University of Qom
Engineering Management and Soft Computing
JEMSC.QOM
Online ISSN: 2538-2675

Localization of Global Forecasting Models with a Clustering Approach

Hossein Abbasimehr[1] and Mohammad Khodizadeh-Nahari[2]

1 .Associate Professor, Faculty of Information Technology and Computer Engineering, Azarbaijan Shahid Madani University, Tabriz, Iran, Email: abbasimehr@azaruniv.ac.ir
2. Corresponding Author, Assistant Professor, Faculty of Information Technology and Computer Engineering, Azarbaijan Shahid Madani University, Tabriz, Iran, Email: m.khodizadeh@azaruniv.ac.ir

Article Info

Article type:
Research Article

Article history:
Received 5 November 2024
Received in revised form 10 February 2025
Accepted 4 August 2025
Published online 21 August 2025

Keywords:
Time series forecasting,
Global forecasting model,
Time series clustering,
Long short term memory network.

ABSTRACT

With the increasing generation of time series data, forecasting models trained on a set of time series, known as global forecasting models, outperform univariate forecasting models trained on individual series. However, the performance of global models may decrease when faced with heterogeneous data sets of time series with different lengths. In this study, a new method for localization of clustering-based global forecasting models is presented. The main steps of the proposed method include (1) extracting relevant features from each time series (2) clustering time series based on features extracted using K-Medoids and spectral clustering algorithms (3) implementing a global forecasting model using a combination of Temporal Convolution Network and its training for each cluster. To evaluate the prediction accuracy of the proposed approach, experiments were conducted on the M3 dataset that contains 1426 time series with unequal-length. The results of the experiments show the superior performance of the proposed clustering-based models compared to the baseline models and the benchmark models. The proposed model has 0.57 less error in terms of SMAPE metric.

Cite this article: Abbasimehr, H. & Khodizadeh-Nahari, M. (2025)., Localization of Global Forecasting Models with a Clustering Approach. *Enginiiring Management and Soft Computing, 11* (1). 134-155. **DOI**: https://doi.org/10.22091/jemsc.2025.11595.1218

DOI: 10.22091/jemsc.2025.11595.1218
Publisher: University of Qom

1) INTRODUCTION

A global forecasting model is a model that is obtained by considering samples from several time series. Compared to separate models for each series, these models perform better in predicting related time series. As a result of this evolution, the tendency to use deep learning methods as predictive models of time series has become more popular than other statistical methods. Considering that most time series are not homogeneous, the model obtained using the whole series usually does not have good accuracy. For this reason, localization of global models is a method to increase the accuracy of global forecasting models. The motivation behind this study is to propose a new method for the localization of global forecast models based on clustering.

2) Solution Methid

In this research, we use a clustering-based approach to localize global prediction models. The proposed method generally includes three stages: pre-processing, global forecasting, and post-processing. The pre-processing step comprises extracting the features of the time series for clustering, clustering the time series based on the extracted features. The global forecasting stage includes special pre-processing operations, such as mean normalization, logarithmic transformation, and de-seasonalization, as well as training a forecasting model based on the TCN-CNN model. The topology of the model is represented in Table 1. In the post-processing stage, the obtained forecasts are returned to their original scale by considering the parameters calculated in the pre-processing. In this research, we use the M3 dataset, which includes 1426 time series, to evaluate the proposed method. For clustering time series, we employed K-Medoids (CL.KM) and Spectral clustering (CL. SP) algorithms. The results of experiments indicated that clustering-based models have better results than their baseline counterparts in sMAPE and MASE error measures. Also, the results show the superiority of the proposed models over some benchmarks.

Table1. Topology of TCN-CNN

Layers
Input (shape: (1, Lag (input window))
TCN (filters: 64, kernel size: 3, dilations: (1, 2, 4, 8))
1D Convolutional (filters: 64, kernel size: 4, activation: linear)
1D Convolutional (filters: 16, kernel size: 4, activation: linear)
Flatten
Dense (units: Forecast horizon, activation: linear)
Dense

3) Equations

We use relations 1- 4 for normalization, logarithmic transformation, de-seasonalization.

$$X_{i,\text{normalized}} = \frac{X_i}{\frac{1}{k}\sum_{t=1}^{k} X_{i,k}} \tag{1}$$

k: number of points in time series X_i.

$$T_{i,\text{normalized \& log \_trans}} = \begin{cases} \log\left(T_{i,\text{normalized}}\right), & \min(TS) > 0; \\ \log\left(T_{i,\text{normalized}} + 1\right), & \min(TS) = 0, \end{cases} \tag{2}$$

$$T_{i,\text{normalized \& log \_trans}} = s_i + t_i + r_i \tag{3}$$

$$T_{i,\text{normalized \& log}_{\text{trans}}\text{ \& deseasonalized}} = T_{i,\text{normalized \& log \_trans}} - s_i \tag{4}$$

s_i: seasonal components, t_i: trend components and r_i residual components

For model evaluation we used two major metric: sMAPE (relation 5) and MASE (relation 6)

$$\text{sMAPE} = \frac{200}{h}\sum_{i=1}^{h}\left(\frac{|A_i - F_i|}{|A_i| + |F_i|}\right) \tag{5}$$

$$\mathrm{MASE} = \frac{1}{h} \frac{\sum_{i=1}^{h}(|A_i - F_i|)}{\frac{1}{n-M}\sum_{i=S+1}^{n}(|A_i - A_{i-S}|)} \quad (6)$$

Ai: Actual value Fi: Forecasted value

4) Conclusion

The results of this study indicated that clustering-based global models (CL. KM and CL.SP models in Table 2) that train a separate model per cluster outperform their baseline counterparts, in which a single model is trained using all time series data. Time series clustering identifies homogenous clusters, which leads to accuracy improvement. Also, we investigated the effect of de-seasonalization and showed that it positively impacts the performance of both clustering-based models and baseline ones.

Table 2. Comparison of the proposed model with competing models

MASE	sMAPE	Model
0.894	13.94	CL.KM
0.898	13.99	CL.SP
0.914	14.51	FFORMA
1.065	16.41	CatBoost
1.167	15.74	DeepAR
0.934	14.76	N-BEATS

REFERENCES

Bandara, K., C. Bergmeir, and S. Smyl, Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach. Expert Systems with Applications,140: p. 112896, (2020), https://doi.org/10.1016/j.eswa.2019.112896

Godahewa, R., et al., Ensembles of localised models for time series forecasting. Knowledge-Based Systems, 233: p. 107518, (2021), https://doi.org/10.1016/j.knosys.2021.107518

Bandara, K., Forecasting with Big Data Using Global Forecasting Models, in Forecasting with Artificial Intelligence: Theory and Applications, M. Hamoudia, S. Makridakis, and E. Spiliotis, Editors. Springer Nature Switzerland: Cham. p. 107-122, (2023), https://doi.org/10.1007/978-3-031-35879-1.

JEMSC.QOM
شاپا الکترونیکی: ۲۶۷۵-۲۵۳۸
مدیریت مهندسی و رایانش نرم
دانشگاه قم

محلی‌سازی مدل‌های پیش‌بینی سراسری با یک رویکرد خوشه‌بندی

حسین عباسی‌مهر [۱] و محمد خودی‌زاده نهاری [۲]✉

۱. دانشیار، دانشکده فناوری اطلاعات و مهندسی کامپیوتر، دانشگاه شهید مدنی آذربایجان، تبریز، ایران، رایانامه: abbasimehr@azaruniv.ac.ir
۲. نویسنده مسئول، استادیار، دانشکده فناوری اطلاعات و مهندسی کامپیوتر، دانشگاه شهید مدنی آذربایجان، تبریز، ایران، رایانامه : m.khodizadeh@azaruniv.ac.ir

اطلاعات مقاله

نوع مقاله: مقاله پژوهشی

تاریخ دریافت: ۱۴۰۳/۰۸/۱۵

تاریخ بازنگری: ۱۴۰۳/۱۱/۲۲

تاریخ پذیرش: ۱۴۰۴/۰۵/۱۳

تاریخ انتشار: ۱۴۰۴/۰۵/۳۰

کلیدواژه‌ها:
پیش‌بینی سری زمانی، خوشه‌بندی سری زمانی، شبکه حافظه کوتاه‌مدت طولانی، مدل پیش‌بینی سراسری

چکیده

با تولید روزافزون داده‌های سری زمانی، مدل‌های پیش‌بینی که بر روی مجموعه‌ای از سری‌های زمانی آموزش داده می‌شوند و به‌عنوان مدل‌های پیش‌بینی سراسری شناخته می‌شوند، نسبت به مدل‌های پیش‌بینی تک‌متغیره که بر روی سری‌های جداگانه آموزش می‌بینند، عملکرد بهتری دارند. با این حال، عملکرد مدل‌های سراسری ممکن است در مواجهه با مجموعه داده‌های ناهمگن سری‌های زمانی با طول‌های متفاوت، کاهش یابد. در این مطالعه، یک روش جدید برای محلی‌سازی مدل‌های پیش‌بینی سراسری مبتنی بر خوشه‌بندی ارائه می‌شود. مراحل اصلی روش ارائه‌شده شامل: ۱) استخراج ویژگی‌های مرتبط از هر سری زمانی، ۲) خوشه‌بندی سری‌های زمانی بر پایه ویژگی‌های استخراج‌شده با استفاده از الگوریتم‌های K-Medoids و خوشه‌بندی طیفی، ۳) پیاده‌سازی دو مدل پیش‌بینی سراسری با استفاده از مدل مبتنی بر شبکه کانولوشنی زمانی و مدل مبتنی بر مکانیزم توجه چندگانه می‌باشد. برای ارزیابی دقت پیش‌بینی، آزمایشاتی بر روی مجموعه داده M3 شامل ۱۴۲۸ سری زمانی با طول‌های غیریکسان انجام شد. نتایج آزمایشات، نشان‌دهنده عملکرد برتر مدل‌های پیشنهادی در مقایسه با مدل‌های پایه و مدل‌های محک است. در معیار SMAPE، مدل پیشنهادی ۰.۵۷ خطای کمتری دارد.

استناد: عباسی‌مهر، حسین و محمد، خودی‌زاده نهاری. (۱۴۰۴). «محلی‌سازی مدل‌های پیش‌بینی سراسری با یک رویکرد خوشه‌بندی». *مدیریت مهندسی و رایانش نرم*، دوره ۱۱ (۱). صص: ۱۵۵-۱۳۴. https://doi.org/10.22091/jemsc.2025.11595.1218

ناشر: دانشگاه قم

۱) مقدمه

با توجه به رشد روزافزون تولید داده در صنایع مختلف ازجمله انرژی، خرده‌فروشی و گردشگری، مدل‌های پیش‌بینی سری‌های زمانی نقش بسیار حیاتی و حساسی را در این صنایع ایفا می‌کنند (گوداهیوا[۱] و همکاران، ۲۰۲۱). به‌همین‌دلیل و با توجه به پیچیدگی روزافزون مسائل پیش‌بینی سری‌های زمانی، نیاز است این مدل‌ها از پیش‌بینی سری‌های زمانی به‌صورت جداگانه به پیش‌بینی مجموعه‌ای از سری‌های زمانی مرتبط‌به‌هم تکامل یابند (باندرا[۲] و همکاران، ۲۰۲۰). در صنایع مختلف، انجام پیش‌بینی‌های دقیق نقش بسیار حیاتی در تصمیم‌گیری‌های تجاری ایفا می‌کند. برای مثال: در صنعت خرده‌فروشی مانند فروشگاه‌های بزرگی چون آمازون[۳] و یا والمارت[۴]، پیش‌بینی میزان تقاضا برای تنظیم صحیح موجودی کالاها، نقش بسزایی دارد. در حوزه تامین انرژی، پیش‌بینی میزان مصرف در میزان تخصیص منابع ارزشمند سوختی و مدیریت بازار انرژی، بسیار حیاتی است. همچنین در صنعت گردشگری و خدمات حوزه سلامت، پیش‌بینی میزان تقاضا برای خدمات در مناطق جغرافیایی مختلف از اهمیت بالایی برخوردار است (هیوامالاگ[۵] و همکاران، ۲۰۲۲ و مارتینز[۶]). و همکاران، ۲۰۱۸)

تاکنون مطالعات بسیاری در پیش‌بینی سری‌های زمانی انجام شده‌است. در این زمینه، اکثر تحقیقات بر پیش‌بینی سری‌های زمانی تکی متمرکز بوده‌اند. بطورکلی این تحقیقات دو دسته از روش‌ها را مورد استفاده قرار داده‌است. دسته اول شامل: روش‌های سنتی آماری نظیر میانگین متحرک[۷]، هموارسازی نمایی[۸] و روش ARIMA [۹] (هیندمن[۱۰] و همکاران، ۲۰۰۲ و باکس و همکاران ۲۰۱۵) که مهمترین مشکلی که روش‌های سنتی دارند، این است که برای پیش‌بینی سری‌های زمانی حاوی الگوهای خطی مناسب هستند و دسته دوم که برای مدل‌سازی الگوها و روابط غیرخطی موجود در داده‌ها مناسب هستند، اخیرا در روش‌های یادگیری ماشین و یادگیری عمیق کاربردهای موفقی داشته‌اند (پارمیزان[۱۱] و همکاران، ۲۰۱۹). در دسته روش‌های یادگیری ماشین، شبکه‌های عصبی مصنوعی (ANN)، ماشین‌های بردار پشتیبان (SVM) و K نزدیک‌ترین همسایه (KNNs) جزء پرکاربردترین روش‌ها هستند. همچنین گونه‌های مختلف روش‌های شبکه عصبی بازگشتی (RNN) نظیر شبکه حافظه کوتاه‌مدت طولانی (LSTM)، واحد گیتی بازگشتی (GRU)، عملکرد موفقی در پیش‌بینی سری زمانی از خود نشان داده‌اند (هیوامالاگ و همکاران، ۲۰۲۱). شبکه‌های عصبی بازگشتی، توانایی پردازش و مدل‌سازی وابستگی بین داده‌ها در طول زمان را دارند. اغلب مطالعات پیشین روی ساخت مدل بر پایه یک سری زمانی تکی متمرکز بوده‌اند (باندرا و همکاران، ۲۰۲۰). بزرگترین چالش استفاده از روش‌های یادگیری ماشین و یادگیری عمیق با اتخاذ این رویکرد، کافی‌نبودن نمونه‌های داده جهت آموزش بهینه مدل‌ها است.

یک مدل پیش‌بینی سراسری، مدلی است که با درنظرگرفتن نمونه‌هایی از چندین سری زمانی به‌دست می‌آید. این

---

[1] Godahewa
[2] Bandara
[3] www.amazon.com
[4] www.walmart.com
[5] Hewamalage
[6] Martínez
[7] Moving average
[8] Exponential Smoothing
[9] Autoregressive integrated moving average
[10] Hyndman
[11] Parmezan

مدل‌ها، در مقایسه با مدل‌های جداگانه، عملکرد بهتری در پیش‌بینی سری‌های زمانی مرتبط‌به‌هم دارند (گوداهیوا و همکاران، ۲۰۲۱، هیوامالاگ و همکاران، ۲۰۲۲ و هیوامالاگ و همکاران، ۲۰۲۱). درنتیجه این تکامل، تمایل به استفاده از روش‌های یادگیری عمیق به‌عنوان مدل‌های پیش‌بینی‌کننده سری‌های زمانی نسبت به سایر روش‌های آماری، محبوبیت بیشتری یافته‌است. با توجه به اینکه در یک مجموعه سری زمانی، اغلب سری‌های زمانی همگن نیستند، معمولا مدل به‌دست‌آمده با استفاده از کل سری‌ها، از دقت خوبی برخوردار نیست. به‌همین‌دلیل، محلی‌سازی مدل‌های سراسری یک روش برای افزایش دقت مدل‌های پیش‌بینی سراسری می‌باشد (گوداهیوا و همکاران، ۲۰۲۱ و باندرا و همکاران، ۲۰۲۰).

در این مطالعه، یک روش جدید برای محلی‌سازی مدل‌های پیش‌بینی سراسری مبتنی بر خوشه‌بندی ارائه می‌شود. با استفاده از خوشه‌بندی سری‌های زمانی و افراز آنها به خوشه‌هایی شامل سری‌های زمانی مشابه، قادر خواهیم بود راه‌حلی برای مشکل محلی‌سازی مدل‌های پیش‌بینی سراسری ارائه دهیم. زیرا با استفاده از این روش می‌توان یک مدل سراسری پیش‌بینی برای هر خوشه آموزش داد که بتواند جزئیات دقیق‌تری از سری‌های زمانی مشابه را فرا گیرد.

مراحل اصلی روش ارائه‌شده شامل: ۱) استخراج ویژگی‌های مفید و مرتبط از هر سری زمانی، ۲) خوشه‌بندی سری‌های زمانی بر پایه ویژگی‌های استخراج‌شده با استفاده از الگوریتم‌های K-Medoids (شوبرت و روزیو[12]، ۲۰۲۱) و خوشه‌بندی طیفی[13] (یی و ساکورای[14]، ۲۰۱۶)، ۳) پیاده‌سازی دو مدل پیش‌بینی سراسری با ترکیب شبکه پیچشی زمانی[15] و شبکه پیچشی[16] و مدل مبتنی بر مکانیزم توجه چندگانه[17] است.

با توجه به خاصیت ناهمگونی موجود در مجموعه داده‌های سری زمانی، پیش‌پردازش سری‌ها و همچنین پس‌پردازش نتایج پیش‌بینی از اهمیت فوق‌العاده‌ای برخوردار است. در روش پیشنهادی در مرحله پس‌پردازش نتایج پیش‌بینی از مدل هموارسازی نمایی[18] جهت تخمین مولفه‌های فصلی استفاده می‌شود.

روش پیشنهادی روی مجموعه‌داده M3 پیاده‌سازی شد. نتایج آزمایشات، نشان‌دهنده برتری مدل‌های پیش‌بینی حاصل از رویکرد پیشنهادی می‌باشد. روش‌های مبتنی بر خوشه‌بندی نسبت به روش پایه دارای عملکرد پیش‌بینی بهتری هستند. همچنین مقایسه دقت مدل‌ها نسبت به مدل‌های موجود در تحقیقات فعلی حاکی از برتری مدل پیشنهادی است. بطور خلاصه نوآوری‌های این تحقیق به شرح زیر است:

۱. ارائه رویکردی جهت محلی‌سازی مدل‌های پیش‌بینی سراسری به‌منظور افزایش دقت پیش‌بینی در مجموعه داده‌های ناهمگن سری زمانی

۲. ارائه یک روش خوشه‌بندی مبتنی بر ویژگی جهت خوشه‌بندی سری‌های زمانی با طول متغیر

۳. پیاده‌سازی دو مدل پیش‌بینی مبتنی بر شبکه‌های عصبی عمیق

۴. ارزیابی رویکرد پیشنهادی با استفاده از ۱۴۲۸ سری زمانی مجموعه‌داده M3

[12] Schubert & Rousseeuw
[13] Spectral clustering
[14] Ye & Sakurai
[15] Temporal Convolutional Networks (TCN)
[16] Convolutional neural network
[17] Multi-Head Attention
[18] Exponenial Smoothing

ساختار ادامه این مقاله بدین شرح است: در بخش ۲ به مرور ادبیات پیش‌بینی سراسری پرداخته می‌شود، در بخش ۳ روش پیشنهادی با جزئیات کامل توصیف خواهد شد. بخش ۴ شامل توصیف مجموعه‌داده، تنظیمات آزمایشات و تشریح معیارهای ارزیابی عملکرد، نتایج پیش‌بینی و تحلیل عملکرد مدل‌ها خواهد بود. در بخش ۵ نتیجه‌گیری بیان خواهد شد.

۲) پیشینه تحقیق

در سال‌های اخیر، مطالعات بسیاری در حوزه پیش‌بینی سری‌های زمانی مبتنی بر یادگیری ماشین و مخصوصا یادگیری عمیق صورت گرفته‌است (هیوامالاگ و همکاران، ۲۰۲۱). بطورکلی می‌توان مدل‌های پیش‌بینی سری زمانی مبتنی بر رویکرد یادگیری ماشین را به دو دسته اصلی سراسری و محلی تقسیم کرد (گوداهیوا و همکاران، ۲۰۲۱) که در میان آنها، مدل‌های پیش‌بینی سراسری (جانوفسکی [۱۹]و همکاران ۲۰۲۰)، یک رویکرد نسبتا جدید هستند و پس از استفاده به‌عنوان پایه و اساس حل مسئله در رقابت‌های M4 بسیار محبوب شدند (اسمیل [۲۰]، ۲۰۲۰). برخلاف رویکردهای محلی مانند هموارسازی نمایی (ETS) (هیندمن و همکاران، ۲۰۰۸) و مدل‌های خود رگرسیون میانگین متحرک انباشته (ARIMA) (باکس [۲۱] و همکاران، ۲۰۱۵) که در آنها یک مدل پیش‌بینی به ازای هر سری زمانی ساخته می‌شود، در رویکرد مدل‌های پیش‌بینی سراسری [۲۲] یک مدل واحد با استفاده از تمام سری‌های زمانی ساخته می‌شوند (جانوفسکی و همکاران ۲۰۲۰). این مدل‌ها پتانسیل بسیار بالایی در ایجاد دقت بالای پیش‌بینی از خود نشان داده‌اند. برای مثال: در مسابقات معتبر M4 و M5 (ماکریداکیس [۲۳] و همکاران، ۲۰۲۲ و ماکریداکیس و همکاران، ۲۰۱۸) با ارائه دقت پیش‌بینی بالا، این مدل‌ها توانستند برنده مسابقات شوند.

با داشتن یک مجموعه‌داده با p سری زمانی $X = \{X_1, X_2, \dots, X_p\}$، در پیش‌بینی سراسری هدف، ساخت مدل F با درنظرگرفتن تمامی داده‌های مجموعه داده است. مدل سراسری F به این صورت تعریف می‌شود:

$$X_i^M = F(X, \theta) \quad (۱)$$

که در آن M نشان‌دهنده افق پیش‌بینی است. i نشان‌دهنده سری زمانی موردنظر است که مقادیر آینده آن توسط مدل F پیش‌بینی خواهد شد و θ پارامترهای مدل F است که با اطلاعات تمامی سری‌های زمانی تعیین می‌شود. برای مثال: در یک شبکه عصبی، θ نشان‌دهنده وزن‌ها و بایاس است. بنابراین براساس مدل پایه‌ای که در ساخت مدل پیش‌بینی استفاده می‌شود، θ تغییر می‌کند. درصورت استفاده از رویکرد سنتی، پیش‌بینی سری زمانی شامل: ساخت مدل‌های جداگانه به ازای هر سری خواهد شد. درنتیجه دارای p مدل $F = \{F_1, F_2, \dots, F_p\}$ خواهیم بود که پارامترهای هر مدل با داده‌های هر سری به‌دست می‌آیند.

مدل‌های سراسری این قابلیت را دارند که در مجموعه‌ای از داده‌های سری زمانی با طول‌های متفاوت اعمال شوند. در مواردی که سری‌های زمانی طول‌های متفاوتی داشته باشند، تعداد نمونه‌های تولیدشده برای هر سری زمانی ممکن است

[19] Januschowski
[20] Smyl
[21] Box
[22] Global Forecasing Models (GFM)
[23] Makridakis

متفاوت می‌شود. با این حال، چنین تناقضی، مانعی در روند آموزش مدل‌های سراسری به‌وجود نمی‌آورد. این یک تمایز قابل توجه بین GFM ها و مدل‌های پیش‌بینی تک‌متغیره است.

شبکه‌های عصبی بازگشتی (RNN) به‌دلیل توانایی طبیعی آنها در پردازش داده‌های متوالی برای پیش‌بینی سری‌های زمانی نتایج خوبی داشته‌اند. در حوزه مدل‌های سراسری، RNN ها جزء مدل‌های یادگیری عمیق پُرکاربرد می‌باشند (هیواماالاگ و همکاران، ۲۰۲۱). اسمیل و کوبر[۲۴] (۲۰۱۶) معماری جدید چندلایه با شبکه LSTM معرفی کردند که شامل اتصالات پرش برای کاهش مشکل ناپدیدشدن گرادیان است. رویکرد آنها همچنین دارای ویژگی‌های کمکی به‌دست‌آمده از مدل هموارسازی نمایی (ETS) بود که بطور قابل‌توجهی دقت مدل LSTM را افزایش داد. نتایج تجربی آنها نشان داد که مدل LSTM بهبودیافته با ETS، از مدل استاندارد LSTM بهتر عمل می‌کند. در ادامه به تشریح چندین مطالعه مهم دیگر در زمینه مدل‌های پیش‌بینی سراسری می‌پردازیم.

باندرا و همکاران(۲۰۲۰)، از یک معماری LSTM چندلایه[۲۵] در کنار تکنیک‌های خوشه‌بندی سری‌های زمانی برای محلی‌سازی GFM ها و حل مشکل ناهمگونی در داده‌های سری زمانی استفاده کرده‌اند. روش پیشنهادی آنها، عملکرد بهتری نسبت به مدل LSTM که در آن یک مدل واحد روی تمام داده‌های سری زمانی آموزش داده می‌شود، از خود نشان داده‌است.

در مطالعه‌ای، سالیناس[۲۶] و همکاران (۲۰۲۰) تکنیک DeepAR را ارائه کردند، یک مدل دنباله‌به‌دنباله[۲۷] با معماری رمزگشای RNN که برای پیش‌بینی احتمالی طراحی شده‌است و قابلیت پیش‌بینی سری‌های زمانی مرتبط‌به‌هم را دارد.

لاپتیو[۲۸] و همکاران (۲۰۱۷) استفاده از شبکه رمزگذار خودکار برای آموزش یک مدل سراسری در سری‌های زمانی متنوع را پیشنهاد داده‌اند. رمزگذار خودکار ویژگی‌هایی را از سری‌های زمانی استخراج می‌کند که این ویژگی‌ها با ورودی خام برای پیش‌بینی نهایی ترکیب می‌شود.

اسمیل (۲۰۲۰) یک مدل ترکیبی مبتنی بر هموارسازی نمایی و شبکه RNN (ES-RNN) را معرفی کرده‌است. این مدل شامل پارامترهای محلی (خاص هر سری) و سراسری است که امکان یادگیری مدل سراسری با درنظرگرفتن ویژگی‌های منحصربه‌فرد هر سری را فراهم می‌کند. این روش برنده رقابت پیش‌بینی M4 شده‌است.

باندرا و همکاران (۲۰۲۱) افزایش دقت GFM را از طریق داده‌افزایی و یادگیری انتقالی پیشنهاد کردند. آنها از تکنیک‌های مختلف داده‌افزایی سری‌های زمانی برای تکمیل سری‌های زمانی استفاده کردند سپس از داده‌های تولیدشده برای آموزش مدل از طریق رویکردهای یادگیری تلفیقی و انتقالی استفاده کردند. یافته‌های آنها نشان داد که این روش‌ها از مدل‌های سراسری پایه و چندین تکنیک پیش‌بینی تک‌متغیره پیشرفته به‌ویژه در مجموعه‌داده‌های کوچک تا متوسط، بهتر عمل می‌کنند.

گوداهیوا و همکاران، (۲۰۲۱) تکنیک‌های مختلف یادگیری گروهی با مدل‌های مختلف پیشنهاد داده‌اند و نتایج،

[24] Smyl & Kuber
[25] Stacked
[26] Salinas
[27] Seq-to-Seq
[28] Laptev

حاکی از برتری مدل‌های یادگیری گروهی نسبت به مدل‌های سنتی و یادگیری عمیق دارد. باندرا (۲۰۲۳) رویکردهای مختلفی برای پرداختن به رانش مفهومی[۲۹] در پیش‌بینی سری‌های زمانی پیشنهاد داده‌است. چالشی که در مدل‌های پیش‌بینی سراسری وجود دارد این است که این مدل‌ها به اندازه کافی برای سری‌های زمانی خاص، محلی نیستند به‌ویژه زمانی که مجموعه سری‌های زمانی ناهمگن هستند. رویکردی کاربردی‌تر و عمومی‌تر برای پرداختن به این چالش که توانایی استفاده در هرگونه مدل پیش‌بینی سراسری را داشته باشد، می‌تواند توسعه مدل‌هایی برای زیرمجموعه‌هایی از تمام سری‌های زمانی یک مجموعه داده باشد. این رویکرد قادر است بطور همزمان پیچیدگی مدل را کنترل کرده و مدل‌های محلی‌تری بسازد که بر روی گروه‌های خاصی از سری‌های زمانی متمرکز می‌شوند. یک روش پیشنهادی برای اجرای این رویکرد، می‌تواند استفاده از خوشه‌بندی سری‌های زمانی هم به‌عنوان راهی برای ساختن مدل تجمیعی پیچیده کنترل‌شده و هم برای توجه به مسئله ناهمگونی باشد. مونترو و هیندمن[۳۰] (۲۰۲۱) نشان داده شده‌اند که خوشه‌بندی تصادفی سری‌های زمانی می‌تواند با بالا بردن میزان پیچیدگی مدل، دقت پیش‌بینی را بهبود بخشد. برای بررسی ناهمگونی داده‌ها در (باندرا و همکاران، ۲۰۲۰) براساس روشی که پیشنهاد شده‌است، ابتدا سری‌های زمانی خوشه‌بندی می‌شوند سپس مدل‌های پیش‌بینی سراسری جداگانه‌ای برای هر خوشه به‌دست‌آمده، آموزش داده می‌شوند.

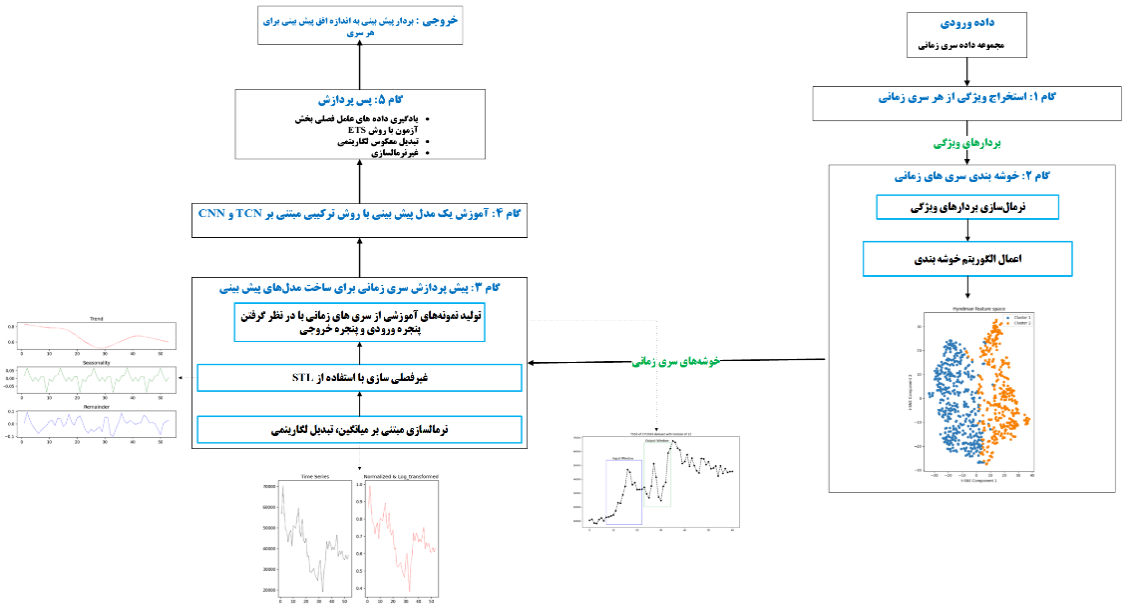
۳) روش تحقیق

در این تحقیق از یک رویکرد مبتنی بر خوشه‌بندی برای محلی‌سازی مدل‌های پیش‌بینی سراسری استفاده می‌کنیم. شکل ۱ مراحل روش پیشنهادی را توصیف می‌کند. همانطورکه در این شکل مشاهده می‌شود، روش پیشنهادی شامل سه مرحله پیش‌پردازش، ساخت مدل پیش‌بینی سراسری و پس‌پردازش است. مرحله پیش‌پردازش شامل استخراج ویژگی سری‌های زمانی برای خوشه‌بندی، خوشه‌بندی سری‌های زمانی مبتنی بر ویژگی‌های استخراج‌شده است. مرحله ساخت مدل پیش‌بینی شامل عملیات پیش‌پردازش خاص این مرحله مانند نرمال‌سازی با میانگین، تبدیل لگاریتمی، حذف مولفه فصلی است. در مرحله پس‌پردازش[۳۱] مقادیر پیش‌بینی به‌دست‌آمده توسط مدل پیش‌بینی با درنظرگرفتن پارامترهای محاسبه‌شده در پیش‌پردازش به مقیاس اولیه خود بازگردانده می‌شوند. در ادامه جزئیات هرکدام از مراحل شرح داده می‌شود.

۳-۱) استخراج ویژگی‌های سری زمانی

با توجه به اینکه اغلب در مجموعه داده‌های سری زمانی، با سری‌های دارای طول‌های غیریکسان سروکار داریم، برای گروه‌بندی سری‌های شبیه‌به‌هم خوشه‌بندی مبتنی بر ویژگی را می‌توان درنظر گرفت. ویژگی‌های استخراج‌شده از هر سری زمانی به‌صورت یک بردار درنظر گرفته می‌شود. در این مطالعه از ویژگی معرفی‌شده توسط هیندمن و همکاران (۲۰۱۹) استفاده می‌گردد (جدول ۱).

[29] Concept drift
[30] Montero & Hyndman
[31] Post processing



شکل ۱. رویکرد پیشنهادی برای محلی‌سازی مدل‌های پیش‌بینی سراسری

جدول ۱. ویژگی‌های مستخرج از سری‌های زمانی برای خوشه‌بندی

ویژگی	توضیحات
Mean	میانگین سری زمانی
Var	واریانس سری زمانی
ACF-1x	مرتبه اول همبستگی
Trend	شدت روند در سری زمانی
Linearity	شدت خطی‌بودن سری زمانی
Curvature	شدت انحنای سری زمانی
Entropy	شدت آنتروپی
Lumpiness	تغییر واریانس در باقیمانده
Spikiness	مقادیر غیرمنتظره سری زمانی
Lshift	تغییر سطح با استفاده از پنجره لغزان
Vchange	تغییر واریانس
Fspots	تعداد نقاط مسطح
Cpoints	تعداد دفعاتی که سری زمانی از خط مبدا عبور می‌کند
Klscore	امتیاز Kullback-Leibler
Change.idx	محل داده با بیشترین امتیاز Kullback-Leibler

۳-۱-۱) خوشه‌بندی سری زمانی

هدف از این گام شناسایی خوشه‌های سری‌های زمانی همگن است. در این مرحله بردارهای ویژگی محاسبه‌شده برای سری‌های زمانی به‌عنوان ورودی الگوریتم خوشه‌بندی سری‌ها مورد استفاده قرار می‌گیرد. شناسایی خوشه‌های دارای سری‌های زمانی شبیه‌به‌هم، می‌تواند عملکرد مدل پیش‌بینی را بهبود بخشد.

الگوریتم ۱ . خوشه‌بندی سری‌های زمانی براساس بردار ویژگی $\mathcal{F}$

FeaturesClustering $(\boldsymbol{\mathcal{F}}, n)$
$\boldsymbol{\mathcal{C}} \leftarrow$ Cluster $\boldsymbol{\mathcal{F}}$ using a clustering algorithm into n groups
Return $\boldsymbol{\mathcal{C}}$

۳-۱-۲) مرحله پیش‌بینی

پس از شناسایی خوشه‌ها، به ازای هر خوشه، یک GFM با استفاده از داده‌های سری زمانی آن خوشه آموزش داده می‌شود. مرحله پیش‌بینی شامل پیش‌پردازش سری‌های زمانی، استفاده از مدل پیش‌بینی برای آموزش GFM و برخی مراحل پس از پردازش است. علاوه بر الگوریتم‌های ۴ -۲ که مراحل اصلی مرحله پیش‌بینی را نشان می‌دهد، جزئیات بیشتری را در زیر ارائه می‌دهیم.

الگوریتم ۲ . مرحله پیش‌پردازش از پیش‌بینی مبتنی بر خوشه‌بندی

Forecasting $(\mathbf{D}, \boldsymbol{\mathcal{C}}, \text{lag, look forward, frequency})$
{Time series preprocessing stage}
Foreach time series $X_i \in \mathbf{D}$ **do**
$m_i \leftarrow \text{Mean}(X_i)$
$X_i \leftarrow \frac{X_i}{m_i}$ # divide by mean
$X_i^{transformed} \leftarrow \text{LogTransformation}(X_i)$
$D^{transformed}.\text{Append}(X_i^{transformed})$ # Add transformed data to list
End
$\text{Seasonalities} \leftarrow \text{DecomposeSeasonalities}(X_i^{transformed} \in D^{transformed})$

الگوریتم ۳ . مرحله ساخت مدل پیش‌بینی

{Model building stage}
$\text{HyperparametersGrid} \leftarrow$ Define Hyperparameter Grid for predictors θ_p: $[\theta_p^1, \theta_p^2, \dots, \theta_p^n]$
Initializing the predictors corresponding to hyperparameters M
Foreach cluster C_j in $\boldsymbol{\mathcal{C}}$, j in {1,2, … , n} do
Foreach $X_i^{transformed}$ in C_j do
multi-input-multi-output (MIMO)
$\text{GeneratedSamples} \leftarrow \text{GeneratingMIMO}(X_i^{transformed}, \text{InWindow, OutWindow})$
End
$D_{train}, D_{validation}, D_{test} \leftarrow \text{Splitting(GeneratedSamples)}$
$\text{Best\_GFM}_j \leftarrow \text{Train}(M, D_{train}, D_{validation})$
End

الگوریتم ٤ . مرحله پس‌پردازش

{Time series postprocessing stage}
Foreach cluster C_j in $\boldsymbol{\mathcal{C}}$, j in {1,2, … , n}
Foreach $X_i^{transformed}$ in C_j do
$r_i \leftarrow \text{Predict}(\text{Best\_GFM}_j, X_i^{transformed})$
$r_i^{reseasonalize} \leftarrow \text{Reseasonalization}(r_i)$
$r_i^{denormalize} \leftarrow \text{Denormalization}(r_i^{reseasonalize})$
$r_i^{InverseLog} \leftarrow \text{InverseLogTransform}(r_i^{denormalize})$
$\mathcal{R}_{C_j}.\text{Append}(r_i^{InverseLog})$ #Add forecasting results to R^k
End
$\mathcal{R}.\text{Append}(\mathcal{R}_{C_j})$
End
Return $\mathcal{R}$

به جهت اینکه سری‌های موجود در یک مجموعه داده سری زمانی، دارای مقیاس و روند مختلفی هستند، از این رو به‌منظور آموزش مناسب GFM روی $TS = \{X_i\}_{i=1}^{N}$ نیاز است تا پیش‌پردازش‌هایی روی داده‌ها صورت گیرد. پیش‌پردازش‌های صورت‌گرفته به این شرح است:

۳-۱-۳) نرمال‌سازی

برای یکنواخت‌کردن مقیاس در سری‌های زمانی در ابتدا مقادیر هر سری بر میانگین آن تقسیم می‌شوند(رابطه (۲)).

$$X_{i,normalized} = \frac{X_i}{\frac{1}{k}\sum_{t=1}^{k} X_{i,k}} \quad (۲)$$

که در آن K نشان‌دهنده تعداد نقاط سری زمانی X_i و $i \in \{1,2,\dots,N\}$ است.

۳-۱-۴) تبدیل لگاریتمی

به‌منظور تثبیت واریانس هر سری زمانی، تبدیل لگاریتمی روی سری نرمال‌شده با استفاده از رابطه (۳) صورت می‌گیرد.

$$T_{i,normalized\ \&\ log\ \_trans} = \begin{cases} \log(T_{i,normalized}), & \min(TS) > 0; \\ \log(T_{i,normalized} + 1), & \min(TS) = 0, \end{cases} \quad (۳)$$

۳-۱-۵) غیرفصلی‌سازی

در این مرحله عمل غیرفصلی‌سازی روی سری زمانی انجام می‌شود. روش استخراج الگوی فصلی که از آن بهره خواهیم برد، معروف به تجزیه الگوی فصلی- روند با استفاده از الگوریتم LOESS موسوم به STL خواهد بود (کلولند[۳۲] و همکاران، ۱۹۹۰). بطور خلاصه این روش به‌عنوان یک روش قوی برای تجزیه یک سری زمانی به سه بخش الگوی فصلی، الگوی روند و سری زمانی باقیمانده شناخته می‌شود. برای غیرفصلی‌سازی از رابطه (۴) برای تجزیه سری به مولفه‌های فصلی، روند، باقیمانده استفاده کرده و سپس با استفاده از رابطه (۵) مولفه فصلی را حذف می‌کنیم.

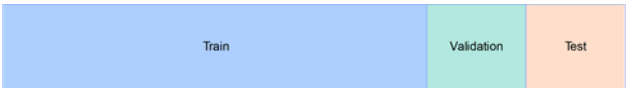
$$T_{i,normalized\ \&\ log\ \_trans} = s_i + t_i + r_i \quad (۴)$$

$$T_{i,normalized\ \&\ log_{trans}\ \&\ deseasonalized} = T_{i,normalized\ \&\ log\ \_trans} - s_i \quad (۵)$$

در روابط (۴) و (۵) t_i مولفه فصلی، $t_i, r_i\ s_i$ مولفه روند و r_i باقی مانده را نشان می‌دهد.

۳-۱-۶) تقسیم داده‌ها

در مسئله پیش‌بینی سری زمانی، هدف پیش‌بینی h نقطه آینده یک سری است. بدین‌منظور در ابتدا داده‌های هر سری زمانی به سه بخش آموزشی، اعتبارسنجی و آزمون تقسیم می‌شود. بخش آموزشی و اعتبارسنجی در ساخت مدل پیش‌بینی مورد استفاده قرار می‌گیرند و بخش آزمون برای ارزیابی عملکرد مدل استفاده می‌شود (شکل ۲) .

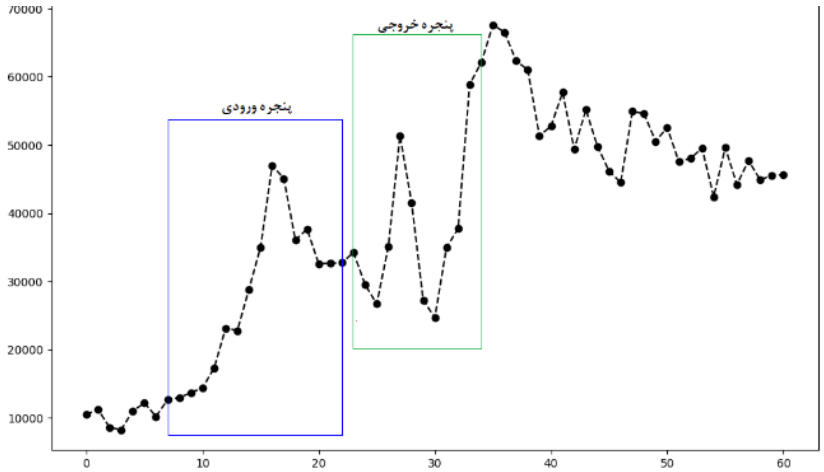


شکل ۲. مثالی از تقسیم داده‌های یک سری زمانی به سه بخش آموزشی، اعتبارسنجی و آزمون

[32] Cleveland

۳-۱-۷) نمونه‌سازی با روش پنجره لغزان

بسیاری از مطالعات، مدل‌های پیش‌بینی را بر مبنای استراتژی پیش‌بینی تک‌نقطه طراحی می‌کنند. در این تحقیق برای پیش‌بینی h نقطه آینده، از استراتژی پیش‌بینی نقاط چندگانه MIMO استفاده می‌کنیم که در آن h نقطه آینده بطور همزمان توسط مدل پیش‌بینی می‌شود. همچنین برای تبدیل سری‌های زمانی به فرمت قابل‌استفاده توسط شبکه‌های عصبی از روشی که تحت عنوان پنجره لغزان است، استفاده می‌شود. در ابتدا یک سری زمانی با طول مشخص tsLength با استفاده از این روش و به‌صورت یک گام در هر لغزش پنجره، به مجموعه‌ای از تکه‌هایی به تعداد tsLength، inputLength و outputLength تبدیل می‌شود. در اینجا outputLength نشان‌دهنده اُفق پیش‌بینی خواهد بود و inputLength نشانگر تعداد مقادیری است که به‌عنوان ورودی استفاده خواهند شد. در شکل ۳ نمونه‌ای از اعمال پنجره لغزان بر روی یک سری زمانی آورده شده‌است.



شکل ۳. نمونه‌ای از اعمال پنجره لغزان بر روی یک سری زمانی

۳-۱-۸) آموزش مدل پیش‌بینی

در این مرحله به ازای هر خوشه به‌دست‌آمده در مرحله پیش‌پردازش، یک مدل پیش‌بینی سراسری ساخته می‌شود. در مطالعات پیشین، گونه‌های مختلف شبکه عصبی بازگشتی مانند شبکه حافظه کوتاه‌مدت طولانی بطور گسترده مورد استفاده قرار گرفته‌است. برای بررسی اثربخشی روش پیشنهادی، ما مدل‌های پیش‌بینی مبتنی بر یادگیری عمیق را با استفاده از شبکه‌های کانولوشنی زمانی (TCN) پیشنهاد می‌کنیم (بای[۳۳] و همکاران، ۲۰۱۸). TCN یک توسعه‌یافته از CNN است که در مقایسه با RNNها عملکرد بهتری نشان داده‌است. همچنین TCN قادر به پردازش موازی داده‌ها است که منجر به آموزش سریع‌تر مدل می‌شود. معماری شبکه TCN-CNN مورد استفاده در جدول ۲ توصیف شده‌است. همچنین برای ارزیابی بیشتر رویکرد پیشنهادی، یک روش مبتنی بر مکانیزم توجه چندگانه (وسوانی و همکاران[۳۴]، ۲۰۱۷) را پیاده‌سازی می‌کنیم. این مدل شامل لایه‌های ورودی، مکانیزم توجه چندگانه، مسطح‌سازی[۳۵]، کاملا متصل[۳۶] و خروجی است.

[33] Bai
[34] Vaswani
[35] Flatten
[36] Dense

جدول ۲. توپولوژی مدل TCN-CNN

توپولوژی مدل TCN-CNN
Input (shape: (1, Lag (input window))
TCN (filters: 64, kernel size: 3, dilations: (1, 2, 4, 8))
1D Convolutional (filters: 64, kernel size: 4, activation: linear)
1D Convolutional (filters: 16, kernel size: 4, activation: linear)
Flatten
Dense (units: Forecast horizon, activation: linear)
Dense

۳-۲) پس پردازش

بعد از آموزش مدل پیش‌بینی، مدل به‌دست‌آمده روی بخش آموزش هر سری زمانی در مجموعه داده، اعمال می‌شود. مقادیر پیش‌بینی به‌دست‌آمده باید پس‌پردازش شوند تا به مقیاس اولیه داده‌ها برگردند. مراحل پس‌پردازش، معکوس مراحل پیش‌پردازش سری زمانی می‌باشد.

۳-۲-۱) فصلی‌سازی

به جهت اینکه در مرحله پیش‌پردازش، مقادیر مولفه‌های فصلی برای بخش آزمون هر سری زمانی (h نقطه پیش‌بینی) قابل محاسبه نبود، برای به‌دست‌آوردن مولفه‌های فصلی بخش آزمون، یک مدل هموارسازی نمایی را روی داده‌های مولفه فصلی بخش آموزشی هر سری زمانی آموزش می‌دهیم و از آن مدل برای پیش‌بینی مولفه‌های فصلی بخش آزمون استفاده می‌کنیم. در فصلی‌سازی، مولفه‌های فصلی به‌دست‌آمده به مقادیر پیش‌بینی هر سری زمانی افزوده می‌شوند.

۳-۲-۲) تبدیل معکوس لگاریتمی

در این مرحله روی مقادیر پیش‌بینی به‌دست‌آمده، تبدیل معکوس لگاریتمی انجام می‌شود.

۳-۲-۳) غیرنرمال‌سازی

در این مرحله، مقادیر پیش‌بینی به‌دست‌آمده در مقدار میانگین هر سری زمانی ضرب می‌شود و مقادیر پیش‌بینی نهایی به‌دست می‌آید.

۴) یافته‌ها

در این بخش ابتدا مجموعه‌داده مورد استفاده را توصیف می‌کنیم سپس به بیان تنظیمات آزمایشات، ابرپارامترها و تعریف معیارهای خطا پرداخته می‌شود. همچنین گونه‌های مختلف مدل‌های پیش‌بینی و مدل‌های محک را توصیف می‌کنیم. درنهایت عملکرد مدل‌های بکارگرفته‌شده را در مقایسه با مدل‌های معیار ارزیابی می‌کند.

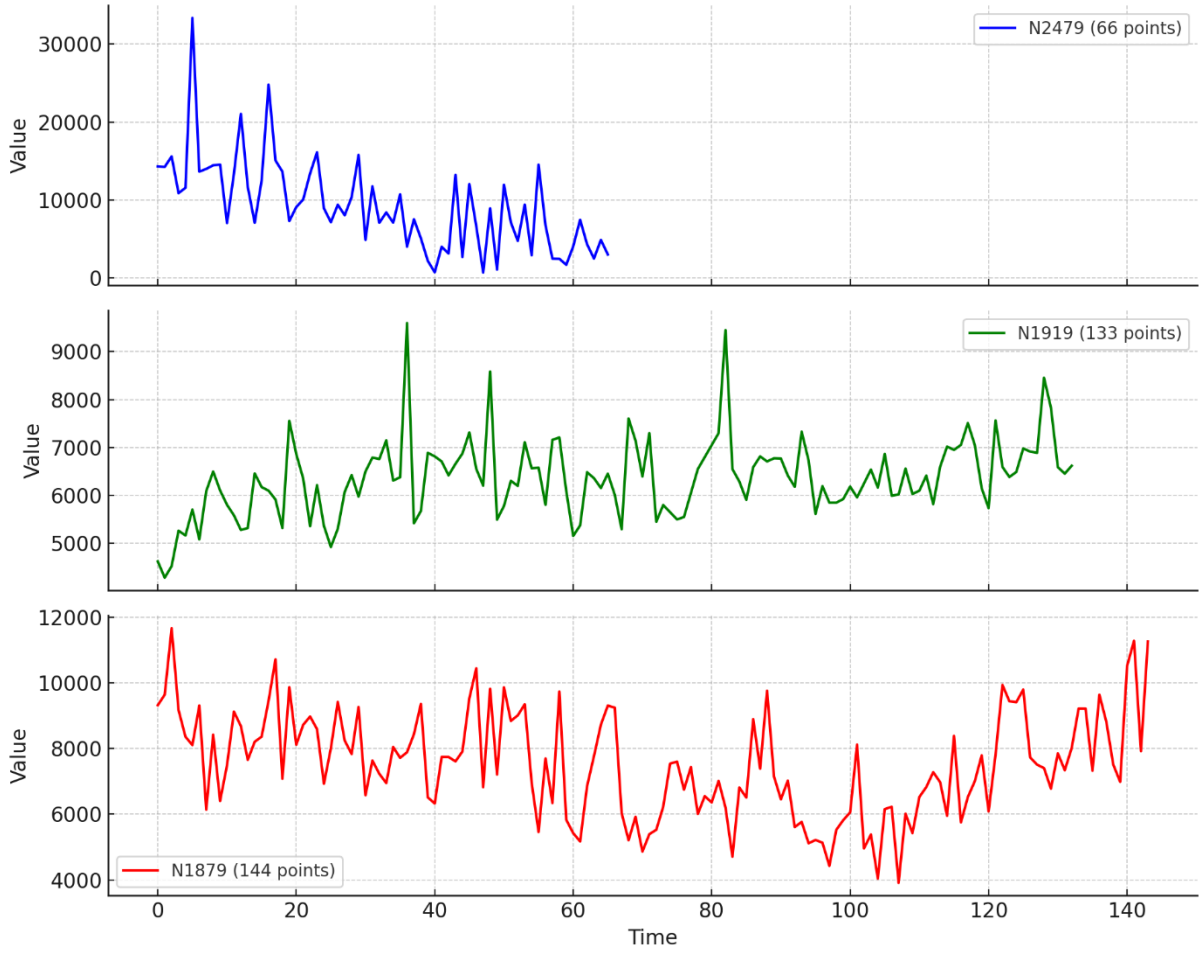
۴-۱) مجموعه‌داده

در این تحقیق، ما از مجموعه‌داده M3 برای ارزیابی روش پیشنهادی استفاده می‌کنیم. این مجموعه‌داده یکی از داده‌های مورداستفاده در مسابقات پیش‌بینی سری زمانی است. سه سری زمانی از این مجموعه‌داده با طول‌های متفاوت در شکل ۴ به تصویر کشیده شده‌است. مجموعه‌داده M3، دارای ۶ مجموعه‌داده است که مشخصات آنها در جدول ۳ نمایش داده

شده‌است. این مجموعه‌داده دارای سری‌های زمانی با طول‌های غیریکسان است چنانکه طول کوتاهترین سری و بلندترین سری در این مجموعه به‌ترتیب برابر با ۶۶ و ۱۴۴ است. در تمام مدل‌های مورداستفاده، ما یک استراتژی پیش‌بینی چندگامی را اجرا کردیم که امکان پیش‌بینی برای h نقاط زمانی را بطور همزمان فراهم می‌کند. جاییکه h با افق پیش‌بینی در هر مجموعه‌داده مطابقت دارد. ویژگی‌های این مجموعه‌داده در جدول ۳ توصیف شده‌است.

جدول ۳. توصیف مجموعه‌داده M3

مجموعه داده	تعداد سری	اُفق پیش‌بینی	فرکانس
M3-demo	111	18	ماهانه
M3-finance	145	18	ماهانه
M3-industry	334	18	ماهانه
M3-macro	312	18	ماهانه
M3-micro	474	18	ماهانه
M3-other	52	18	ماهانه



شکل ٤. مصورسازی سه سری زمانی از مجموعه‌داده M3 با طول‌های متفاوت

۴-۲) تنظیمات آزمایشات و ابرپارامترها:

روش پیشنهادی با زبان پایتون و با استفاده از کتابخانه یادگیری عمیق Keras پیاده‌سازی شده‌است. همچنین برای استخراج ویژگی از سری‌های زمانی، از کتابخانه tsfeature در زبان R استفاده شده‌است. در مرحله پیش‌بینی، مهمترین ابرپارامترها اندازه پنجره ورودی، اندازه دسته، تعداد دور و نرخ یادگیری است. در جدول ۴ لیست محدوده مقادیر برای این ابرپارامترها نشان داده شده‌است. با توجه به غیریکسان‌بودن طول سری‌های زمانی، اندازه پنجره ورودی برای هر مجموعه‌داده طوری انتخاب گردید که امکان استفاده حداکثری از اطلاعات گذشته سری زمانی وجود داشته باشد.

جدول ٤. محدوده مقادیر مورد استفاده ابرپارامترها

مجموعه‌داده	اندازه دسته	تعداد دورها	نرخ یادگیری
M3-Demo	[20-100]	[5-50]	0.01-0.0001
M3-Finance	[20-100]	[5-50]	0.01-0.0001
M3-Industry	[20-100]	[5-10]	0.01-0.0001
M3-Macro	[20-100]	[5-50]	0.01-0.0001
M3-Micro	[40-100]	[5-35]	0.01-0.0001
M3-Other	[20-100]	[5-50]	0.01-0.0001

۴-۳) معیارهای ارزیابی

در این مطالعه از دو معیار اصلی و معروف در ارزیابی مدل‌های پیش‌بینی سری‌های زمانی شامل sMAPE و MASE استفاده می‌کنیم (باندرا، ۲۰۲۳). با درنظرگرفتن اینکه در هرکدام از سری‌های زمانی h نقطه آخر هر سری به‌عنوان بخش آزمون درنظر گرفته می‌شود، این معیارها به‌صورت زیر تعریف می‌شوند.

$$\mathrm{sMAPE} = \frac{200}{h}\sum_{i=1}^{h}\left(\frac{|A_i - F_i|}{|A_i| + |F_i|}\right) \quad (۶)$$

$$\mathrm{MASE} = \frac{1}{h}\frac{\sum_{i=1}^{h}(|A_i - F_i|)}{\frac{1}{n-M}\sum_{i=S+1}^{n}(|A_i - A_{i-S}|)} \quad (۷)$$

در روابط (۶) و (۷) A_i و F_i نشان‌دهنده مقادیر واقعی و پیش‌بینی‌شده در نقطه i هستند. پارامتر h نشان‌دهنده افق پیش‌بینی است. n در رابطه (۷) نشان‌دهنده تعداد نقاط بخش آموزشی هر سری زمانی است. S دوره فصلی هر سری را نشان می‌دهد.

برای یک مجموعه‌داده سری زمانی، هرکدام از این معیارهای خطا در سطح هر سری محاسبه شده و درنهایت میانگین خطاها به‌عنوان معیار نهایی کل مجموعه‌داده محاسبه می‌گردد و در نتایج گزارش ارائه می‌شود.

۴-۴) مدل‌های پیاده‌سازی‌شده:

با توجه به روش پیشنهادی و نوع الگوریتم خوشه‌بندی و همچنین نوع پیش‌پردازش سری‌ها، ۳ مدل پیش‌بینی مبتنی بر روش TCN-CNN و همچنین ۳ مدل مبتنی بر روش مکانیزم توجه چندگانه پیاده‌سازی گردید (جدول ۵). دو الگوریتم K-Medoids و خوشه‌بندی طیفی در مرحله خوشه‌بندی مورد استفاده قرار گرفتند. پارامتر تعداد خوشه‌ها برابر ۲ درنظر گرفته

شد. در تمامی مدل‌های پیش‌بینی از تابع زیان[۳۷] MSE و همچنین بهینه‌ساز Adam[۳۸] (کینگما[۳۹] و همکاران) استفاده گردید. تعداد تکرارهای شبکه[۴۰] با استفاده از روش توقف زودهنگام[۴۱] مشخص می‌شود.

جدول ۵. مدل‌های پیاده‌سازی‌شده

نام مدل	الگوریتم خوشه‌بندی
TC-Baseline	عدم خوشه‌بندی
TC-CL-KM	K-Medoids
TC-CL-CP	خوشه‌بندی طیفی
MHA-Baseline	عدم خوشه‌بندی
MHA-CL-KM	K-Medoids
MHA-CL-CP	خوشه‌بندی طیفی

۴-۵) نتایج و تحلیل

نتایج مدل‌های پیاده‌سازی‌شده بر مبنای معیارهای میانگین sMAPE و MASE به‌ترتیب در جدول ۶ و ۷ آورده شده‌است. مطابق جدول ۷، تحلیل نتایج مدل‌های مبتنی بر TCN-CNN نشان می‌دهد که در معیار MASE مدل TC-CL-KM عملکرد بهتری را نسبت به مدل پایه و مدل TC-CL-CP از خود نشان داده‌است. همچنین بطورکلی میانگین خطای MASE برای مدل‌های مبتنی بر خوشه‌بندی کمتر از مدل پایه است که نشان‌دهنده اثربخشی بکارگیری روش‌های خوشه‌بندی در شناسایی سری‌های مرتبط‌به‌هم و ساخت مدل‌های خاص هر خوشه است. همچنین مطابق جدول ۷، تحلیل نتایج مدل‌های مبتنی بر مکانیزم توجه چندگانه بر مبنای معیار MASE حاکی از آن است که هر دو مدل مبتنی بر خوشه‌بندی، نتایج بهتری نسبت به مدل پایه متناظر خود (MHA-Baseline) کسب کرده‌اند. همچنین مقایسه کلی دو مدل TCN-CNN و مدل توجه چندگانه حاکی از برتری TCN-CNN در تمامی حالت‌ها است.

جدول همانطورکه در جدول ۶ مشخص است، با بکارگیری مدل TCN-CNN به‌عنوان مدل پیش‌بینی، هر دو مدل مبتنی بر خوشه‌بندی در میانگین کلی نتایج بهتری نسبت به مدل پایه (TC-Baseline) در معیار خطای SMAPE دارند. مدل TC-CL-KM که از روش خوشه‌بندی K-Medoids بهره می‌گیرد، در معیار SMAPE در میانگین کلی بهتر از مدل TC-CL-CP عمل می‌کند. این مدل در دو مجموعه‌داده M3-macro و M3-Micro بهترین عملکرد را دارد و می‌توان گفت با توجه به اندازه این مجموعه‌داده، الگوریتم K-Medoids برای مجموعه‌داده‌های با اندازه بزرگتر بهتر از الگوریتم خوشه‌بندی طیفی عمل کرده‌است.

همچنین نتایج جدول ۶ نشان می‌دهد که با بکارگیری مدل توجه چندگانه (MHA) به‌عنوان مدل پیش‌بینی، هر دو مدل مبتنی بر خوشه‌بندی نسبت به مدل پایه متناظر خود (MHA-Baseline) در معیار SMAPE نتایج بهتری از خود نشان داده‌اند. مقایسه نتایج روش‌های مبتنی بر TCN-CNN و روش مکانیزم توجه چندگانه نشان می‌دهد که روش TCN-CNN

[37] Loss function
[38] Adaptive Moment Estimation
[39] Kingma
[40] Epoch
[41] Early stopping

دارای عملکرد بهتری نسبت به روش مکانیزم توجه چندگانه در تمامی مدل‌ها دارد. به‌عبارت‌دیگر، هم در رویکرد بدون خوشه‌بندی و هم در رویکرد با خوشه‌بندی، روش TCN-CNN نتایج بهتری از خود نشان می‌دهد. دلیل این تفاوت می‌تواند ناشی از ماهیت رویکرد پیش‌بینی سراسری باشد که در آن نمونه‌های داده از مجموعه‌ای از سری های زمانی با خصوصیات متفاوت استخراج می‌شوند و این کار مانع یادگیری موثر روش مبتنی بر توجه می‌شود. شایان ذکر است که طبق تحقیقات قبلی، در پیش‌بینی سری‌های زمانی به‌صورت مجزا، مدل‌های پیش‌بینی مبتنی بر مکانیزم توجه، نتایج بهتری از خود نشان داده‌اند.

جدول ٦. نتایج مدل‌های پیاده‌سازی‌شده بر مبنای معیارهای SMAPE

مدل	M3-Demo	M3-Finance	M3-Industry	M3-Macro	M3-Micro	M3-Other	میانگین کل
نتایج مدل TCN-CNN							
TC-Baseline	8.62	14.43	13.03	6.88	21.20	10.04	14.09
TC-CL-KM	9.02	14.09	13.08	6.47	21.02	9.79	13.94
TC-CL-CP	8.58	13.76	13.13	6.67	21.21	9.73	13.99
نتایج مدل MHA							
MHA-Baseline	8.81	16.84	13.51	6.91	21.72	11.55	14.69
MHA-CL-KM	8.99	16.61	13.16	6.83	21.61	10.59	14.51
MHA-CL-SP	8.60	15.90	13.17	6.46	21.67	10.43	14.35

مطابق جدول ۷، تحلیل نتایج مدل‌های مبتنی بر TCN-CNN نشان می‌دهد که در معیار MASE مدل TC-CL-KM عملکرد بهتری را نسبت به مدل پایه و مدل TC-CL-CP از خود نشان داده‌است. همچنین بطورکلی میانگین خطای MASE برای مدل‌های مبتنی بر خوشه‌بندی کمتر از مدل پایه است که نشان‌دهنده اثربخشی بکارگیری روش‌های خوشه‌بندی در شناسایی سری‌های مرتبط‌به‌هم و ساخت مدل‌های خاص هر خوشه است. همچنین مطابق جدول ۷، تحلیل نتایج مدل‌های مبتنی بر مکانیزم توجه چندگانه بر مبنای معیار MASE حاکی از آن است که هر دو مدل مبتنی بر خوشه‌بندی، نتایج بهتری نسبت به مدل پایه متناظر خود (MHA-Baseline) کسب کرده‌اند. همچنین مقایسه کلی دو مدل TCN-CNN و مدل توجه چندگانه حاکی از برتری TCN-CNN در تمامی حالت‌ها است.

جدول ۷. نتایج مدل‌های پیاده‌سازی‌شده بر مبنای معیار MASE

مدل	M3-Demo	M3-Finance	M3-Industry	M3-Macro	M3-Micro	M3-Other	میانگین کل
نتایج مدل TCN-CNN							
TC-Baseline	0.916	1.238	1.034	1.003	0.715	0.619	0.918
TC-CL-KM	0.896	1.210	1.037	0.953	0.690	0.591	0.894
TC-CL-CP	0.833	1.162	1.051	0.988	0.698	0.594	0.898
نتایج مدل MHA							
MHA-Baseline	1.434	1.484	1.142	1.075	0.705	0.723	1.21
MHA-CL-KM	1.253	1.454	1.067	1.065	0.710	0.694	0.99
MHA-CL-CP	1.042	1.393	1.095	0.981	0.712	0.676	0.95

۴-۶) مقایسه با مدل‌های محک

در بخش قبلی به این نتیجه رسیدیم که مدل‌های مبتنی بر خوشه‌بندی نسبت به مدل‌های بدون خوشه‌بندی، عملکرد بهتری با هر دو مدل پیش‌بینی از خود نشان می‌دهند. در این بخش به‌منظور ارزیابی کامل روش پیشنهادی به مقایسه برترین مدل‌های به‌دست‌آمده در این تحقیق، یعنی مدل‌های TCN-CNN با مدل‌های محک می‌پردازیم. مدل‌های محک مورد استفاده شامل: مدل‌های گروهی FFORMA (مونترو و همکاران، ۲۰۲۰)، مدل CatBoost (پروخورنکوا[۴۲] و همکاران ۲۰۱۸) هستند. همچنین مدل‌های مبتنی بر یادگیری عمیق شامل: مدل DeepAR (سالیناس و همکاران، ۲۰۲۰) و مدل N-BEATS (اورشکین[۴۳] و همکاران، ۲۰۱۹) هستند. به جهت اینکه در مطالعات گذشته، یک خطای کلی برای مجموعه‌داده M3 ارائه شده‌است، برای مقایسه صحیح مدل‌های ارائه‌شده، در این بخش میانگین خطاها برای کل سری‌های موجود در مجموعه‌داده M3 که شامل ۱۴۲۶ مورد می‌باشد، در جدول ۸ و شکل‌های ۵ و ۶ گزارش می‌شود. در معیار sMAPE مدل‌های ارائه‌شده در این تحقیق نتایج بهتری نسبت به مدل‌های تحقیقات پیشین دارند. از منظر معیار MASE هر دو مدل مبتنی بر خوشه‌بندی شامل مدل TC-CL-KM و همچنین مدل TC-CL-CP نتایج بهتری نسبت مدل‌های FFORMA، CatBoost، Deep AR و N-BEATS دارند.

جدول ۸. مقایسه بهترین مدل به‌دست‌آمده با مدل‌های موجود استخراج‌شده از (گوداهیوا و همکاران، ۲۰۲۱)

مدل	معیار sMAPE	معیار MASE
TC-CL-KM	13.94	0.894
TC-CL-CP	13.99	0.898
FFORMA	14.51	0.914
CatBoost	16.41	1.065
DeepAR	15.74	1.167
N-BEATS	14.76	0.934

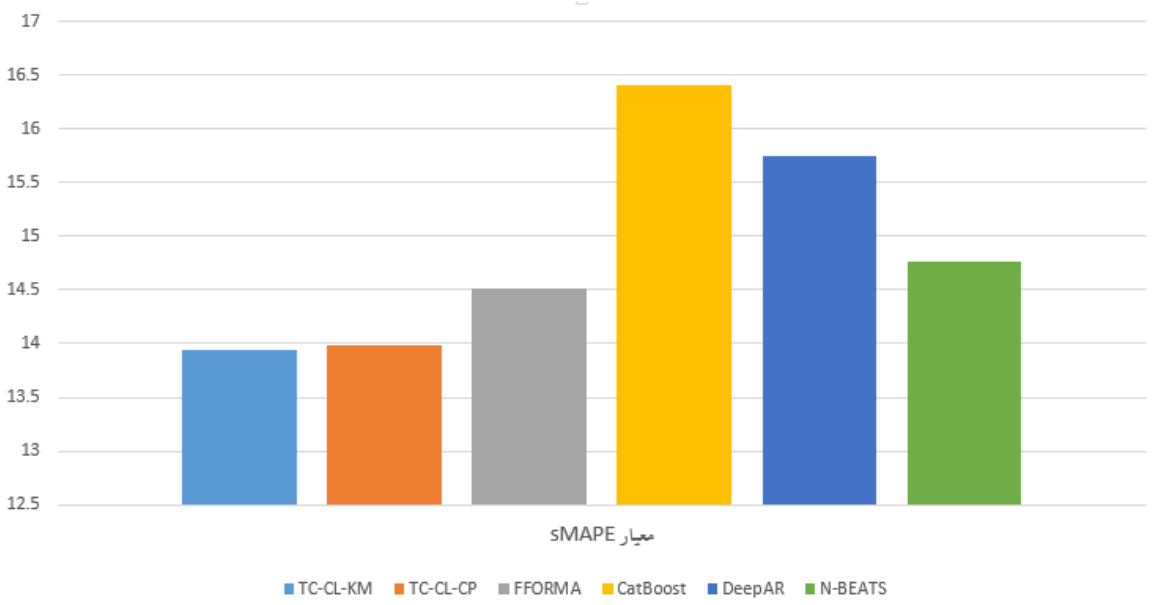
۴-۷) پیچیدگی زمانی روش پیشنهادی

با توجه به اینکه در رویکرد پیشنهادی، سری‌های زمانی به خوشه‌های مختلف افراز شده و به ازای هر خوشه یک مدل سراسری خاص آموزش داده می‌شود، امکان موازی‌سازی آموزش مدل‌ها به‌وجود می‌آید. در جدول ۹ زمان اجرای هر مرحله از رویکرد پیشنهادی برای مدل پایه TC-Baseline و همچنین مدل مبتنی بر خوشه‌بندی TC-CL-CP بیان شده‌است. محیط محاسباتی مورد استفاده برای اجرای آزمایشات شامل منابع CPU و GPU است که شامل پردازنده Intel Xeon با ۲ هسته مجازی (vCPUs) و ۱۳ گیگابایت RAM و همچنین کارت گرافیک NVIDIA Tesla P100 با ۱۶ گیگابایت حافظه می‌باشد. همانطورکه در این جدول مشخص است، زمان کل روش مبتنی بر خوشه‌بندی به‌دلیل موازی‌سازی کمتر از مدل پایه است که در آن یک مدل پیش‌بینی سراسری به ازای تمامی سری‌های زمانی در یک مجموعه آموزش داده می‌شود. بنابراین با درنظرگرفتن نتایج جداول ۶ و ۷ و همچنین تحلیل‌های محاسباتی، می‌توان به این نتیجه رسید که روش پیشنهادی موجب افزایش دقت و کاهش زمان اجرا می‌شود.

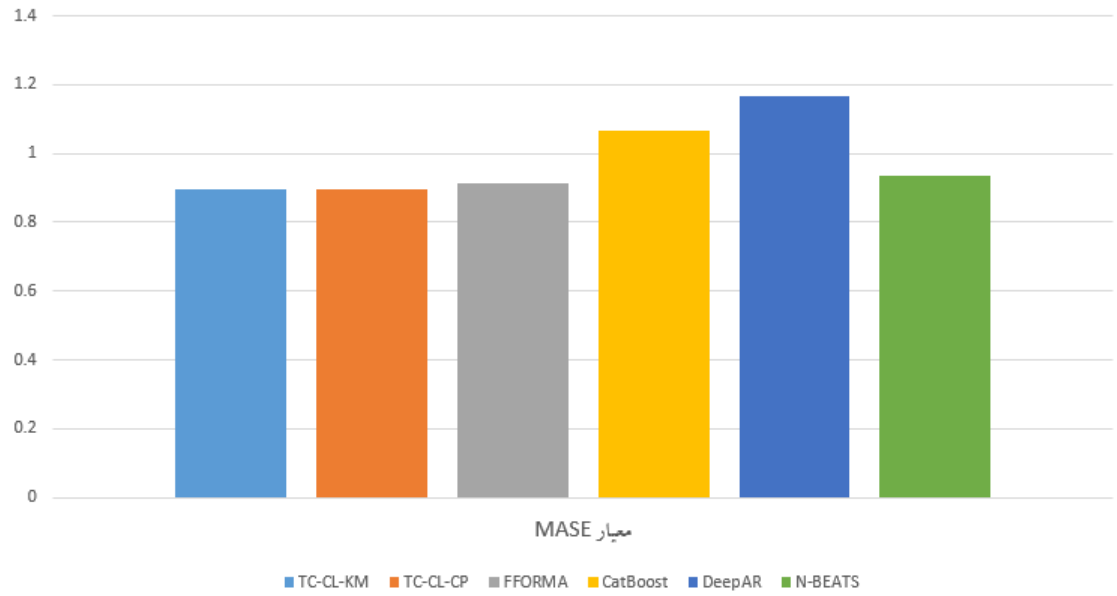
[42] Prokhorenkova
[43] Oreshkin

جدول ۹ . زمان محاسباتی مدل مبتنی بر خوشه‌بندی و مدل پایه

مراحل رویکرد پیشنهادی	مدل	زمان (ثانیه)
استخراج ویژگی	TC-Baseline	-
	TC-CL-CP	27.15
خوشه‌بندی	TC-Baseline	-
	TC-CL-CP	0.2
پیش‌پردازش (مرحله پیش‌بینی)	TC-Baseline	59.9
	TC-CL-CP	25.44
آموزش مدل پیش‌بینی	TC-Baseline	325.78
	TC-CL-CP	122.76
زمان کل	TC-Baseline	385.68
	TC-CL-CP	175.55



شکل ۵. مقایسه با مدل‌های محک از نظر معیار sMAPE



شکل ٦. مقایسه با مدل‌های محک از نظر معیار MASE

۵) نتیجه‌گیری و پیشنهادها

به‌دلیل تولید گسترده داده‌ها در قالب سری زمانی، مدل‌های پیش‌بینی سراسری نسبت به روش‌های پیش‌بینی سری‌های زمانی تک‌متغیره مورد توجه واقع شده‌اند. در این مطالعه، یک رویکرد مبتنی بر خوشه‌بندی، برای بهبود دقت مدل‌های سراسری به‌منظور مقابله با ناهمگونی داده‌ها در مجموعه‌های سری زمانی پیشنهاد داده شد. نتایج آزمایشات انجام‌شده بر روی مجموعه M3 نشان داد که در اکثر موارد، مدل‌های ایجادشده با استفاده از روش پیشنهادی، عملکرد بهتری نسبت به مدل‌های پایه و همچنین مدل‌های محک از خود نشان داده‌اند. این امر قابلیت روش پیشنهادی را در شناسایی خوشه‌های همگن سری‌های زمانی و ساخت مدل خاص هر خوشه را نشان می‌دهد. علاوه بر این، تحلیل نتایج نشان داد که مدل‌های به‌دست‌آمده از رویکرد پیشنهادی و به‌ویژه مدل مبتنی بر خوشه‌بندی k-Medoids، میانگین sMAPE و میانگین MASE کمتری نسبت به مدل پایه که در آن خوشه‌بندی صورت نمی‌گیرد، دارند. همچنین مقایسه نتایج نشان داد مدل‌های حاصل از روش پیشنهادی، عملکرد بهتری نسبت به مدل‌های پیشرفته مانند DeepAR، N-BEATS، FFORMA و CatBoost دارند. مقایسه بیشتر با رویکردهای مبتنی بر خوشه‌بندی موجود در ادبیات موضوع نیز نشان داد مدل‌های پیشنهادی این تحقیق عملکرد بهتری از نظر میانگین sMAPE و میانگین MASE داشته‌اند.

آزمایشات ما نشان داد که خوشه‌بندی k-Medoids عملکرد بهتری نسبت به خوشه‌بندی طیفی داشت. عملکرد رویکرد پیشنهادی به استخراج ویژگی‌های معنی‌دار، خوشه‌بندی باکیفیت و همچنین مدل پیش‌بینی دقیق و مستحکم وابسته است. در تحقیقات آینده، قصد داریم روش پیشنهادی را با استفاده از شبکه‌های خودرمزنگار و روش‌های خوشه‌بندی عمیق توسعه دهیم.

منابع

Bai, S., Kolter, J. Z., & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint arXiv:1803.01271. https://doi.org/10.48550/arXiv.1803.01271

Bandara, K., Bergmeir, C., & Smyl, S. (2020). Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach. Expert systems with applications, 140, 112896. https://doi.org/10.1016/j.eswa.2019.112896

Bandara, K., Hewamalage, H., Liu, Y. H., Kang, Y., & Bergmeir, C. (2021). Improving the accuracy of global forecasting models using time series data augmentation. Pattern Recognition, 120, 108148. https://doi.org/10.1016/j.patcog.2021.108148

Bandara, K. (2023). Forecasting with Big Data Using Global Forecasting Models. In Forecasting with Artificial Intelligence: Theory and Applications (pp. 107-122). M. Hamoudia, S. Makridakis, and E. Spiliotis, Editors. 2023, Cham: Springer Nature Switzerland: https://doi.org/10.1007/978-3-031-35879-1

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: forecasting and control. John Wiley & Sons. https://doi.org/10.1111/jtsa.12194

Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A seasonal-trend decomposition. J. off. Stat, 6(1), 3-73.

Godahewa, R., Bandara, K., Webb, G. I., Smyl, S., & Bergmeir, C. (2021). Ensembles of localised models for time series forecasting. Knowledge-Based Systems, 233, 107518. https://doi.org/10.1016/j.knosys.2021.107518

Hewamalage, H., Bergmeir, C., & Bandara, K. (2021). Recurrent neural networks for time series forecasting: Current status and future directions. International Journal of Forecasting, 37(1), 388-427. https://doi.org/10.1016/j.ijforecast.2020.06.008

Hewamalage, H., Bergmeir, C., & Bandara, K. (2022). Global models for time series forecasting: A simulation study. Pattern Recognition, 124, 108441. https://doi.org/10.1016/j.patcog.2021.108441

Hyndman, R. J., Koehler, A. B., Snyder, R. D., & Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. International Journal of forecasting, 18(3), 439-454. https://doi.org/10.1016/S0169-2070(01)00110-8

Hyndman, R., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). Forecasting with exponential smoothing: the state space approach. Springer Science & Business Media. https://doi.org/10.1007/978-3-540-71918-2

Hyndman, R., Kang, Y., Montero-Manso, P., Talagala, T., Wang, E., Yang, Y., & O'Hara-Wild, M. (2019). tsfeatures: Time series feature extraction. R package version, 1(0).

Januschowski, T., Gasthaus, J., Wang, Y., Salinas, D., Flunkert, V., Bohlke-Schneider, M., & Callot, L. (2020). Criteria for classifying forecasting methods. International Journal of Forecasting, 36(1), 167-177. https://doi.org/10.1016/j.ijforecast.2019.05.008

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Laptev, N., Yosinski, J., Li, L. E., & Smyl, S. (2017, August). Time-series extreme event forecasting with neural networks at uber. In International conference on machine learning (Vol. 34, pp. 1-5). sn.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). The M4 Competition: Results, findings, conclusion and way forward. International Journal of forecasting, 34(4), 802-808. https://doi.org/10.1016/j.ijforecast.2018.06.001

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). M5 accuracy competition: Results, findings, and conclusions. International Journal of Forecasting, 38(4), 1346-1364. https://doi.org/10.1016/j.ijforecast.2021.11.013

Martínez, F., Frías, M. P., Pérez-Godoy, M. D., & Rivera, A. J. (2018). Dealing with seasonality by narrowing the training set in time series forecasting with kNN. Expert systems with applications, 103, 38-48. https://doi.org/10.1016/j.eswa.2018.03.005

Montero-Manso, P., Athanasopoulos, G., Hyndman, R. J., & Talagala, T. S. (2020). FFORMA: Feature-based forecast model averaging. International Journal of Forecasting, 36(1), 86-92. https://doi.org/10.1016/j.ijforecast.2019.02.011

Montero-Manso, P., & Hyndman, R. J. (2021). Principles and algorithms for forecasting groups of time series: Locality and globality. International Journal of Forecasting, 37(4), 1632-1653. https://doi.org/10.1016/j.ijforecast.2021.03.004

Oreshkin, B. N., Carpov, D., Chapados, N., & Bengio, Y. (2019). N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. arXiv preprint arXiv:1905.10437. https://doi.org/10.48550/arXiv.1905.10437

Parmezan, A. R. S., Souza, V. M., & Batista, G. E. (2019). Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model. Information sciences, 484, 302-337. https://doi.org/10.1016/j.ins.2019.01.076

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. Advances in neural information processing systems, 31. https://doi.org/10.48550/arXiv.1706.09516

Salinas, D., Flunkert, V., Gasthaus, J., & Januschowski, T. (2020). DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International journal of forecasting*, *36*(3), 1181-1191. https://doi.org/10.1016/j.ijforecast.2019.07.001

Schubert, E., & Rousseeuw, P. J. (2021). Fast and eager k-medoids clustering: O (k) runtime improvement of the PAM, CLARA, and CLARANS algorithms. Information Systems, 101, 101804. https://doi.org/10.1016/j.is.2021.101804

Smyl, S., & Kuber, K. (2016, June). Data preprocessing and augmentation for multiple short time series forecasting with recurrent neural networks. In 36th international symposium on forecasting.

Smyl, S. (2020). A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. International journal of forecasting, 36(1), 75-85. https://doi.org/10.1016/j.ijforecast.2019.03.017

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser and I. Polosukhin (2017). Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, California, USA, Curran Associates Inc.: 6000–6010.

Ye, X., & Sakurai, T. (2016). Robust similarity measure for spectral clustering based on shared neighbors. ETRI journal, 38(3), 540-550. https://doi.org/10.4218/etrij.16.0115.0517