

УДК 519.7:004.8

МЕТОДИ КЛАСТЕРИЗАЦІЇ ЧАСОВИХ РЯДІВ ТРЕНДУ ЗА ДЕТЕКТУВАННЯ ТРЕНДУ

Слепцов О.М.

Науковий керівник – д-р техн. наук, проф. Кіріченко Л.О.
Харківський національний університет радіоелектроніки, каф. ПМ
м. Харків, Україна

тел. +38 (057) 702-14-36, email: oleksandr.slieptsov@nure.ua

The article discusses the utilization of machine learning techniques for clustering time series data based on predetermined criteria, enabling the identification of trend components. Various clustering methods, including k-means, are employed in the article. Additionally, the article explores multiple criteria for trend detection in short time series, which are often characterized by noise. Experimental findings demonstrate that using these criteria as features yields high clustering accuracy, as evidenced by quantitative and qualitative metrics. However, the results are contingent upon several factors, such as the length of the time series, the type of trend, and the choice of clustering method.

Визначення тренду в часових рядах є важливим інструментом аналізу в різних галузях, таких як фінансова аналітика, аналіз даних, виробничий аналіз та прогнозування. Використання методів машинного навчання для кластеризації часових рядів є ефективним підходом для аналізу великих обсягів часових даних, який дозволяє автоматично групувати ряди на основі їх подібності, виявляти причинно-наслідкові зв'язки та передбачати майбутні зміни.

Критерій тренду використовується для визначення наявності та напрямку тренду в часових рядах, що дозволяє аналізувати динаміку змін в часі та робити висновки про тенденції та зміни в досліджуваних даних.

Один з підходів до кластеризації часових рядів на основі ознак наявності або відсутності тренду - це використання методів машинного навчання, таких як метод кластеризації k-середніх або ієрархічна кластеризація. Кластеризація є процесом групування схожих об'єктів разом в одні кластери на основі їхніх спільних властивостей. В даному випадку, часові ряди розглядаються як об'єкти, і їхні ознаки на основі наявності або відсутності тренду використовуються для кластеризації.

Проте, визначення тренду в коротких часових рядах може бути складним завданням. Короткі часові ряди можуть бути зашумленими, тобто містити велику кількість випадкових відхилень від основного тренду, і водночас мати обмежену кількість даних для виявлення довгострокових тенденцій. Це може ускладнити процес виявлення тренду і точність результатів.

Тому, одним з важливих аспектів в кластеризації часових рядів є формування відповідних ознак, які дозволяють детектувати тренд. Це можуть бути різноманітні ознаки, такі як рухомі середні, експоненційно зважені

середні, відсоткові зміни, зміни періодичності, амплітудні характеристики та інші. Вибір та використання правильних ознак може визначити ефективність кластеризації та здатність виявлення тренду в коротких часових рядах.

У дослідженні були розглянуті різні підходи для виявлення тренду в коротких часових рядах, складаючихся з від 8 до 30 значень. Використані методи включали критерій серій, критерій інверсій, критерій Фостера-Стюарта, критерій Фішера та критерій на основі дискретного вейвлет-перетворення. Кожен з цих методів надавав статистику, яка вказувала на наявність тренду.

Для проведення експериментів використовувалися модельні часові ряди, що склалися з суми різних типів трендів (лінійних, поліноміальних, експоненційних і т. д.) та випадкових значень величин (нормальних, рівномірних і т. д.). Кожен часовий ряд був представлений у векторному вигляді, а його значення використовувалися для кластеризації.

Отримані висновки підтверджують високу ефективність використання описаних критеріїв як ознак для кластеризації коротких часових рядів. Це підтверджується як кількісними, так і якісними показниками, які свідчать про чітку розподільність кластерів та зрозумілість отриманих результатів. Проте, слід врахувати, що результати кластеризації можуть бути вплинуті кількома факторами, такими як довжина часового ряду, тип тренду та відношення сигнал/шум. Ці фактори мають бути враховані при оцінці достовірності результатів та їхній інтерпретації.

Список використаних джерел:

1. Aghabozorgi, S., Shirkhorshidi, A.S., & Wah, T.J. (2015). Time-series clustering. *A Decade Review Information systems*, 53, 16-38
2. J Aggarwal, C., & Reddy, C. (2013). *Data Clustering: Algorithms and Applications*. CRC Press.
3. Liao, T.W. (2005). Clustering of time series data – a survey. *Pattern Recognition*, 38 (11), 1857-1874.
4. Cowan, G. (1998). *Statistical Data Analysis*. Clarendon Press.