**Анализ точечных аномалий во временных рядах временных рядов геофизических величин**

**Караваева К.Э.1**

1студент,

Московский государственный университет имени М.В.Ломоносова,физический факультет, Москва, РоссияE–mail: karavaeva.ke21@physics.msu.ru

Современные технологии позволяют проводить мониторинг текущей климатической ситуации путём регулярного сбора данных на метеорологических станциях в различных точках планеты. Для большинства технологических систем результаты мониторинга можно представить в виде временного ряда. Временной ряд – это последовательность наблюдений за определенным параметром в разные моменты времени. Одной из задач анализа временных рядов является поиск аномальных данных, которые возникают при неисправности измерительных приборов.

Принято классифицировать аномалии на точечные, коллективные и контекстуальные [1]. В данной работе мы будет рассматривать детектирование точечных аномалий (выбросов).

Временной ряд можно представить в виде суммы слабо или медленно меняющей части $m\_{t}$ (тренда), периодически изменяющейся циклической части $S\_{t}$ (сезонность) и остатка
$ε\_{t}$.

$f\left(t\right)=m\_{t}+S\_{t}+ε\_{t}$$f\left(t\right)=m\_{t}+S\_{t}+ε\_{t}$ (1)

 Некоторую информацию о сезонности можно получить при анализе автокорреляционной и частичной автокорреляционной функций. Основной вклад вносят годовые и дневные колебания исследуемых геофизических величин. Тренд можно аппроксимировать линейной зависимостью.

На рис. 1 и рис. 2 представлены соответственно трендовая и сезонная составляющие временного ряда температуры, измеренных на высоте 30 м на метеостанции “ASIA-FLUX” во Вьетнаме. Данные были предоставлены ИПЭЭ им. А.Н. Северцова РАН.



***Рис. 1.*** Трендовая составляющая временного ряда температуры на высоте 30 м



***Рис. 2.*** Сезонная составляющая временного ряда температуры на высоте 30 м

Для анализа и сравнения различных алгоритмов детектирования аномалий во временных рядах были рассмотрены синтетические ряды, подобные реальным рядам рассматриваемых величин. В синтетических рядах мы можем генерировать необходимые для изучения аномалии с разными параметрами (например, можно изменять частоту и величину выбросов). На рис. 3 приведен пример синтетического ряда.



***Рис. 3.*** Пример синтетического ряда

 Выделяют два подхода к поиску аномалий – с учителем и без учителя [1]. В первом подходе каждое измерение имеет метку класса (0 – нормальное измерение, 1 – аномалия), во втором разметки данных нет, но есть априорная информация, что аномалий нет. В реальности данные чаще всего неразмеченные, поэтому предпочтителен подход без учителя.

 В подходе без учителя выделяют следующие методы: статистические, основанные на близости данных, основанные на предсказании ряда и основанные на восстановлении ряда [2]. В этой работе применяются как статистические, так и методы прогнозирования.

 Целью данной работы является разработка эффективного метода для анализа и поиска точечных аномалий во временном ряду в режиме реального времени. Для решения задач такого типа предлагается использовать комбинацию статистических методов и алгоритмов машинного обучения.

**Литература**

1. Banerjee A., Chandola V. and Kumar V. Anomaly detection: A survey // ACM Computing Surveys. - 2009. - Т. 41. - No 15. - С. 1-58.
2. Geiger A., Liu D., Alnegheimish S., Cuesta-Infante A., Veeramachaneni K. TadGAN: Time Series Anomaly Detection Using Generative Adversarial Networks // Материалы конференции IEEE International Conference on Big Data. - 2020.
3. Leys, C., et al. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median // Journal of Experimental Social Psychology. - 2013.- Т. 49. - No 4. - С. 764-766.
4. Suomela J. Median Filtering is Equivalent to Sorting: <https://doi.org/10.48550/arXiv.1406.1717>