Математические структуры и моделирование
2015. № 3(35). С. 50–55

УДК 519.237.07

МЕТОД ВЕРЛЕ ДЛЯ АППРОКСИМАЦИИ НАБОРА ТОЧЕК УПРУГОЙ КАРТОЙ

В.А. Шовин
научный сотрудник, e-mail: v.shovin@mail.ru
В.В. Гольтяпин
доцент, к.ф.-м.н., e-mail: goltyapin@mail.ru

Институт математики им. С.Л. Соболева СО РАН

Аннотация. Для решения задачи аппроксимации набора точек упругой картой предлагается использовать метод Верле и физическую интерпретацию точек данных в многомерном пространстве как гравитационные центры. Приводится частный случай упругой карты для реализации нелинейных главных компонент.

Ключевые слова: метод Верле, упругая карта, факторный анализ, нелинейные главные компоненты.

Введение

Упругая карта — это наиболее общий вид структур для аппроксимации данных. Это набор узлов и упругих связей между ними. В качестве таких связей могут выступать пружинная связь между парой точек с равновесным расстоянием между точками и ребра жёсткости тройки узлов с равновесным углом между узлами.

Для аппроксимации набора точек упругой картой предлагается использовать физическую интерпретацию точек данных как гравитационные центры, притягивающие узлы упругой карты.

Для расчёта движения точек упругой карты в гравитационном поле и учёта связей между узлами упругой карты предлагается использовать метод численного интегрирования Верле.

Метод Верле — это итерационный метод вычисления следующего местоположения точки по текущему и прошлому.

Частным случаем упругой карты являются нелинейные главные компоненты. Это набор упругих цепочек с общей точкой пересечения. При большой жёсткости упругих связей нелинейные главные компоненты переходят в классические главные компоненты факторного анализа.

1. Метод Верле

Алгоритм Верле используется для вычисления следующего положения точки по текущему и прошлому:

$$\bar{x}_j^i = \bar{x}_j^{i-1} + \bar{v}_j + \sum_{k=1}^{n} \bar{d}_k,$$

$\bar{x}_j^i = \left(x_{j1}^i, x_{j2}^i, \ldots, x_{jm}^i\right)$ — вычисляемые координаты j-ой точки на i-ой итерации,

m — размерность пространства,

$\bar{v}_j = \bar{x}_j^{i-1} - \bar{x}_j^{i-2}$ — вектор скорости j-ой точки.

$\bar{d}_k = \alpha \frac{\bar{D}_k - \bar{x}_j^{i-1}}{\left|\bar{D}_k - \bar{x}_j^{i-1}\right|}$ — вектор влияния k-го гравитационного центра, представленного точкой данных $\bar{D}_k$,

$$\alpha = 0.01.$$

На систему точек накладываются ограничения: некоторые из точек связаны упругими стержнями заданной длины.

Алгоритм работает следующим образом:

1. Вычисляются новые положения точек.
2. Для каждой связи удовлетворяется соответствующее условие.
3. Шаг 2 повторяется s раз.

Например, $s = 16$.

Процедура релаксации связи описывается следующими формулами:

Если связь представлена точками $\bar{a}$ и $\bar{b}$ с равновесным расстоянием между ними t, то

$$\bar{a}^i = \bar{a}^{i-1} + \bar{r},$$

$$\bar{b}^i = \bar{b}^{i-1} - \bar{r},$$

$$\bar{r} = f \cdot q \cdot \frac{t - \left|\bar{a}^{i-1} - \bar{b}^{i-1}\right|}{\left|\bar{a}^{i-1} - \bar{b}^{i-1}\right|} \left(\bar{a}^{i-1} - \bar{b}^{i-1}\right),$$

$f = 0.7$ — коэффициент упругости связи,

$q = \dfrac{1}{s}$ — коэффициент, зависящий от числа s повторений шага 2.

2. Главные компоненты и факторная модель

Модель главных компонент описывается следующими формулами

$$\overrightarrow{z_i} = a_{i1}\overrightarrow{p_1} + a_{i2}\overrightarrow{p_2} + \ldots + a_{ig}\overrightarrow{p_g} + d_i\overrightarrow{u_i},$$

m — число переменных,

g — число факторов,

$\overrightarrow{z_i}$ — исходные переменные,

$\overrightarrow{p_i}$ — общие факторы,

$\overrightarrow{u_i}$ — специфичные факторы.

В качестве системы точек для метода Верле используется координатная система, представленная осями $H^j = \left\{\bar{h}_i^j\right\}$,

$i = 1 \ldots w$, где

w — число точек, образующих факторную ось,

$$j = 1 \ldots g,$$

g — количество факторных осей.

Например, при $w = 11$ (число w должно быть нечётным, большим 1)

$$\bar{h}_i^j = \begin{cases} i \cdot l, i = j, \\ c \cdot l, i \neq j. \end{cases}$$

l — расстояние между парой соседних точек, образующих факторную ось.

$c = \dfrac{w+1}{2} = 6$ — индекс центральной точки факторной оси.

Пары точек $\left\{\bar{h}_i^j, \bar{h}_{i+1}^j\right\}$ образуют связь с равновесным расстоянием l.

Пары точек $\left\{\bar{h}_1^j, \bar{h}_w^j\right\}$ образуют связь с равновесным расстоянием $(w-1) \cdot l$.

Все пары точек $\left\{\bar{h}_c^i, \bar{h}_c^j\right\}$ различных осей i и j образуют связь с нулевым равновесным расстоянием.

Пары точек $\left\{\bar{h}_1^i, \bar{h}_1^j\right\}$, $\left\{\bar{h}_1^i, \bar{h}_w^j\right\}$, $\left\{\bar{h}_w^i, \bar{h}_1^j\right\}$, $\left\{\bar{h}_w^i, \bar{h}_w^j\right\}$ различных осей i и j образуют связь с равновесным расстоянием, равным начальному расстоянию между точками. Эти связи задают прямые углы между различными факторными осями i и j.

Факторная структура

Элементы факторной структуры a_{ij} могут быть определены как коэффициент корреляции между j-ой факторной осью и i-ой осью исходной системы координат:

$$a_{ij} = \frac{\bar{z}_i \cdot \bar{p}_j}{|\bar{z}_i| \cdot |\bar{p}_j|},$$

$$\bar{z}_i = (\delta_{i1}, \ldots, \delta_{im}),$$

$\delta_{ij} = \begin{cases} 1, i = j, \\ 0, i \neq j \end{cases}$ — символ Кронекера,

$\bar{p}_j = \bar{h}_1^j - \bar{h}_c^j$ — вектор направления j-ой факторной оси.

3. Численный эксперимент

В качестве исходных параметров были взяты 15 биофизических показателей для 131 лица с артериальной гипертензией начальной стадии:

1) *вес,*
2) *индекс массы тела (ИМТ),*
3) *частота дыхания (ЧД),*

4) *сегментоядерные нейтрофилы (С),*
5) *лимфоциты (Л),*
6) *конечно-систолический размер левого желудочка (КСР),*
7) *конечно-систолический объем левого желудочка (КСО),*
8) *конечно-диастолический размер левого желудочка (КДР),*
9) *конечно-диастолический объем левого желудочка (КДО),*
10) *ударный объем (УО),*
11) *минутный объем сердца (МОС),*
12) *общее периферическое сосудистое сопротивление (ОПСС),*
13) *индекс Хильдебрандта (ИХ),*
14) *фракция выброса левого желудочка (ФВ),*
15) *фракция укорочения левого желудочка (ФУ).*

Программная реализация

Метод Верле был реализован программно с использованием общедоступной JavaScript библиотеки Verlet.js, которая была усовершенствована для многомерного случая. Web-приложение факторного анализа на базе метода Верле доступно по адресу `http://svlaboratory.org/application/pca` после регистрации нового пользователя. Приложение позволяет визуализировать процесс сходимости метода Верле в заданной плоскости координат (рис. 1).

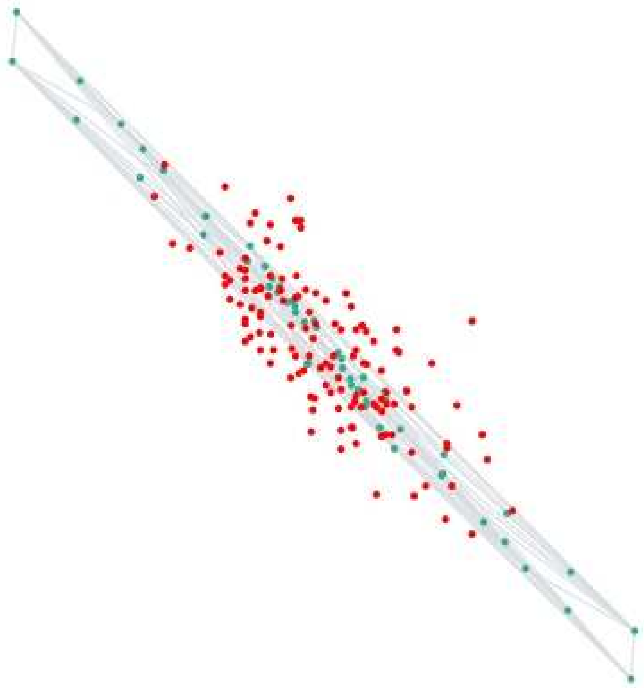

Рис. 1. Визуализация метода Верле в web-приложении.

Результирующая факторная структура для данных артериальной гипертензии представлена в таблице 1.

Таблица 1. Исходное факторное решение (метод Верле)

	$F1$	$F2$	$F3$	$F4$	$F5$
Вес	0,077	-0,658	-0,312	0,387	**-0,56**
ИМТ	0,027	-0,598	-0,355	0,294	**-0,655**
ЧД	-0,181	-0,161	0,483	**-0,689**	-0,484
С	-0,054	-0,191	**-0,726**	-0,576	0,317
Л	0,091	0,273	**0,705**	0,588	-0,275
КСР	**-0,818**	0,383	-0,232	0,101	-0,348
КСО	**-0,821**	0,405	-0,231	0,124	-0,305
КДР	**-0,971**	-0,081	0,012	0,212	0,072
КДО	**-0,975**	-0,051	-0,016	0,21	0,051
УО	**-0,78**	-0,44	0,194	0,21	0,34
МОС	**-0,833**	-0,333	0,183	0,147	0,375
ОПСС	**0,8**	0,431	-0,231	-0,015	-0,348
ИХ	0,073	0,249	-0,439	**0,74**	0,439
ФВ	0,159	**-0,732**	0,391	-0,019	0,535
ФУ	0,199	**-0,682**	0,42	0,174	0,537

Таблица 2. Факторная структура по критерию интерпретабельности (косоугольный случай)

	$F1$	$F2$	$F3$	$F4$	$F5$
Вес	0,021	-0,044	-0,003	-0,059	**0,996**
ИМТ	0,010	0,079	0,052	0,023	**0,995**
ЧД	-0,015	0,0204	0,000	**0,999**	-0,038
С	-0,028	-0,013	**0,994**	-0,089	-0,032
Л	-0,014	0,042	**-0,998**	0,030	-0,029
КСР	**0,60**	0,793	0,023	0,039	0,035
КСО	**0,618**	0,785	0,015	0,000	0,003
КДР	**0,978**	0,204	0,006	-0,002	0,012
КДО	**0,968**	0,248	0,016	-0,010	0,015
УО	**0,948**	-0,313	0,003	-0,001	0,021
МОС	**0,966**	-0,240	0,038	0,008	-0,088
ОПСС	**-0,925**	0,321	-0,095	-0,147	0,074
ИХ	0,101	0,081	-0,081	**-0,988**	0,007
ФВ	0,148	**-0,987**	0,009	0,064	-0,003
ФУ	0,146	**-0,976**	-0,141	-0,067	0,026

Факторное решение после факторного вращения по критерию интерпретабельности, предложенное в работе [1], представлено в таблице 2. Данные факторные структуры подтверждаются предыдущими работами [2].

4. Заключение

В качестве упругой карты была использована многомерная факторная структура, содержащая факторные оси и связи между различными точками факторных осей. Для аппроксимации набора точек упругой картой был использован метод Верле и физическая интерпретация точек данных в многомерном пространстве как гравитационные центры. Для 15 биофизических показателей артериальной гипертензии начальной стадии была получена факторная структура на базе метода Верле. Полученная факторная структура подтверждается результатами ранних работ. В качестве многомерной упругой карты может быть использована произвольная структура, тем самым предложенный метод аппроксимации на базе метода Верле имеет обобщённый характер.

Литература

1. Шовин В.А., Гольтяпин В.В. Методы вращения факторных структур // Математические структуры и моделирование. 2015. № 2 (34). С. 75–83.
2. Гольтяпин В.В., Шовин В.А. Косоугольная факторная модель артериальной гипертензии первой стадии // Вестник Омского университета. 2010. № 4. С. 120–128.

VERLET METHOD FOR APPROXIMATION OF A SET OF POINTS OF THE ELASTIC MAP

V.A. Shovin
researcher, e-mail: v.shovin@mail.ru
V.V. Goltyapin
Ass. Prof., Ph.D.(Math), e-mail: goltyapin@mail.ru

Omsk Branch of the Institution of the Russian Academy of Sciences Institute of Mathematics. S. Siberian Branch of RAS

Abstract. To solve the problem of approximation of a set of points of the elastic map it is proposed to use the Verlet method and the physical interpretation of the data points in a multidimensional space as gravitational centers. A special case of elastic map for the implementation of non-linear principal components is given.

Keywords: Verlet method, elastic map, factor analysis, nonlinear principal components.