

ДОСЛІДЖЕННЯ КЛАСТЕРНОЇ СТРУКТУРИ ЧАСОВИХ РЯДІВ

Башкатов Є.О.

Науковий керівник – д.т.н., проф. Кіріченко Л.О.

Харківський національний університет радіоелектроніки
61166, Харків, просп. Науки, 14, каф. прикладної математики,

тел. (057) 702-14-36, e-mail: yevhen.bashkatov@nure.ua

These abstracts propose an algorithm for isolating the cluster structure in time series. This will allow to process samples of large volume, by reducing the dimensionality of the original sample for further effective statistical analysis in order to establish the properties of the original time series, such as fractality. The algorithm is universal for constructing dynamic visibility graphs of DVG for arbitrary time intervals of measurement reflection.

Однією з цілей аналізу часових рядів є знаходження певних закономірностей в їх структурі. Новим і цікавим підходом дослідження структури часового ряду є виявлення особливостей за допомогою побудови графа динамічної видимості [1].

У даній роботі пропонується модифікація алгоритму побудови графа динамічної видимості, в якій виключається попарний аналіз загальної кількості вершин, що відповідають значенням даного часового ряду, що веде до істотного скорочення обчислювальних ресурсів і зменшення часу обробки вихідної інформації. Запропонований алгоритм носить універсальний характер для побудови графів динамічної видимості для довільних проміжків часу відображення вимірювань і дозволяє уявити тимчасової ряд у вигляді набору непересічних кластерів.

Кластером ми будемо називати сукупність пов'язаних між собою вузлів графа, не пов'язаних з іншими кластерами і розташованими між двома вершинами графа видимості. При цьому одиничний вузол, не пов'язаний з іншими вузлами, також буде кластером. Вершина, яка є початком кластера, належить цьому кластеру. Таким чином, кожен часовий ряд нам характеризується зв'язком між кількістю кластерів і загальному кількістю вузлів. У роботі в якості вихідної інформації розглянуто тимчасової ряд, що містить в собі вимірювання температури через рівні проміжки часу, рівні одному дню протягом ряду років [2].

На рисунку 1 відображені результати роботи запропонованого алгоритму для розбиття на кластери частини часового ряду температур зафіксованих з 01.01.1981р. по 12.01.1981р. З вершини в точці 0 виходять промені зору і вказують на наступну вершину початку кластера, беручи до уваги факт видимості всіх попередніх значень. В ході роботи програми, реалізованої на алгоритмічній мові Python, виділені кластери з вершинами [0,1], [2, 3, 4, 5], [6], [7], [8, 9, 10, 11] з довжинами 2, 4, 1, 1, 4 відповідно.

При цьому, сума довжин виділених кластерів дорівнює кількості значень досліджуваного ряду.

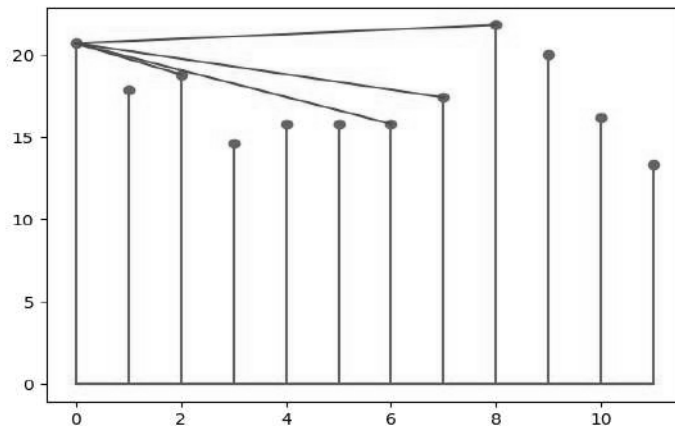


Рисунок 1 – Результат роботи алгоритму пошуку кластерів

В результаті реалізації запропонованого в роботі алгоритму кластеризації вдається виділити кластерну структуру часового ряду без побудови графа взаємної видимості. Виділення кластерів дозволить отримати кількісні характеристики для ефективного статистичного аналізу часових рядів, а також їх якісного порівняння.

Список використаних джерел:

1. Снарский А. А. Безсуднов И. В. Критические явления в графе динамической видимости. URL: <http://old.bulletin.kpi.ua/files/2013-4-25.pdf>.
2. Brabban P. Daily Minimum Temperatures in Melbourne URL: <https://www.kaggle.com/paulbrabban/daily-minimum-temperatures-in-melbourne>.