МЕТОД РАСПОЗНАВАНИЯ В ИНТЕРАКТИВНОМ РЕЖИМЕ АНОМАЛИЙ ГРАДИЕНТОВ СКАЛЯРНЫХ ПОЛЕЙ НАБЛЮДЕНИЙ

Ю.Е. Шишкин^{1,2}, А.В. Скатков¹

¹ ФГАОУ ВО «Севастопольский государственный университет»,
РФ, г. Севастополь, ул. Университетская, 33

² Институт природно-технических систем,
РФ, г. Севастополь, ул. Ленина, 28
E-mail: yourockpro@gmail.com

Предлагается подход к решению задачи оперативного распознавания аномалий в данных скалярных полей мониторинга объектов и процессов с использованием градиентного метода и системы специфичных информационных метрик среднеквадратического отклонения модуля и направления разности градиентных полей. Разработан метод, обеспечивающий работу системы бинарного классификатора, построена математическая модель и процедура распознавания, работающая в интерактивном режиме. Проиллюстрирована работа предложенного подхода на примере выборки данных вертикального пространственного зондирования температуры акватории вод города Севастополя в районе бухты Круглая.

Ключевые слова: мониторинг, математическое моделирование, Большие Данные, цифровая фильтрация, градиент, скалярное поле, детектирование аномалий, кластеризация, критические системы, интеллектуальный анализ данных.

Поступила в редакцию: 24.04.2018.

Введение. В настоящее время оперативное распознавание аномалий в градиентах скалярных полей наблюдений динамических сложных систем, объектов и процессов является актуальной научной задачей и лежит в основе многих инженерных, экологических задач, прогноза погоды, моделирования изменения климата и др. [1].

Во многих работах рассматривалась задача определения аномалий скалярных полей [2, 3]. Для повышения качества принимаемых решений и повышения оперативности важно контролировать и градиент, который значительно более чувствительный к изменению самой скалярной величины и позволяет повысить оперативность выявления значимых изменений абсолютных значений и динамики скалярных полей [4, 5].

Технология Больших Данных особенно способствует анализу градиентов полей, поскольку более-менее точное определение градиентов представляет собой объемную вычислительную задачу [6]. Если для регистрации скалярных полей наблюдений успешно используются стандартные технологии, то определение градиентов этих полей с достаточно малым шагом сетки требует применение технологии Больших Данных [7]. По-

скольку оценивание градиента при тех же объемах исходных данных дает ошибку на порядок выше, для обеспечения той же точности шаг сетки должен быть на порядок меньше [8].

Следует отметить, что наиболее энергичное развитие средств и методов измерения скалярных полей температуры акватории стало возможным благодаря технологическому прогрессу в области гидрологических средств автоматизированного мониторинга ключевых параметров, развитию элементной базы и информационных систем, позволяющих осуществлять сбор, синхронизацию и анализ больших массивов данных в реальном времени [9]. Важной особенностью скалярного поля температур является его вариационная изменчивость особенно в верхних слоях акватории, поэтому применение градиентных методов позволяет частично решить вызванные этим проблемы существенно снизив влияние вариаций [10].

Постановка задачи. Разработать метод, который позволяет построить систему бинарного классификатора, с заданной чувствительностью, к возникающим аномалиям (управляемым возмущениям) по данным, представленным в виде множества значений замеров в не-

которых точках скалярного поля, с использованием ортогонального градиентного фильтра.

Определить чувствительность численных метрик среднеквадратического отклонения модуля и направления разностей градиентов эталонных и фактических скалярных полей температуры к изменениям динамики течений водной системы и изменению общего температурного фона. Выявить условия, при которых эти метрики обладают избирательными характеристиками чувствительности.

Материалы и методы. Для иллюстрации работы предложенного подхода в качестве исходного входного потока используется выборка данных вертикального пространственного зондирования температуры акватории вод города Севастополя в районе бухты Круглая [11], при этом задан набор данных $A(x, y, t)$, однозначно задающий значение температуры t в точках x – ширина бухты и y – горизонт (глубина) замера.

Суть метода заключается в том, что градиент измеренного скалярного поля температур сравнивается с координатно привязанным эталоном скалярного поля, и по среднеквадратическому отклонению разности компонентов сравниваемых градиентных полей делается вывод о наличии аномалии [12, 13].

Математическая модель, лежащая в основе предлагаемого метода. Зададим скалярное поле в двумерном пространстве следующим образом:

Предполагается, что заданы область $D \subset R_2$ и правило, задающее отображение $M(x, y) \in D \xrightarrow{f} w \in R_1$, согласно которому каждой точке M по правилу f ставится в соответствие число w . Тогда скалярное поле выражается функцией $w = f(M) = f(x, y)$.

На практике определение значений параметров мониторинга w в каждой точке области наблюдения D не представляется возможным, зачастую используется достаточно большие значения Δx и Δy в связи с высокими стоимостными и временными затратами на проведение замеров. Поэтому, в таких случаях, имеют дело с аппроксимирующей математической моделью, на основе, например, линейной или сплайновой интерполяции, заданной в виде системы регрессионных уравнений, итерационного сглаживания и др. Эта задача является практической, за нее брались самые высокие умы начиная от Эйлера, Лагранжа и Чебышева и до сих пор она удовлетворительно в общем виде не разрешена. Тем более оправдано использование интерактивных режимов, в которых исследователь будет выбирать тот или иной способ аппроксимации либо их комбинацию ориентируясь на сущность решаемой задачи и на свои предпочтения.

Представим структурную схему бинарного классификатора, реализованного в системе поддержки принятия решений (рис. 1).

Рис. 1. Структура математического обеспечения бинарного классификатора

Пусть заданы множества X и Y такие что определяют правило $M(x \in X, y \in Y) \xrightarrow{f} w$ только в некоторых точках, а в остальных точках $M(x, y) \in D$ функция не определена,

$$f(M(x, y)) = f_m(M(x, y)) \forall x \in X, y \in Y, \quad (1)$$

при этом f_m определена для всех $M(x, y) \in D$.

Для задания такого правила в общем случае используются различные методы аппроксимации функции двух и более переменных, при этом для сложных систем вместо условия (1) может производиться интерполяция с использованием критерия минимума среднеквадратической ошибки [14, 15].

Рассмотрим пример применения метода определения аномалий на скалярном поле температур, полученных путем осуществления вертикального зондирования по всей ширине бухты Круглая г. Севастополя, в летний период.

Пусть имеется матрица $A(x, y, t)$, однозначно задающая значение температуры t в точках $M(x \in X, y \in Y)$ на плоскости L скалярного поля $D \subset R_2$, точки M – станции наблюдений. Обозначим A^i как i -й элемент матрицы A , а $N = |A|$ – число строк матрицы A .

Будем считать, что заданная плоскость L соответствует вертикальному

тогда мы имеем дело со скалярным дискретно заданным полем. Зададим математическую модель в виде правила f_m такого, что удовлетворяет условию:

среза температуры водной среды по всей ширине бухты.

На основании данных матрицы A можно задать функцию скалярного поля в виде $f: (A_x, A_y) \rightarrow A_t$, при этом функция задается однозначно в случае выполнения условия [16]:

$$\forall i, j \in [0, |A|] \quad \exists (A_x^i, A_y^i) = (A_x^j, A_y^j). \quad (2)$$

Таким образом заданное скалярное поле – это поле температур, при этом известны значения этого поля в некоторых точках, заданное в виде системы функций:

$$\begin{cases} f(1, 1) = 24.7 \\ f(1, 5) = 23.8 \\ \dots \\ f(180, 20) = 21.6 \end{cases}.$$

Для реализации функции f осуществим матричное преобразование (3), которое удовлетворяет условиям (2) и (4):

$$A = \begin{bmatrix} A_x^0 & A_y^0 & A_t^0 \\ A_x^1 & A_y^1 & A_t^1 \\ \dots & \dots & \dots \\ A_x^N & A_y^N & A_t^N \end{bmatrix} \rightarrow B = \begin{bmatrix} B_{\min(A_x), \min(A_y)} & \dots & B_{\max(A_x), \min(A_y)} \\ \vdots & \ddots & \vdots \\ B_{\min(A_x), \max(A_y)} & \dots & B_{\max(A_x), \max(A_y)} \end{bmatrix}, \quad (3)$$

$$\forall i \in [0, |A|] \quad B_{A_x^i, A_y^i} = A_t^i. \quad (4)$$

Вообще говоря, условие (2) не является обязательным, однако в случае если оно не было выполнено, преобразование будет недетерминированным и как следствие необратимым.

Определим поведение скалярного поля во всех точках пространства D , для этого разобьем это пространство на множество плоскостей T , заданных треугольниками, значения скаляров в вершинах которых определены по условию.

$$T = \left[\langle B_{x_0, y_0}, B_{x_1, y_0}, B_{x_0, y_1} \rangle, \langle B_{x_1, y_1}, B_{x_1, y_0}, B_{x_0, y_1} \rangle, \dots, \langle B_{x_{\max}, y_{\max}}, B_{x_{\max-1}, y_{\max}}, B_{x_{\max-1}, y_{\max}} \rangle \right]$$

Сделаем предположение, что значение температуры между точками $T_i \in T$ изменяется линейно. Тогда каждая тройка точек плоскости $T_i \in T$ определяет свою единственную плоскость, уравнение этой плоскости (5) определяет скалярное поле в областях, ограниченных этими точками.

$$\begin{vmatrix} x - P_x^1 & x - P_y^1 & t - P_t^1 \\ P_x^2 - P_x^1 & P_y^2 - P_y^1 & P_t^2 - P_t^1 \\ P_x^3 - P_x^1 & P_y^3 - P_y^1 & P_t^3 - P_t^1 \end{vmatrix} = 0, \quad (5)$$

где P_a^b это a -й компонент b -й точки, определенных в плоскости $T_i \in T$.

Построив систему уравнений плоскостей для каждой T_i получим результирующее уравнение f_m , определяющее скалярное поле в каждой его точке. Построим график поверхности (рис. 2) по результирующей функции f_m на всей области $D \subset R_2$.

Рис. 2. График поверхности функции скалярного поля температуры по координатам x ширины и y глубины (горизонта) зондирования

Таким образом была произведена линейная интерполяция скалярного поля температуры.

Градиентом этого скалярного поля $M(x, y) \in D \xrightarrow{\text{grad}} \vec{g} \in R_2$ является вектор, который соответствует точке $M \in D$ и выражается функцией:

$$\text{grad}(x, y) = (f_x', f_y') = \frac{df}{dx} \vec{i} + \frac{df}{dy} \vec{j},$$

где $\xrightarrow{\text{grad}}$ это оператор, отображающий точку скалярного поля с координатами x, y в соответствующий ей вектор градиентного поля.

Модуль вектора градиента в каждой точке определяется отношением:

$$|\text{grad}(x, y)| = \sqrt{\left(\frac{df}{dx}\right)^2 + \left(\frac{df}{dy}\right)^2}.$$

Направление вектора градиента в каждой точке определяется отношением:

$$\text{dir}(x, y) = \begin{cases} 0, \text{ если } \frac{df}{dx} = \frac{df}{dy} = 0 \\ \arccos \frac{\frac{df}{dx}}{|\text{grad}(x, y)|} \\ 2\pi - \arccos \frac{\frac{df}{dx}}{|\text{grad}(x, y)|} \end{cases}.$$

Таким образом, на множестве M определено векторное поле градиента.

Построим векторный график функции f_m (рис. 3) на всей области $D \subset R_2$.

Рис. 3. Векторный график функции градиента скалярного поля температуры

Для упрощения дальнейших расчетов и уменьшения числа опорных точек произведем нормализацию значений координат x и y осуществив линейное преобразование вида:

$$x_n = ax, \quad x_n \in [0, 5] \quad \forall x \in X,$$

$$y_n = by, \quad y_n \in [0, 4] \quad \forall y \in Y.$$

Для полученного градиентного поля $\vec{g} \in G$ определим матрицы и построим график поверхности модуля вектора градиента (рис. 4) $V^m_{x,y} = |\vec{g}_{x,y}|$ и направления вектора градиента, выраженного в градусах (рис. 5) $V^d_{x,y} = \text{dir}(\vec{g}_{x,y})$ в точках, заданных в матрице B .

Очевидно, что для точного расчета информационных метрик требуется произвести расчет в большем количестве точек, чем задано в исходной матрице, для этого определим сетку с меньшим шагом и определим значение скаляра в каждой ячейке этой сетки в соответствии с соотношениями:

$$x_m = ax, \quad x_m \in [0, 50] \quad \forall x \in X,$$

$$y_m = by, \quad y_m \in [0, 40] \quad \forall y \in Y.$$

Рис. 4. График поверхности модуля градиента скалярного поля температуры

Рис. 5. График поверхности направления градиента скалярного поля температуры

$$S(x, y) = \begin{cases} M(x, y), & \text{если } M(x, y) \neq \emptyset \\ S(x \pm 1, y \pm 1), & \text{если } \begin{cases} x = 0 \\ y = 0 \end{cases} \\ \text{Mean}(\text{Neighbor}(S(x, y))) & \Delta \text{dist}(S) \leq \text{limit} \end{cases}$$

где функция $\text{Mean}(\text{Neighbor}(S(x, y)))$ возвращает среднее арифметическое значений всех точек, лежащих в окрестности $S(x, y)$. Очевидно, что на каждой из итераций вычислений значения матрицы S будут меняться до тех пор, пока приращение среднеквадратического отклонения $\Delta \text{dist}(S)$ не станет достаточно мало $\Delta \text{dist}(S) \leq \text{limit}$.

В таком случае преобразованная матрица температур – S примет вид (рис. 6).

Рис. 6. График линий уровня скалярного поля температуры по координатам x ширины и y глубины зондирования

Градиент скалярного поля наблюдений обладает свойством инвариантности к изменению математического ожидания скалярного поля, что имеет место в случае, когда рассматривается система, находящаяся в состоянии динамического равновесия. Например, если рассматриваемое скалярное поле отражает поле температур вертикального среза водной среды, то градиент этого поля позволяет выделить три класса состояний:

- эталонное состояние (статистически устойчивое состояние для заданного времени года);
- аномалии модуля градиента температур (которые могут свидетельствовать о изменении скорости течения перпендикулярной плоскости наблюдаемого поля);
- аномалии направления градиента температур (которые могут быть вызваны значительными изменениями характера подводных течений, которые сами по себе также являются аномальными).

Примем полученные соотношения, описывающие градиент скалярного поля наблюдений температуры в летний период за эталон. Внесем искажения в исходный набор данных соответствующие модели данных на этом же участке в зимний и летний периоды, но полученные за год до эталонного и произведем

сравнение модуля и направления градиента этого поля. Сравнение производится путем вычитания матриц модуля градиента (рис. 7) и матриц направления градиента (рис. 8).

Рис. 7. График поверхности модуля разности градиентов скалярного поля температуры

Рис. 8. График поверхности направления разности градиентов скалярного поля температуры

В качестве численной метрики использовалось среднеквадратическое отклонение матриц разностей градиентов скалярного поля, которое для представленного примера составило 0,134 и 0,281 для модуля градиента, 28,954 и 34,376 для направления в летний и зимний периоды соответственно.

Заключение. Предложен подход к решению задачи оперативного обнаружения аномалий скалярных полей наблюдений с использованием градиентного метода.

Установлено что для рассматриваемого случая среднеквадратическое отклонение разности градиентов скалярных полей температуры обладает избирательной чувствительностью к изменениям динамики водной системы и общего температурного фона. В частности, среднеквадратическое отклонение модулей разности градиентов скалярного поля отражает меру сезонных изменений общего температурного фона, особенно на небольших глубинах, прогреваемых в летний период. При этом среднеквадратическое отклонение направления разности градиентов скалярного поля напротив обладает слабой чувствительностью к сезонным изменениям общего температурного фона, но отражает меру изменения динамики распространения подводных течений.

Разработанные в процессе исследования программные модули позволяют применять предложенный подход для произвольного набора данных скалярных полей и обеспечивают возможность пользователю задавать пороговые значения среднеквадратических отклонений модуля и направления разности градиентных полей в параметрах информационных метрик системы поддержки принятия решений в интерактивном режиме, на основании значений которых осуществляется работа бинарного классификатора. Использование метода на основе градиента, который значительно более чувствительный к изменению самой скалярной величины чем ее абсолютное значение, позволяет повысить качество принимаемых решений и оперативность выявления значимых изменений всего скалярного поля.

Разработанный метод может быть использован на практике как компонент в системе оценки и прогнозирования объема биомассы промысловых морских рыб в Азово-Черноморском и других регионах с целью оптимизации процесса

осуществления промысла и формирования практических рекомендаций по рациональному выбору времени и места вылова.

Предложенный подход разработан для использования в СППР, реализующийся в виде бинарного классификатора, блоков задания порога α и эталонного скалярного поля. Предназначен для нахождения пограничных зон неуверенного распознавания аномалий, поскольку целью мониторинга является не только сбор информации об объекте и процесса, но и обнаружение критических событий для оперативного принятия неотложных мер по ликвидации критических состояний.

Работа выполнена при частичной поддержке Российского фонда фундаментальных исследований (грант № 15-29-07936/17).

СПИСОК ЛИТЕРАТУРЫ

1. *Unsupervised* real-time anomaly detection for streaming data / S. Ahmad, A. Lavin, S. Purdy [et al.] // *Neurocomputing*. 2017. Vol. 262. P. 134–147. DOI: j.neucom.2017.04.070.
2. *Derkahov S.E., Manashov A.N.* Anomalous dimensions of composite operators in scalar field theories // *Journal of Mathematical Sciences*. 2010. Vol. 168: P. 837–855. DOI: 10.1007/s10958-010-0032-9.
3. *Fursaev D.V., Miele G.* Finite-temperature scalar field theory in static de Sitter space // *Phys. Rev. D*. 1993. Vol. 49. P. 987-998. DOI: 10.1103/PhysRevD.49.987.
4. *Скатков А.В., Шишкин Ю.Е.* Кластеризация данных в задачах обнаружения аномалий на основе ортогональных фильтров // *Системы контроля окружающей среды*. 2018. № 11 (31). С. 36–43.
5. *Скатков А.В., Шишкин Ю.Е.* Модель обнаружения аномалий в наблюдениях параметров полей окружающей среды с использованием систем мониторинга // *Системы контроля окружающей среды*. 2017. № 10 (30). С. 48–53.
6. *Pääkkönen P., Pakkala D.* Reference Architecture and Classification of Technologies, Products and Services for Big Data Systems // *Big Data Research*. 2015.

Vol. 2. P. 166–186. DOI: 10.1016/j.bdr.2015.01.001.

7. Скатков А.В., Шишкин Ю.Е. Мониторинг качественных изменений состояний сетевого трафика в облачных вычислительных средах // Автоматизация и приборостроение: проблемы, решения: материалы междунар. науч.-техн. конф. / под науч. ред. В.Я. Коппа. Севастополь, 11–15 сентября 2017 г. Севастополь: СевГУ, 2017. С. 137–138.

8. Shishkin Y.E. Big Data visualization in decision making // Science in Progress: тез. Всерос. науч.-практ. конф. магистрантов и аспирантов. Новосибирск, 20 октября 2016 г. Новосибирск: НГТУ, 2016. С. 203–205. ISBN 978-5-7782-3094-1.

9. Скатков А.В., Шишкин Ю.Е. Разработка модели оптимизации мониторинга при частичной неопределенности в виде системы массового обслуживания // Развитие методологии современной экономической науки и менеджмента: материалы I Междисциплинарной Всероссийской науч.-практ. конф. Севастополь, 4-5 мая 2017 г. Севастополь: СевГУ, 2017. С. 611–618.

10. Bâki H. Is the global sea surface temperature rise accelerating? // Geodesy and Geodynamics. 2018. Vol. 2. P. 109–116. DOI: 10.1016/j.geog.2018.04.002.

11. Бурмистрова Н.В., Жук В.Ф., Мельникова Е.Б. Термохалинная структура вод на траверсе бухты Круглая и её

влияние на интенсивность поля биолюминесценции // Природничий альманах. 2011. Вып. 15. С. 14–25.

12. Скатков А.В., Шишкин Ю.Е. Модель системы мониторинга в условиях риска // Экологическая, промышленная и энергетическая безопасность – 2017: сборник статей по материалам науч.-практ. конф. с международным участием / под ред. Ю.А. Омельчук, Н.В. Ляминной, Г.В. Кучерик. 2017. С. 1239–1242.

13. Спектральный анализ аномального магнитного поля земли для разновысотных съемок / В.И. Одинцов, Н.М. Ротанова, Ю.П. Цветков [и др.] // Геомагнетизм и аэрономия. 2000. Т. 40. № 2. С. 59–66.

14. Dorit S. Complexity and approximations for submodular minimization problems on two variables per inequality constraints // Discrete Applied Mathematics. 2018. Vol. 1. P. 12–20 DOI: 10.1016/j.dam.2018.04.012.

15. Брюховецкий А.А., Скатков А.В., Шишкин Ю.Е. Моделирование процессов обнаружения аномалий в сложно-структурированных данных мониторинга // Системы контроля окружающей среды. 2017. № 9 (29). С. 45–49.

16. A lock-free approach to parallelizing stochastic gradient descent / B. Recht, W. Re, S. Wright [et al.] // Advances in Neural Information Processing Systems – 24. Curran Associates, Inc. 2011. P. 693–701.

METHOD FOR DETECTING ANOMALIES IN AN INTERACTIVE MODE IN OBSERVATIONS SCALAR FIELDS GRADIENTS

Y.E. Shishkin^{1,2}, A.V. Skatkov¹

¹ Federal State Educational Institution of Higher Education «Sevastopol State University», Russian Federation, Sevastopol, Universitetskaya St., 33

² Institute of Natural and Technical Systems, Russian Federation, Sevastopol, Lenin St., 28

An approach to solve the problem of operative anomalies recognition in scalar fields data for monitoring objects and processes using the gradient method and the system of specific information metrics is proposed. Metrics of module and direction standard deviation of the gradient fields difference are used. A binary classifier system method and a mathematical model and a recognition procedure that works in an interactive mode are constructed. The process of proposed approach operation is illustrated using the example of water temperature vertical spatial sounding data sample of Kruglaya Bay area of the city of Sevastopol.

Keywords: monitoring, mathematical modeling, Big Data, digital filtering, gradient, scalar field, detection of anomalies, clustering, critical systems, data mining.