

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М.В. Ломоносова

ФАКУЛЬТЕТ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И КИБЕРНЕТИКИ

ISSN 2411-1473

**Современные
информационные технологии
И
ИТ-образование**

Научный журнал

Том 2 (№ 11)

Москва
2015

УДК [004:377/378](063)
ББК 74.5(0)я431+74.6(0)я431+32.81(0)я431
С 56

**Современные информационные технологии и ИТ-образование. Т. 2 (№ 11),
2015. - 614 с. (ISSN 2411-1473)**

В данном выпуске журнала представлены доклады X Юбилейной международной научно-практической конференции «Современные информационные технологии и ИТ-образование», прошедшей в Московском государственном университете имени М.В. Ломоносова 20-22 ноября 2015 года.

Журнал «Современные информационные технологии и ИТ-образование» включен в наукометрическую базу «Российский индекс научного цитирования» с размещением полнотекстовых версий в научной электронной библиотеке eLIBRARY.RU. URL: http://elibrary.ru/title_about.asp?id=52785



*Издание осуществлено при финансовой поддержке
Российского фонда фундаментальных исследований
(Грант РФФИ № 15-07-20760_з)*

Учредитель:

Фонд содействия развитию интернет-медиа, ИТ-образования, человеческого потенциала «Лига интернет-медиа»

Издатель:

Фонд содействия развитию интернет-медиа, ИТ-образования, человеческого потенциала «Лига интернет-медиа»

Адрес редакции:

119991, г. Москва, ГСП-1, Ленинские горы, д. 1, стр. 52, факультет ВМК МГУ имени М.В. Ломоносова, каб. 375. E-mail: sukhomlin@mail.ru, тел./факс: (495) 939-46-26.

Журнал зарегистрирован Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций (Роскомнадзор).

Свидетельство о регистрации средства массовой информации ПИ № ФС77-61433 от 10 апреля 2015 г.

Издается с 2005 года. Выходит 1 раз в год.

Редакционная коллегия журнала:

Главный редактор:

Сухомлин В.А. - доктор технических наук, профессор, заведующий лабораторией ОИТ факультета ВМК МГУ имени М.В. Ломоносова, Президент Фонда «Лига интернет-медиа»;

Члены редакционной коллегии:

Веремей Е.И. - доктор физ.-мат. наук, профессор, СПбГУ;

Гергель В.П. - доктор физ.-мат. наук, профессор, ННГУ им. Н.И. Лобачевского;

Самуйлов К.Е. - доктор физ.-мат. наук, профессор, РУДН;

Калиниченко Л.А. - доктор физ.-мат. наук, профессор, вед. н.с. ИПИ РАН ФИЦ ИУ РАН;

Лугачев М.И. - доктор экономических наук, профессор, МГУ имени М.В. Ломоносова;

Любецкий В.А. - доктор физ.-мат. наук, профессор, ИППИ РАН им. А.А. Харкевича;

Нечаев В. В. - доктор технических наук, профессор, МИРЭА;

Посыпкин М.А. - доктор физ.-мат. наук, вед. н. с. ИППИ РАН им. А.А. Харкевича;

Язенин А.В. - доктор физ.-мат. наук, декан факультета ПМиК, профессор, ТвГУ;

Намиот Д.Е. - кандидат физ.-мат. наук, с.н.с. факультета ВМК МГУ имени М.В. Ломоносова;

Зубарева Е.В. - кандидат пед. наук, доцент, н.с. факультета ВМК МГУ имени М.В. Ломоносова;

Сотникова М.В. - кандидат физ.-мат. наук, доцент СПбГУ.

Статьи, поступающие в редакцию, рецензируются. За достоверность сведений, изложенных в статьях, ответственность несут авторы публикаций. Мнение редакции может не совпадать с мнением авторов материалов. При перепечатке ссылка на журнал обязательна.

Материалы публикуются в авторской редакции. При перепечатке и цитировании материалов ссылка на журнал «Современные информационные технологии и ИТ-образование» обязательна.

Многоуровневые модели окружающей среды в мегаполисах.....	267
<i>Васильев А.Н., Тархов Д.А., Шемякина Т.А.</i>	
Модель неизотермического химического реактора на основе параметрических нейронных сетей. Гибридный метод.....	271
<i>Васильев А.Н., Тархов Д.А., Шемякина Т.А.</i>	
Мезо-уровневая нейросетевая модель загрязнения атмосферного воздуха Санкт-Петербурга по данным мониторинга.....	279
<i>Зыкина А.В., Запорожец Д.Н.</i>	
Пакет прикладных программ для моделирования и решения процессов с использованием аппарата вариационных неравенств.....	284
<i>Бабичева Т.С.</i>	
Транспортные потоки: математическое и имитационное моделирование	290
<i>Идрисова Д.И., Каверзнева Т.Т., Тархов Д.А., Лазовская Т.В.</i>	
Моделирование распределения опасного вещества в тупиковом тоннеле с использованием нейросетевого подхода.....	297
<i>Пишкинас А.О., Оносов И.А., Корчагин С.А., Романчук С.П., Терин Д.В.</i>	
Разработка программных средств моделирования композитных наноматериалов.....	301
<i>Семенов М.Г.</i>	
Модель Марковица: математические аспекты и компьютерная реализация.....	306
<i>Сенчилов В.В.</i>	
О решении одной видоизмененной краевой задачи типа Неймана в классах метааналитических в круге функций с применением системы компьютерной математики Maple.....	310
<i>Скрипачев В.О., Пирхавка А.П., Полушковский Ю.А., Суровцева И.В., Жуков А.О., Яковлев О.В.</i>	
Аспекты создания информационной системы для обработки ионосферных данных.....	316
<i>Юмагулов М.Г., Беликова О.Н., Исанбаева Н.Р.</i>	
Моделирование областей устойчивости точек либрации ограниченной задачи трех тел с помощью систем компьютерной математики.....	321
<i>Лесников С.В., Булыгина Д.С., Лесников А.В., Лесников Г.С.</i>	
Конструирование гипертекстового информационно-поискового тезауруса метаязыка лингвистики.....	326
<i>Понятский В.М., Егоров Д.Б.</i>	
Программный комплекс моделирования последовательности методов видеообработки для задач управления.....	334
<i>Тархов Д.А., Симакина А.А., Суднева А.И.</i>	
Обработка данных методом треугольных приближений.....	341
<i>Тарасенко Ф.Д., Тархов Д.А.</i>	
Сравнительный анализ применения различных базисных функций в алгоритмах последовательного сглаживания данных.....	350
<i>Тархов Д.А., Шаньшин И.К., Шаханов Д.О.</i>	
Сравнение нейросетевого и классического подхода к задаче идентификации миграционных процессов.....	355
<i>Васильев А.Н., Кузнецов Е.Б., Леонов С.С.</i>	
Нейросетевой метод идентификации и анализа модели деформирования металлических конструкций в условиях ползучести.....	360
<i>Бурцев А.А., Сидоров С.А.</i>	
Программный комплекс ДССП-ТВМ для структурированного программирования троичной [виртуальной] машины.....	371
ПАРАЛЛЕЛЬНОЕ И РАСПРЕДЕЛЕННОЕ ПРОГРАММИРОВАНИЕ, ГРИД-ТЕХНОЛОГИИ, ПРОГРАММИРОВАНИЕ НА ГРАФИЧЕСКИХ ПРОЦЕССОРАХ.....	380
<i>Курочкин И.И., Гуменный Д.Г.</i>	
Безопасность сетей SDN. Классификация атак.....	381
<i>Захаров В.Н., Мунерман В.И.</i>	
Параллельный алгоритм умножения многомерных матриц.....	384
<i>Мунерман В.И., Мунерман Д.В.</i>	

Тарасенко Ф.Д.¹, Тархов Д.А.²

¹Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Россия, студент института прикладной математики и механики, oudi@mail.ru

²Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Россия, д.т.н., профессор, dtarkhov@gmail.com

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ПРИМЕНЕНИЯ РАЗЛИЧНЫХ БАЗИСНЫХ ФУНКЦИЙ В АЛГОРИТМАХ ПОСЛЕДОВАТЕЛЬНОГО СГЛАЖИВАНИЯ ДАННЫХ

КЛЮЧЕВЫЕ СЛОВА

Последовательные алгоритмы, обработка данных, RBF-сети, сплайн.

АННОТАЦИЯ

Исследованы алгоритмы построения зависимости по экспериментальным данным, построенные на основе последовательной обработки точек по одной. Рассмотрены четыре алгоритма обработки экспериментальных данных и проведен их сравнительный анализ для разных исходных данных, базисных функций и наборов параметров. Исходные данные имели не только статический, но и динамический характер, когда искомая модель менялась в процессе работы алгоритма.

Проведены численные эксперименты, позволившие на основе результатов сравнительного анализа сделать выводы об эффективности алгоритмов в тех или иных случаях, и дать рекомендации по их использованию.

Введение. В данной работе, вслед за [1-3] рассматриваются методы нахождения зависимости $y=f(x)$ по экспериментальным данным $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$ в ситуации, когда точки (x_i, y_i) поступают и обрабатываются по одной, что может быть связано с необходимостью обрабатывать данные в режиме реального времени. Будем искомым зависимость представлять в виде $y = \sum_{j=1}^n c_j \varphi_j(x)$, где $\varphi_j = \varphi(\alpha_j |x - z_j|)$ или $\varphi_j = \varphi(\alpha_j (x - z_j)^2)$. Функции такой конструкции принято называть RBF-сетями [3,4]. Подбор параметров c_j , α_j и z_j называется обучением сети. В данной статье рассматривается случай, когда при обучении подбираются только коэффициенты c_j .

Как известно, любую кусочно-линейную функцию можно представить в виде суммы $\sum_{j=1}^n c_j \varphi(\alpha_j (x - z_j))$, если выбрать в качестве базисной функции треугольную шапочку. Аналогичным образом можно любой сплайн разложить по базисным функциям, соответствующим его степени и гладкости [6].

В численных экспериментах использовались следующие базисные сплайны [6]: парабола, кубическая парабола, гауссиан.

Нами исследовано несколько алгоритмов сглаживания данных:

- 1) Обработка точек по одной с адаптацией веса базисной функции с ближайшим центром.
- 2) То же, но с адаптацией весов двух ближайших функций.
- 3) то же, что и в 2) но предполагается линейная зависимость скорости адаптации веса от близости абсциссы добавляемой точки к центру базисной функции [1].
- 4) Нахождение оптимальных коэффициентов разложения по базисным функциям решением линейной системы.

Для подходов 1 и 3 исследовалась также вариация алгоритма с заданным заранее законом изменения шага.

В подходах 1-3 основой для адаптации коэффициентов является минимизация

квадратичного функционала ошибки.

Подход 1.

Начальные значения коэффициентов $c_j=0$. При переходе от $N-1$ наблюдений к N изменение весового коэффициента базисной функции с ближайшим центром, обозначенное $\Delta_k(N)$, получается равным

$$\Delta_k(N) = \frac{Q_k}{S_k}, \quad (1)$$

где

$$S_k = \sum_{i=1}^N \varphi_k^2(x_i), \quad Q_k = \sum_{i=1}^N \varphi_k(x_i) \delta_i(N). \quad (2)$$

При переходе к следующему шагу ошибки нужно пересчитать по формуле

$$\delta_i(N+1) = \delta_i(N) - \Delta_k \varphi_k(x_i). \quad (3)$$

Подход 2.

В отличие от первого алгоритма выбираются две базисные функции, между центрами которых находится поступившее наблюдение.

Изменение весовых коэффициентов производится по формулам

$$\Delta_k = \frac{Q_k S_{k+1} - Q_{k+1} P_k}{S_k S_{k+1} - P_k^2}, \quad \Delta_{k+1} = \frac{Q_k S_{k+1} - Q_{k+1} P_k}{S_k S_{k+1} - P_k^2} \quad (4)$$

где

$$P_k = \sum_{i=1}^N \varphi_k(x_i) \varphi_{k+1}(x_i). \quad (5)$$

При обращении в ноль знаменателя (2) используются формулы (7) и (8).

При переходе к новому шагу слагаемые в суммы (2) и (5) добавляются для всех базисных функций, для которых $\varphi_k(x_i) \neq 0$.

Ошибки δ_i при переходе к следующему шагу нужно пересчитать по формуле

$$\delta_i(N+1) = \delta_i(N) - \Delta_k \varphi_k(x_i) - \Delta_{k+1} \varphi_{k+1}(x_i). \quad (6)$$

Подход 3.

Пусть x_N находится между центрами функций с номерами k и $k+1$. В прежних обозначениях используем формулы [1]:

$$\Delta_k(N) = \Delta(N) \lambda \quad \text{и} \quad \Delta_{k+1}(N) = \Delta(N) (1-\lambda), \quad (7)$$

где в [1] рекомендуется выбирать $\lambda = \frac{z_{k+1} - x_N}{z_{k+1} - z_k}$.

Число $\Delta(N)$ находится минимизацией функционала ошибки и получается равным

$$\Delta(N) = \frac{\lambda Q_k + (1-\lambda) Q_{k+1}}{\lambda^2 S_k + 2\lambda(1-\lambda) P_k + (1-\lambda)^2 S_{k+1}}. \quad (8)$$

Ошибки δ_i при переходе к следующему шагу нужно пересчитать по формуле

$$\delta_i(N+1) = \delta_i(N) - \Delta(N) (\lambda \varphi_k(x_i) + (1-\lambda) \varphi_{k+1}(x_i)). \quad (9)$$

Для алгоритмов 1 и 3 можно вместо формул (1), (7) и (8) использовать убывающие по заданному закону $\Delta(N)$, например $\Delta(N) = \left(1 - \frac{2}{T}\right)^N \delta_k(N)$, где $T = 5n + N_{max}$ [1]. Это существенно сокращает объём вычислений.

Подход 4.

Этот подход предусматривает построение линейной регрессии и предполагает решение системы линейных уравнений, которое можно получать рекуррентным образом при появлении нового наблюдения [3,4,7].

Результаты численных экспериментов. Проведено сравнительное тестирование алгоритмов, реализующих подходы 1-4 для указанных выше базисных функций для разных величин ошибки «измерений», количества точек и числа базисных функций. Также, проанализированы случаи динамических (переменных во времени) данных при различном наборе

параметров и выборе базисных функций. Интервал изменения аргумента [0;1].

Ниже приведены иллюстрации работы алгоритмов, некоторые наиболее характерные данные численных экспериментов вы и выводы, полученные при анализе результатов этих экспериментов.

На первом рисунке представлен результат обработки экспериментально зависимости, полученной генерацией точек около функции $\sin(\pi x) + 0.1 \sin(10 \pi x)$, с использованием первого алгоритма и формулы из [1] для 20-ти базисных функций, 400-сот экспериментальных точек, ошибки «измерений» – 0,1 и 10-ти прогонов алгоритма.

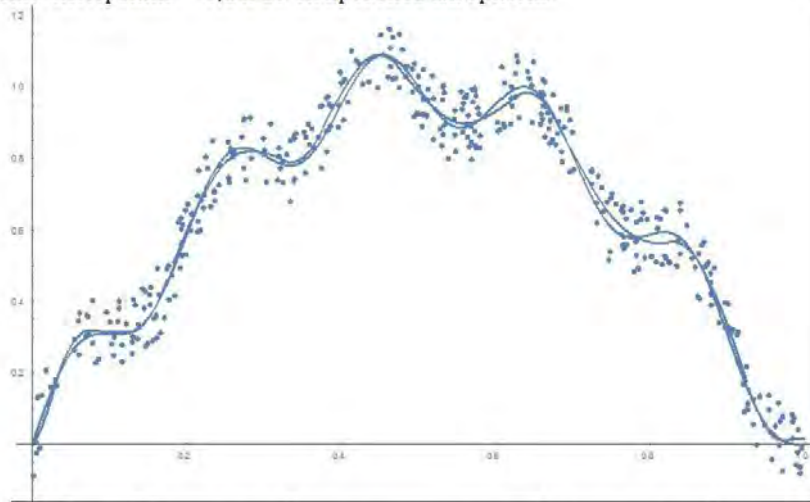


Рисунок 1. Результат работы первого алгоритма для статических данных

На втором рисунке представлены результаты применения второго подхода для динамических данных. Приближаемая функция – $\sin(w\pi x)$, где w меняется за время вычислений от 1 до 4,5. Базисная функция – кубическая парабола. 300 экспериментальных точек. 10 базисных функций. Ошибка измерений – 0,1. Изображены три момента в которые исходная функция имеет вид соответственно: $\sin(\pi x)$, $\sin(2,25\pi x)$, $\sin(4,5\pi x)$.

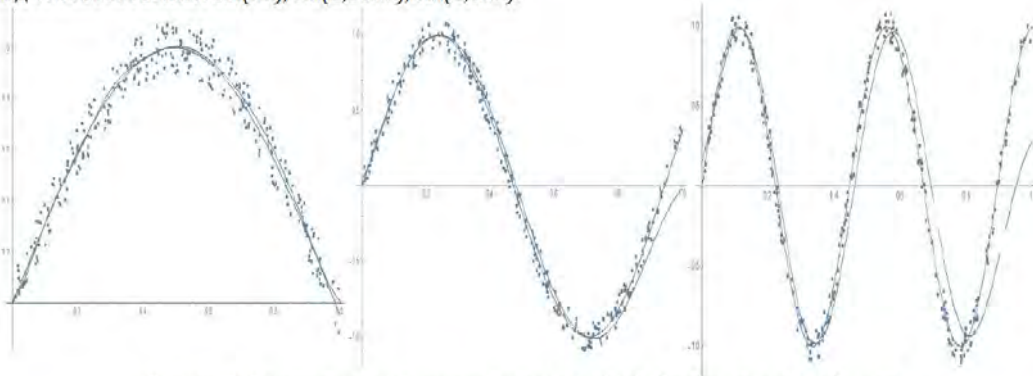


Рисунок 2: Результаты применения второго подхода для динамических данных

Таблица 1: Математическое ожидание ошибки для восьми тестов алгоритма при 50 точках наблюдений. Статические данные. Приближаемая функция - $\sin(\pi x)$. Число базисных функций – 10. Число экспериментальных точек – 50.

ϵ	m	Подход	Треугольная шапочка	Парабола	Кубическая функция	Гауссиан
0.01	10	1 Т.	0.0045	0.0359	0,0251	0,0337
1	10		0.2880	0.3115	0,0404	0,0308
0.01	1		0.0247	0.0363	0,0536	0,0209

0.01	10	1 P.	0,0047	0,0050	0,0056	0,3342
1	10		0,0320	0,0406	0,0361	0,0355
0.01	1		0,1888	0,1126	0,1130	0,1313
0.01	10	2	0,0039	0,0039	0,0054	0,0038
1	10		0,0331	0,0330	0,0281	0,0343
0.01	1		0,0057	0,0150	0,0195	0,0146
0.01	10	3 T.	0,0132	0,0125	0,0238	0,0217
1	10		0,0292	0,0378	0,0387	0,0398
0.01	1		0,0464	0,0403	0,0365	0,0487
0.01	10	3 P.	0,0044	0,0044	0,0053	0,0422
1	10		0,0278	0,0342	0,0358	0,0579
0.01	1		0,1016	0,0975	0,0945	0,1041
0.01		4	0,0031	0,0035	0,0040	0,0080
1			0,2838	0,2542	0,2633	0,2901

Здесь и дальше, ϵ - погрешность наблюдений, m - число прогонов алгоритма, n - кол-во точек наблюдений, P - регулируемый шаг в соответствии с формулой из [1], T - точный в соответствии с формулами, приведёнными выше для каждого алгоритма.

Таблица 2: Математическое ожидание ошибки для четырех тестов алгоритма. Динамические данные. Приближаемая функция - $\sin(wpx)$, где w меняется за время вычислений от 1 до 2 и до 4,5. Число базисных функций - 10. Ошибка измерений - 0,1.

m	n	w	Подход	Треугольная шапочка	Парабола	Кубическая функция	Гауссиан
100	300	4,5	1	0,4252	0,4084	0,4060	0,4360
50	100	2		0,1908	0,1910	0,1755	0,1811
100	300	4,5	2	0,2880	0,2624	0,2481	0,2348
50	100	2		0,1454	0,1361	0,1321	0,1233
100	300	4,5	3	0,4176	0,4296	0,4422	0,4767
50	100	2		0,1854	0,1815	0,1779	0,1817

Выводы

- Рассмотренные алгоритмы и базисные функции показали хорошую работоспособность в рассматриваемых задачах.
- Ни один из алгоритмов не имеет решающего преимущества над другими. Это позволяет рекомендовать самый простой из них - первый.
- Точное вычисление шага не имеет преимуществ над формулой из [1], которые могли бы компенсировать существенное увеличение вычислительной сложности, если выборка достаточно велика или продублирована необходимое число раз.
- При больших ошибках и малом числе экспериментальных точек все методы работают неудовлетворительно. Несколько лучше других получаются результаты сглаживания при использовании кубической БФ для большого количества точек и гауссиана для малого.
- Результаты, полученные при использовании разных базисных функций примерно одинаковы. Выбор конкретной функции диктуется условиями на гладкость. Если таких условий нет, то предпочтительной является треугольная функция в силу минимальности своего носителя (что приводит к минимальной вычислительной сложности). Если такие условия заранее неизвестны или достаточно жестки, то предпочтение следует гауссиан, как имеющий бесконечную гладкость.
- Наилучшим из рассмотренных алгоритмов для сглаживания динамических данных является подход 2, так как он значительно лучше обрабатывает конец промежутка.

Благодарность: Статья подготовлена по результатам исследования, выполненного при финансовой поддержке гранта Российского Научного Фонда (проект 14-38-00009) «Программно-целевое управление комплексным развитием Арктической зоны РФ (Санкт-Петербургский

политехнический университет Петра Великого).

Литература

1. Хакимов Б.В. Моделирование корреляционных зависимостей сплайнами на примерах в геологии и экологии. М.: МГУ; СПб: Нева, 2003. 144 с.
2. Хакимов Б.В., Михеев И.М. Нелинейная модель нейрона – многомерный сплайн // Нейрокомпьютеры: разработка, применение. 2012. № 7. С. 36–40.
3. Тархов Д.А. Последовательные алгоритмы сглаживания данных// Нейрокомпьютеры: разработка, применение. 2015. № 3. С. 11–18.
4. Тархов Д.А. Нейросетевые модели и алгоритмы. – М.: Радиотехника, 2014. – 352 с.
5. Хайкин С. Нейронные сети: полный курс, 2-е издание. Пер. с англ. – М.: Изд. дом «Вильямс», 2006. – 1104 с.
6. Свиньян С.Ф. Базисные сплайны в теории отсчётов сигналов. СПб., Наука, 2003. – 118с.
7. Алберт А. Регрессия, псевдоинверсия и рекуррентное оценивание. - М.: Наука, 1977. 224с.