная карта Кохонена: размер карты для Dataset for ADL Recognition with Wrist-worn Accelerometer — 40×40 , для Character Trajectories Data Set — 16×16 .

Из множества методов объединения классификаторов [18] мы выбрали следующие, показывающие хорошее качество классификации на данных наборах и классификаторах: метод голосования (Vote), вероятностный метод по правилу Байеса (Bayes), вероятностный метод по правилу Байеса с упрощением по правилу умножения (Bayes PRO), суммы (Bayes SUM), максимума (Bayes MAX), минимума (Bayes MIN), среднего (Bayes AVG). В табл. 2 и 3 выделены методы объединения, показывающие наилучшее качество классификации [19, 20].

В результате анализа оценки качества классификации можно отметить, что метод SAUDM имеет приемлемую дискриминантную способность и значительно увеличивает итоговое качество классификации при объединении его с другими методами.

На наборе Dataset for ADL Recognition with Wrist-worn Accelerometer:

- HMM & SAUDM лучше HMM на 2,43 %;
- HCRF & SAUDM лучше HCRF на 2,86 %;
- NPM-PGM & SAUDM лучше NPM-PGM на 4,14 %.

На наборе Character Trajectories Data Set:

- HMM & SAUDM лучше HMM на 4,2 %;
- HCRF & SAUDM лучше HCRF на 6,8 %;
- NPM-PGM & SAUDM лучше NPM-PGM на 5,2 %.

Увеличение качества классификации при объединении метода SAUDM с другими методами говорит о том, что он показывает лучшие результаты классификации на отдельно взятых классах или экземплярах классов благодаря использованию информации о структуре сигнала при построении решающего правила.

Заключение

Разработан метод классификации сигналов на основе спектрального анализа графа кластерного разбиения. В процессе достижения этой цели были решены следующие задачи:

- предложен метод формирования графа модели с использованием отображения обучающих данных на самоорганизующуюся карту Кохонена;

- на основе упорядоченности карты Кохонена предложен способ формирования графа данных, заключающийся в выполнении операции пересечения множества узлов карты, соответствующих наблюдаемым данным, и множества вершин графа модели;

- на основе упорядоченности карты Кохонена выбран способ построения матрицы соответствия по принципу «один к одному». В качестве функции сравнения атрибутов ребер использовано отношение равенства.

Проведен эксперимент по классификации сигналов на примере решения практических задач

распознавания рукописных символов по траектории пера и распознавания действия человека по показаниям акселерометра наручного браслета. В результате анализа оценки качества классификации можно отметить то, что метод SAUDM имеет приемлемую дискриминантную способность и значительно увеличивает итоговое качество классификации при объединении его с другими методами в среднем на 4 %.

Увеличение качества классификации методов при объединении их с методом SAUDM говорит

о его отличительной особенности, характеризующейся учетом структурной информации, содержащейся в графе кластерного разбиения, путем ее анализа на основе методов спектральной теории графов.

В дальнейшем метод, описанный в работе, может быть обобщен до случая выполнения кластерного разбиения произвольными методами кластеризации путем введения дополнительных эвристических правил формирования графа модели, графа данных, а также матрицы их соответствия.

Литература

1. Рабинер Л. Р. Скрытые марковские модели и их применение в избранных приложениях при распознавании речи: обзор // ТИИЭР. 1989. Т. 77. № 2. С. 86–120.
2. Gunawardana A., Mahajan M., Acero A., Platt J. C. Hidden Conditional Random Fields for Phone Classification // Intern. Conf. on Speech Communication and Technology. 2005. P. 1117–1120.
3. Koller D., Friedman N. Probabilistic Graphical Models: Principles and Techniques. — MIT Press, 2009. — 1265 p.
4. Кузнецов Л. А., Бугаков Д. А. Разработка меры оценки информационного расстояния между графическими объектами // Информационно-управляющие системы. 2013. № 1(62). С. 74–79.
5. Bengoetxea E. Inexact Graph Matching Using Estimation of Distribution Algorithms: PhD thesis/ Ecole Nationale Supérieure des Télécommunications. — Paris, 2002. — 187 p.
6. Leordeanu M., Hebert M. A Spectral Technique for Correspondence Problems Using Pairwise Constraints // Tenth IEEE Intern. Conf. on Computer Vision (ICCV). 2005. Vol. 2. P. 1482–1489.
7. Nan Hu, Rustamov R., Guibas L. Stable and Informative Spectral Signatures for Graph Matching // IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2014. P. 2313–2320.
8. Cour T., Srinivasan P., Shi J. Balanced Graph Matching // Advances in Neural Information Processing Systems 19 (NIPS 2006). 2006. P. 313–320.
9. Цветкович Д., Дуб М., Захс Х. Спектры графов. Теория и применение. — Киев: Наукова думка, 1984. — 384 с.
10. Кохонен Т. Самоорганизующиеся карты. — М.: Бинном, 2008. — 655 с.
11. Gorban A., Kegl B., Wunsch D., Zinovyev A. Principal Manifolds for Data Visualisation and Dimension Reduction. — N. Y.: Springer, 2008. — 340 p.
12. Cottrell M., Fort J.-C. Étude d'un Processus D'auto-organisation // Annales de l'institut Henri Poincaré. 1987. Vol. 23. P. 1–20.
13. Bruno B., Mastrogiovanni F., Sgorbissa A., Vernazza T. Analysis of Human Behavior Recognition Algorithms Based on Acceleration Data // IEEE Intern. Conf. on Robotics and Automation (ICRA). 2013. P. 1602–1607.
14. Паламарь И. Н., Юлин С. С. Порождающая графическая вероятностная модель на основе главных многообразий // Тр. СПИИРАН. 2014. № 2. С. 227–247.
15. Palamar I. N., Yulin S. S. Probabilistic Graphical Model Based on Growing Neural Gas for Long Time Series Classification // Modern Applied Science. 2015. Vol. 9. N 2. P. 108–115.
16. UCI Machine Learning Repository. Dataset for ADL Recognition with Wrist-worn Accelerometer Data Set, 2014. <http://archive.ics.uci.edu/ml/datasets/Dataset+for+ADL+Recognition+with+Wrist-worn+Accelerometer> (дата обращения: 01.01.2015).
17. UCI Machine Learning Repository. Character Trajectories Data Set, 2008. <http://archive.ics.uci.edu/ml/datasets/Character+Trajectories> (дата обращения: 01.01.2015).
18. Городецкий В. И., Серебряков С. В. Методы и алгоритмы коллективного распознавания: обзор // Тр. СПИИРАН. 2006. № 3. С. 139–171.
19. Ogorzalek M., Merkwirth C. Statistical Learning and Ensembling Techniques for Time-series-based Model Building, 2006. http://www.eie.polyu.edu.hk/~maciej/pdf/hong_kong_seminar-14_03_2006.pdf (дата обращения: 01.01.2015).
20. Dietterich T. Ensemble Methods in Machine Learning // Multiple Classifier Systems. 2000. Vol. 1857. P. 1–15.

UDC 004.931:519.176

doi:10.15217/issn1684-8853.2015.2.23

The Method of Time-Series Classification Based on Spectral Analysis of Self-Organizing MapYulin S. S.^a, Programmer Engineer, julin.serg@gmail.comPalamar I. N.^b, PhD, Tech., Professor, irina.palamar@mail.ru^aLutch, JSC, 25, blvd. Pobedy, 152920, Rybinsk, Russian Federation^bP. A. Solovyov Rybinsk State Aviation Technical University (RSATU), 53, Pushkin St., 152934, Rybinsk, Russian Federation

Purpose: Most time-series classification methods take into account the internal structure of objects, resting only upon Markov assumptions. This leads to a significant loss of the discriminant information contained in the time-series data. The purpose of this work is improving the quality of time-series classification methods by analyzing the information about the structure of SOM (Self-Organizing Map) nodes. **Results:** A classification method is proposed, based on spectral analysis of graphs built by a supervised learning model. A time-series classification method is developed through analysing the structure of clusters obtained as a result of mapping data on a nonlinear principal manifold by SOM algorithm. This set of nodes (clusters) is represented as a graph. A data graph and a model graph are defined, being specified on a topological ordered SOM structure. The integer score of matching between the data graph and the model graph is calculated using the method of spectral graph theory. It is experimentally proved that classification quality is higher when the method proposed in the paper is combined with the state-of-art methods, such as HMM (Hidden Markov Model) and HCRF (Hidden Conditional Random Fields), or with NPM-PGM (Nonlinear Principal Manifolds – Probabilistic Graphical Model) which we previously developed. **Practical relevance:** The developed method can be used for the recognition of time-series data, for example, handwriting recognition or human motion primitives recognition.

Keywords — Time-Series Classification, Self-Organizing Map, Spectral Graph Theory, Graph Matching.

References

- Rabiner L. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proc. of the IEEE*, 1989, vol. 77, no. 2, pp. 257–286.
- Gunawardana A., Mahajan M., Acero A., Platt J. C. Hidden Conditional Random Fields for Phone Classification. *Intern. Conf. on Speech Communication and Technology*, 2005, pp. 1117–1120.
- Koller D., Friedman N. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, 2009. 1265 p.
- Kuznetsov L. A., Bugakov D. A. Development of Measures of Assessing Information Distance Between Graphical Objects. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2013, no. 1(62), pp. 74–79 (In Russian).
- Bengoetxea E. *Inexact Graph Matching Using Estimation of Distribution Algorithms*. PhD thesis. Paris, Ecole Nationale Supérieure des Télécommunications, 2002. 187 p.
- Leordeanu M., Hebert M. A Spectral Technique for Correspondence Problems Using Pairwise Constraints. *Tenth IEEE Intern. Conf. on Computer Vision (ICCV)*, 2005, vol. 2, pp. 1482–1489.
- Nan Hu, Rustamov R., Guibas L. Stable and Informative Spectral Signatures for Graph Matching. *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 2313–2320.
- Cour T., Srinivasan P., Shi J. Balanced Graph Matching. *Advances in Neural Information Processing Systems 19 (NIPS 2006)*, 2006, pp. 313–320.
- Cvetkovic D., Doob M., Sachs H. *Spectra of Graphs. Theory and Application*. Berlin, VEB Deutscher Verlag der Wissenschaften, 1980. 384 p.
- Kohonen T. *Self-Organizing Maps*. New York, Springer, 2000. 521 p.
- Gorban A., Kegl B., Wunsch D., Zinovyev A. *Principal Manifolds for Data Visualisation and Dimension Reduction*. New York, Springer, 2008. 340 p.
- Cottrell M., Fort J.-C. Étude d'un Processus D'auto-organisation. *Annales de l'institut Henri Poincaré*, 1987, vol. 23, pp. 1–20 (In French).
- Bruno B., Mastrogiovanni F., Sgorbissa A., Vernazza T. Analysis of Human Behavior Recognition Algorithms Based on Acceleration Data. *IEEE Intern. Conf. on Robotics and Automation (ICRA)*, 2013, pp. 1602–1607.
- Palamar I. N., Yulin S. S. Generative Probabilistic Graphical Model Base on the Principal Manifolds. *Trudy SPIIRAN*, 2014, no. 2, pp. 227–247 (In Russian).
- Palamar I. N., Yulin S. S. Probabilistic Graphical Model Based on Growing Neural Gas for Long Time Series Classification. *Modern Applied Science*, 2015, vol. 9, no. 2, pp. 108–115.
- UCI Machine Learning Repository. Dataset for ADL Recognition with Wrist-worn Accelerometer Data Set, 2014. Available at: <http://archive.ics.uci.edu/ml/datasets/Dataset+for+ADL+Recognition+with+Wrist-worn+Accelerometer> (accessed 1 January 2015).
- UCI Machine Learning Repository. Character Trajectories Data Set, 2008. Available at: <http://archive.ics.uci.edu/ml/datasets/Character+Trajectories> (accessed 1 January 2015).
- Gorodetsky V. I., Serebryakov S. V. Methods and Algorithms of the Collective Recognition. *Trudy SPIIRAN*, 2006, no. 3, pp. 139–171 (In Russian).
- Ogorzalek M., Merkwirth C. *Statistical Learning and Ensemble Techniques for Time-Series-Based Model Building*, 2006. Available at: http://www.eie.polyu.edu.hk/~maciej/pdf/hong_kong_seminar-14_03_2006.pdf (accessed 1 January 2015).
- Dietterich T. Ensemble Methods in Machine Learning. *Multiple Classifier Systems*, 2000, vol. 1857, pp. 1–15.