

ПОИСК АНОМАЛИЙ ВО ВРЕМЕННЫХ РЯДАХ НА ОСНОВЕ ОЦЕНИВАНИЯ ИХ ПАРАМЕТРОВ

Мельникова П. А.

Научный руководитель – д.т.н., проф. Кириченко Л. О.
Харьковский национальный университет радиоэлектроники
61166, Харьков, просп. Науки, 14, каф. прикладной математики,
тел. (057) 702-14-36, e-mail: polina.melnykova@nure.ua

Anomaly detection in every domain is an important topic of discussion and the knowledge about the effective ways to perform anomaly detection is an important skill to have for any Data Scientist. A problem of anomaly detection in time series is considered. A method is proposed, which allows to detect anomalies in time series efficiently by their parameters.

В последнее время большое внимание уделяется анализу временных рядов, которые используются в различных областях и описывают длительные процессы, протекающие во времени. Возможность найти аномалии в некотором наборе данных важна при анализе работы сложных технических систем, анализе сетевого трафика, в медицине, в банковском деле и в других предметных областях [1].

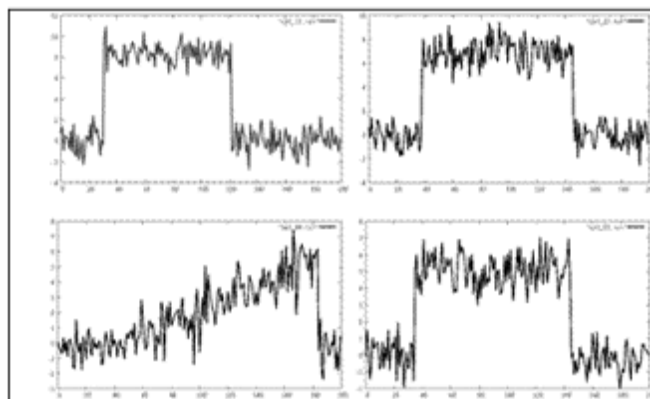


Рисунок 1 – Примеры временных рядов, где аномальным является левый нижний.

Так как временной ряд – это упорядоченная последовательность чисел, то его можно характеризовать статистическими параметрами. Таким образом для каждого ряда можно высчитать вектор числовых характеристик. В рамках данной работы был выбран список параметров, который включает статистические, корреляционные, энтропийные и вероятностные характеристики, посчитанные по временному ряду. Такое соответствие приравнивает аномальность временного ряда к аномальности параметров этого ряда. Целью представленной работы является поиск аномалий во временных рядах на основе оценивания их параметров [2].

Поиском аномалий (выбросов) называется проблема обнаружения объектов не похожих на остальные данные, не подчиняющихся ожидаемым, «нормальным» паттернам поведения. В большинстве практических приложений разметка нормальных и аномальных классов данных отсутствует, в связи с чем проблема выявления аномалий рассматривается как задача обучения без учителя [3].

В данной статье рассмотрен поиск аномалий во временных рядах методом кластеризации DBSCAN (Density-based spatial clustering of applications with noise). Кластеризация – это автоматическое разбиение элементов некоторого множества на группы в зависимости от их схожести. Сами же группы принято называть кластерами.

DBSCAN – плотностный алгоритм кластеризации, выделяющий «внутренние», «пограничные» и «шумовые» точки в данных. В его основе лежит идея о том, что плотность точек внутри каждого кластера существенно выше плотности точек снаружи, в том числе выше плотности шумовых областей.

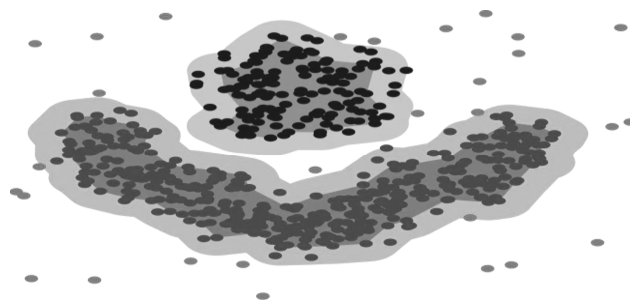


Рисунок 2 – Результат кластерного анализа методом DBSCAN

В рамках данной работы объектами кластеризации являются наборы параметров, характеризующие временные ряды. Элементы, находящиеся близко друг к другу, объединяются в кластеры. Аномальными будут считаться те объекты, которые не попали ни в один кластер (область шума).

Список використаних джерел:

1. Проблема обнаружения аномалий в наборах временных рядов URL: <http://www.swsys.ru/index.php?page=article&id=3117>
2. Center for Machine Learning and Intelligent Systems URL: <https://archive.ics.uci.edu/ml/datasets/SmartphoneBased+Recognition+of+Human+Activities+and+Postural+Transitions>
3. Chandola V., Banerjee A., Kumar V. Anomaly detection: A survey /ACM computing surveys (CSUR). 2009. Т. 41.