

20. *Karpenko A.P.* Metody povysheniya effektivnosti populyatsionnykh algoritmov global'noy optimizatsii [Methods for increasing the efficiency of population algorithms for global optimization], *Mater. V mezhhregional'noy nauchno-prakticheskoy konferentsii* [Proceedings of the V interregional scientific and practical conference]. Sevastopol': Izd-vo: Sevastopol'skiy gosudarstvennyy universitet, 2019, pp. 87-88.
21. *Kuchuganov V.N., Kuchuganov A.V., Kaksimov D.R.* Algoritm klasterizatsii mnozhestva detaley po chertezham [Algorithm for clustering a set of parts according to drawings], *Programmirovaniye* [Programming]. Moscow: Izd-vo: Rossiyskaya akademiya nauk, 2020, pp. 29-38.

Статью рекомендовал к опубликованию д.т.н., профессор А.Г. Коробейников.

Лебедев Борис Константинович – Южный федеральный университет; e-mail: lebedev.b.k@gmail.com; 347928, г. Таганрог, пер. Некрасовский, 44; тел.: 89282897933; кафедра систем автоматизированного проектирования; профессор.

Лебедев Олег Борисович – e-mail: lebedev.ob@mail.ru; тел.: 89085135512; кафедра систем автоматизированного проектирования; доцент.

Лебедев Владимир Борисович – Московский государственный технический университет имени Н.Э. Баумана; e-mail: lebedev.vlad.bor@mail.ru; 105005, г. Москва, ул. Бауманская 2-я, д. 5, стр. 1; тел.: 89287775005; Научно-производственное объединение «Новые технологии»; с.н.с.

Lebedev Boris Konstantinovich – Southern Federal University; e-mail: lebedev.b.k@gmail.com; 44, Nekrasovsky, Taganrog, 347928, Russia; phone: +79282897933; the department of computer aided design; professor.

Lebedev Oleg Borisovich – e-mail: lebedev.ob@mail.ru; phone: +79085135512; the department of computer aided design; associate professor.

Lebedev Vladimir Borisovich – Moscow State Technical University named after N.E. Bauman; e-mail: lebedev.vlad.bor@mail.ru; 105005, Moscow, st. Baumanskaya 2-nd, 5, build. 1; phone: +79287775005; Research and Production Association "New Technologies"; senior researcher.

УДК 004.041

DOI 10.18522/2311-3103-2020-4-93-107

А.А. Сорокин, И.М. Бородинский, А.В. Дагаев

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ ВОССТАНОВЛЕНИЯ ПРОПУЩЕННЫХ ДАННЫХ

В последние десятилетия качественно развиваются методы системного анализа, что связано с увеличением скорости технического развития, уплотнением временных процессов, быстрым ростом накапливаемой информации и новыми возможностями вычислительной техники. К этим методам относятся методы анализа большого объема данных, методы добычи данных, методы аналитического моделирования, методы параллельной обработки данных, нейросетевые методы, методы прогнозирования и другие. Представленные методы позволяют быстро и качественно обрабатывать разнородные кластеры информации, аккумулировать и синтезировать данные, обобщать и классифицировать информацию. К последним из представленных методов относятся методы интерполяции и экстраполяции потерянной, поврежденной или неполученной информации. Данные методы позволяют структурировать, восстанавливать и моделировать информацию на основе статистических данных, математических и алгоритмических методов. Таким образом в статье рассматривается проблема восстановления пропущенных данных в графических и сложных объектах. Приводятся литературные источники по рассматриваемым задачам. В них приводится обширная информация по рассматриваемой тематике: представлены генетические алгоритмы используемые для пространственной интерполяции; рассмотрено решение задач неоднородности интерполяции сейсмических данных; описано использо-

вание сплайн-аппроксимации для расчета характеристик нелинейных электронных компонентов; разобран метод построения модели трехмерных параметрических рациональных тел с помощью обобщенной интерполяции Безье, что позволяет моделировать форму тела и анизотропное пространство; описаны методы применяющие нечеткие линейные уравнения, которые широко распространены в компьютерном зрении; исследован метод адаптивной интерполяции на основе градиента учитывающий локальный градиент исходного изображения. В статье выполняется сравнение нескольких распространенных методов интерполяции и реставрации данных, таких как: билинейная интерполяция, поверхность Безье. Кратко описывается каждый метод и особенности его применения в рамках проведенного эксперимента. Приводится результат серии экспериментов с представленными методами с различным количеством испытаний. В заключении делаются выводы о рациональности выбора одного из предложенных методов без применения длительного натурного эксперимента в каждом случае.

Двумерные экспериментальные данные; интерполяция данных; реставрация данных; анализ методов интерполяции.

A.A. Sorokin, I.M. Borodyansky, A.V. Dagaev

COMPARATIVE ANALYSIS OF MISSING DATA RECOVERY METHODS

In recent decades, the methods of system analysis have been developing qualitatively. It is associated with an increase in the rate of technical development, the densification of time processes, the rapid growth of accumulated information and new capabilities of computer technology. These include methods for analyzing large amounts of data, methods of data mining, methods of analytical modeling, methods of parallel data processing, neural network methods, forecasting methods, and others. The presented methods make it possible to quickly and efficiently process heterogeneous clusters of information, accumulate and synthesize data, generalize and classify information. The last of the presented methods are methods of interpolation and extrapolation of lost, damaged or missing information. These methods allow to structure, restore and model information based on statistical data, mathematical and algorithmic methods. Thus, the article deals with the problem of recovering missing data in graphic and complex objects. Literary sources on the problems under consideration are given. They provide extensive information on the topic under consideration: present genetic algorithms used for spatial interpolation; the solution of problems of heterogeneity of interpolation of seismic data is considered; it is described the use of spline approximation to calculate the characteristics of nonlinear electronic components; the method of constructing a model of three-dimensional parametric rational bodies using generalized Bezier interpolation is analyzed, which allows modeling the shape of a body and anisotropic space; methods using fuzzy linear equations are described, which are widespread in computer vision; the method of adaptive interpolation based on the gradient and taking into account the local gradient of the original image is investigated. It is made comparing several common methods of interpolation and data restoration, in article, such as: bilinear interpolation, Bezier surface. Each method and features of its application within the framework of the experiment are briefly described. The result of a series of experiments with the presented methods with different numbers of tests is presented. In conclusion, summary is drawn about the rationality of choosing one of the proposed methods without the use of a long field experiment in each case.

Two-dimensional experimental data; data interpolation; data restoration; analysis of interpolation methods.

1. Введение. Сегодня анализ информации используется во всех областях знания человека. Моделирование и эксперименты создают базу для качественного развития науки. Анализ данных становится возможен только при наличии статистической информации, которая может быть получена физическим и математическим экспериментальным путём. Однако на практике ученые часто сталкиваются с проблемой нехватки информации или погрешности и неточности полученных данных. Кроме того, все чаще возникают задачи предсказания поведения выходной функции по имеющейся статистике. Таким образом, при сборе статистической

информации часто приходится дополнять недостающие данные модельными данными, представляющими характеристики состояния системы в требуемый промежуток времени. Задачи аппроксимации и интерполяции данных стали появляться с развитием компьютерной техники, а также алгоритмических и математических методов. Известны работы с шестидесятых годов прошлого столетия [22], за последние десятилетия появилось много методов данного направления исследования [1–15], но вопрос применимости методов при различном объеме и качестве исходных данных стоит и сейчас. В [16–18] представлены современные нелинейные методы фильтрации на основе PDE (Partial Differential Equations) и методы непрозрачного изображения, представлены основы теории интерполяции, приведена классификация методов, рассматриваются точные и приближенные, однопараметрические и многопараметрические методы. В источниках [19–21] дано описание теории интерполяции линейных операторов. Излагаются основные методы построения интерполяционных пространств, изучаются их свойства, рассматриваются различные методы многомерной аппроксимации и интерполяции в конечномерных вещественных и комплексных векторных пространствах. Как видно из литературных источников, теме интерполяции уделено большое внимание, однако вопросам применимости к различным наборам данных и оптимальности применения методов посвящено немного источников [17–19]. Таким образом основными аспектами при анализе статистических данных является проблема пропуска и противоречивости данных. В большинстве случаев эксперименты нужны, чтобы собрать достаточный для формирования математической модели набор данных. При таком подходе, изучаемый процесс имеет конечное математическое описание, которое на данный момент неизвестно и возникает необходимость заполнения пропущенных данных условно случайными значениями, влияние которых на разрабатываемую модель не должно быть существенным.

Целью текущего исследования было определение оптимального метода интерполяции или реставрации экспериментальных данных в двумерном пространстве, при известном допустимом пороге среднестатистического отклонения от исходных значений с учетом времени выполнения этой операции.

Следует выделить основные варианты решения задачи:

- ◆ повторные эксперименты с целью получить пропущенные данные;
- ◆ использование шаблонов для восстановления данных (на основе других экспериментов уже сформированы шаблоны возможных решений, производится выбор наиболее адекватного шаблона);
- ◆ использование некоторого метода интерполяции (на области вокруг пропущенных данных выбираются опорные точки и формируется некоторая поверхность, проходящая через них; полученные с этой поверхности значения в дальнейшем используются для заполнения пропусков).

В первом случае возможно систематическое повторение пропусков, связанное либо с отсутствием возможных значений, либо с проблемами сбора информации. Второй случай требует наличия повторяющихся или шаблонных областей значений в рамках исследованной области. Третий метод балансирует между возможным отклонением и сложностью и временем выполнения этой операции.

Следует отметить, что теория восстановления пропущенных данных постоянно развивается, появляются новые алгоритмы и модернизируются старые. Это связано с невозможностью разработки универсального алгоритма, выдающего наилучшие результаты во всех случаях. Чаще всего оптимальность отдельного метода решения приводится в источниках на конкретном примере, что априори делает алгоритм неуниверсальным [23].

В итоге, задачи исследования свелись к поиску адекватного метода для случайного эксперимента, который будет реже отклоняться от реальных значений (при условии, что все законы распределения результатов эксперимента заранее известен) в определенных временных рамках при определенном допуске объема пропущенных данных.

Общий набор методов (алгоритмов) заполнения пропусков можно представить в виде схемы, приведенной на рис. 1. Каждый конечный элемент на ней отражает группу похожих методик. Но отдельные методы внутри группы могут кардинально отличаться по результатам их применения на одних и тех же наборах данных.

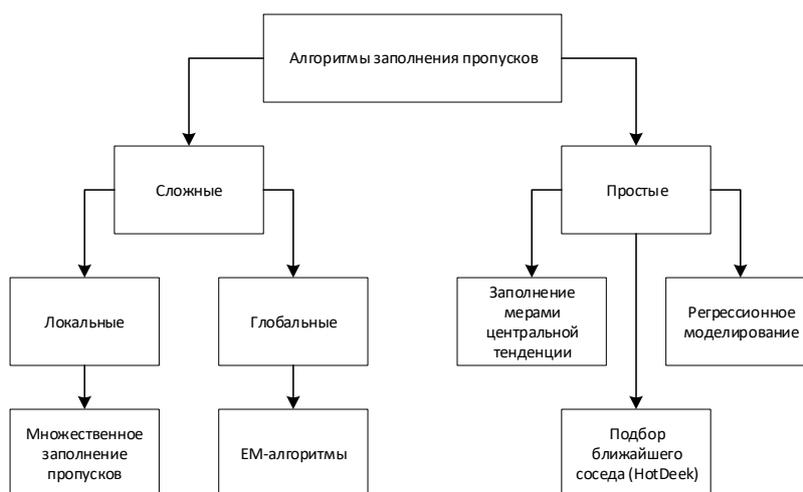


Рис. 1. Классификация алгоритмов

2. Интерполяция и реставрация двумерных данных. Одним из ключевых вопросов определения метода восстановления данных является понимание, допустимости применения существующих методов в различных ситуациях. Перед применением метода необходимо понять, что:

- ◆ при сборе исходных данных неизбежно возникла ошибка (точно измерительных приборов, наличие шумов и т.д.);
- ◆ данные были нанесены на двумерную сетку с неким шагом, т.е. фактически была произведена аппроксимация, а закон изменения между двумя соседними точками при выбранном шаге аппроксимации может кардинально различаться со случаем более мелкого шага (например, замеры с шагом 0,01 вместо 0,1);
- ◆ отсутствующие данные могут располагаться как внутри исследуемой области, так и с краю.

Последний случай связан с неполнотой классического понимания понятия интерполяция (методы нахождения точек в пределах заданного интервала), пропущенные значения расположенные на краю исследуемой области, относятся к ней, хотя, фактически, их получение с математической точки зрения должно опираться на понятие экстраполяция (методы нахождения точек за пределами заданного интервала). Тут и приходит на помощь понятие реставрация данных, по аналогии с классической реставрацией, например, старых фотографий, где, опираясь на известные фрагменты, можно попробовать частично, а иногда и полностью, воссоздать исходный вид.

Следует отметить, что оптимальный метод восстановления данных, расположенных на контурах исследуемой области, может кардинально отличаться от метода восстановления данных внутри нее. С учетом этого, в рамках данного исследования, вопрос реставрации пограничных значений не рассматривается.

3. Определение применяемых методов и условий эксперимента. Для построения адекватных моделей были определены несколько исходных неоднородных законов распределения данных в двумерном пространстве. В качестве основных отправных моментов использовались сложные тригонометрические функции с неравноценным влиянием на общую область исследования, например, существующие карты высот, карты распределения температуры и давления. Перед каждым экспериментом для соблюдения достоверности на данные в пределах области был наложен шум (случайные значения в пределах $\pm 5\%$ от разницы между минимальным и максимальным значением):

$$random\left(-\frac{|\max(A) - \min(A)|}{20}, +\frac{|\max(A) - \min(A)|}{20}\right),$$

где A – матрица значений. Это позволило в рамках исследования сделать закон распределения еще более неявным и смоделировать ситуации более близкие к реальным измерениям.

Для упрощения эксперимент был разбит на несколько итераций:

- 1) моделирование задачи для области 10×10 – выявление допустимого порога потери значений, после которого эффективность всех методов сводится на нет;
- 2) моделирование задачи для области 100×100 – общая цель та же самая, но предельное количество итераций заранее определено на основе предыдущего эксперимента;
- 3) моделирование конечной задачи с областью 1000×1000 – детализация двух предыдущих экспериментов.

В каждом случае удалялось 10% случайных значений (при условии сохранности двух крайних линий по обеим осям). Соответственно, допустимый порог доверительного интервала так же был задан 10% .

Для анализа были определены следующие методы:

- 1) заполнение нулевыми значениями;
- 2) заполнение минимальными значениями из известных;
- 3) заполнение максимальными значениями из известных;
- 4) заполнение средними значениями из известных [24];
- 5) заполнение минимальными значениями из окружающих;
- 6) заполнение максимальными значениями из окружающих;
- 7) заполнение средними значениями из окружающих;
- 8) билинейная интерполяция (4 опорные точки, образующие прямоугольную область);
- 9) квадратичная поверхность Безье (16 опорных точек, образующих прямоугольную область);
- 10) бикубическая интерполяция полиномом Лагранжа (16 опорных точек, образующих прямоугольную область);
- 11) бикубическая интерполяция натуральными сплайнами (16 опорных точек, образующих прямоугольную область);
- 12) метод итерационной взвешенной реставрации по соседним значениям [25].

Первый метод фактически не вносит корректировок (кроме случаев, когда известные значения находятся по разные стороны условной оси Oz). Следующие три метода строятся на простом принципе – заполняем одним из известных значе-

ний (в случае со средним – его необходимо сначала высчитать). Эти методы, после минимального набора действий, являются самыми быстрыми, т.к. требуют всего не более двух проходов, правда, по всей матрице.

Следующие три метода можно было бы назвать самыми быстрыми, если бы не случаи, когда окружающие, заполненные значениями ячейки матрицы, необходимо искать не для отдельной точки, а для целой области. В таких случаях число переборов возможных для решения значений может превысить разовый обход всей матрицы для первого метода. Кроме этого, необходимо определить общее минимальное количество известных значений для принятия конечного решения.

Билинейная интерполяция строится на выявление минимальной квадратной (прямоугольной) области вокруг пропущенного значения или области. Для этой операции нам потребуется четыре точки (рис. 2) – вершины нашего прямоугольника – $Q_{11}(x_1, y_1)$, $Q_{12}(x_2, y_1)$, $Q_{21}(x_1, y_2)$, $Q_{22}(x_2, y_2)$. Интерполяция производится пошагово. На первом шаге выявляются значения вспомогательных точек – R (здесь и далее будем считать, что сначала производится интерполяция в горизонтальном направлении, т.е. нахождение значений над и под пропущенными – x_r , а уже затем в вертикальном – y_r):

$$f(R_1) = \frac{x_r - x_1}{x_2 - x_1} \times (f(Q_{12}) - f(Q_{11}));$$

$$f(R_2) = \frac{x_r - x_1}{x_2 - x_1} \times (f(Q_{21}) - f(Q_{22})).$$

На втором – значение искомой точки – P :

$$f(P) = \frac{y_r - y_1}{y_2 - y_1} \times (f(R_2) - f(R_1)).$$

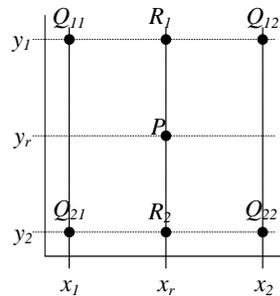


Рис. 2. Билинейная интерполяция

Интерполяция с помощью *поверхности Безье* (рис. 3) фактически можно считать идентичной билинейной линейной по сложности ее построения. Для ее построения будем брать 16 точек – четыре ряда по четыре точки (Q_{11} – Q_{44}). Вторая и третья точки будут крайними опорными (Bi_0 и Bi_2). Первая и вторая пара точек будут образовывать две прямых, на пересечении которых будет находиться промежуточная опорная точка (Bi_1) квадратичной кривой Безье:

$$\begin{cases} y_{\bar{i}} = \frac{(Q_{i1} - Q_{i2})x_{\bar{i}} + (x_1 Q_{i2} - x_2 Q_{i1})}{x_1 - x_2}; \\ y_{\bar{i}} = \frac{(Q_{i3} - Q_{i4})x_{\bar{i}} + (x_3 Q_{i4} - x_4 Q_{i3})}{x_3 - x_4}; \end{cases}$$

$$f(R_i) = \left(1 - \frac{x_r - x_2}{x_3 - x_2}\right)^2 B_{i0} + 2 \frac{x_r - x_2}{x_3 - x_2} \left(1 - \frac{x_r - x_2}{x_3 - x_2}\right) B_{i1} + \left(\frac{x_r - x_2}{x_3 - x_2}\right)^2 B_{i2}.$$

После определения ключевых значений каждой квадратичной кривой аналогичным образом определим значение конечной точки P :

$$\begin{cases} x_p = \frac{(f(R_1) - f(R_2))y_p + (y_1 f(R_2) - y_2 f(R_1))}{y_1 - y_2} \\ x_p = \frac{(f(R_3) - f(R_4))y_p + (y_3 f(R_4) - y_4 f(R_3))}{y_3 - y_4} \end{cases}$$

$$f(P) = \left(1 - \frac{y_r - y_2}{y_3 - y_2}\right)^2 f(R_2) + 2 \frac{y_r - y_2}{y_3 - y_2} \left(1 - \frac{y_r - y_2}{y_3 - y_2}\right) P + \left(\frac{y_r - y_2}{y_3 - y_2}\right)^2 f(R_3).$$

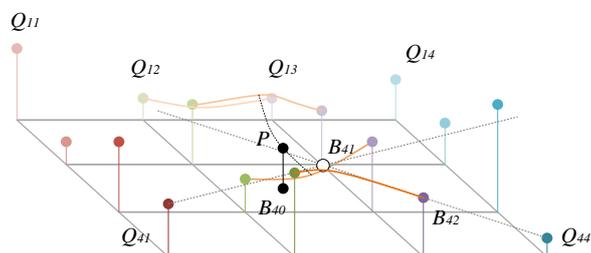


Рис. 3. Интерполяция с помощью поверхности Безье

При интерполяции с помощью *полинома Лагранжа* был выбран полином третьей степени, формирующий поверхность. Общая методика аналогична предыдущим – сначала строятся четыре опорных полинома и на каждом находится ключевая точка, потом по полученным точкам формируется конечный полином и определяется искомое значение (рис. 4):

$$L_i(x_r) = \sum_{j=1}^4 Q_j l_{ij}(x_r);$$

$$l_{ij}(x_r) = \prod_{k=1, k \neq j}^4 \frac{x_r - x_k}{x_j - x_k};$$

$$L(y_p) = \sum_{i=1}^4 L_i l_i(y_p);$$

$$l_i(y_p) = \prod_{j=1, j \neq i}^4 \frac{y_p - y_j}{y_i - y_j}.$$

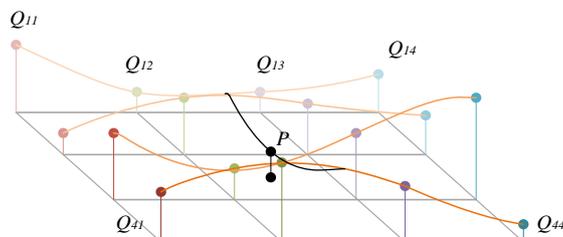


Рис. 4. Интерполяция с помощью полинома Лагранжа

Бикубическая интерполяция сплайнами один из самых популярных методов интерполяции двумерных массивов данных (например, в компьютерной графике). Существует множество методов построения сплайнов (отрезков функций, образующих общую поверхность). В нашем случае будем применять натуральные сплайны (рис. 5):

$$\begin{cases} f_{i1}(x_1) = Q_{i1} \\ f_{i1}(x_2) = Q_{i2} \\ f_{i2}(x_2) = Q_{i2} \\ f_{i2}(x_3) = Q_{i3} \\ f_{i3}(x_3) = Q_{i3} \\ f_{i3}(x_4) = Q_{i4} \\ f_{i1}(x_2) = f_{i2}(x_2); \\ f_{i2}(x_3) = f_{i3}(x_3) \\ f_{i1}(x_2) = f_{i2}(x_2) \\ f_{i2}(x_3) = f_{i3}(x_3) \\ f_{i1}(x_1) = 0 \\ f_{i3}(x_4) = 0 \end{cases};$$

$$R_i(x_r, f_{i2}(x_r)).$$

В общем виде методика сводится к получению все тех же четырех промежуточных значений (R_i) и нахождению искомого (P) на построенном по ним аналогичным образом конечном сплайне:

$$P(y_p, f_2(y_p)).$$

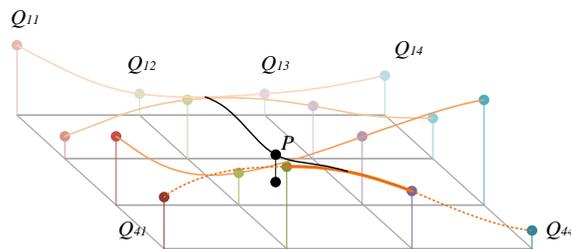


Рис. 5. Бикубическая интерполяция сплайнами

Последний метод, подробно описанный в [25], является разновидностью ZET-алгоритма [23, 26] или метода 7. Отличием от метода 7 является появление понятия вес значения, определяющийся на каждом шаге линейно относительно количества шагов, прошедших от появления значения в матрице данных. Стоит отметить, что сам метод работает лишь в комплексе с алгоритмом устранения выбросов и является частью комплекса методов поиска решения.

4. Анализ методов. В ходе эксперимента замерялось среднее отклонение конечного набора данных от исходного. Для эксперимента были взяты матрицы размерами 10×10 , 100×100 и 1000×1000 , каждая последующая размерность которых была детализацией предыдущей. На рис. 6, 7 отображено поведение алгоритмов при первых 10 и 100 экспериментах, соответственно.

На основе полученных данных можно сделать выводы, что чаще всего наиболее адекватны исходным данным результаты методов 7, 8 и 12. Методы 7 (средние значения из известных окружающих) и 12 (итерационная взвешенная реставрация по соседним значениям) достаточно близки по реализации, метод 8 (билинейная интерполяция) использует малое значение (4 точки) вокруг искомого

го значения, что уменьшает искажение закона распределения данных по сравнению с поиском полностью описанной прямоугольной области, используемой в методах 9–11.

При этом стоит отметить, что при увеличении количества итераций и общего количества значений в исследуемом пространстве (матрице) положительно себя показывают методы, основанные на кубической интерполяции – менее 3 % отклонения. В то же время, метод 9 (квадратичная поверхность Безье), показавший в отдельных испытаниях как лучший, так и худший результат, показывает сложность построения адекватной поверхности для решения поставленной задачи, если она опирается на достаточно далеко разнесенные точки.

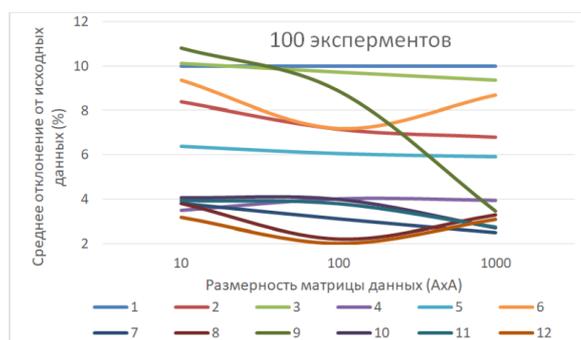


Рис. 6. Моделирование при 100 экспериментах

5. Формирование комплексного метода оценки оптимальных способов реставрации данных. Одной из целей проводимого исследования было исключение наименее точных методов интерполяции из перечня. Соответственно, результатом эксперимента стало ожидаемое полное исключение следующих методов:

- ◆ заполнение нулевыми значениями;
- ◆ заполнение минимальными значениями из известных;
- ◆ заполнение максимальными значениями из известных;
- ◆ заполнение максимальными значениями из окружающих.

Построение квадратичной поверхности Безье показало свою эффективность лишь на больших наборах значений. Остальные методы были достаточно эффективны.

Следующим этапом являлось построение метода выбора способа интерполяции для частных случаев и формирование обобщенного алгоритма. Комплексный подход заключался в анализе имеющихся фрагментов данных, поиске в них достаточно непрерывных участков, моделировании потери имеющихся данных и попытка их реставрации каждым из методов. За небольшое количество итераций (не более 10) набирается достаточно информации об отклонениях, чтобы определить, каким методом произвести полную реставрацию набора данных. В ходе экспериментов было определено, что: – для наборов данных, подчиняющихся, в целом, гладким законам (градиентные переходы, карты местности, тепловые карты и т.д.) наиболее успешными стали интерполяционные методы (8, 10, 11);

- ◆ для зашумленных и стохастических наборов данных успешнее работали методы, основанные на соседних значениях (4, 5, 7);
- ◆ для достаточно больших наборов данных ряде случаев с выгодной стороны себя показали методы 9 и 12;
- ◆ результат применения любого метода более гладкий, чем исходные данные.

Полученный в ходе эксперимента обобщенный алгоритм представлен на рис. 7. N и M – размеры матрицы значений; в случае, если обрабатывается не прямоугольная область, блок с ними можно модифицировать.

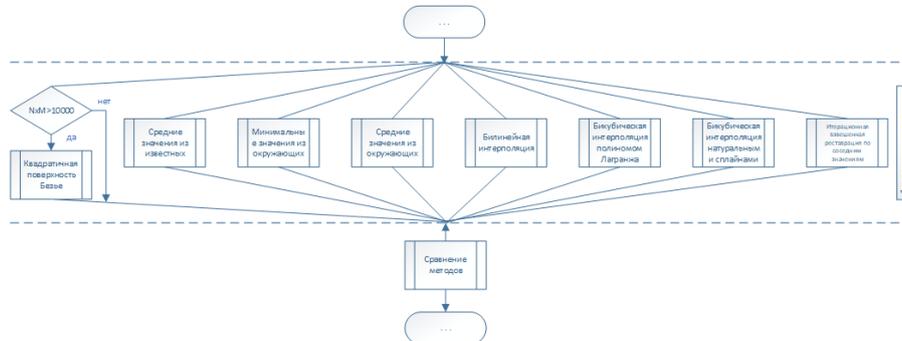


Рис. 7. Обобщенный алгоритм

Рассмотрим для примера вариант реставрации данных статистических данных. Имеется тепловая карта (рис. 9, слева), на которой присутствует дефект – в ходе замера температуры часть датчиков не работала (черная область). Попробуем провести реставрацию данных. Для этого:

- ◆ добавим контрольную область для проверки “качества” работы алгоритма (рис. 9., справа – второе черное пятно);
- ◆ выделим 3 достаточно большие прямоугольные области без поврежденных данных (рис. 9, справа);
- ◆ выполним поиск наиболее оптимального метода по обобщенному алгоритму (рис. 8).

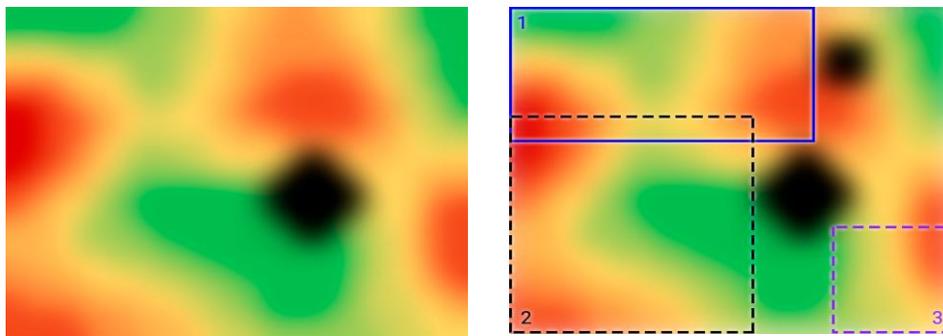


Рис. 9. Поврежденные данные: слева – исходные данные, черная область – отсутствующие значения, справа – добавлено контрольное повреждение данных и выделены 3 области для анализа

Общий объем данных слишком мал, чтобы применять метод 9, поэтому исключим его из списка анализируемых алгоритмов. Результат анализа приведен на рис. 10. Наиболее стабильно вели себя алгоритмы 7 и 4, но наименьшее среднее значение отклонения показал алгоритм 11 – бикубическая интерполяция натуральными сплайнами.



Рис. 10. Поиск оптимального алгоритма на областях для анализа

После реставрации с помощью выбранного алгоритма (подробнее было написано в пункте 3) получается полная температурная карта (рис. 11, слева). Достоверность работы алгоритма определим с помощью контрольного повреждения. Среднее отклонение значений в контрольной области получилось 2,79 % (рис. 11, справа – верхняя область). На рис. 11 отклонение показано с градациями серого. Для исходного повреждения разница считалась с 0.

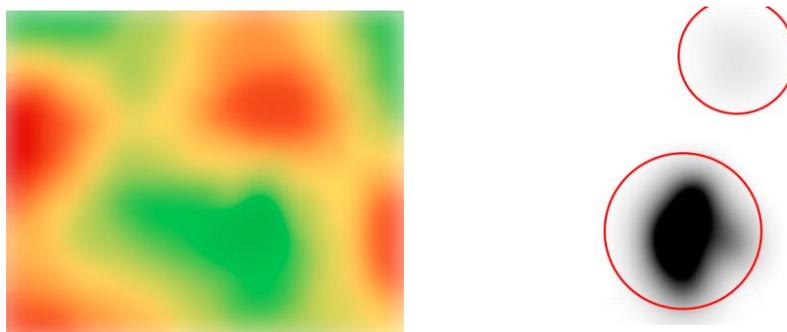


Рис. 11. Результат работы алгоритма: слева – восстановленные данные, справа – детализация разницы

Дальнейшие эксперименты с применением этого метода показали, что наиболее эффективным является выявление двух-трех наиболее оптимальных методов (чем больше исследуемых значений, тем больше методов стоит брать), реставрация с их помощью и усреднение конечных результатов. В таком случае показатель отклонения по отдельным случаям становится робастным, а среднее отклонение укладывается в пределах 2,25 %. В то же самое время попытка восстановления растровых изображений (JPEG со сжатием до 75 %) не дает желаемого результата с точки зрения визуального восприятия.

Заключение. Результатом поставленного эксперимента стало общее представление адекватности результатов некоторых из существующих методов восстановления потерянных экспериментальных данных исходным данным и получение комплексного метода оценки оптимальных способов реставрации данных. В качестве наиболее оптимальных методов для интерполяции (реставрации) потерянных данных, опираясь на результаты исследования, можно порекомендовать методы, основанные на определении среднего значения соседних с потерянным. В случае длительных экспериментов и обработки больших объемов данных, возможно, лучше себя покажут бикубические методы.

Независимо от результатов данного эксперимента, для каждого случая правильным будет провести базовый анализ, например, по приведенной выше методике. На основе его результатов можно будет определить один или несколько методов, с помощью которых можно провести восстановление данных и дальнейшую их обработку. Также в дальнейшем можно будет разработать кворум функцию, которая будет: оптимизировать существующие алгоритмы, создавать смешанные алгоритмы, выбирать оптимальный алгоритм в зависимости от начальных условий и объема статистических данных. Кроме того, в дальнейшем возможна модификация алгоритма с учетом градиентного и SVR методов.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Wei Liu, Dongmei Zhang, Ao Wang*. Research of Spatial Data Interpolation Algorithm Based on SVR Optimization by GA // *Computational Intelligence and Intelligent Systems*. – 2010. – P. 137-145.
2. *Xihua Yang, Xiaojin Xie, De Li Liu, Fei Ji, Lin Wang*. Spatial Interpolation of Daily Rainfall Data for Local Climate Impact Assessment over Greater Sydney Region // *Advances in Meteorology*. – 2015.
3. *Zhihua Tian, Bin Xie, Yong Pan, Ningbo Wang, Hao Cheng*. Uermainaiti Wumaietjiang. Data Interpolation Method Based on Minimum Total Variation Method // *Proceedings of the 2019 International Conference on Artificial Intelligence and Computer Science*. – 2019. – P. 682-685.
4. *Зайцева Е.В.* Выбор метода интерполяции при решении задачи статистического восстановления непрерывных распределений геологических параметров при построении информационных систем экологического мониторинга // *Горный информационно-аналитический бюллетень*. – 2002. – № 2. – С. 197-198.
5. *Мехова Г.А., Руденко Н.В., Трофименко В.Н. Трофименко Е.Н.* Интерполяция экспериментальных данных при определении характеристик и параметров нелинейных электронных компонентов // *Тр. северо-кавказского филиала московского технического университета связи и информатики*. – 2014. – № 1. – С. 278-280.
6. *Аюшев Т.В., Булычев Р.Н.* Моделирование параметрических твердых тел с применением обобщенной линейной интерполяции // *26-я международная конференция "Graphicson 2016" АНО научного общества «Графикон» и Нижегородского государственного архитектурно-строительного университета*. – 2016. – С. 72-76.
7. *Жаринов И.О., Жаринов О.О.* Интерполяция линии спектральных цветностей цветового локуса на основе сплайнов Безье // *Известия высших учебных заведений. Приборостроение*. – 2015. – Т. 58 (12). – С. 985-992.
8. *Abd Wahab, Rozaimi Zakaria, Jamaludin Ali*. Fuzzy Interpolation Rational Bezier Curve // *Ninth International Conference on Computer Graphics, Imaging and Visualization*. – 2012. – P. 63-67.
9. *Аюшев Т.В., Булычев Р.Н.* Моделирование параметрических рациональных тел с использованием обобщенной интерполяции Безье // *Вестник Бурятского государственного университета. Математика, информатика*. – 2018. – № 1. – С. 83-94.
10. *Аюшев Т.В., Павлова С.В., Булычев Р.Н.* Геометрическое моделирование многослойных конструкций с применением кривых и поверхностей безье // *Динамика систем, механизмов и машин*. – 2016. – № 3. – С. 125-128.
11. *Федорук В.А.* Обработка экспериментальных данных на основе "Методики сдвига" при интерполяции кубическими сплайнами // *Вестник сибирской государственной автомобильно-дорожной академии*. – 2016. – № 2 (48). – С. 132-136.
12. *Ромаданова М.М.* Алгоритмы построения монотонного весового кубического сплайна // *Известия тульского государственного университета. Технические науки*. – 2018. – № 9. – С. 180-192.
13. *Yanlin Geng, Tong Lin, Zhouchen Lin, Pengwei Hao*. Refined Exponential Filter with Applications to Image Restoration and Interpolation. *Computer Vision - ACCV 2009*. – P. 33-42. – https://www.researchgate.net/publication/220745293_Refined_Exponential_Filter_with_Applications_to_Image_Restoration_and_Interpolation.
14. *Щерба Е.В.* Анализ применимости методов интерполяции и экстраполяции для решения задачи восстановления изображения // *Компьютерная оптика*. – 2009. – Т. 33 (3). – С. 336-339.

15. *Jinyu Chu, Ju Liu, Jianping Qiao, Xiaoling Wang, Yujun Li.* Gradient-based adaptive interpolation in super-resolution image restoration // International Conference on Signal Processing Proceedings. – 2009.
16. *Tudor Barbu.* Novel Diffusion-Based Models for Image Restoration and Interpolation. – 2019.
17. *Половко А.М., Бутусов П.Н.* Интерполяция. Методы и компьютерные технологии их реализации. – М.: БХВ-Петербург, 2016. – 320 с.
18. *Рассел Джесси.* Бикубическая интерполяция. – М.: Книга по Требованию, 2013. – 814 с.
19. *Асташкин С.В.* Интерполяция операторов и ее приложения. – М.: Книга по требованию, 2013. – 188 с.
20. *Крейн С.Г., Петунин Ю.И., Семенов Е.М.* Интерполяция линейных операторов. – М.: Главная редакция физико-математической литературы издательства "Наука", 2014. – 400 с.
21. *Михеев С.Е.* Многомерная аппроксимация и интерполяция. – СПб.: Изд-во С.-Петерб. ун-та, 2012. – 59 с.
22. *Уолли Дж.Л.* Интерполяция и аппроксимация рациональными функциями в комплексной области. – М.: Изд-во иностранной литературы, 1961. – 508 с.
23. *Мартышенко С.Н.* Методы восстановления пропусков в данных, представленных в различных измерительных шкалах // Территория новых возможностей. Вестник Владивостокского государственного университета экономики и сервиса. – 2013. – № 4 (22). – С. 242-255.
24. *Злоба Е., Яцкив И.* Статистические методы восстановления пропущенных данных // Computer Modelling & New Technologies. – 2002. – Vol. 6, No. 1. – P. 51-61.
25. *Сорокин А.А., Коваленко Р.А., Яковлева Е.А.* Прогнозирование стоимости аренды складских помещений на основе статистических данных // Евразийский союз ученых. – 2018. – № 12 (57). – С. 59-62.
26. *Рыженкова К.В.* Методы восстановления пропуска данных при проведении статистических исследований // Интеллект. Инновации. Инвестиции. – 2012. – № 3. – С. 127-133.

REFERENCES

1. *Wei Liu, Dongmei Zhang, Ao Wang.* Research of Spatial Data Interpolation Algorithm Based on SVR Optimization by GA, *Computational Intelligence and Intelligent Systems*, 2010, pp. 137-145.
2. *Xihua Yang, Xiaojin Xie, De Li Liu, Fei Ji, Lin Wang.* Spatial Interpolation of Daily Rainfall Data for Local Climate Impact Assessment over Greater Sydney Region, *Advances in Meteorology*, 2015.
3. *Zhihua Tian, Bin Xie, Yong Pan, Ningbo Wang, Hao Cheng.* Uermaimaiti Wumaierjiang. Data Interpolation Method Based on Minimum Total Variation Method, *Proceedings of the 2019 International Conference on Artificial Intelligence and Computer Science*, 2019, pp. 682-685.
4. *Zaytseva E.V.* Vybór metoda interpolýatsii pri reshenii zadachi statisticheskogo vosstanovleniya nepreryvnykh raspredeleniy geologicheskikh parametrov pri postroenii informatsionnykh sistem ekologicheskogo monitoringa [Selection of the interpolation method for solving the problem of statistical restoration of continuous distributions of geological parameters in the construction of information systems for environmental monitoring], *Gornyy informatsionno-analiticheskiy byulleten'* [Mining information and analytical Bulletin], 2002, No. 2, pp. 197-198.
5. *Mekhova G.A., Rudenko N.V., Trofimenko V.N., Trofimenko E.N.* Interpolyatsiya eksperimental'nykh dannykh pri opredelenii kharakteristik i parametrov nelineynykh elektronnykh komponentov [Interpolation of experimental data in determining the characteristics and parameters of nonlinear electronic components], *Tr. severo-kavkazskogo filiala moskovskogo tekhnicheskogo universiteta svyazi i informatiki* [Proceedings of the North Caucasus branch of the Moscow technical University of communications and Informatics], 2014, No. 1, pp. 278-280.
6. *Ayushev T.V., Bulychev R.N.* Modelirovanie parametricheskikh tverdykh tel s primeneniem obobshchennoy lineynoy interpolyatsii [Modeling of parametric solids using generalized linear interpolation], *26-ya mezhdunarodnaya konferentsiya "Graphicon 2016" ANO nauchnogo obshchestva «Grafikon» i Nizhegorodskogo gosudarstvennogo arkhitekturno-stroitel'nogo universiteta* [26th International Conference "Graphicon 2016" of the Scientific Society "Grafikon" and Nizhny Novgorod State Architectural and Construction University], 2016, pp. 72-76.

7. Zharinov I.O., Zharinov O.O. Interpolyatsiya linii spektral'nykh tsvetnostey tsvetovogo lokusa na osnove splaynov Bez'e [Interpolation of the spectral chromaticity line of a color locus based on Bezier splines], *Izvestiya vysshikh uchebnykh zavedeniy. Priborostroenie* [Izvestia of higher educational institutions. Equipment], 2015, Vol. 58 (12), pp. 985-992.
8. Abd Wahab, Rozaimi Zakaria, Jamaludin Ali. Fuzzy Interpolation Rational Bezier Curve, *Ninth International Conference on Computer Graphics, Imaging and Visualization*, 2012, pp. 63-67.
9. Ayushev T.V., Bulychev R.N. Modelirovanie parametricheskikh ratsional'nykh tel s ispol'zovaniem obobshchennoy interpolyatsii Bez'e [Modeling of parametric rational bodies using generalized interpolation Bezier], *Vestnik Buryatskogo gosudarstvennogo universiteta. Matematika, informatika* [Vestnik Buryat State University. Mathematics, informatics], 2018, No. 1, pp. 83-94.
10. Ayushev T.V., Pavlova S.V., Bulychev R.N. Geometricheskoe modelirovanie mnogoslonykh konstruktivnykh s primeneniem krivyykh i poverkhnostey bez'e [Geometrical modeling of multi-layered constructions using curves and bezier surfaces], *Dinamika sistem, mekhanizmov i mashin* [Dynamics of systems, mechanisms and machines], 2016, No. 3, pp. 125-128.
11. Fedoruk V.A. Obrabotka eksperimental'nykh dannykh na osnove "Metodiki sdviga" pri interpolyatsii kubicheskimi splaynami [Processing of experimental data on the basis of the "Offset Methodology" for interpolation of cubic splines], *Vestnik sibirskoy gosudarstvennoy avtomobil'no-dorozhnoy akademii* [Bulletin of the Siberian State Road and Road Academy], 2016, No. 2 (48), pp. 132-136.
12. Romadanova M.M. Algoritmy postroeniya monotonnogo vesovogo kubicheskogo splayna [Algorithms of construction of monotonous weight cubic spline], *Izvestiya tul'skogo gosudarstvennogo universiteta. Tekhnicheskie nauki* [Izvestia of Tula State University. Technical sciences], 2018, No. 9, pp. 180-192.
13. Yanlin Geng, Tong Lin, Zhouchen Lin, Pengwei Hao. Refined Exponential Filter with Applications to Image Restoration and Interpolation. *Computer Vision - ACCV 2009*, pp. 33-42. Available at: https://www.researchgate.net/publication/220745293_Refined_Exponential_Filter_with_Applications_to_Image_Restoration_and_Interpolation.
14. Shcherba E.V. Analiz primenimosti metodov interpolyatsii i ekstrapolyatsii dlya resheniya zadachi vosstanovleniya izobrazheniya [Analysis of the applicability of interpolation and extrapolation methods to solve the problem of image recovery], *Komp'yuternaya optika* [Computer optics], 2009, No. 33 (3), pp. 336-339.
15. Jinyu Chu, Ju Liu, Jianping Qiao, Xiaoling Wang, Yujun Li. Gradient-based adaptive interpolation in super-resolution image restoration, *International Conference on Signal Processing Proceedings*, 2009.
16. Tudor Barbu. Novel Diffusion-Based Models for Image Restoration and Interpolation, 2019.
17. Polovko A.M., Butusov P.N. Interpolyatsiya. Metody i komp'yuternye tekhnologii ikh realizatsii [Interpolation. Methods and computer technologies of their implementation]. Moscow: BKhV-Peterburg, 2016, 320 p.
18. Rassel Dzhessi. Bikubicheskaya interpolyatsiya [Bikubic Interpolation]. Moscow: Kniga po Trebovaniyu, 2013, 814 p.
19. Astashkin S.V. Interpolyatsiya operatorov i ee prilozheniya [Interpolation of operators and its applications]. Moscow: Kniga po trebovaniyu, 2013, 188 p.
20. Kreyin S.G., Petunin Yu.I., Semenov E.M. Interpolyatsiya lineynykh operatorov [Interpolation of line operators]. Moscow: Glavnaya redaktsiya fiziko-matematicheskoy literatury izdatel'stva "Nauka", 2014, 400 p.
21. Mikheev S.E. Mnogomernaya approksimatsiya i interpolyatsiya [Multidimensional approximation and interpolation]. Saint Petersburg: Izd-vo S.-Peterb. un-ta, 2012, 59 p.
22. Uolsh Dzh.L. Interpolyatsiya i approksimatsiya ratsional'nymi funktsiyami v kompleksnoy oblasti [Interpolation and approximation of rational functions in the complex area]. Moscow: Izd-vo inostrannoy literatury, 1961, 508 p.
23. Martysenko S.N. Metody vosstanovleniya propuskov v dannykh, predstavlenykh v razlichnykh izmeritel'nykh shkalakh [Methods for recovering gaps in data presented in various measuring scales], *Territoriya novykh vozmozhnostey. Vestnik Vladivostokskogo gosudarstvennogo universiteta ekonomiki i servisa* [Territory of new opportunities. Bulletin of the Vladivostok State University of Economics and Service], 2013, No. 4 (22), pp. 242-255.

24. Zloba E., Yatskiv I. Statisticheskie metody vosstanovleniya propushchennykh dannykh [Statistical methods for recovering missing data], *Computer Modelling & New Technologies*, 2002, Vol. 6, No. 1, pp. 51-61.
25. Sorokin A.A., Kovalenko R.A., Yakovleva E.A. Prognozirovaniye stoimosti arendy skladskikh pomeshcheniy na osnove statisticheskikh dannykh [Forecasting the cost of rental warehouse based on statistical data], *Evraziyskiy soyuz uchenykh* [Eurasian Union of Scientists], 2018, No. 12 (57), pp. 59-62.
26. Ryzhenkova K.V. Metody vosstanovleniya propuska dannykh pri provedenii statisticheskikh issledovaniy [Methods for recovering data gaps in statistical studies], *Intellekt. Innovatsii. Investitsii* [Intelligence. Innovation. Investments], 2012, No. 3, pp. 127-133.

Статью рекомендовал к опубликованию д.т.н. С.В. Колесниченко.

Сорокин Алексей Андреевич – Ивангородский гуманитарно-технический институт (филиал) Санкт-петербургского государственного университета аэрокосмического приборостроения; e-mail: A.A.Sorokin@ifguar.ru; 188491, Ивангород, ул. Котовского, 1; кафедра прикладной математики, информатики и информационных таможенных технологий; старший преподаватель.

Дагаев Александр Владимирович – e-mail: adagaev@list.ru; кафедра прикладной математики, информатики и информационных таможенных технологий; к.т.н.; доцент.

Бородинский Илья Михайлович – Южный федеральный университет; e-mail: iborodyanskiy@sfedu.ru; 347928, г. Таганрог, пер. Некрасовский, 44; тел.: +79185027169; кафедра информационных измерительных технологий и систем; доцент.

Sorokin Aleksei Andreevich – St. Petersburg University of Aerospace Instrumentation; e-mail: A.A.Sorokin@ifguar.ru; 1, Kotovskogo street, Ivangorod, 188491, Russia; the department of applied mathematics, computer science and customs information technologies; senior lecturer.

Dagaev Aleksandr Vladimirovich – e-mail: adagaev@list.ru; the department of applied mathematics, computer science and customs information technologies; cand. of eng. sc.; associate professor.

Borodyansky Ilya Mikhailovich – Southern Federal University; e-mail: iborodyanskiy@sfedu.ru; 44, Nekrasovskiy lane, Taganrog, 347928, Russia; phone: +79185027169; the department of information measuring technologies and systems; associate professor.

УДК 519.27

DOI 10.18522/2311-3103-2020-4-107-117

Г.Ф. Филаретов, З. Бучаала

НЕПАРАМЕТРИЧЕСКИЙ МЕТОД ОБНАРУЖЕНИЯ РАЗЛАДКИ ВРЕМЕННОГО РЯДА С ИСПОЛЬЗОВАНИЕМ МЕХАНИЗМА СЛУЧАЙНЫХ БЛУЖДЕНИЙ

Рассмотрена задача оперативного обнаружения внезапного изменения вероятностных свойств временного ряда, обычно трактуемая как задача обнаружения разладки наблюдаемого стохастического процесса. Отмечается актуальность развития исследований по данной тематике, что обусловлено появлением всё новых прикладных задач, где методы и алгоритмы обнаружения разладки могут успешно использоваться – в частности, при создании мониторинговых систем в промышленности, экологии, медицине и др. Обсуждаются две основные разновидности методов обнаружения разладки: параметрические и непараметрические. Отмечено, что, хотя непараметрические методы при прочих равных условиях уступают параметрическим по эффективности (быстроте обнаружения разладки), но зато обладают рядом преимуществ, не требуя, в частности, контролируемого процесса. Это принципиально важно при построении мониторинговых систем, когда детальная информация об этих свойствах может либо полностью отсутствовать и тогда необходимо проводить достаточно трудоемкое его предварительное исследование, либо быть малодостоверной. Предложен оригинальный последовательный непараметрический алгоритм обнаружения разладки на основе