

DOI: 10.34185/1991-7848.itmm.2023.01.038

## КЛАСТЕРИЗАЦІЯ ЧАСОВИХ РЯДІВ НА ОСНОВІ НАЯВНОСТІ ТРЕНДУ ЗА ДОПОМОГОЮ МАШИННОГО НАВЧАННЯ

Кіріченко Л.О., Яндуков Д.Є., Хандак Д.В., Слепцов О.М.

*Харківський національний університет радіоелектроніки, Україна*

Визначення тренду в часових рядах є важливим аналітичним інструментом у багатьох технічних галузях, таких як фінансова аналітика, аналіз даних, виробничий аналіз, прогнозування і т. д. Кластеризація часових рядів за допомогою машинного навчання є ефективним інструментом для аналізу великих обсягів часових даних. Це дозволяє автоматично групувати ряди на основі їх подібності, може допомогти виявити причинно-наслідкові зв'язки між рядами та передбачити майбутні зміни в даних [1].

Один з підходів до кластеризації часових рядів за ознаками наявності або відсутності тренду - це використання методів машинного навчання, таких як метод k-середніх кластеризація або ієрархічна кластеризація [2, 3]. Визначення тренда в коротких часових рядах є складним завданням, оскільки такі ряди можуть бути дуже зашумленими і мати мало даних для виявлення довгострокових тенденцій. Тому одним із найважливіших завдань є формування ознак, які дозволяють детектувати тренд.

В роботі були розглянуті такі критерії, що дозволяють визначати наявність тренда в коротких часових рядах (від 8 до 30 значень): критерій серій, критерій інверсій, критерій Фостера-Стюарта, критерій Фішера та критерій на основі дискретного вейвлет-перетворення [4]. Кожен із критеріїв, застосований до часового ряду, генерує статистику - число, що відповідає наявності тренду. Таким чином кожен часовий ряд представлений у вигляді вектора, значення якого використовуються для кластеризації.

Були проведені чисельні експерименти з модельними часовими рядами, які являли собою суму складової з функцією тренду різного типу (лінійної, поліноміальної, експоненційної та ін.) та незалежними значеннями випадкової величини (нормальної, рівномірної та др.).

Отримані результати показали, що використання описаних критеріїв як ознаки демонструє високу точність кластеризації. Це підтверджується як кількісними, так і якісними показниками, які говорять про високий ступінь поділу кластерів і хорошу інтерпретованість результатів. Проте слід зазначити, що результати кластеризації залежать від кількох факторів. Зокрема, довжина ряду, вид тренду і відношення сигнал/шум суттєво впливають на результати.

### Література

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

### TREND-BASED CLUSTERING OF TIME SERIES USING MACHINE LEARNING

Kirichenko Liudmyla, Yandukov D., Khandak D., Slietsov O.

**Abstract.** The article discusses the use of machine learning methods for clustering time series according to specified criteria, allowing to determine the presence of a trend component. Several clustering methods are used including k-means . The article explores several criteria for detecting trends in short time series, which are often noisy . Experimental results indicate that using these criteria as features demonstrates high clustering accuracy, with both quantitative and qualitative metrics. The results depend on various factors, such as the length of the time series, the type of trend, and the choice of clustering method.

**Keywords:** trend, clustering, hour series, machine learning, trend criterion.

### References

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