



Международный журнал информационных технологий и энергоэффективности

Сайт журнала: <http://www.openaccessscience.ru/index.php/ijcse/>



УДК 004.67

МЕТОДЫ ВЫДЕЛЕНИЯ АНОМАЛЬНЫХ ЗНАЧЕНИЙ ВО ВРЕМЕННЫХ РЯДАХ СПУТНИКОВЫХ ДАННЫХ

Ильясов Р.Р.

ФГБОУ ВО "УФИМСКИЙ УНИВЕРСИТЕТ НАУКИ И ТЕХНОЛОГИЙ", Уфа, Россия (450076, Республика Башкортостан, г Уфа, ул Заки Валиди, д. 32), e-mail: rustam.iljasov2015@yandex.ru

Данная статья посвящена определению аномальных значений во временных рядах спутниковых данных. Наиболее значимым является рассмотрение различных подходов к выявлению необычных и выбивающихся значений в наборах данных, полученных в результате наблюдений. В статье рассматриваются некоторые из методов, используя которые, есть возможность выявить аномальные значения.

Ключевые слова: Аномальные значения, временной ряд, спутниковые данные, анализ данных, диагностирование, метод исследования.

METHODS FOR DETECTING ANOMALOUS VALUES IN TIME SERIES OF SATELLITE DATA

Ilyasov R.R.

UFA UNIVERSITY OF SCIENCE AND TECHNOLOGY, Ufa, Russia (450076, Republic of Bashkortostan, Ufa, Zaki Validi st., 32), e-mail: rustam.iljasov2015@yandex.ru

This article is devoted to the definition of anomalous values in the time series of satellite data. The most significant is the consideration of various approaches to identifying unusual and out-of-place values in datasets obtained as a result of observations. The article discusses some of the methods using which it is possible to identify abnormal values.

Keywords: anomalous values, time series, satellite data, data analysis, diagnostics, the research method.

Введение

В современной научной среде анализ временных рядов спутниковых данных играет ключевую роль в понимании динамических процессов на Земле и в околоземном пространстве. Спутниковые наблюдения предоставляют обширные объемы информации о климатических, экологических и геологических явлениях, что делает их бесценным ресурсом для научных исследований и прикладных задач.

С увеличением объема собираемых данных, возникает необходимость в разработке эффективных методов обработки и анализа, включая выделение аномальных значений.

Полученные результаты могут предложить ценные практические рекомендации для повышения достоверности анализа и интерпретации спутниковых данных в различных областях науки и приложений [1].

Методы выделения аномальных значений во временных рядах спутниковых данных

С увеличением объемов спутниковых данных возникает необходимость в разработке эффективных методов выделения аномалий для обеспечения точности и достоверности результатов исследований. Различные факторы, такие как технические помехи, атмосферные условия или природные катаклизмы, могут внести изменения в наблюдаемые временные ряды. Следовательно, актуальной задачей является разработка методов, способных автоматически выделять аномальные значения и отделять их от нормальных трендов.

Временные ряды, полученные из спутниковых данных, обладают рядом особенностей, которые необходимо учитывать при разработке методов выделения аномалий. Прежде всего, данные могут содержать пропуски из-за различных технических проблем или ограничений в сборе информации. Это представляет дополнительную сложность в анализе, так как пропущенные значения могут исказить общую картину временного ряда и вносить дополнительные искажения при выделении аномалий [2].

Метод межквартильного размаха для выделения аномалий

Принцип работы решения заключается в использовании межквартильного размаха для определения границ выбросов и выделения аномальных значений.

В контексте определения выбросов, межквартильный размах используется для определения "нормального" диапазона значений данных и выявления потенциальных выбросов за пределами этого диапазона. Выбросами считаются значения, лежащие за пределами установленных границ, которые обычно определяются как:

$$\text{Нижняя граница} = Q_1 - 1.5 \cdot \text{IQR}$$

$$\text{Верхняя граница} = Q_3 + 1.5 \cdot \text{IQR}$$

Где Q_1 – первый квартиль (25-й процентиль), Q_3 – третий квартиль (75-й процентиль) и IQR – межквартильный размах.

Одним из преимуществ данного метода является простота реализации с использованием стандартных статистических функций, а также применимость к временным рядам для выделения аномальных значений, например, скорости изменения спутниковых данных [3].

Метод порогового значения для выделения аномалий

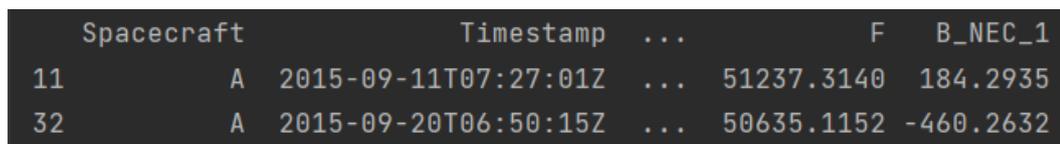
Принцип работы решения заключается в использовании порогового значения для определения аномалий в данных. Вычисляется разница между последовательными значениями спутниковых данных для каждой строки с использованием языка программирования Python. Устанавливается пороговое значение. Все значения, превышающие по модулю это пороговое значение, считаются аномальными. Отфильтровываются строки с аномальными значениями.

Таким образом, данный метод основан на установке порога и выделении всех значений, которые превышают этот порог как аномалий. Он может быть полезен в ситуациях, когда есть

явное представление о том, какие значения параметра считаются аномальными, и пороговое значение может быть установлено на основе экспертного знания или предварительного анализа данных.

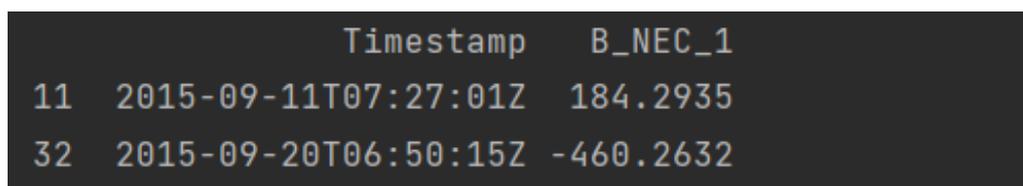
Оценка результатов исследования

Крайне важно, чтобы результаты исследования были наиболее точны. Реализация методов для определения аномальных значений спутниковых данных производилось с помощью языка программирования Python и библиотеки Pandas. На Рисунках 1–2 представлены результаты исследования с использованием методов межквартильного размаха и порогового значения.



	Spacecraft	Timestamp	...	F	B_NEC_1
11	A	2015-09-11T07:27:01Z	...	51237.3140	184.2935
32	A	2015-09-20T06:50:15Z	...	50635.1152	-460.2632

Рисунок 1 – Результат определения аномальных значений с использованием метода межквартильного размаха



	Timestamp	B_NEC_1
11	2015-09-11T07:27:01Z	184.2935
32	2015-09-20T06:50:15Z	-460.2632

Рисунок 2 – Результат определения аномальных значений с использованием метода порогового значения

Полученные результаты отображают выявленные аномальные значения скорости изменения одного из параметров спутниковых данных.

Как видно из результатов исследования, использованные методы показывают идентичные результаты, что говорит о точности и достоверности результатов.

Выводы

Таким образом, в ходе исследования были определены аномальные значения во временных рядах спутниковых данных с использованием различных методов. Использование нескольких методов увеличивает уверенность в результате, адаптируется к различным типам аномалий и компенсирует ограничения каждого отдельного из них, что является важным аспектом в обработке данных. Исходя из результатов выявления и оценки аномальных значений, впоследствии, предоставляется возможность их прогнозирования.

Список литературы

1. Обнаружение аномалий: основы, методы и применение в современном мире / Научные статьи.
2. Поиск аномалий во временных рядах / Хабр (habr.com).
3. Interquartile Range to Detect Outliers in Data / GeeksforGeeks.
4. Anomaly Detection in Time Series Data / GeeksforGeeks.

References

1. Anomaly Detection: Basics, Methods and Application in the Modern World / Scientific Articles.
 2. Search for anomalies in time series / Habr (habr.com).
 3. Interquartile Range to Detect Outliers in Data / GeeksforGeeks.
 4. Anomaly Detection in Time Series Data / GeeksforGeeks.
-