

## ВИЯВЛЕННЯ КОЛЕКТИВНИХ АНОМАЛІЙ У ЧАСОВИХ РЯДАХ

Федоренко Є.Д., Яндуков Д.Є.

Наукові керівники – проф. Кириченко Л.О., доц. Кобзев В.Г.

Харківський національний університет радіоелектроніки  
61166, Харків, просп. Науки, 14, каф. прикладної математики,  
тел. (057) 702-14-36, e-mail: [yevheniia.fedorenko1@nure.ua](mailto:yevheniia.fedorenko1@nure.ua).

Anomaly detection for time series data has been an important area of research for a long time. The main work on anomaly detection techniques has focused on statistical approaches. In recent years, more and more machine learning algorithms have been developed to detect anomalies in time series. There are three basic types of anomalies, this work we devoted to the study of collective anomalies as the most complex and frequently encountered type of anomalies. The implementation of this type of anomalies was carried out by the k-means method as the most universal method in our opinion.

Інтелектуальний аналіз даних (Data Mining) - це сукупність методів виявлення в даних раніше невідомих, нетривіальних, практично корисних і доступних інтерпретації знань, необхідних для прийняття рішень в різних сферах людської діяльності.

Одним із завдань інтелектуального аналізу є виявлення аномалій - ідентифікація рідкісних і незвичайних елементів, подій або спостережень, що викликають підозри, оскільки вони значно відрізняються від більшості даних.

Часовий ряд представляє собою упорядковану в часі послідовність значень якого-небудь процесу (наприклад, значення деякого датчика). Необхідність виявлення незвичайних спостережень (викидів, або аномалій) у часових рядах часто виникає в таких ситуаціях, як моніторинг стану обладнання, відстеження несподіваних коливань на ринку цінних паперів, облік показників стану здоров'я пацієнтів і т.д.

Існує три основних типи аномалій у часових рядах: точкові, контекстуальні і колективні аномалії. У даній роботі ми будемо розглядати колективні аномалії, як найбільш поширені для часових рядів.

Колективні аномалії виникають, коли послідовність пов'язаних примірників даних (наприклад, ділянку часового ряду) є аномальною по відношенню до цілого набору даних.

Окремий екземпляр даних в такій послідовності може не бути відхиленням, проте спільна поява таких екземплярів є колективною аномалією.

На рисунку 1 нижче жирним виділена область А, яка істотно відрізняється від напрямку кривої.

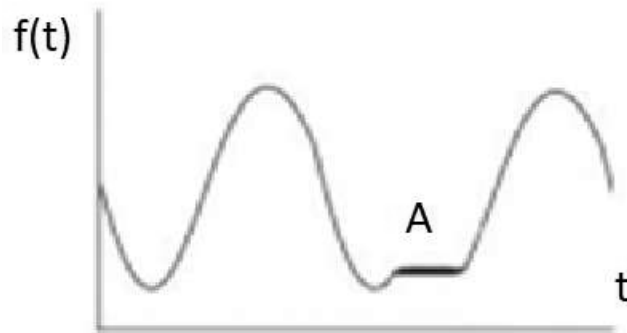


Рисунок 1. Приклад колективних аномалій в часовому ряді.

Це - приклад колективних аномалій. Їх основна ідея полягає в тому, що такі згруповані точки не можуть бути аномаліями окремо. Для ідентифікації такого роду аномалій існує багато методів, в тому числі методи кластеризації, які дозволяють виділити аномальні значення в вигляді окремого кластера.

Якщо ми розглядаємо велику кількість спостережень, то буде раціонально використовувати метод  $k$ -середніх. Цей метод є одним з найпростіших і популярних методів кластеризації. Алгоритм розбиває множину елементів векторного простору на заздалегідь відоме число кластерів  $k$ . Загальна концепція даного методу полягає в наступному: на кожній ітерації переобчислюють центр мас для кожного кластера, отриманого на попередньому кроці, потім вектори розбиваються на кластери знову відповідно до того, який з нових центрів виявився ближчим за обраною метрикою. Алгоритм завершується, коли на якісь ітерації не відбувається зміни внутрішньо-кластерної відстані. Це відбувається за кінцеве число ітерацій, так як кількість можливих розбиттів кінцевої множини кінцева, а на кожному кроці сумарна квадратичне відхилення зменшується, тому зациклення неможливо.

У роботі розглянуто підхід до виявлення колективних аномалій у часових рядах, заснований на використанні методів кластеризації, зокрема методу  $k$ -середніх.

#### **Список використаних джерел:**

1. Jiang Y., Zeng C., Xu J., Li T. «Real time contextual collective anomaly detection over multiple data streams», 2014.
2. Anguilli F., Fassetti F. «Detecting distance-based outliers in streams of data». In CIKM, 2007.
3. Mohammad Braei, Dr.-Ing. Sebastian Wagner. «Anomaly detection in univariate time-series: a survey on the state -of-the-art», 2020.
4. Шкодырев В.П., Ягафаров К.И., Баштовенко В.А., Ильин Е.Э. «Обзор методов обнаружения аномалий в потоках данных».