*Математические структуры и моделирование*
*2016. № 2(38). С. 66–71*

УДК 519.6

# АППРОКСИМАЦИЯ ДАННЫХ НА БАЗЕ МЕТОДА ВЕРЛЕ

**В.А. Шовин**
научный сотрудник, e-mail: v.shovin@mail.ru

Институт математики им. С.Л. Соболева Сибирского отделения РАН

**Аннотация.** Одним из методов аппроксимации данных сложной топологии является построение древовидных структур. В работе предложен эвристический метод построения древовидного графа на базе метода Верле и физической интерпретации точек данных в многомерном пространстве как центров притяжения. Данный метод построения древовидной структуры протестирован на примере древовидных данных.

**Ключевые слова:** метод Верле, топологические грамматики, упругая энергия, аппроксимация.

## Введение

Известны несколько способов построения древовидных графов без циклов, наилучшим образом аппроксимирующих данные (топологические грамматики, минимизация упругой энергии графа) [1, 2]. Древовидный граф представляет собой 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 древовидной структурой предлагается использовать физическую интерпретацию точек данных как центров, притягивающих узлы графа.

Для расчёта движения точек древовидного графа в поле притяжения и учёта связей между узлами графа предлагается использовать метод численного интегрирования Верле.

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

## 1. Метод Верле

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

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

$\bar{x}_j^i = \left(x_{j1}^i, x_{j2}^i, \ldots, x_{jm}^i\right)$ — 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.

$\bar{d}_{kj} = \alpha \frac{\bar{D}_k - \bar{x}_j^{i-1}}{\left|\bar{D}_k - \bar{x}_j^{i-1}\right|}$ — вектор влияния $k$-го центра притяжения, представленного точкой данных $\bar{D}_k$, на $j$-ую точку,

$$\alpha = 0.01.$$

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

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

$f = 0.7$ — коэффициент упругости связи,

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

Тройки узлов могут образовывать ребра жёсткости с равновесным углом. Такие связи предлагается эмулировать связями в виде упругих стержней между крайними точками. Равновесное расстояние между крайними точками при этом задаётся из равенства треугольника. Если связь представлена точками $\bar{a}$, $\bar{b}$, $\bar{c}$ с равновесным углом $\angle abc = \beta$, тогда равновесное расстояние $ac = t$ между точками $\bar{a}$ и $\bar{c}$ вычисляется по формуле:

$$t = \sqrt{(ab)^2 + (bc)^2 - 2\,(ab)\,(bc)\cos\beta}.$$

## 2. Древовидная структура

Древовидный граф не имеет циклов. Точки данных представлены как центры притяжения. При этом используется следующий алгоритм роста графа:

1. Задаётся число $R$ — радиус поглощения центров притяжения узлами графа. Под поглощением понимается то, что те центры притяжения, которые находятся на расстоянии меньше $R$ к какому либо узлу графа, не притягивают других узлов графа, а оказывают влияние только на этот узел.

2. В граф добавляется новая узловая точка. Начальное положение узловой точки выбирается равным положению ближайшего центра притяжения, который никому не принадлежит, т. е. находится на расстоянии большем $R$ от любого узла графа.

3. Узел графа совершает изменение своего положения в поле притяжения до момента остановки, задаваемого следующим условием. Расстояния между предыдущими и текущими положениями всех узлов графа должны быть меньше заданного числа $\varepsilon$.

4. Узел $\bar{x}$ включается в граф двумя возможными способами: добавлением упругой связи к ближайшему узлу карты по специальной метрике или заменой ближайшей связи графа на две новых, в центр которых помещается данный узел. Равновесные расстояния между узлами графа для новых связей определяются равными начальным расстояниям между узлами этих связей.

   Ближайшая связь к узлу $\bar{x}$ определяется как любая из связей между узлами $\bar{a}$ и $\bar{b}$, удовлетворяющая условиям:

   $f_1x + f_2x < ab$, где

   $f_1$ и $f_2$ — фокусы эллипса с большой осью $ab$ и малой осью длиной равной $\frac{1}{2}ab$.

   Если найдена хотя бы одна ближайшая связь, то ближайшие узлы не рассматриваются. Метрика $d$ для определения ближайшего узла задаётся следующим образом:

   $d(p,q) = e(p,q) - \eta N/e(p,q)$, где

   $e$ — евклидова метрика,

   $N$ — число центров притяжения $g$, удовлетворяющих условиям:

$$f_1g + f_2g < pq,$$

   $f_1$ и $f_2$ — фокусы эллипса с большой осью $pq$ и малой осью длиной равной $k = 10$, при этом центр притяжения $g$ входит в число $N$, если расстояния до остальных центров притяжения из этого числа по евклидовой метрике $e$ не меньше $\theta = 5$;

$$\eta = 100.$$

   Специальная метрика имеет эвристический характер и предназначена для того, чтобы связи между узлами графа проходили преимущественно в

областях наибольшего скопления точек данных, а не в пустых областях, пусть и представляющих кратчайшее расстояние по евклидовой метрике.

5. На тройки узлов, не образующих разветвлений графа, может накладываться угловая связь с развёрнутым равновесным углом. Это имеет смысл в случае необходимости построения прямых ветвей покрывающего графа.

6. Переход в пункт 2. Когда новых узлов, удовлетворяющих пункту 2, не найдено, выполняют операцию $R := h \cdot R$, $h = 0.9$ до тех пор, пока не появится узел, удовлетворяющий пункту 2.

7. Условие остановки роста графа:

Выбирается один из критериев качества аппроксимации:

$$v_1 = \frac{1}{n}\sum_{i=1}^{n} dist(\bar{D}_i, nearestGraphNode(\bar{D}_i))$$

или

$$v_2 = \max_{i=1}^{n}\left(dist(\bar{D}_i, nearestGraphNode(\bar{D}_i))\right)$$

$n$ — объём выборки данных,

$dist$ — евклидово расстояние,

$nearestGraphNode(\bar{D}_i)$ — ближайший по евклидовой метрике узел графа к точке данных $\bar{D}_i$.

Задаётся величина точности аппроксимации $precision = 0.05$.

Рост графа останавливается при $p = \frac{v}{d} \leqslant precision = 0.05$, где

$$d = \sqrt{m \cdot (b-a)^2}$$

$m$ — размерность пространства данных;

$[a, b]$ — интервал разброса данных вдоль координатных осей.

## 3. Линейная структура

Если требуется построить линейную структуру без ветвлений, то в алгоритме построения древовидной структуры изменяется пункт 4:

Если определён ближайший для узла $\bar{x}$ узел $\bar{a}$, не являющийся краевым узлом линейной структуры, то рассматривается парный узел $\bar{b}$ для узла $\bar{a}$, являющийся ближайшим для узла $\bar{x}$. Связь $ab$ заменяется на две связи $ax$ и $xb$, в центр которой помещается узел $\bar{x}$. В остальном пункт 4 не изменяется.

## 4. Численный эксперимент

Работа алгоритма роста графа на базе метода Верле была протестирована на примере древовидных данных (рис. 1).

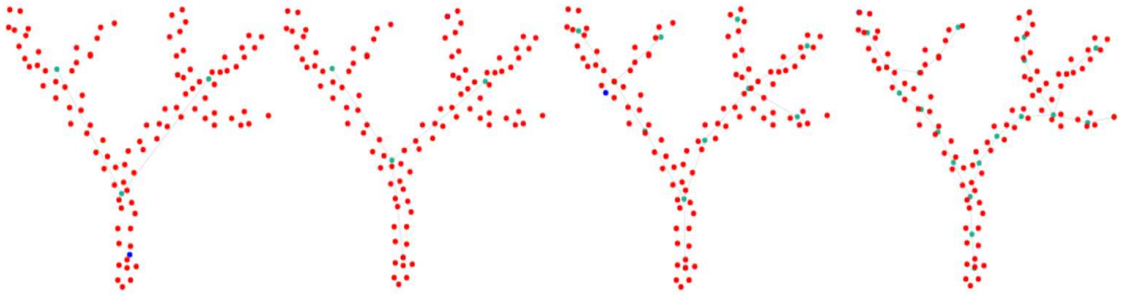

Рис. 1. Визуализация процесса роста древовидного графа

Рис. 2. Визуализация процесса роста аппроксимирующей структуры

Также алгоритм построения аппроксимирующей структуры был протестирован на данных без разветвлений (рис. 2). Данный метод является альтернативой методам построения главных кривых [3–5].

**Программная реализация**

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

## 5. Заключение

Предложен эвристический метод построения древовидной структуры для аппроксимации данных на базе метода Верле и физической интерпретации точек данных в многомерном пространстве как центров притяжения. Алгоритм роста древовидного графа протестирован на данных древовидной структуры, а также на данных нелинейной природы без разветвлений.

## Литература

1. Gorban A.N., Sumner N.R., Zinovyev A.Y. Topological grammars for data approximation // Applied Mathematics Letters. 2007. № 20. P. 382–386.
2. Francis R. Bach, Michael I. Jordan Beyond Independent Components: Trees and Clusters // Journal of Machine Learning Research. 2003. № 4. P. 1205–1233.
3. Hastie T. Principal Curves and Surfaces // Ph.D Dissertation, Stanford Linear Accelerator Center, Stanford University, Stanford, California, US, November 1984.
4. Kegl B. Principal curves: learning, design, and applications // Ph. D. Thesis, Concordia University, Canada, 1999.
5. Der R., Steinmetz U., Balzuweit G., Schuurmann G. Nonlinear Principal Component Analysis // University of Leipzig, Institute of Informatics, 1998. 619 p.

## APPROXIMATION OF VERLET METHOD FOR DATA

**V.A. Shovin**
Scientist Researcher, e-mail: v.shovin@mail.ru

Omsk Branch of Sobolev Institute of Mathematics SB RAS

**Abstract.** One of the methods of approximation for the data with complex topology is building tree structures. In this paper we propose heuristic method for constructing tree graph based on Verlet method and physical interpretation of the data points in a multidimensional space as the attraction centers. This method of constructing tree structure is tested on the example of tree data.

**Keywords:** Verlet method, topological grammars, elastic energy, approximation.