



Autoregressive Modeling of Fuzzy Data Based on the Support Vector Machine

Ranjbar, E.¹ , Akbari, M. G.²  Zarei, R.³ 

¹Department of Statistics, University of Birjand, Birjand, Iran.

²Department of Statistics, University of Guilan, Rasht, Iran.

Corresponding author: M. G. Akbari, g_z_akbari@birjand.ac.ir

Received: ۲۸/۹/۲۰۲۴ **Revised:** ۲۳/۱/۲۰۲۵ **Accepted and Published Online:** ۲۵/۱/۲۰۲۵.

Introduction

Time series analysis is widely used for vast applied statistics such as sales, economics, astronomy, and weather forecasting. The classical time series models rely on exact data. In real-world applications, data frequently contains inherent uncertainties due to measurement errors, human factors, or inherent variability. This imprecision can significantly impact the accuracy and reliability of traditional time series models. To address this challenge, we propose a novel fuzzy autoregressive time series model that incorporates the robust and flexible framework of Support Vector Machines (SVM).

Material and Methods

The proposed model leverages the SVM approach to handle the fuzzy nature of the data effectively. Specifically, a kernel function is employed within the SVM framework to enhance model stability and flexibility by mapping the data into a higher-dimensional space. Constraints are integrated into the model to control the influence of data points and improve the robustness of the predictions. The performance of the proposed fuzzy autoregressive model is evaluated using a combination of goodness-of-fit criteria, including measures such as mean absolute error, root mean squared error, and mean absolute percentage error.

Results and Discussion

This study introduces a novel fuzzy autoregressive time series model based

on the SVM approach to address the inherent uncertainty often encountered in time series data. With its kernel-based approach, the SVM framework gives the model significant flexibility in capturing complex, non-linear relationships within the time series. This flexibility is crucial in accurately representing the dynamic behavior of real-world systems, which often exhibit non-linear trends and unpredictable fluctuations. The incorporation of constraints within the SVM framework allows for effective control over the influence of individual data points, mitigating the impact of potential outliers or noisy observations. This is particularly beneficial in time series analysis, where outliers can significantly degrade the performance of traditional forecasting models. The empirical results, based on both simulated and real-world datasets, demonstrate the superior performance of the proposed model compared to existing methods. This result suggests that integrating fuzzy logic and SVM techniques offers a promising approach to improving the accuracy and robustness of time series forecasting.

Conclusion

This research introduces a novel fuzzy autoregressive time series model based on the SVM approach, effectively addressing the challenges posed by imprecise observations in time series analysis. Integrating kernel functions and constraints enhances the model's flexibility, stability, and robustness. The promising results from simulated and real-world data suggest that the proposed model offers a valuable tool for accurately modeling and forecasting time series with inherent uncertainties. However, the proposed method can only be applied to triangular fuzzy numbers. Thus, further investigations are necessary to extend the proposed method to any LR fuzzy numbers. Future studies should be focused on detecting outliers and extending the proposed method to a robust fuzzy time series model based on in-sample and out-sample forecasting criteria.

Keywords: Autoregressive model, Fuzzy number, Kernel function, Lagrange function, Support vector machine.

Mathematics Subject Classification (2010): 62J12, 62H30.



©The Author(s). The Publisher is Iranian Statistical Society.

This is an open access article distributed under the terms and conditions of [CC BY-NC 4.0](https://creativecommons.org/licenses/by-nc/4.0/)

مدل سازی اتو رگرسیو داده های نادقیق بر اساس ماشین بردار پشتیبان

الهام رنجبر^۱، محمدقاسم اکبری^۱، رضا زارعی^۲

^۱گروه آمار، دانشکده علوم ریاضی و آمار، دانشگاه بیرجند

^۲گروه آمار، دانشکده علوم ریاضی، دانشگاه گیلان

نویسنده مسئول: محمد قاسم اکبری، g_z_akbari@birjand.ac.ir

تاریخ دریافت: ۱۴۰۳/۷/۷ تاریخ بازنگری: ۱۴۰۳/۱۱/۴ تاریخ پذیرش و انتشار: ۱۴۰۳/۱۱/۶

چکیده: در تحلیل سری های زمانی ممکن است با وضعیت هایی روبرو شده باشیم که در آن برخی از ارکان مدل، کمیت های نادقیق باشند. یکی از متداول ترین این وضعیت ها، نادقیق بودن مشاهدات تحت بررسی است که معمولا در اثر خطای اندازه گیری یا اشتباهات انسانی رخ می دهد. در این مقاله، یک مدل جدید سری زمانی اتو رگرسیو فازی مبتنی بر رویکرد ماشین بردار پشتیبان پیشنهاد می شود. برای این منظور، از تابع هسته برای استواری و انعطاف مدل و از قیود لحاظ شده در مدل برای کنترل نقاط استفاده شده است. به منظور بررسی عملکرد و اثر بخشی مدل سری زمانی اتو رگرسیو فازی پیشنهادی، برخی معیارهای نیکویی برازش استفاده می شوند. نتایج به دست آمده بر اساس یک مثال از داده های سری زمانی فازی شبیه سازی شده و دو مثال واقعی، نشان داد که روش پیشنهادی در مقایسه با سایر روش های موجود دارای عملکرد بهتری بوده است.

واژه های کلیدی: تابع هسته، تابع لاگرانژ، عدد فازی، ماشین بردار پشتیبان، مدل اتو رگرسیو. کد موضوع بندی ریاضی (۲۰۱۰): 62J12، 62H30.

۱ مقدمه

سری های زمانی، مشاهداتی هستند که در طول زمان، گردآوری می شوند. به عنوان مثال متوسط درجه حرارت، رطوبت، سرعت باد، قیمت یک سهام خاص، تعداد گردشگران، طول عمر و میزان غلظت هوا نمونه هایی از سری های زمانی

هستند، هدف اصلی در تحلیل سری زمانی یک پدیده، ایجاد یک مدل آماری برای داده های وابسته به زمان بر اساس اطلاعات گذشته آن است. با این کار امکان پیش بینی در آینده ایجاد می شود. در بسیاری از اتفاقات واقعی زندگی با مجموعه مشاهدات نادقیق نظیر حدود رانندگی در طول روز، مبلغ خسارت بیمه نامه شخص ثالث در طول ماه، قیمت زمین در طول ماه روبرو خواهیم شد که ممکن است یک یا چند فرض از فرض هایی که در مورد مدل سازی داده های فازی مبتنی بر زمان در نظر گرفته می شود برقرار نباشد. در چنین شرایطی استفاده از مدل های کلاسیک منجر به ارائه نتایجی نادرست برای تحلیل سیستم های تحت بررسی خواهد شد. پس از معرفی نظریه مجموعه های فازی زاده (۱۹۵۶) برای صورت بندی مفاهیم نادقیق، این نظریه نقشی کلیدی در بیشتر زمینه های علمی از جمله احتمال و آمار پیدا کرده است.

مطالعات متنوعی در زمینه مدل سازی داده های سری زمانی با رویکرد فازی توسط محققین مختلفی انجام شده است که در ادامه برخی از مهم ترین این مطالعات مرور می شوند. سانگ و چیسام (۱۹۹۳) مدل سری زمانی داده های دقیق برای پیش بینی پذیرش دانشگاه آلاباما بر اساس مجموعه های فازی ارائه داده اند. سانگ و همکاران (۱۹۹۵) از طریق تعریف یک عملگر جمع و تفاضل فازی، مدل سری زمانی ارائه کرده اند که برای مدل بندی سری زمانی فازی ایستا کاربرد دارد. چن (۱۹۹۶) با ارائه مجموعه عملگرهای جبری اصلاح شده، باعث بهبود و افزایش دقت محاسبات در مدل سانگ و چیسام (۱۹۹۳) گردید. هونگ (۲۰۰۵) نتایج مربوط به مدل سری زمانی فازی را به سری زمانی نا ایستا و سری زمانی در حالت کلی تعمیم داد. روان جونهو و همکاران (۲۰۱۳) یک روش ترکیبی مبتنی بر خوشه بندی، فازی سازی داده ها و رویکرد ماشین بردار پشتیبان را برای تحلیل داده های سری زمانی فازی ارائه دادند. گویی بن و همکاران (۲۰۱۵) و درودیان و نیاکی (۲۰۲۱) با استفاده از روش بردار پشتیبان به مدل سازی و پیش بینی داده های وابسته به زمان در حالت کلاسیک و فازی پرداختند. حسامیان و اکبری (۲۰۱۸) یک رویکرد نیمه پارامتری مبتنی بر تابع هموارساز با روش هسته نادریا-واتسون را برای مدل سری زمانی فازی پیشنهاد دادند. زارعی و همکاران (۲۰۲۰) مدل اتو رگرسیو مرتبه اول و مرتبه دوم نیمه پارامتری بر اساس داده های سری زمانی فازی را بررسی و با استفاده از یک تابع هسته این مدل را با مدل های دیگر مقایسه نموده اند. اسدالهی و همکاران (۲۰۲۱) در مدل ارائه شده خود اساساً از تابع هسته به صورت مستقیم در معادله هدف استفاده کرده اند و هیچ گونه شرطی برای آن در نظر نگرفته اند. حسامیان و همکاران (۲۰۲۲) یک روش جدید ناپارامتری جمعی را برای مدل بندی داده های سری زمانی فازی بر اساس توابع هموارساز فازی ارائه نمودند. محمدی و همکاران (۱۴۰۳) با استفاده از مفهوم تابع تکیه گاه و روابط همبستگی بین داده ها به مدل بندی و پیش بینی الگوی اتورگرسیو برای داده های فازی پرداخته اند. طاهری و همکاران (۱۴۰۳) بر اساس متغیر فازی تعمیم یافته و ارائه تعریفی جدید برای توابع اتو کواریانس و خود همبستگی بین متغیرهای تصادفی فازی، یک مدل میانگین متحرک مرتبه q را برای تحلیل داده های سری زمانی فازی پیشنهاد دادند.

در این مقاله، یک روشی مبتنی بر به کارگیری رویکرد ماشین بردار پشتیبان برای مراکز، پهنای چپ و پهنای راست، به منظور مدل سازی داده های سری زمانی اتورگرسیو فازی پیشنهاد شده است. یکی از ویژگی های روش پیشنهادی در نظر گرفتن روند سری زمانی بوده که با توجه به تابع هسته (هموارساز) در معادلات گنجانده شده است

که سبب شده تا مانا بودن یا مانا نبودن داده‌های تحت بررسی خللی در مدل‌سازی ایجاد نکند. علاوه بر این، در روش پیشنهادی مشاهدات ذاتا فازی بوده و با استفاده از شرایطی که در معادله هدف آمده، مدل‌سازی صورت گرفته است. ادامه مقاله به این‌صورت سازماندهی شده است: مفاهیم مقدماتی مجموعه‌های فازی و رویکرد ماشین بردار پشتیبان در بخش دوم ارائه شده است. در بخش سوم، یک روش جدید بر پایه ماشین بردار پشتیبان برای مدل‌بندی و تحلیل داده‌های سری‌زمانی فازی بر پایه مدل اتورگرسیو ارائه شده است. به‌منظور تشریح جزئیات محاسباتی مدل پیشنهادی و بررسی عملکرد آن با سایر روش‌های موجود، یک مثال شبیه‌سازی شده و دو مثال با داده‌های واقعی در بخش چهارم مورد تحلیل قرار گرفته است. جمع‌بندی و نتیجه‌گیری در بخش پنجم ارائه شده است.

۲ مفاهیم مقدماتی

مجموعه فازی $\tilde{A}$ از مجموعه مرجع $\mathbb{X}$ به‌صورت $\{(x, \mu_{\tilde{A}}(x)) : x \in \mathbb{X}\}$ نمایش داده می‌شود که در آن $\mu_{\tilde{A}}(x) : \mathbb{X} \rightarrow [0, 1]$ تابع عضویت $\tilde{A}$ است. برای هر $\alpha \in (0, 1]$ ، زیرمجموعه α $\{x \in \mathbb{X} \mid \mu_{\tilde{A}}(x) \geq \alpha\}$ -برش $\tilde{A}$ نامیده شده و به‌صورت $[A_{[\alpha]}^L, A_{[\alpha]}^U]$ نشان داده می‌شود که در آن $A_{[\alpha]}^L = \inf \{x : x \in \tilde{A}_{[\alpha]}\}$ و $A_{[\alpha]}^U = \sup \{x : x \in \tilde{A}_{[\alpha]}\}$ هستند.

تعریف ۱. مجموعه فازی $\tilde{A}$ از $\mathbb{R}$ را یک عدد فازی گویند، هرگاه $\tilde{A}$ تک‌نمائی باشد و به‌ازای هر $\alpha \in (0, 1]$ ، α -برش‌های $\tilde{A}$ ، بازه‌های بسته و کراندار باشند.

تعریف ۲. فرض کنید $\tilde{A}$ یک عدد فازی از مجموعه مرجع $\mathbb{X}$ باشد. در این صورت $\tilde{A}$ را یک عدد فازی LR گویند، هرگاه تابع عضویت آن به‌صورت

$$\mu_{\tilde{A}}(x) = \begin{cases} L(\frac{m-x}{\ell}) & x \leq m, \\ R(\frac{x-m}{r}) & x > m, \end{cases}$$

تعریف شده باشد، که در آن L و R توابعی غیر صعودی از $\mathbb{R}^+$ به $[0, 1]$ هستند و $L(0) = R(0) = 1$. یک عدد فازی LR با نماد $(m; \ell, r)_{LR}$ نمایش داده می‌شود. عدد m میانه و اعداد مثبت ℓ و r به‌ترتیب پهنای چپ و پهنای راست $\tilde{A}$ می‌نامند. علاوه بر این، عدد فازی $\tilde{A}$ را یک عدد فازی مثلثی نامیده و با نماد $(m; \ell, r)_T$ نشان داده می‌شود است، هرگاه $R(x) = L(x) = \max\{0, 1 - x\}$.

تعریف ۳. برای دو عدد فازی LR ، $\tilde{A} = (m; \ell, r)$ و $\tilde{B} = (n; \ell', r')$ و $\lambda \in \mathbb{R}$ عملگرهای جبری بین اعداد فازی عبارتند از

$$\tilde{A} \oplus \tilde{B} = (m; \ell, r)_{LR} \oplus (n; \ell', r')_{LR} = (m + n; \ell + \ell', r + r')_{LR},$$

$$\begin{aligned}\lambda + \tilde{A} &= \lambda + (m; \ell, r)_{LR} = (\lambda + m; \ell, r)_{LR}, \\ \lambda \otimes \tilde{A} &= \begin{cases} (\lambda m; \lambda \ell, \lambda r)_{LR} & \lambda \geq 0, \\ (\lambda m; -\lambda r, -\lambda \ell)_{RL} & \lambda < 0. \end{cases}\end{aligned}\quad (1)$$

۲.۱ ماشین بردار پشتیبان

ماشین بردار پشتیبان^۱ SVM یکی از روش‌های یادگیری نظارت‌شده است که از آن برای طبقه‌بندی و رگرسیون استفاده می‌شود. اولین بار **وونیک** (۱۹۹۵) ماشین بردار پشتیبان را برای مدل‌بندی مسایل رگرسیون تعمیم داد. مبنای کاری دسته‌بندی بردار پشتیبان، دسته‌بندی خطی داده‌ها است. برای این منظور سعی می‌شود خطی انتخاب شود که حاشیه اطمینان بیشتری داشته باشد. پیدا کردن خط بهینه برای داده‌ها به وسیله روش‌های حل مسأله بهینه‌سازی درجه دو که روش‌های شناخته شده‌ای در حل مسائل محدودیت‌دار هستند، صورت می‌گیرد. قبل از دسته‌بندی خطی، برای اینکه ماشین بتواند داده‌های با پیچیدگی بالا را دسته‌بندی کند، داده‌ها به وسیله تابع ϕ به فضای با ابعاد خیلی بالاتر برده می‌شود. یک روش حل مسأله بهینه‌سازی در SVM استفاده از تابع لاگرانژ و حل مسأله بهینه‌سازی دوگان لاگرانژ است (سایکنز و همکاران، ۲۰۰۲؛ ایزنمن، ۲۰۰۸).

تعریف ۴. تابع حقیقی مقدار، نامنفی و انتگرال‌پذیر $K(\cdot) : \mathbb{R} \rightarrow \mathbb{R}^+$ را یک تابع هسته^۲ گویند اگر دارای خواص $K(u) \geq 0$ و $K(u) = K(-u)$ به‌ازای هر u و $\int K(u)du = 1$ ، $\int uK(u)du = 0$ و $\int u^2 K(u)du = \sigma_K^2 < \infty$ باشد.

در این مقاله از هسته نرمال، $K(u) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}u^2)$ ، $u \in \mathbb{R}$ استفاده می‌شود، که دامنه توابع هسته به غیر از هسته نرمال، همسایگی به مرکز صفر و شعاع یک بوده و مقادیر $K(u)$ وزن‌های این همسایگی هستند. برای هسته نرمال این همسایگی به‌عنوان پراکندگی در نظر گرفته می‌شود.

تعریف ۵. (اسدالهی و همکاران، ۲۰۲۱) فرض کنید $\phi : Z \rightarrow F$ نگاشت غیرخطی از فضای $Z \subset \mathbb{R}^d$ به فضای چند بعدی F باشد. برای هر $\mathbf{x}$ و $\mathbf{z}$ عضو فضای ورودی Z ، تابع $\Phi : \mathbb{R}^{2d} \rightarrow \mathbb{R}^+$ به‌صورت $\Phi(\mathbf{x}, \mathbf{z}) = \phi(\mathbf{x})' \cdot \phi(\mathbf{z})$ تعریف می‌شود. قابل ذکر است که در فضای برداری، تابع هسته تعریف شده را می‌توان به‌صورت $\Phi(\mathbf{x}, \mathbf{z}) = K(\frac{\|\mathbf{x} - \mathbf{z}\|}{h})$ ، $h > 0$ نوشت.

تعریف ۶. فرض کنید $f : \mathbb{R}^n \rightarrow \mathbb{R}$. اگر مسأله بهینه‌سازی $\min f(\mathbf{x})$ مسأله اولیه نامیده شود، به‌طوری‌که $\mathbf{x} \in \mathbb{R}^n$ و $g_j(\mathbf{x}) \geq 0$ $j = 1, \dots, m$ و $h_i(\mathbf{x}) = 0$ $i = 1, \dots, k$ در این صورت مسأله دوگان لاگرانژ متناظر با آن برای تمام $\boldsymbol{\mu} = (\mu_1, \dots, \mu_m)' \in \mathbb{R}^m$ و $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_k)' \in \mathbb{R}^k$ که در آن $D(\boldsymbol{\mu}, \boldsymbol{\lambda}) = \max D(\boldsymbol{\mu}, \boldsymbol{\lambda})$ به‌طوری‌که $\boldsymbol{\mu} \geq 0$ بیان می‌شود (بازارا و همکاران، ۲۰۱۰)، که در آن

¹Support Vector Machines

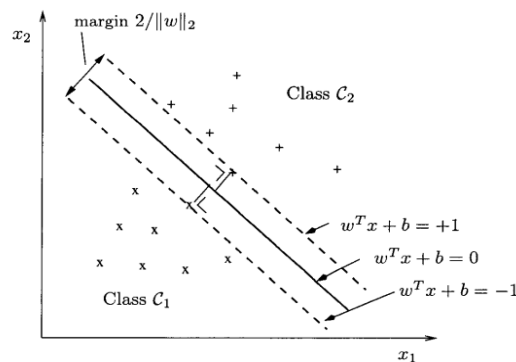
²Kernel Function

رنجبر، ا. و همکاران ۱۰۳

$$L(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \inf \{ f(\mathbf{x}) + \sum_{i=1}^k \lambda_i h_i(\mathbf{x}) - \sum_{j=1}^m \mu_j g_j(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^n \}$$

و به تابع $\sum_{i=1}^k \lambda_i h_i(\mathbf{x}) - \sum_{j=1}^m \mu_j g_j(\mathbf{x})$ تابع لاگرانژ می‌گویند.

مجموعه داده‌های $\{\mathbf{x}_i, y_i\}$ $i = 1, \dots, n$ را در نظر بگیرید که در آن $\mathbf{x}_i = (x_1, \dots, x_p)^T$ بردار ویژگی‌ها و $y_i \in \{-1, +1\}$ برچسب هستند. معادله ابرصفحه جداکننده برای i امین داده به صورت $\mathbf{w}^T \mathbf{x}_i + b$ در نظر گرفته شده است. اکنون دو ابر صفحه حاشیه‌ای طوری در نظر گرفته شده‌اند که اگر $y_i = +1$ آن‌گاه $\mathbf{w}^T \mathbf{x}_i + b > 1$ و اگر $y_i = -1$ آن‌گاه $\mathbf{w}^T \mathbf{x}_i + b < -1$ است. هدف ماشین بردار پشتیبان یافتن ابرصفحه‌ای است که داده‌های جدایی‌پذیر خطی را جدا کند. در واقع لازم است برای هر $\mathbf{x}_i$ برچسب y_i به درستی پیش‌بینی شود. چیزی که در اینجا اهمیت دارد یافتن مقدار $\mathbf{w}$ و b است و به دنبال ابرصفحه‌ی بهینه خواهیم بود. برای مقایسه‌ی دو ابرصفحه، ابرصفحه‌ای انتخاب شود که حاشیه اطمینان بزرگتری داشته باشد. همانطور که در شکل ۱ دیده می‌شود، ایزنمن (۲۰۰۸) با استفاده از ابر صفحه جدا کننده که دقیقاً وسط بین دو ابرصفحه حاشیه‌ای قرار دارد، فاصله بین این دو ابر صفحه حاشیه‌ای را برابر با $\frac{2}{\|\mathbf{w}\|}$ گرفت. از آن جایی که هدف پیدا کردن ابرصفحه جداکننده



شکل ۱. فاصله بردار پشتیبان‌ها از ابرصفحه بهینه

با بیشترین فاصله از نقاط حاشیه‌ای است کفایت $\frac{\|\mathbf{w}\|}{2}$ کمینه شود. به مشاهداتی که روی ابر صفحات حاشیه‌ای قرار می‌گیرند، بردارهای پشتیبان و به فاصله بردارهای پشتیبان تا ابرصفحه‌ی جداکننده (مرز تصمیم‌گیری) حاشیه می‌گویند. به این ترتیب مسئله بهینه‌سازی مقید

$$\min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 \quad y_i (\mathbf{w}^T \mathbf{x}_i + b) \geq 1 \quad (2)$$

ارائه و با روش دوگان حل می‌شود. با استفاده از ضرایب نامنفی لاگرانژ می‌توان مسئله (۲) را به صورت

$$\max_{\alpha} \min_{\mathbf{w}, b} : L_P = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_i \alpha_i [y_i (\mathbf{w}^T \mathbf{x}_i + b) - 1]. \quad (3)$$

بازنویسی نمود و برای کمینه کردن آن خواهیم داشت

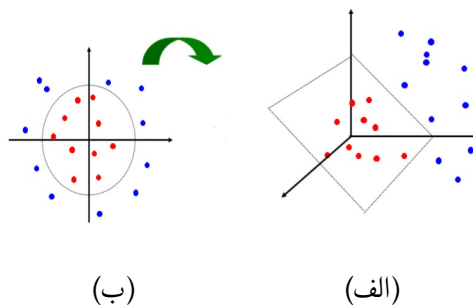
$$\begin{cases} \frac{\partial L_P}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w} = \sum_i \alpha_i y_i \mathbf{x}_i, \\ \frac{\partial L_P}{\partial b} = 0 \Rightarrow \sum_i \alpha_i y_i = 0. \end{cases} \quad (4)$$

پس از جایگذاری روابط (۴) در (۳) دوگان آن به صورت

$$L_D = \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j$$

محاسبه می‌شود. بنابراین با محاسبه کمینه L_P یا بیشینه L_D با شرط $\alpha_i \geq 0$ ، می‌توان به جواب نهایی مسئله دست یافت.

تفکیک کننده‌های خطی در شرایطی که داده‌های تحت بررسی دارای پیچیدگی بوده و الگوی مشخصی ندارند دارای عملکرد نسبتاً ضعیفی هستند. یک رویکرد پیشنهادی در چنین وضعیت‌هایی، انتقال داده‌ها به فضایی با بعد بالاتر است. همان‌طور که در شکل ۲ ملاحظه می‌شود، کلاس بندی داده‌ها به مراتب ساده‌تر شده است. در حالت‌های ذکر شده در بخش‌های گذشته در واقع از حاصل ضرب داخلی داده‌های آموزشی ورودی در بردارهای پشتیبان برای تشکیل جدا کننده خطی به شکل ابر صفحه استفاده شده است، در اینجا ابتدا داده‌ها را در یک فضای اقلیدسی با ابعاد بالاتر نگاشت نموده سپس از ضرب داخلی عناصر بدست آمده استفاده می‌شود.



شکل ۲. نمایشی از انتقال داده از فضای دو بعدی (ب) به فضای سه بعدی (الف)

وینیک (۱۹۹۵) نشان داد اگر تابع هسته به صورت $K(\mathbf{x}_i, \mathbf{x}_j) = \varphi^T(\mathbf{x}_i) \varphi(\mathbf{x}_j)$ تعریف شود، تابع

هسته مورد استفاده مناسب بوده و داده‌های جدایی ناپذیر در فضای اصلی در فضای نگاشته شده جدایی پذیر خواهند بود. از این‌رو، جایگزینی تابع هسته می‌تواند الگوریتمی غیرخطی را از الگوریتم خطی ذکر شده ایجاد نماید (اسدالهی و همکاران، ۲۰۲۱).

۳ مدل پیشنهادی

مدل امکانی داده‌های فازی $\tilde{Z}_t$ ، $t = 0, 1, \dots, T$ را به صورت

$$\tilde{Z}_t = \bigoplus_{j=1}^T W_j \tilde{Z}_{t-j} + W_0 \quad t = P+1, P+2, \dots, T$$

در نظر بگیرید، که در آن $\tilde{Z}_t = (Z_t, \ell_{Z_t})_{LL}$ و ضرایب W_j و W_0 مقادیری حقیقی هستند که در مدل‌سازی به دنبال برآورد آن‌ها با روش بردار پشتیبان خواهیم بود. با توجه به شکل ۳، معادله هدف را به صورت

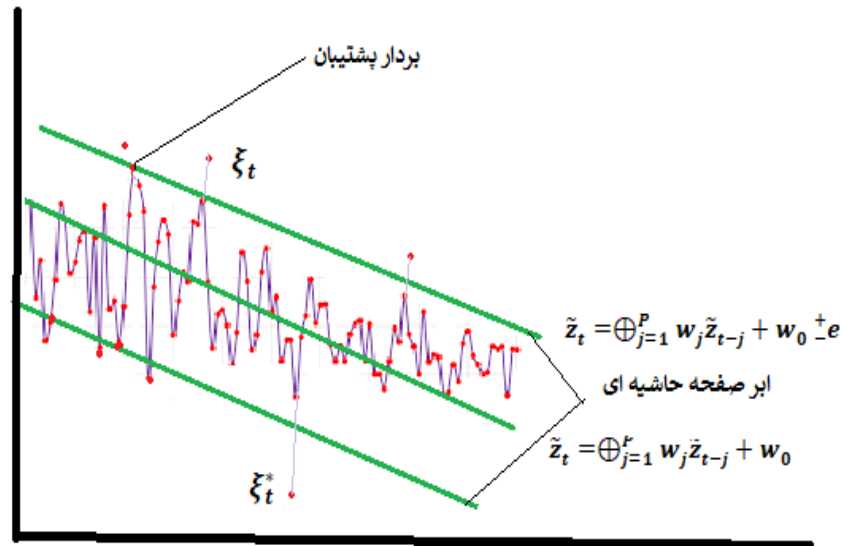
$$J_p(\mathbf{W}, \xi)(\mathbf{W}, W_0, \xi_{t1}, \xi_{t1}^*, \xi_{t2}, \xi_{t2}^*) = \min \left\{ \frac{\|\mathbf{W}\|^2}{2} + C \sum_{k=1}^Y \sum_{t=P+1}^T (\xi_{tk} + \xi_{tk}^*) \right\} \quad (5)$$

در نظر بگیرید، به طوری که

$$\begin{cases} Z_t - (\sum_{j=1}^P W_j Z_{t-j} + W_0) \leq e + \xi_{t1} \rightarrow \alpha_{t1} \\ \sum_{j=1}^P W_j Z_{t-j} + W_0 - Z_t \leq e + \xi_{t1}^* \rightarrow \alpha_{t1}^* \\ (Z_t - \ell_{Z_t}) - (\sum_{j=1}^P W_j Z_{t-j} + W_0 - \sum_{j=1}^P |W_j| \ell_{Z_{t-j}}) \leq e + \xi_{t2} \rightarrow \alpha_{t2} \\ (\sum_{j=1}^P W_j \ell_{t-j} + W_0 - \sum_{j=1}^P |W_j| \ell_{Z_{t-j}})(Z_t - \ell_{Z_t}) \leq e + \xi_{t2}^* \rightarrow \alpha_{t2}^* \\ \xi_{t1}, \xi_{t1}^*, \xi_{t2}, \xi_{t2}^* \geq 0 \end{cases}$$

با استفاده از روش ضرایب لاگرانژ، معادله (۵) به صورت

$$\begin{aligned} L(\mathbf{W}, \xi, \alpha) = & J_p(\mathbf{W}, W_0, \xi_{t1}, \xi_{t1}^*, \xi_{t2}, \xi_{t2}^*) - \sum_{t=P+1}^T \alpha_{t1} (\xi_{t1} - Z_t + \mathbf{W}_t^T \mathbf{Z}_{t_P} \\ & + W_0 + e) - \sum_{t=P+1}^T \alpha_{t1}^* (\xi_{t1}^* - \mathbf{W}_t^T \mathbf{Z}_{t_P} - W_0 + e + Z_t) \\ & - \sum_{t=P+1}^T \alpha_{t2} (\xi_{t2} - Z_t + \ell_{Z_t} + \mathbf{W}_t^T \mathbf{Z}_{t_P} + W_0 - |\mathbf{W}|^T \mathbf{1}_{\mathbf{Z}_{t_P}} + e) \\ & - \sum_{t=P+1}^T \alpha_{t2}^* (\xi_{t2}^* - \mathbf{W}_t^T \mathbf{Z}_{t_P} - W_0 + |\mathbf{W}|^T \mathbf{1}_{\mathbf{Z}_{t_P}} + e + Z_t - \ell_{Z_t}) \\ & - \sum_{k=1}^Y \sum_{t=P+1}^T \eta_{tk}^* \xi_{tk}^* - \sum_{k=1}^Y \sum_{t=P+1}^T \eta_{tk} \xi_{tk} \end{aligned} \quad (6)$$



شکل ۳. بردار پشتیبان داده وابسته به زمان در حالت خطی با خطا

بدست می‌آید، که در آن $\alpha_{tk}, \alpha_{tk}^*, \eta_t, \eta_t^*$ ضرایب لاگرانژ هستند. بنابراین

$$\left\{ \begin{array}{l} \frac{\partial L_P}{\partial \mathbf{W}} = 0 \Rightarrow \mathbf{W} = \sum_{t=P+1}^T (\alpha_{t1} - \alpha_{t1}^*) \mathbf{Z}_{tP} - \sum_{t=P+1}^T (\alpha_{t2} - \alpha_{t2}^*) (\mathbf{Z}_{tP} - \mathbf{1}_{\mathbf{Z}_{tP}}) \\ \frac{\partial L_P}{\partial \mathbf{W}^*} = 0 \Rightarrow \sum_{k=1}^V \sum_{t=P+1}^T (\alpha_{tk} - \alpha_{tk}^*) = 0 \\ \frac{\partial L_P}{\partial \xi_{tk}} = 0 \Rightarrow C - \alpha_{tk}^* - \eta_{tk}^* = 0 \Rightarrow 0 < \alpha_{tk}^* < C \\ \frac{\partial L_P}{\partial \xi_{tk}} = 0 \Rightarrow C - \alpha_{tk} - \eta_{tk} = 0 \Rightarrow 0 < \alpha_{tk} < C \end{array} \right. \quad (7)$$

که پس از جایگذاری معادلات (۷) در (۶) دوگان آن به صورت

$$\begin{aligned} -\frac{\|\mathbf{W}\|^2}{2} - e \sum_{k=1}^V \sum_{t=P+1}^T (\alpha_{tk} + \alpha_{tk}^*) + \sum_{t=P+1}^T z_t (\alpha_{t1} + \alpha_{t1}^*) \\ + \sum_{t=P+1}^T (z_t - \ell_{z_t}) (\alpha_{t2} + \alpha_{t2}^*) \end{aligned} \quad (8)$$

رنجبر، ا. و همکاران ۱۰۷

بدست می‌آید. حال دوگان مساله برنامه‌ریزی شده قبل برای به‌دست آوردن ضرایب منجر به حل مساله

$$\begin{aligned} \min_{\alpha_{tk}, \alpha_{tk}^*} & -\frac{1}{\Psi} \left(\sum_{i=P+1}^T \sum_{j=P+1}^T (\alpha_{i1} - \alpha_{i1}^*) (\alpha_{j2} - \alpha_{j2}^*) \mathbf{Z}_{iP}^T \mathbf{Z}_{jP} \right. \\ & + \sum_{i=P+1}^T \sum_{j=P+1}^T (\alpha_{i1} + \alpha_{i1}^*) (\alpha_{j2} - \alpha_{j2}^*) \mathbf{Z}_{iP}^T (\mathbf{Z}_{jP} - \mathbf{1}_{\mathbf{Z}_{jP}}) \\ & + \sum_{i=P+1}^T \sum_{j=P+1}^T (\alpha_{i2} + \alpha_{i2}^*) (\alpha_{j1} - \alpha_{j1}^*) (\mathbf{Z}_{iP} - \mathbf{1}_{\mathbf{Z}_{iP}})^T \mathbf{Z}_{jP} \\ & + \sum_{i=P+1}^T \sum_{j=P+1}^T (\alpha_{i2} + \alpha_{i2}^*) (\alpha_{j2} - \alpha_{j2}^*) (\mathbf{Z}_{iP} - \mathbf{1}_{\mathbf{Z}_{iP}})^T (\mathbf{Z}_{jP} - \mathbf{1}_{\mathbf{Z}_{jP}}) \Big) \\ & - e \sum_{k=1}^2 \sum_{t=P+1}^T (\alpha_{tk} + \alpha_{tk}^*) + \sum_{t=P+1}^T (\alpha_{t1} - \alpha_{t1}^*) Z_t \\ & + \sum_{t=P+1}^T (\alpha_{t2} - \alpha_{t2}^*) (Z_t - \ell_{Z_t}) \end{aligned}$$

می‌گردد. به‌طوری‌که

$$\begin{cases} \sum_{t=P+1}^T (\alpha_{tk} + \alpha_{tk}^*) = 0 & k = 1, 2 \\ \alpha_{tk}, \alpha_{tk}^* \in [0, 2] & k = 1, 2 \\ \mathbf{Z}_{iP} = (Z_{i-1}, \dots, Z_{i-P}) & i = P+1, \dots, T \end{cases}$$

در نتیجه

$$\begin{aligned} \tilde{z}_t = (\hat{z}_t, \ell_{z_t}) &= \hat{\mathbf{w}}^T \tilde{\mathbf{Z}}_{tP} + w_* = (\hat{\mathbf{w}}^T \mathbf{Z}_{tP} + w_*, |\hat{\mathbf{w}}|^T \ell_{\mathbf{Z}_{tP}})_T \\ &= \left(\begin{array}{c} \sum_{\ell=P+1}^T (\hat{\alpha}_{\ell 1} - \hat{\alpha}_{\ell 1}^*) \mathbf{Z}_{\ell P}^T \mathbf{Z}_{tP} + \sum_{\ell=P+1}^T (\hat{\alpha}_{\ell 2} - \hat{\alpha}_{\ell 2}^*) (\mathbf{Z}_{\ell P}^T \mathbf{Z}_{tP} - \ell_{\mathbf{Z}_{\ell P}}^T \mathbf{Z}_{tP}) + \hat{W}_*, \\ \sum_{\ell=P+1}^T |(\hat{\alpha}_{\ell 1} - \hat{\alpha}_{\ell 1}^*)| |\mathbf{Z}_{\ell P}^T \mathbf{Z}_{tP}| + \sum_{\ell=P+1}^T |(\hat{\alpha}_{\ell 2} - \hat{\alpha}_{\ell 2}^*)| |(\mathbf{Z}_{\ell P}^T \mathbf{Z}_{tP} - \ell_{\mathbf{Z}_{\ell P}}^T \ell_{\mathbf{Z}_{tP}})| \end{array} \right)_T \end{aligned}$$

و مقدار w_* با روش KKT (اسدالهی و همکاران، ۲۰۲۱) با توجه به معادلات

$$\begin{cases} W_* = Z_t - \sum_{j=1}^P W_j Z_{t-j} - e = 0 & \alpha_{t1} \in [0, 2], \xi_{t1} = 0 \\ W_* = Z_t - \sum_{j=1}^P W_j Z_{t-j} + e = 0 & \alpha_{t1}^* \in [0, 2], \xi_{t1}^* = 0 \end{cases}$$

محاسبه می‌شود یا

$$\begin{cases} W_* = Z_t - \mathbf{w}_j^T \mathbf{Z}_{tP} - e = 0 & \alpha_{t1} \in [0, 2], \xi_{t1} = 0 \\ W_* = Z_t - \mathbf{w}_j^T \mathbf{Z}_{tP} + e = 0 & \alpha_{t1}^* \in [0, 2], \xi_{t1}^* = 0 \end{cases}$$

به طور کلی

$$\hat{W}_* = \frac{1}{|S|} \sum_{t \in S} (Z_t - \mathbf{w}_j^T \mathbf{Z}_{tP} - \text{Sign}(\alpha_{t\downarrow} - \alpha_{t\downarrow}^*) e)$$

$$S = \{t | 0 < \alpha_{t\downarrow} - \alpha_{t\downarrow}^* < C\}$$

که منظور از $|S|$ میانگین است. در حالت غیر خطی برای جلوگیری از بیان مطالب اضافی به بیان تابع دوگان مساله خطی محدب بخش قبل رابطه (۸) با استفاده از تابع هسته معرفی شده $K(x, y)$ به صورت

$$\begin{aligned} \min_{\alpha_{tk}, \alpha_{tk}^*} & -\frac{1}{\gamma} \left(\sum_{i=P+1}^T \sum_{j=P+1}^T (\alpha_{i\downarrow} - \alpha_{i\downarrow}^*)(\alpha_{j\downarrow} - \alpha_{j\downarrow}^*) \varphi^T(\mathbf{Z}_{iP}) \cdot \varphi(\mathbf{Z}_{jP}) \right. \\ & + \sum_{i=P+1}^T \sum_{j=P+1}^T (\alpha_{i\downarrow} + \alpha_{i\downarrow}^*)(\alpha_{j\uparrow} - \alpha_{j\uparrow}^*) \varphi^T(\mathbf{Z}_{iP}) \cdot \varphi(\mathbf{Z}_{jP} - \mathbf{1}_{\mathbf{Z}_{jP}}) \\ & + \sum_{i=P+1}^T \sum_{j=P+1}^T (\alpha_{i\uparrow} + \alpha_{i\uparrow}^*)(\alpha_{j\downarrow} - \alpha_{j\downarrow}^*) \varphi^T(\mathbf{Z}_{iP} - \mathbf{1}_{\mathbf{Z}_{iP}}) \cdot \varphi(\mathbf{Z}_{jP}) \\ & \left. + \sum_{i=P+1}^T \sum_{j=P+1}^T (\alpha_{i\uparrow} + \alpha_{i\uparrow}^*)(\alpha_{j\uparrow} - \alpha_{j\uparrow}^*) \varphi^T(\mathbf{Z}_{iP} - \mathbf{1}_{\mathbf{Z}_{iP}}) \cdot \varphi(\mathbf{Z}_{jP} - \mathbf{1}_{\mathbf{Z}_{jP}}) \right) \\ & - e \sum_{k=\downarrow}^{\uparrow} \sum_{t=P+1}^T (\alpha_{tk} + \alpha_{tk}^*) + \sum_{t=P+1}^T (\alpha_{t\downarrow} - \alpha_{t\downarrow}^*) Z_t \\ & + \sum_{t=P+1}^T (\alpha_{t\uparrow} - \alpha_{t\uparrow}^*) (Z_t - \ell_{Z_t}) \end{aligned} \quad (9)$$

بسنده خواهیم کرد. پس از حل معادله دوگان (۹) مقادیر $\alpha_{tk}, \alpha_{tk}^*$ ها مقدار برآورد $\hat{Z}_t$ به صورت

$$\begin{aligned} \tilde{z}_t = (\hat{z}_t, \hat{\ell}_{z_t}) &= \hat{\mathbf{w}}^T \tilde{\mathbf{Z}}_{tP} + w_* = (\hat{\mathbf{w}}^T \mathbf{Z}_{tP} + w_*, |\hat{\mathbf{w}}|^T \ell_{\mathbf{Z}_{tP}})_T \\ &= \left(\frac{\sum_{\ell=P+1}^T (\hat{\alpha}_{\ell\downarrow} - \hat{\alpha}_{\ell\downarrow}^*) K_h(\mathbf{Z}_{\ell P}, \mathbf{Z}_{tP}) + \sum_{\ell=P+1}^T (\hat{\alpha}_{\ell\uparrow} - \hat{\alpha}_{\ell\uparrow}^*) K_h(\mathbf{Z}_{\ell P} - \ell_{\mathbf{Z}_{tP}}, \mathbf{Z}_{tP}) + \hat{W}_*,}{|\sum_{\ell=P+1}^T (\hat{\alpha}_{\ell\downarrow} - \hat{\alpha}_{\ell\downarrow}^*) k_h(\mathbf{Z}_{\ell P}, \ell_{\mathbf{Z}_{tP}}) + \sum_{\ell=P+1}^T (\hat{\alpha}_{\ell\uparrow} - \hat{\alpha}_{\ell\uparrow}^*) k_h(\mathbf{Z}_{\ell P} - \ell_{\mathbf{Z}_{tP}}, \ell_{\mathbf{Z}_{tP}})|} \right)_T \end{aligned}$$

به دست خواهد آمد. پارامتر h موجود در تابع هسته محقق را قادر می سازد تا یک مدل مناسب به داده ها برازش دهد. به عبارتی دقیق تر، مقدار بهینه این پارامتر باعث تخمین یک مدل هموار نسبت به داده ها شده به طوریکه نسبت به داده های دور افتاده هم استوار باشد. مقدار زیاد این پارامتر باعث اریبی و مقدار کم آن باعث واریانس زیاد است. بنابراین، مقدار بهینه آن مقداری خواهد بود که ریسک را کمتر کند.

۴ تحلیل داده ها

در این بخش به منظور تشریح جزئیات روش پیشنهادی و مقایسه عملکرد پان با سایر روش های موجود، یک مثال شبیه سازی شده و دو مثال با داده های واقعی مورد تحلیل قرار گرفته است. لازم به ذکر است که برای حل مسائل بهینه سازی در روش پیشنهادی از برنامه نویسی در نرم افزار *Mathematica* استفاده شده است. علاوه بر این،

در این بخش از معیارهای نیکویی برازش که تعریف آن‌ها در زیر ارائه شده است، برای ارزیابی و مقایسه توانایی پیش‌بینی مدل‌ها استفاده شده است.
-اندازه مشابهت:

$$MSM = \frac{1}{n} \sum_{i=1}^n S_{UI}(\tilde{x}_i, \tilde{x}_i),$$

به‌طوری‌که $S_{UI}(\tilde{x}_i, \tilde{x}_i) = \frac{Card(\tilde{x}_i \cap \tilde{x}_i)}{Card(\tilde{x}_i \cup \tilde{x}_i)} = \frac{\int \min(\tilde{x}_i(y), \tilde{x}_i(y)) dy}{\int \max(\tilde{x}_i(y), \tilde{x}_i(y)) dy}$ اعداد فازی دلخواه هستند.

- میانگین مربعات خطا:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n D^*(\tilde{x}_i, \tilde{x}_i)}{n}}$$

به‌طوری‌که $D^*(\tilde{x}_i, \tilde{x}_i) = \frac{1}{3} (|m_{\tilde{x}_i} - m_{\tilde{x}_i}| + c_1 \times |\ell_{\tilde{x}_i} - \ell_{\tilde{x}_i}| + c_2 \times |r_{\tilde{x}_i} - r_{\tilde{x}_i}|)$ که در آن $c_2 = \int_0^1 R^{-1}(\alpha) d\alpha$ و $c_1 = \int_0^1 L^{-1}(\alpha) d\alpha$
- میانگین قدرمطلق مقیاسی خطا:

$$MASE = \frac{1}{T-P} \sum_{i=p+1}^T \frac{D^*(\tilde{x}_i, \tilde{x}_i)}{\frac{1}{T-P} \sum_{j=p+1}^T D^*(\tilde{x}_j, \tilde{x}_{j-1})}$$

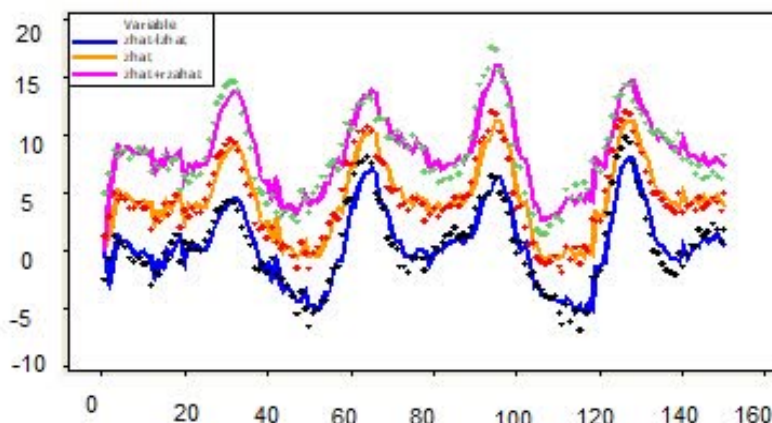
مثال ۱. (داده‌های شبیه‌سازی) تعداد ۱۵۰ داده از سری‌زمانی با مشخصات

$$Z_t = te^{\gamma t} + 0.5 \log(t) \sin(0.1t) + 2 \log(2t) \cos(0.2t) + \varepsilon_t \quad \varepsilon_t \sim N(0, 0.1^2)$$

$$\ell_{Z_t} = 4 - 0.3 \log(2t) \cos(0.3t) + \epsilon'_t \quad \epsilon'_t \sim N(0, 0.05^2)$$

شبیه‌سازی شده است، که در آن ورودی‌های t به‌صورت تصادفی از $U(-10, 10)$ مشاهده شده‌اند و $L(x) = \max(0, 1-x)$. لازم به‌ذکر است که علت انتخاب چنین قالبی برای تولید داده‌های فازی، رونددار بودن داده‌ها به گونه‌ای بوده است که پهنای اعداد فازی شبیه‌سازی شده نیز دارای روند باشند. شبیه‌سازی به تعداد $s = 10$ بار انجام شده و در هر بار به‌منظور ایجاد داده دور افتاده، به‌صورت تصادفی، مقدار نه به مراکز ده عدد فازی تولید شده، و مقدار هشت به پهنای ده عدد فازی دیگر اضافه گردید. پس از آن، مطابق مراحل ذکر شده در بخش سوم و با در نظر گرفتن مقادیر $c = 100$ ، $e = 1$ و $T = 150$ ، $h = 0.5$ برای هسته نرمال و نمونه ۱۵۰ تایی، نمودار برآورد مدل سری‌زمانی داده‌های فازی برای مرکز، ابهام چپ و ابهام راست در شکل ۴ نمایش داده شده است.

نتایج معیارهای نیکویی برازش برای مدل پیشنهادی و سایر مدل‌های تحت بررسی در این مطالعه، در جدول



شکل ۴. نمودار برازش شده برای داده‌های تحت بررسی در مثال ۱.

۴ گزارش شده است که بر اساس نتایج به‌دست آمده در هر سه شاخص ارزیابی مورد استفاده در این مطالعه، روش پیشنهادی مبتنی بر ماشین بردار پشتیبان دارای بهترین عملکرد در بین روش‌های مذکور بوده است ($\overline{MSM} = ۷,۴۳$, $\overline{RMSE} = ۱۷,۰۲$, $\overline{MASE} = ۰,۴۹$). هم‌چنین، نمودارهای ارائه شده در شکل ۴ نشان می‌دهند که مدل‌های ارائه شده بر اساس روش پیشنهادی نسبت به داده‌های دور افتاده موجود در مدل استوار بوده و این داده‌ها تاثیری در روند مدل نداشته‌اند.

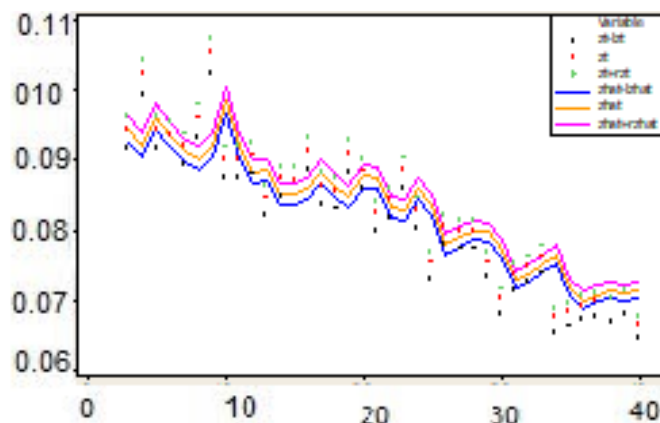
جدول ۱. نتایج معیارهای نیکویی برازش برای مدل پیشنهادی و سایر مدل‌ها

مدل	$\overline{MASE}$	$\overline{RMSE}$	$\overline{MSM}$
مدل پیشنهادی	۷,۴۳	۱۷,۰۲	۰,۴۹
حسامیان و همکاران (۲۰۲۲)	۸,۸۴	۲۱,۳۶	۰,۴۱
زارعی و همکاران (۲۰۲۰)	۱۰,۸۴	۲۴,۳۵	۰,۳۵
محمدی و همکاران (۱۴۰۳)	۱۰,۳۸	۲۲,۱۷	۰,۳۸

مثال ۲. داده‌های فازی مثلثی $\tilde{Z}(t) = (Z(t); ۰,۹۷Z(t), ۱,۰۲Z(t))_T$ گزارش شده در جدول ۲ بر اساس گزارش سازمان محیط زیست آمریکا در خصوص متوسط غلظت سالانه ازن در مناطق مختلف طی سال‌های ۱۹۸۰ تا ۲۰۱۹ در نظر گرفته شده است (<https://www.epa.gov/air-trends/ozone-trends>). با در نظر گرفتن مقادیر $e = ۱$, $c = ۱۰۰$, $T = ۱۵۰$ و $h = ۰,۰۵$ بر اساس مراحل در نظر گرفته شده برای روش پیشنهادی در بخش سوم، نمودار برآورد مدل سری‌زمانی داده‌های فازی برای مرکز، ابهام چپ و ابهام راست در شکل ۵ نمایش داده شده است.

جدول ۲. داده‌های فازی مثلثی مربوط به غلظت ازن

r_{Z_t}	$Z(t)$	ℓ_{Z_t}	Year	t	r_{Z_t}	$Z(t)$	ℓ_{Z_t}	Year	t
۰/۰۸۴۰	۰/۰۸۲۴	۰/۰۷۹۹	۲۰۰۰	۲۱	۰/۱۰۳۶	۰/۱۰۱۵	۰/۰۹۸۵	۱۹۸۰	۱
۰/۰۸۵۹	۰/۰۸۴۲۹	۰/۰۸۱۸	۲۰۰۱	۲۲	۰/۰۹۷۹	۰/۰۹۵۹	۰/۰۹۳۱	۱۹۸۱	۲
۰/۰۹۰۳	۰/۰۸۸۵	۰/۰۸۵۹	۲۰۰۲	۲۳	۰/۰۹۶۳	۰/۰۹۴۴	۰/۰۹۱۶	۱۹۸۲	۳
۰/۰۸۴۵	۰/۰۸۲۹	۰/۰۸۰۴	۲۰۰۳	۲۴	۰/۱۰۴۳	۰/۱۰۲۲	۰/۰۹۹۲	۱۹۸۳	۴
۰/۰۷۶۹	۰/۰۷۵۴	۰/۰۷۳۱	۲۰۰۴	۲۵	۰/۰۹۶۳	۰/۰۹۴۴	۰/۰۹۱۵	۱۹۸۴	۵
۰/۰۸۲۱	۰/۰۸۰۵	۰/۰۷۸۱	۲۰۰۵	۲۶	۰/۰۹۵۴	۰/۰۹۳۵	۰/۰۹۰۷	۱۹۸۵	۶
۰/۰۸۱۳	۰/۰۷۹۷	۰/۰۷۷۳	۲۰۰۶	۲۷	۰/۰۹۳۷	۰/۰۹۱۹	۰/۰۸۹۲	۱۹۸۶	۷
۰/۰۸۱۴	۰/۰۷۹۸	۰/۰۷۷۴	۲۰۰۷	۲۸	۰/۰۹۷۹	۰/۰۹۶۰	۰/۰۹۳۲	۱۹۸۷	۸
۰/۰۷۷۰	۰/۰۷۵۴	۰/۰۷۳۱۸	۲۰۰۸	۲۹	۰/۱۰۷۴	۰/۱۰۵۳	۰/۱۰۲۲	۱۹۸۸	۹
۰/۰۷۱۵	۰/۰۷۰۲	۰/۰۶۸۱	۲۰۰۹	۳۰	۰/۰۹۱۸	۰/۰۹۰۱	۰/۰۸۷۳	۱۹۸۹	۱۰
۰/۰۷۵۱	۰/۰۷۳۶	۰/۰۷۱۴	۲۰۱۰	۳۱	۰/۰۹۲۰	۰/۰۹۰۱	۰/۰۸۷۴	۱۹۹۰	۱۱
۰/۰۷۶۲	۰/۰۷۴۸	۰/۰۷۲۵	۲۰۱۱	۳۲	۰/۰۹۲۵	۰/۰۹۰۷	۰/۰۸۸۰	۱۹۹۱	۱۲
۰/۰۷۷۸	۰/۰۷۶۲	۰/۰۷۳۹	۲۰۱۲	۳۳	۰/۰۸۶۲	۰/۰۸۴۶	۰/۰۸۲۰	۱۹۹۲	۱۳
۰/۰۶۸۸	۰/۰۶۷۵	۰/۰۶۵۵	۲۰۱۳	۳۴	۰/۰۸۹۱	۰/۰۸۷۴	۰/۰۸۴۷	۱۹۹۳	۱۴
۰/۰۶۹۶	۰/۰۶۸۲	۰/۰۶۶۲	۲۰۱۴	۳۵	۰/۰۸۸۹	۰/۰۸۷۲۰	۰/۰۸۴۵	۱۹۹۴	۱۵
۰/۰۷۰۷	۰/۰۶۹۳	۰/۰۶۷۲	۲۰۱۵	۳۶	۰/۰۹۳۳	۰/۰۹۱۴	۰/۰۸۸۷	۱۹۹۵	۱۶
۰/۰۷۱۲	۰/۰۶۹۸	۰/۰۶۷۷	۲۰۱۶	۳۷	۰/۰۸۸۱	۰/۰۸۶۴	۰/۰۸۳۸	۱۹۹۶	۱۷
۰/۰۷۰۴	۰/۰۶۹۰	۰/۰۶۷۰	۲۰۱۷	۳۸	۰/۰۸۷۴	۰/۰۸۵۶	۰/۰۸۳۱	۱۹۹۷	۱۸
۰/۰۷۱۴	۰/۰۷۰۱	۰/۰۶۸۰	۲۰۱۸	۳۹	۰/۰۹۲۸	۰/۰۹۱۰	۰/۰۸۸۳	۱۹۹۸	۱۹
۰/۰۶۷۸	۰/۰۶۶۵	۰/۰۶۴۵	۲۰۱۹	۴۰	۰/۰۹۰۲	۰/۰۸۸۴	۰/۰۸۵۸	۱۹۹۹	۲۰



شکل ۵. نمودار برازش شده به داده‌های فازی.

نتایج معیارهای نیکویی برازش برای مدل پیشنهادی و سایر مدل‌های تحت بررسی در این مطالعه، در جدول ۳ گزارش شده است که بر اساس نتایج به دست آمده در هر سه شاخص، روش پیشنهادی در این مطالعه عملکردی بهتر نسبت به سایر روش‌ها را دارا بوده است. همچنین، نمودارهای ارائه شده در شکل ۵ نشان می‌دهند که مدل‌های ارائه شده بر اساس روش پیشنهادی نسبت به داده‌های دور افتاده موجود در مدل استوار بوده و این داده‌ها تأثیری در روند مدل نداشته‌اند.

جدول ۳. نتایج معیارهای نیکویی برازش برای مدل پیشنهادی و سایر مدل‌ها

مدل	MSM	RMSE	MASE
مدل پیشنهادی	۰/۵۲	۱/۰۸	۰/۸۱
حسامیان و همکاران (۲۰۲۲)	۰/۴۸	۱۳/۸۵	۰/۹۱
زارعی و همکاران (۲۰۲۰)	۰/۴۴	۱۶/۱۲	۱/۲۱
محمدی و همکاران (۱۴۰۳)	۰/۴۷	۱۵/۰۳	۱/۰۰

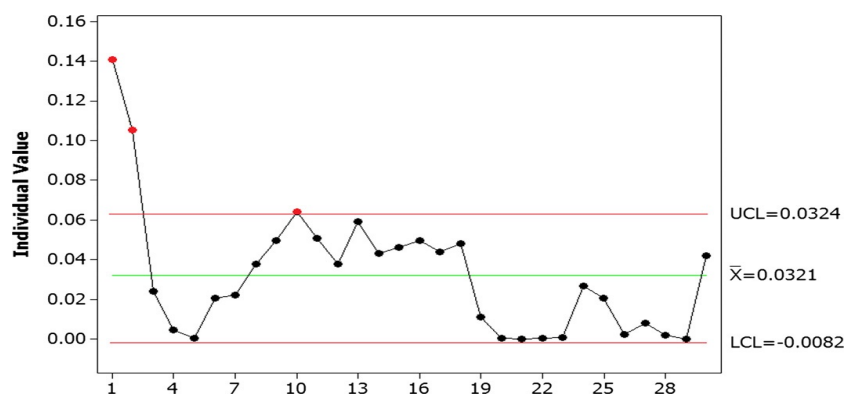
مثال ۳. در یک بررسی برای ارزیابی قابلیت اطمینان نرم‌افزار، یک تیم ارزیاب هفت نفره در یک بازه زمانی به طول سی ماه، مجموعه داده‌های گزارش شده در جدول ۴ را در قالب اعداد فازی مثلثی گزارش داده‌اند (دورسو و گاستالدی، ۲۰۰۲). در نخستین گام، وجود داده‌های دور افتاده با استفاده از نمودار ارائه شده در شکل ۶ مورد ارزیابی قرار گرفت که بر اساس آن می‌توان گفت دو داده دور افتاده در بین مجموعه داده تحت بررسی وجود دارد. با در نظر گرفتن مقادیر $e = 1$ ، $c = 100$ ، $T = 150$ و $h = 0.5$ و بر اساس مراحل در نظر گرفته شده برای روش پیشنهادی در بخش سوم، نمودار برآورد مدل سری‌زمانی داده‌های فازی برای مرکز، ابهام چپ و ابهام راست در شکل ۷ نمایش داده شده است. خلاصه نتایج حاصل از بررسی معیارهای نیکویی برازش در مقایسه روش پیشنهادی با برخی روش‌های دیگر در جدول ۵ ارائه شده است که نشان دهنده عملکرد بهتر مدل پیشنهادی در مدل‌سازی داده‌های فازی در این مثال است.

جدول ۴. داده‌های فازی مثلثی ارزیابی قابلیت اطمینان

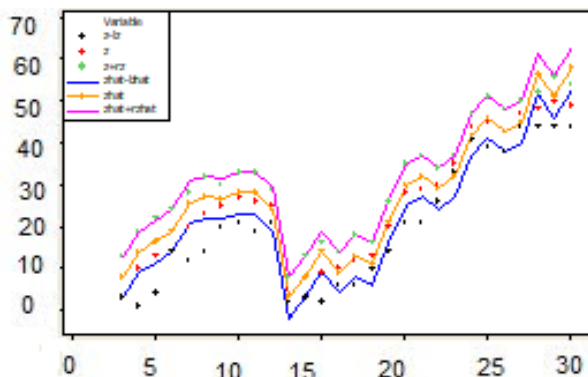
$\tilde{z}_t$	t	$\tilde{z}_t$	t	$\tilde{z}_t$	t
$(29; 8, 9)_T$	۲۱	$(26; 7, 6)_T$	۱۱	$(6; 2, 8)_T$	۱
$(30; 4, 7)_T$	۲۲	$(25; 4, 6)_T$	۱۲	$(5; 3, 6)_T$	۲
$(35; 2, 3)_T$	۲۳	$(5; 3, 2)_T$	۱۳	$(8; 5, 8)_T$	۳
$(44; 3, 5)_T$	۲۴	$(8; 5, 3)_T$	۱۴	$(10; 9, 9)_T$	۴
$(45; 6, 6)_T$	۲۵	$(9; 7, 8)_T$	۱۵	$(13; 9, 8)_T$	۵
$(43; 5, 4)_T$	۲۶	$(10; 4, 3)_T$	۱۶	$(19; 5, 7)_T$	۶
$(47; 3, 9)_T$	۲۷	$(12; 6, 8)_T$	۱۷	$(20; 8, 5)_T$	۷
$(48; 4, 5)_T$	۲۸	$(13; 3, 9)_T$	۱۸	$(23; 9, 5)_T$	۸
$(50; 6, 4)_T$	۲۹	$(20; 6, 7)_T$	۱۹	$(25; 5, 7)_T$	۹
$(49; 5, 4)_T$	۳۰	$(28; 7, 9)_T$	۲۰	$(27; 6, 8)_T$	۱۰

جدول ۵. نتایج معیارهای نیکویی برازش برای مدل پیشنهادی و سایر مدل‌ها

مدل	MSM	RMSE	MASE
مدل پیشنهادی	۰/۴۸	۵/۳۳	۲/۱۳
حسامیان و همکاران (۲۰۲۲)	۰/۴۶	۶/۵۳	۲/۹۱
زارعی و همکاران (۲۰۲۰)	۰/۴۴	۶/۷۲	۳/۰۱
محمدی و همکاران (۱۴۰۳)	۰/۴۳	۶/۸۴	۳/۱۸



شکل ۶. نمودار $\bar{X}$ برای تشخیص داده دور افتاده.



شکل ۷. نمودار برازش شده برای داده‌های فازی.

۵ بحث و نتیجه‌گیری

در این مقاله یک روش جدید مبتنی بر رویکرد ماشین بردار پشتیبان برای مدل‌سازی داده‌های سری زمانی اتورگرسیو فازی پیشنهاد گردید. استفاده از مزایای روش ماشین بردار پشتیبان، با رویکرد مبتنی بر هسته، انعطاف‌پذیری قابل‌توجهی برای مدل‌سازی روابط پیچیده و غیرخطی در سری‌های زمانی فراهم نمود. علاوه بر این، در مقایسه با روش‌های موجود، ادغام محدودیت‌ها در چارچوب رویکرد ماشین بردار پشتیبان، امکان کنترل تأثیر داده‌های پرت را فراهم می‌کند و تأثیر چنین مشاهداتی در مدل‌سازی را کاهش می‌دهد. نتایج تجربی، بر اساس یک مجموعه داده شبیه‌سازی شده و مجموعه داده‌های واقعی، عملکرد بهتر مدل پیشنهادی را در مقایسه با روش‌های موجود نشان داد. با این وجود، روش پیشنهادی محدود به داده‌های فازی مثلثی می‌باشد و توسعه آن برای مجموعه داده‌های فازی از نوع LR می‌تواند یکی از موضوعات مورد علاقه در مطالعات آینده باشد.

تقدیر و تشکر

نویسندگان از داوران، سردبیر، و ویراستار محترم که پیشنهادات ارزشمند ایشان موجب بهبود مقاله گردید کمال تشکر و قدردانی را دارند.

مراجع

طاهری، س.، اکبری، م. ق. و حسامیان، غ. ر. (۱۴۰۳)، مدل سازی میانگین متحرک بر اساس α - شک متغیرهای تصادفی فازی، *مجله علوم آماری*، ۱۸، ۱۲۷-۱۰۳.

محمدی، ح.، اکبری، م. ق. و حسامیان، غ. ر. (۱۴۰۳)، مدل سازی اتو رگرسیو بر اساس تابع تکیه گاه متغیرهای تصادفی فازی، *مجله علوم آماری*، ۱۸، ۱۹۲-۱۷۳.

Asadolahi, M., Akbari, M. G., Hesamian, G. and Arefi, M. (2021), A Robust Support Vector Regression with Exact Predictors and Fuzzy Responses. *International Journal of Approximate Reasoning*, **132**, 206–225.

Bazaraa, M. S., Jarvis, J. J. and Sherali, H. D. (2010), *Linear Programming and Network Flows*, 4th Edition. New York: Wiley.

Chen, S. (1996), Forecasting Enrollments Based on Fuzzy Time Series, *Fuzzy Sets and Systems*, **81**, 311–319.

D'Urso, P. and Gastaldi, T. (2002), an Orderwise Polynomial Regression Procedure for Fuzzy Data, *Fuzzy Sets and Systems*, **130**, 1–19.

Doroudyan, M. H. and Niaki, S. T. A. (2021), Pattern Recognition in Financial Surveillance with The ARMA-GARCH Time Series Model Using Support Vector Machine, *Expert Systems with Applications*.

Gui, B., Wei, X., Shen, Q., Qi, J. and Guo, L. (2015), Financial Time Series Forecasting using Support Vector Machine, *IEEE*.

Hesamian, G. and Akbari, M. G. (2018), A Semi-Parametric Model for Time Series based on Fuzzy Data, *Iranian Journal of Fuzzy Systems*, **26**, 2953-2666.

- Hesamian, G., Torkian, F. and Yarmohammadi, M. (2022), A Fuzzy Non-parametric Time Series Model Based on Fuzzy Data, *Iranian Journal of Fuzzy Systems*, **19**, 61-72.
- Hong, D. (2005), a Note on Fuzzy Time-Series Model, *Fuzzy Sets and Systems*, **155**, 309–316.
- Izenman, A. J. (2008), Modern Multivariate Statistical Techniques (Regression, Classification, and Manifold Learning), *International Statistical Review*.
- Ruan, J., Wang, X. and Shi, Y. (2013), Developing Fast Predictors for Large-scale Time Series using Fuzzy Granular Support Vector Machines , *Applied Soft Computing Journal*, **13**, 3981-4000
- Song, Q. and Chissom, B. S. (1993), Fuzzy Time Series and Its Models, *Fuzzy Sets and Systems*, **54**, 269–277.
- Song, Q., Leland, R. P. and Chissom, B. S. (1995), A New Fuzzy Time-series Model of Fuzzy Number Observations, *Fuzzy Sets and Systems*, **73**, 341–348.
- Suykens, J. A. K., Gestel, J. V., Brabanter, J. D., Moor, B. D. and Vandewalle, J. (2002), Least Squares Support Vector Machines, World Scientific Pub. Co., Singapore, 29-55.
- Vapnik, V. (1995), The Nature of Statistical Learning Theory, *Springer-Verlag* , New York.
- Zadeh, L. A. (1956), Fuzzy Sets, *Information and Control*, **8**, 338-353.
- Zarei, R., Akbari, M. G. and Chaci, J. (2020), Modeling Autoregressive Fuzzy Time Series Data Based on Semi-parametric Methods, *Soft Computing*, **24**, 7295–7304.