

Обнаружение аномальных компонентов в пространственных обзорах на основе многомерной модели пуассоновских потоков и их когнитивная визуализация

В. Л. Горохов, И. А. Брусакова

Санкт-Петербургский государственный электротехнический университет

«ЛЭТИ» им. В.И. Ульянова (Ленина)

E-mail: vlgorohov@mail.ru

УДК 519.1

Аннотация. В статье предлагается методика обнаружения аномальных компонентов в пространственных сканах многомерных данных для многомерных обзоров в ГИС технологиях. Обнаружение аномальных компонентов осуществляется на основе несмещенных равномерно наиболее мощных алгоритмов в условиях глубокой априорной неопределенности в отношении параметров распределений данных обзоров. Результаты обнаружения контролируются средствами когнитивной компьютерной графики. Методика применяется для обработки многомерных данных астрономических наблюдений. Эти методы весьма успешно применяются в астрофизике и могут быть использованы для широкого круга задач в BIG DATA. Методика такого сочетания может быть также сориентирована на выявление и прогнозирование чрезвычайных ситуаций в сложных системах.

Ключевые слова: несмещенные алгоритмы обнаружения и когнитивные методы визуализации; многомерные временные ряды; 3-D машинная графика

I. ВВЕДЕНИЕ

Многомерный по совокупности характеристик мониторинг объектов в киберфизических системах наталкивается на серьезные трудности. Эти трудности связаны прежде всего с петабайтным количеством объектов мониторинга и с значительным набором типов измеряемых характеристик объектов мониторинга. Сочетание этих признаков приводит к, так называемому, «проклятию размерности». Это сочетание обуславливает значительное торможение скорости вычислений и в настоящее время обозначается как проблема BIG DATA.

Арсенал методов, алгоритмов и методик, направленных на преодоление этих трудностей, непрерывно пополняется. Здесь и современные искусственные нейронные сети, сверточные архитектуры, концепции глубокого обучения и т. д.

Кроме этих перспективных тенденций активно начинают использовать и последние достижения теории устойчивого обнаружения сигналов [1] и практики применения алгоритмов когнитивной компьютерной графики [3].

В статье предлагаются алгоритмы и методики устойчивого обнаружения аномальных компонентов в

пространственных сканах многомерных данных мониторинга. Обнаружение аномальных компонентов осуществляется на основе равномерно наиболее мощных подобных несмещенных алгоритмов в условиях глубокой априорной неопределенности в отношении параметров распределений данных [1]. Результаты использования алгоритмов и методик иллюстрируются их применением для обработки многомерных данных астрономических наблюдений [2]. Результаты обнаружения подтверждаются средствами когнитивной компьютерной графики [3]. Предлагаемые алгоритмы и методики весьма успешно применяются в астрофизике и могут быть использованы для широкого круга задач в BIG DATA.

Методика такого сочетания может быть также сориентирована на выявление и прогнозирование чрезвычайных ситуаций в сложных системах. Здесь возможно сочетать возможности максимально эффективных устойчивых (робастных) автоматических устройств обнаружения аномальных явлений и возможности интуитивных механизмов человеческого сознания. Возникает возможность гармоничного сочетания искусственного и естественного интеллекта для управления киберфизическими системами.

II. КОНЦЕПЦИЯ МНОГОМЕРНЫХ СЛУЧАЙНЫХ ПОТОКОВ ДЛЯ ОПИСАНИЯ МНОГОМЕРНЫХ ДАННЫХ

В реальной практике обработки многомерных данных отдельные пространственные сканы $\{\Delta x_i\}_k$ (где k – номер скана) разбиваются на фиксированные фрагменты Δx_i , в каждом из которых измеряется число m_i выявленных объектов мониторинга (например – галактик [2]).

Факт выявления объекта зависит от некоторых характеристик μ_j этого объекта (например – светимость галактик, это могут быть энергетические характеристики гамма барстеров, как характеристики родительских галактик). Эти характеристики определяют интенсивность пространственного пуассоновского потока. Таким образом, объекты могут быть представлены в виде пуассоновского потока, который характеризуется вероятностью появления m_i объектов в пространственном интервале Δx_i k -го скана. При этом эта вероятность зависит от характеристики μ . Величины μ_i в рамках данной модели являются априорно неопределенными и могут отличаться между собой.

Модель пуассоновского потока может быть обобщена в рамках теории случайных потоков, где рассматриваются потоки кратных точек, парнокоррелированные потоки, потоки Бернулли и т. д.

В настоящей работе предлагаемый алгоритм позволяет синхронно анализировать весь набор пространственных сканов и фиксировать факт синхронного обнаружения аномалий в соответствующих фрагментах скана.

Наличие аномалии во фрагменте можно трактовать как сигнал. Отсутствие аномалий во фрагменте можно трактовать, как фоновую составляющую. Пусть во фрагменте Δx_i измерено m_i объектов мониторинга. Их вероятностное описание задается совместным распределением вероятностей в форме: $p(m_1, m_n) = \{ \exp(-\sum \mu_i \Delta x_i) \exp\{m_i \log(\mu_i \Delta x_i)\} / \prod \mu_i!\}$.

Распределение принадлежит экспоненциальному семейству и обладает достаточными статистиками $U = m_1$ и $X = \{x_2, \dots, x_n\}$, где $x_i = m_i = (m_1 / (n-1))$. Распределение характеризуется одномерным полезным параметром $\Theta = \log\{m_1 / (\gamma \sqrt{\prod m_i})\}$ и фоновым $(n-1)$ -мерным параметром $\log(\mu_i \Delta x_i)$.

Коэффициент γ показывает, во сколько раз интенсивность пуассоновского потока в проверяемом интервале, при наличии аномалии, должна превышать среднее по остальным фоновым компонентам. Параметры данного распределения вероятностей нелинейно зависят от характеристик μ_i .

Однако для варианта, когда $n=2$ и когда осуществляется последовательный просмотр всего скана $\{\Delta x_i\}$ этой сканирующей контрастной парой, можно получить решающее правило для проверки статистических гипотез обнаружения аномалий.

Задача обнаружения аномалий (сигнала) может быть сформулирована как задача проверки сложных статистических гипотез относительно параметров введенного выше распределения вероятностей: $H_0: \Theta < 0$, $\mu_j, j=2, \dots, n$ (сигнала нет); $H_1: \Theta > 0$, $\mu_j, j=2, \dots, n$ (сигнала есть).

Вид решающего правила согласно [1] таков: $\phi(U, X) = \{1, \text{ если } U > C(\alpha, X) \text{ и } 0 \text{ если } U \leq C(\alpha, X)\}$. Здесь $C(\alpha, X)$ пороговая функция, зависящая от уровня вероятности ложной тревоги α и достаточной статистики X . Пороговая функция $C(\alpha, X)$ определяется из уравнения $E\{\phi(U, X)\}$ при $\Theta=0$ и $X_j=0$, где E – условное математическое ожидание при $\Theta=0$. Данное решающее правило согласно [1] является равномерно наиболее мощным несмещенным алгоритмом обнаружения.

Вариант $n=2$ можно использовать в качестве сканирующей пары фрагментов по всему скану, выявляя таким контрастным методом аномалии по всему скану.

Следует особо отметить, что предлагаемый алгоритм обнаружения сохраняет постоянной вероятность ложной тревоги α при любых значениях мешающих параметрах. Кроме этого он обеспечивает вычисление вероятности правильного обнаружения для предполагаемых соотношений сигнал-шум $q = (\mu_1 / \mu_i)$ при различных значениях вероятности ложной тревоги и интенсивности

фоновой (шумовой) компоненты. Все это показывает высокую устойчивость (робастность) предлагаемых алгоритмов.

Учитывая наличие k -набора пространственных сканов, появляется возможность обнаружения синхронных аномалий в совокупности сканов на основе алгоритмов обнаружения стохастического возрастания в бинарно квантованной последовательности [1].

Следует специально заметить, что данная концепция синхронного обнаружения аномалий может быть легко распространена как на случай набора сканов, так и на случай многомерных случайных потоков более широкого класса.

III. КОГНИТИВНАЯ ПОДДЕРЖКА И КОНТРОЛЬ ВЫЯВЛЕННЫХ АНОМАЛИЙ

Когнитивная наука в настоящее время предлагает графические инструменты визуализации многомерных данных и многомерных распределений этих данных для контроля результатов работы алгоритмов автоматического обнаружения аномалий в многомерных данных [3]. Подобные графические инструменты позволяют мобилизовать когнитивные возможности человеческого сознания для контроля работы автоматических устройств. Ниже приводится пример когнитивной визуализации многомерных сканов демонстрирующей возможность визуально контролировать факт обнаружения аномалий в ситуации, когда объекты мониторинга характеризуются целым набором характеристик этих объектов (рис. 1).

Рассматривается совокупность пространственных сканов для фиксированных фрагментов, в которых измеряется число объектов (в данном случае это галактики) мониторинга m_i . Каждый скан содержит число объектов измеренных с помощью разных физических характеристик этих объектов. Это и светимости в разных спектральных диапазонах, красные смещения, геометрические особенности этих объектов, энергетические характеристики гамма барстеров родительских галактик E_{iso} , E_{peak} , и т. д.

Методами когнитивной компьютерной графики можно визуализировать весь комплекс связей между этими характеристиками, можно фиксировать многомерные корреляции между этими характеристиками, включая и пространственные корреляции аномалий, которые объективируются методами описанными выше.

В данном примере [4–7] рассматривается анализ данных космологического обзора (например, Pantheon). Данный каталог содержит, как набор измеренных характеристик, так и пространственные характеристики объектов. Сочетание алгоритмов пространственного анализа объектов с методами когнитивными визуализации многомерных корреляционных характеристик позволяет выявлять принципиально новые и неожиданные закономерности. Так применение когнитивного анализа данных обзора показывает, что маргинальная регрессионная зависимость $E_{iso} - E_{peak}$ имеет место быть, но с многомерной точки зрения, эта зависимость имеет характер многомерной поверхности. Это подтверждает результаты оценки коэффициентов регрессии, полученных ранее средствами обнаружения

аномалий средствами статистического обнаружения (рис. 1).

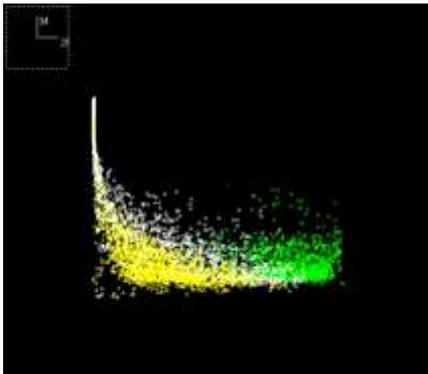


Рис. 1. Результат когнитивной динамической проекции многомерной регрессионной зависимости характеристик объекта мониторинга E_{iso} и E_{peak} .

На этом рисунке показано, как можно средствами когнитивной графики контролировать обнаружение синхронного обнаружения аномалий в пространственных сканах для набора разнотипных характеристик объектов мониторинга. В данном случае это: E_{iso} и E_{peak} энергетические характеристики гамма барстеров родительских галактик. При этом статистически фиксируется влияние еще целого набора разнотипных характеристик. Их влияние визуализируется в динамике, как псевдотрехмерные поверхности.

Геометрия этих поверхностей отражает, как многомерные корреляционные связи, так и потенциальные возможности для многомерной классификации и распознавания объектов и образов. Это в свою очередь открывает новые возможности при обучении нейронных сетей (восходящих алгоритмов) и рациональной «обратной» интерпретации решений принимаемых этими сетями. Известно, что именно логическая «обратная» интерпретация решений порождаемых нейронными сетями и создает серьезные проблемы. Преодоление этих проблем видится, в том числе, и средствами когнитивной машинной графики.

IV. ЗАКЛЮЧЕНИЕ

Подобная концепция устойчивого обнаружения аномалий в сочетании с когнитивным контролем многомерных пространственных или временных характеристик объектов киберфизических и природных

систем позволяет системному аналитику более продуктивно использовать возможности своего образного мышления и возможности устойчивых алгоритмов обнаружения аномалий в задачах многомерной статистики. При этом открываются принципиально новые возможности изучения, контроля и управления многомерными процессами, характерными для в киберфизических систем.

Важно обратить внимание на возможности предлагаемых методик сочетать дискретное и континуальное описание многомерных данных. При этом неопределенность в отношении всех типов измеряемых характеристик не влияет на устойчивость работы алгоритмов обнаружения! Это важное достижение современной теории обнаружения, позволяющее принимать решения в условиях глубокой априорной неопределенности присущей киберфизическим системам.

СПИСОК ЛИТЕРАТУРЫ

- [1] Богданович В.А., Вострецов А.Г. Теория устойчивого обнаружения, различия и оценивания сигналов. М.: ФИЗМАТЛИТ, 2003. 320 с. ISBN 5-9221-0359-8
- [2] Baryshev Y., Teerikorpi P. Fundamental Questions of Practical Cosmology: Exploring the 212 Realm of Galaxies; Vol. 383, Astrophysics and Space Science Library, Springer: Berlin, 2012. 213 doi:10.1007/978-94-007-2379-5
- [3] Горохов В.Л., Муравьев И.П. Когнитивная машинная графика. Методы динамических проекций и робастная сегментация многомерных данных: Монография. / Под ред. проф. А.И. Михайлушкина. СПб. СПбГИЭУ, 2007. 172 с.
- [4] Maxim Nikonov, Mikhail Chekal, Stanislav Shirokov, Andrey Baryshev, Vladimir Gorokhov. The line-of-sight analysis of spatial distribution of galaxies in the COSMOS2015 catalogue /Version October 28, 2020 submitted to Universe (2020 by the authors. Submitted to Universe for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>))
- [5] Shirokov S.I., Sokolov I.V., Lovyagin N.Yu., Amati L., Baryshev Yu.V., Sokolov V.V., Gorokhov V.L. High Redshift Long Gamma-Ray Bursts Hubble Diagrams a Test of Basic Cosmological Relations, MNRAS 4-18 (2020)
- [6] Барышев Ю.В., Широков С.И., Ловягин Н.Ю., Горохов В.Л. Крупномасштабные флуктуации плотности в независимых обзорах глубоких полей *Астрономический журнал* Т 93, №4, 2016.
- [7] Горохов В.Л., Широков С.И., Витковский В.В. Когнитивный метод динамической визуализации многомерных данных, применяемый к отношению Амати / Statistical Challenges in Modern Astronomy VII. (SCMA VII) will be held virtually June 7–10, WebInfo@outreach.psu.edu T and hosted by the Penn State Center for Astrostatistics Pennsylvania State University © 2021