ارائهٔ مدل یادگیر ترکیب کرنل‌ها برای پیش بینی سری‌های زمانی براساس رگرسیون بردار پشتیبان و جستجوی فراابتکاری

حدیثه پورعلی[1] و حسام عمرانپور[2*]

[1]دانشکده مهندسی برق و کامپیوتر، دانشگاه صنعتی نوشیروانی بابل، بابل، ایران

[2] دانشکده مهندسی برق و کامپیوتر، دانشگاه صنعتی نوشیروانی بابل، بابل، ایران

چکیده

در این مقاله به ارائه روشی برای پیش‌بینی سری‌زمانی پرداخته شده است. مدلی که در این مقاله ارائه شده بر پایهٔ ترکیب کرنل‌ها و رگرسیون بردار پشتیبان است. رگرسیون بردار پشتیبان با استفاده از کرنل هایش توانایی بالایی در حل مسائل تخمین توابع دارد؛ اما این کرنل‌ها پارامترهایی دارند که نیاز به تنظیم دارند. در مدل پیشنهادی کرنل‌های مختلف بر روی داده‌ها اعمال می‌شوند. خروجی کرنل‌ها با اعمال یک ضریب، با هم ترکیب می‌شوند. این ترکیب باعث می‌شود یک فضای ثانویه جدیدی به‌دست آید. دلیل این امر این است که، ممکن است از بین کرنل‌های موجود فقط یک تعدادی از آن‌ها با ضریب خاصی برای صورت مسأله مفید باشد و ما از این‌که کدام کرنل برای صورت مسألهٔ ما کارا است آگاه نیستیم. همچنین هرکدام از کرنل‌ها پارامترهایی دارند که باید مقادیر بهینهٔ آن‌ها برای دست‌یابی به نتیجهٔ بهتر تعیین شوند؛ ازاین‌رو در مدل ارائه‌شده، یادگیری پارامترهای کرنل و وزن‌های آن‌ها توسط بهینه‌ساز گرگ خاکستری انجام می‌شود مدل پیشنهادی روی پنج مجموعه سری‌زمانی استاندارد پیاده‌سازی شده که نتایج تست براساس معیار RMSE برای سری-زمانی DJ، ۱/۵۸، سری‌زمانی Radio، ۰/۱۷۸، سری زمانی Sunspot، ۱/۷۰۹، نسبت به روش‌های دیگر بهتر شده است. همچنین در انتها به تحلیل نتایج، ارزیابی آماری با آزمون ویلکاکسون رتبه علامت‌دار و ارائه رابطه برای یافتن اندازه پنجره در مدل پرداخته شده است.

واژگان کلیدی: پیش‌بینی سری‌زمانی، رگرسیون بردار پشتیبان، ترکیب توابع کرنل، بهینه سازی

Ensemble Kernel Learning Model for Prediction of Time Series Based on the Support Vector Regression and Meta Heuristic Search

Hadiseh pourali [1] & Hesam omranpour [2*]

[1]Faculty of Electrical and Computer Engineering, Babol Noshirvani University of Thechnology, Babol, Iran.

[2]Faculty of Electrical and Computer Engineering, Babol Noshirvani University of Thechnology, Babol, Iran.

Abstract

In this paper, a method is presented for predicting time series. Time series prediction is a process which predicted future system values based on information obtained from past and present data points. Time series prediction models are widely used in various fields of engineering, economics, etc. The main purpose of using different models for time series prediction is to make the forecast with the greatest accuracy. The model presented in this paper is based on the combination of kernels and support vector regression. Support vector regression is highly capable of solving function estimation problems by using its kernels, but kernels' parameters need to be adjusted. First we have preprocessing phase which

* نویسندهٔ عهده‌دار مکاتبات

* Corresponding author

• تاریخ ارسال مقاله: ۱۳۹۹/۵/۱۹ • تاریخ پذیرش: ۱۳۹۹/۱۰/۲۱ • تاریخ انتشار: ۱۴۰۱/۰۳/۳۱ • نوع مطالعه: پژوهشی

[DOI: 10.52547/jsdp.19.1.39]

includes normalizing data and separating data for testing and training. In proposed model, ten different kernels were used. Five kernels were selected as the best kernels by trial and error and these kernels are applied to data. There probably is only a few of the kernels that are useful for the problem, and we are not aware of which kernels are useful for our problem so kernel outputs aggregate by applying a coefficient. This combination creates a new secondary space. The output is given to support vector regression to construct a model that predicts values exactly ε accurate, which means the predicted values do not deviate more than ε from the original data. This model predicts values by using a leave one out model. Each kernel has parameters that need to be set to optimum values in order to get the best results. Hence in the proposed model, the kernel parameters and their weights are learned by the Gray Wolf Optimizer. This optimizer has been able to provide appropriate answers to many problems, especially challenging problems and has a superior ability to solve the high-dimension problems. By running program in consecutive iterations and examining the different values of the parameters, the optimizer learns the best of them which prediction error has been reduced, and finally returns their best value. The proposed model is implemented on five standard time series and compared to other method, test based on the RMSE criterion for DJ time series, improved by 1.58 point, Radio time series, improved by 0.178 point, and Sunspot time series, improved by 1.709 point. Finally, we analyzed the results, Statistical evaluation by Wilcoxon Signed-Rank Test where the p value is very low compared to the proposed method and CNN-FCM, AR_ model per scale, Multiresolution AR model and ANN methods, slightly lower for Wavelet-HFCM and ANFIS methods and slightly lower than one for SAE-FCM method and at the end provide a relation to find the window size in the model by obtaining the average of peak differences, valley differences, and consecutive peak, and valley differences for the actual values of the training data in exchange for their sequence number in time series.

Keywords- Time series prediction, Support vector regression, Ensemble kernel model, Optimization.

۱-مقدمه

یک سری‌زمانی مجموعه‌ای از نقاط داده است که در فواصل زمانی خاصی نمونه برداری می‌شوند. پیش‌بینی سری‌زمانی فرایندی است که با استفاده از آن مقادیر آینده سامانه براساس اطلاعات به‌دست آمده از نقاط داده‌های گذشته و فعلی پیش‌بینی می‌شوند. مدل‌های پیش‌بینی سری‌زمانی به‌طور گسترده در حوزه‌های مختلف مهندسی، اقتصادی و... مورد استفاده قرار می‌گیرند [1].

پیش‌بینی پیشرفت آینده براساس دانش گذشته سرفصل پیش‌بینی سری‌های زمانی است که کاربردهای گسترده‌ای در زمینه‌های فیزیک، جامعه‌شناسی، پزشکی، مهندسی، امور مالی و سایر موارد دارد [2]. مدل‌های پیش‌بینی تصمیمات تجاری را در سطوح مختلف درون شرکت‌ها شکل می‌دهند. پیش‌بینی شاخص سهام همچنان یک روند چالش‌برانگیز در پیش‌بینی سری‌زمانی مالی است. نوسانات تصادفی در شاخص بورس، پیش‌بینی را دشوار می‌کند. به‌طورکلی، منحنی داده‌های سری‌زمانی دارای یک قسمت خطی و یک بخش غیرخطی است. پیش‌بینی قسمت خطی کار دشواری نیست، اما پیش‌بینی بخش‌های غیرخطی دشوار است. اگرچه مدل‌های مختلف پیش‌بینی غیرخطی در حال استفاده هستند، اما دقت پیش‌بینی آنها فراتر از یک سطح مشخص نیست [3, 4]. در ادامه برخی از روش‌هایی که پژوهش‌گران برای این امر از آن استفاده کرده‌اند، مرور می‌شوند.

در پژوهشی که توسط ***liu*** انجام شده بیان شده است که داده‌های سری‌زمانی به‌طورمعمول غیر ثابت هستند و با گذشت زمان تکامل می‌یابند و عنوان شده است که حتی اگر یادگیری عمیق در رفتار با داده‌های متوالی، مؤثر شناخته شده باشد، پایداری شبکه‌های عصبی عمیق در کنار آمدن با حالت‌های دیده‌نشده در مرحله آموزش نیز مهم است. این مقاله بر اساس یک بلوک شناختی فازی[1] که یادگیری نقشه‌های شناختی فازی[2] مرتبه بالا را در معماری یادگیری عمیق جاسازی می‌کند با این مسأله برخورد می‌کند. و به همین ترتیب، یک شبکه عصبی عمیق تحت عنوان CNN[3]-FCM طراحی می‌کند که شبکه کانولوشن موجود را با FCB ترکیب کرده است. نتایج تجربی این مقاله نشان می‌دهد که عملکرد بسیاری از معماری‌های یادگیری عمیق رایج در هنگام بررسی داده‌های پرت از مجموعه‌آموزش کاهش می‌یابد و FCB نقش مهمی در ارتقا عملکرد CNN-FCM در آزمایش‌های مربوطه دارد. و درنهایت این مقاله نتیجه می‌گیرد که مدل‌سازی سامانه می‌تواند ثبات یادگیری عمیق را در پیش‌بینی سری‌های زمانی ارتقا دهد [5].

در پژوهشی که توسط Javedani انجام شده یک روش ترکیبی مبتنی بر سری‌زمانی فازی و شبکه‌های عصبی پیچشی[4] برای پیش‌بینی بار کوتاه‌مدت[1] ارائه شده

[1] FCB=Fuzzy Cognitive Block
[2] FCM=Fuzzy Cognitive Maps
[3] Convolutional Neural Networks
[4] CNN=Convolutional Neural Networks

[DOI: 10.52547/jsdp.19.1.39]

است. در روش پیشنهادی این مقاله، داده‌های سری‌زمانی چندمتغیره که شامل داده‌های بار ساعتی، سری‌زمانی دمای ساعتی و نسخه فازی سری‌زمانی بار است، به تصاویر چند کاناله تبدیل شده تا به مدل یادگیری عمیق CNN پیشنهادی با معماری مناسب داده شود. با استفاده از تصاویری که از مقادیر متوالی سری‌زمانی چندمتغیره ایجاد شده است، مدل پیشنهادی CNN می‌تواند پارامترهای مهم مربوطه را با روش ضمنی و خودکار، بدون نیاز به تعامل انسانی و دانش تخصصی، به‌تنهایی تعیین و استخراج کند. در این مقاله، نشان داده شده است که استفاده از این روش پیشنهادی از برخی مدل‌های سنتی STLF آسان‌تر است و می‌توان آن را یکی از تفاوت‌های بزرگ بین روش پیشنهادی و برخی از روش‌های پیشرفته STLF دانست [6].

در پژوهشی که توسط hu انجام شده است، یک روش پیش‌بینی سری‌زمانی مبتنی بر شبکه عصبی بازگشتی حافظه طولانی کوتاه‌مدت[2] متفاوت پیشنهاد شده است. در این روش پیشنهادی، با ادغام دروازه فراموشی[3] و دروازه ورودی[4] در یک دروازه به‌روزرسانی[5] و استفاده از لایه سیگموید[6] برای کنترل به‌روزرسانی اطلاعات، ماژول[7] حافظهٔ شبکه عصبی بازگشتی LSTM بهبود داده می‌شود. این مقاله با استفاده از شبکه عصبی بازگشتی LSTM بهبودیافته، یک مدل پیش‌بینی سری‌زمانی را توسعه می‌دهد. نتایج این مقاله، در مقایسه با چندین مدل پیش‌بینی سری‌زمانی معمول، عملکرد بهتری برای پیش‌بینی داده‌های متوالی طولانی دارد [7].

در پژوهشی که توسط yuan انجام شده از نقشه‌های شناختی فازی برای پیش‌بینی سری‌زمانی استفاده شده و بیان شده که بیشتر پژوهش‌های موجود به طراحی یک روش مؤثر برای استخراج سری‌های زمانی ویژه از سری‌های زمانی اصلی اختصاص یافته است که برای ساخت FCM و پیش‌بینی سری‌زمانی استفادهمی‌شود. در این مقاله، یک فریم ورک[8] مبتنی بر نگاشت کرنل[9] و FCMهای مرتبه بالا[10] برای پیش‌بینی سری‌های زمانی با الهام از توابع کرنل و رگرسیون بردار پشتیبان[11] پیشنهاد شده است. که آن را Kernel-HFCM نامیدند. در این مقاله نگاشت کرنل، برای نگاشت سری‌های زمانی یک‌بعدی اصلی به سری‌های زمانی ویژه چندبعدی طراحی شده و سپس الگوریتم انتخاب ویژگی برای انتخاب KFTS[12] از سری‌های زمانی ویژه چندبعدی، برای توسعه HFCM پیشنهاد شده است. و درنهایت، از نگاشت کرنل معکوس برای نگاشت سری‌های زمانی ویژه به سری‌های زمانی یک بعدی پیش‌بینی شده استفاده می‌شود [8].

در پژوهشی که توسط wang انجام شده یک تعمیم جدیدی از FCM به نام Deep FCM برای پیش‌بینی سری‌های زمانی چندمتغیره پیشنهاد شده است تا از مزیت FCM در تفسیر، و از مزیت شبکه‌های عصبی عمیق[13] در پیش‌بینی استفاده کند. در این مقاله به‌طور خاص، برای بهبود قدرت پیش‌بینی، Deep FCM از یک شبکه عصبی قویاً متصل برای مدل‌سازی اتصالات (روابط) بین مفاهیم موجود در یک سامانه و از یک شبکه عصبی بازگشتی برای مدل‌سازی عوامل بیرونی ناشناخته که بر پویایی سامانه تأثیر دارند، استفاده شده است. علاوه‌براین، برای افزایش تفسیرپذیری مدل که به‌وسیلهٔ ساختارهای عمیق تعبیه شده، یک روش مبتنی بر مشتق جزئی برای اندازه‌گیری قدرت اتصال بین مفاهیم در Deep FCM پیشنهاد شده است. Deep FCM درواقع یک سرنخ مهم برای ساخت پیش‌بینی‌های قابل تفسیر برای کاربردهای واقعی فراهم کرده است [9].

در پژوهشی که توسط yang انجام شده است، نقشه‌های شناختی فازی با موفقیت برای مدل‌سازی و پیش‌بینی سری‌زمانی ثابت استفاده شده است. در این مقاله، یک مدل پیش‌بینی سری‌زمانی براساس ترکیبی از FCM های مرتبه بالا با تبدیل موجک زائد[14]پیشنهاد شده است تا سری‌های زمانی بزرگ غیرثابت را در مقیاس بزرگ اداره کند که آن را Wavelet-HFCM نامیدند. تبدیل موجک زائد برای تجزیه سری‌های زمانی اصلی غیرثابت در سری‌های زمانی چندمتغیره اعمال می‌شود، سپس از FCM مرتبه بالا برای مدل‌سازی و پیش‌بینی سری‌های زمانی چندمتغیره استفاده می‌شود. در یادگیری FCM با مرتبه بالا به نمایندگی از سری‌زمانی چندمتغیره در مقیاس بزرگ، یک روش یادگیری سریع مرتبه بالا با استفاده از رگرسیون لبه[15] برای کاهش زمان یادگیری طراحی شده است. سرانجام، جمع‌بندی سری‌های زمانی چندمتغیره سری‌زمانی پیش‌بینی‌شده را در هر مرحله از زمان نتیجه می‌دهد [10].

[1] STLF=Short-Term Load Forecasting
[2] LSTM= Long Short-Term Memory
[3] Forget gate
[4] Input gate
[5] Update gate
[6] Sigmoid
[7] module
[8] framework
[9] kernel mapping
[10] HFCM=High-Order FCMs
[11] SVR=Support Vector Regression
[12] KFTS= Key Feature Time Series
[13] deep neural networks
[14] redundant wavelet
[15] Ridge regression

[DOI: 10.52547/jsdp.19.1.39]

در پژوهشی که توسط Qinkun انجام شده است، یک مدل پیش‌بینی سری‌زمانی چند قدم جلوتر بر اساس ترکیب مدل فیلتر بیزی[1] و شبکه عصبی فازی[2] نوع ۲ ارائه شده است. درهمین‌اواخر، مطالعات نشان می‌دهد که مدل FNN نوع ۲ یک استراتژی امیدوارکننده برای پیش‌بینی سری‌زمانی چند قدم جلو است و BFM یک روش پردازش دنباله اطلاعات بازگشتی است، که به‌طور مؤثری برای پیش‌بینی داده‌های سری‌زمانی استفاده شده است. در این مقاله، از BFM مبتنی بر بازگشت به‌منظور بهبود عملکرد مدل پیش‌بینی مستقیم مبتنی بر FNN استفاده شده‌است. مدل ترکیبی به نام BFM2FNN برای پیش‌بینی داده‌های سری‌زمانی چند قدم جلو ساخته شده است. نتایج شبیه‌سازی و مقایسه نشان داده است که مدل ارائه‌شده اثربخشی و استحکام بیشتری دارد [11].

در پژوهشی که توسط *عمرانپور و آزادیان* انجام شده، یک روش محاسباتی پیش‌بینی مبتنی بر سری‌های زمانی فازی با مرتبه بالا ارائه شده است. عملکرد روش بدین صورت است که پس از فازی‌سازی سری‌زمانی و ایجاد روابط منطقی فازی با استفاده از حد پایین بازهٔ عنصر مورد پیش‌بینی و بازهٔ پس از آن و اختلاف حاصل از عناصر متوالی محاسبات خاصی انجام داده و مجموعه‌ای از ویژگی‌ها را به‌دست آورده است؛ سپس با استفاده از تابع بهینه‌سازی ازدحام ذرات بهترین پارامترها انتخاب می‌شوند. در نهایت پیش‌بینی و غیرفازی سازی انجام شده است [12].

در پژوهشی که توسط bergmeir انجام شده است، یکی از روش‌های استاندارد استفاده‌شده برای ارزیابی مدل‌های طبقه‌بندی و رگرسیون، اعتبارسنجی متقابل k تایی[3] بیان شده است. همچنین استفاده از روش‌های یادگیری ماشین برای پیش‌بینی و اعتبارسنجی متقابل برای کنترل یادگیری بیش از حد[4] متداول بیان شده است. در این مقاله بینش‌های نظری در حمایت از این استدلال‌ها، همراه با یک شبیه‌سازی و مثالی از دنیای واقعی ارائه شده است. و به‌طور تجربی نشان داده شده است که اعتبارسنجی k تایی در مقایسه با سایر روش‌های سری‌زمانی خاص مانند اعتبار سنجی متقابل غیروابسته عملکرد مطلوبی دارد [13].

در پژوهشی که توسط xu انجام شده است، برای رده‌ای از سری‌های زمانی غیرخطی که رفتار پویای آن‌ها با حالت سامانه به آرامی تغییر می‌کند، یک مدل SD-AR[5] ارائه شده است. در این مقاله از مجموعه‌ای از شبکه‌های باور عمیق[6] برای ساخت ضرایب عملکردی وابسته به حالت مدل SD-AR استفاده می‌شود و مدل پیشنهادی مدل DBN-AR نامیده می‌شود که ترکیبی از مزیت DBN در تخمین عملکرد و شایستگی مدل SD-AR در توصیف پویایی غیرخطی است. مدل DBN-AR ارائه شده در این مقاله، به‌وسیلهٔ تغییرات سیگنال حالت با گذشت زمان، هدایت می‌شود و بر اساس راه حل کمینه مربعات با کمینه نُرم و رویکرد شبه ماتریس معکوس، مقادیر اولیه مورد نظر DBNها در مرحله قبل از آموزش تعیین می‌شود و در مرحله تنظیم نهایی، سرانجام تمام پارامترهای مدل DBN-AR توسط الگوریتم انتشار به عقب[7] که برای تنظیم نهایی مدل DBN-AR طراحی شده است، تنظیم می‌شوند. نتایج مقاله نشان داده که مدل DBN-AR در دقت پیش‌بینی از برخی مدل‌ها یا روش‌های موجود برتر است [14].

در روش‌های گفته‌شده، از کرنل استفاده نشده است و آن‌ها از داده‌ها در همین فضا استفاده کرده‌اند و تغییر فضا ندادند و نتایج مناسبی نگرفتند؛ ولی در مدل پیشنهادی که در این مقاله ارائه شده است، فضا را عوض می‌کنیم.

همان‌طور که پیش‌تر گفته شد، هدف از این مقاله پیش‌بینی سری‌زمانی است. هدف اصلی در به‌کارگیری مدل‌های مختلف برای پیش‌بینی سری‌زمانی این است که بتوانیم پیش‌بینی را با بیشترین دقت انجام دهیم. در این مقاله برای رسیدن به این هدف از مدل رگرسیون بردار پشتیبان مبتنی بر ترکیب کرنل‌ها استفاده شده و سعی شده است با نرمال‌سازی داده‌های ورودی و انتقال آن‌ها به فضای دیگر با استفاده از ترکیب کرنل‌های متعدد به دقت بهتری در پیش‌بینی مقادیر برسیم.

در بسیاری از مسائل بهینه‌سازی، الگوریتم‌های فراابتکاری[8] از عملگرهای تصادفی استفاده می‌کنند که برای آن‌ها جست‌وجوی بهتری از رویکردهای قطعی[9] ایجاد می‌کند [15]. یک الگوریتم قطعی[10] [16, 17, 18] با

[1] BFM= Bayesian filtering model
[2] FNN= Fuzzy Neural Network
[3] k-fold cross validation
[4] overfitting

[5] State-Dependent Auto-Regressive
[6] DBN= Deep Belief Networks
[7] BP= back propagation
[8] meta-heuristic
[9] deterministic
[10] deterministic algorithm

[DOI: 10.52547/jsdp.19.1.39]

اطمینان پاسخ مشابهی را برای یک موضوع داده شده با یک نقطه شروع اولیه می‌دهد. که به هرصورت ممکن است، با این رفتار در بهینه محلی گرفتار شود، که می‌تواند به‌عنوان یک اشکال برای الگوریتم‌های بهینه‌سازی قطعی در نظر گرفته شود [19]. گیرافتادن در بهینهٔ محلی به گرفتارشدن یک الگوریتم در راه‌حل‌های محلی و در نتیجهٔ آن به عدم موفقیت در یافتن بهینه واقعی سراسری منجر می‌شود. از آنجا که مسائل واقعی تعداد گسترده‌ای راه‌حل‌های محلی دارند، الگوریتم‌های قطعی کیفیت تعیین‌کننده خود را در یافتن بهینه سراسری از دست می‌دهند [20].

۲- ادبیات پژوهش

در ادبیات ماشین بردار پشتیبان[1]، هنگامی که الگوریتم SVM برای مسائل طبقه‌بندی استفاده می‌شود، به آن طبقه‌بندی بردار پشتیبان[2] گفته می‌شود و هنگامی که از آن برای مسائل رگرسیون استفاده می‌شود، به آن "رگرسیون بردار پشتیبان"گفته می‌شود [21].

SVM یک ماشین جامع است که به‌خوبی در شناخت الگوها و تخمین رگرسیون کاربرد دارد. علاوه‌بر این، SVM با استفاده از ترفندهای کرنل از توانایی تقریب به‌شدت غیرخطی برخوردار است. رگرسیون بردار پشتیبان روشی برای حل مسأله تخمین رگرسیون با استفاده از SVM است. تئوری SVR به شکل زیر بیان می‌شود:

با فرض این‌که داده‌های آموزشی به‌صورت زیر باشد، X فضای الگوهای ورودی را نشان می‌دهد:

$$\{(x_1,y_1),(x_2,y_2),(x_3,y_3),\ldots,(x_i,y_i)\} \subseteq X\times R \qquad (۱)$$

در رگرسیون هدف پیداکردن تابع f(x) است که حداکثر اختلاف خروجی f(x) یعنی y_i به‌ازای x_i، از مقدار y_i واقعی برای تمام داده‌های آموزشی برابر با ε[3] و تا حد ممکن f(x) مانند خطی بدون انحنا باشد:

$$f(x) =< w,x > +b \; with \; w \in x, b \in R \qquad (۲)$$

که $< . , . >$ ضرب داخلی در فضای X را نشان می‌دهد، b انحراف را نشان می‌دهد که باید کم باشد و عدم انحنا در f(x) به معنی w کوچک است. یکی از راه‌ها برای تضمین این امر کمینه‌کردن نرم w است که این موضوع به‌صورت بهینه‌سازی یک مسأله **محدب**[4] قابل حل است:

$$\begin{aligned} &\text{Minimize} \quad \frac{1}{2}\|w\|^2 \\ &\text{Subject to} \quad y_i - < w,x_i > -b \le \varepsilon \end{aligned}$$

$$< w,x_i > +b - y_i \le \varepsilon \qquad (۳)$$

ε یک ثابت مثبت است که اگر تفاوت بین مقدار پیش‌بینی $f(x_i)$ و مقدار y_i واقعی کمتر از ε باشد، نادیده گرفته می‌شود؛ یعنی خطا صفر است. اگر اختلاف بیشتر از ε باشد خطا به مقدار $|f(x_i)-y_i|-\varepsilon$ است. ε خطای مورد انتظار بین مقدار پیش‌بینی‌شده و مقدار واقعی را بیان می‌کند. هرچه ε کمتر باشد، یعنی خطای کمتر مطلوب است و دقت پیش‌بینی نیز بالاتر می‌رود [22]. در بیش‌تر اوقات، انحراف بعضی داده‌ها از مقدار واقعی، بیشتر از ε است و همین امر باعث پیدایش تعمیم جدیدی از تئوری SV شد که با ورود متغیرهای جدید مانند ξ و با استفاده از فرمول‌های جدید، این مسأله را رفع می‌کند:

$$\begin{aligned} &\text{Minimize} \quad \frac{1}{2}\|w\|^2 + c\sum_{i=1}^{N}\xi + \xi^* \\ &\text{Subject to} \quad y_i - < w,x_i > -b \le \varepsilon + \xi_i \\ &\qquad < w,x_i > +b - y_i \le \varepsilon + \xi^* \end{aligned}$$

$$\xi_i, \xi_i^* \ge 0 \quad (i = 1,2,..,n) \qquad (۴)$$

در اینجا N برابر تعداد داده‌های ورودی است و c یک ثابت است که میزان توجه به خطاهای تخمین را مشخص می‌کند. یک c بزرگ، توجه بیشتری را به خطاها اختصاص می‌دهد تا رگرسیون طوری آموزش داده شود که خطاها را با تعمیم کمتر، به کمینه برساند؛ درحالی‌که یک c کوچک، توجه کمتری را به خطاها اختصاص می‌دهد. اگر c به بی نهایت برسد، SVR اجازه نمی‌دهد که خطایی رخ دهد و پیچیدگی مدل زیاد می‌شود؛ درحالی‌که وقتی c به صفر برود، نتیجه می‌تواند مقدار زیادی از خطاها را تحمل کند و مدل پیچیدگی کمتری داشته باشد [23].

با معرفی تابع لاگرانژ[5] و تابع کرنل، فرمول (۴) می‌تواند به فرمول زیر تبدیل شود:

$$\begin{aligned} &\text{Maximize} \quad \sum_{i=1}^{N}(\alpha_i^* - \alpha_i)y_i - \varepsilon\sum_{i=1}^{N}(\alpha_i^* + \alpha_i) \\ &\qquad -\frac{1}{2}\sum_{i,j=1}^{N}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)\,\text{kernel}(x_i,x_j) \\ &\text{Subject to} \quad 0 \le \alpha_i, \alpha_i^* \le c \end{aligned}$$

$$\sum_{i=1}^{N}(\alpha_i^* - \alpha_i) = 0 \quad (i = 1,2,\ldots n) \qquad (۵)$$

بنابراین تابع پیش‌بینی نهایی به‌صورت زیر است:

$$f(x,\alpha_i,\alpha_i^*) = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)\text{kernel}(x_i,x) + b \qquad (۶)$$

که α_i و α_i^* ضرایب لاگرانژ هستند.

[1] SVM=Support Vector Machine
[2] SVC=Support Vector Classification
[3] epsilon
[4] convex
[5] Lagrange

[DOI: 10.52547/jsdp.19.1.39]

در بیش‌تر داده‌های دنیای واقعی داده‌های موجود در صورت مسألهٔ دسته‌بندی به‌صورت خطی جدایی‌پذیر نیستند. همچنین در صورت مسأله‌های تخمین تابع، داده‌ها روی خط قرار نمی‌گیرند. در این صورت ما نیاز داریم داده‌هایمان را وارد فضای جدید کنیم تا بتوانیم با استفاده از یک ابر صفحه مرز موردنظر را تعیین کنیم. این کار به‌وسیلهٔ توابع کرنل انجام می‌شود. درواقع با استفاده از کرنل، داده‌های ورودی در یک فضا با ابعاد بالاتر قرار می‌گیرند و مسائل غیرخطی را به مسائل خطی یا به‌طورتقریبی خطی تبدیل می‌کنند [24]. انتخاب و ساخت توابع کرنل یک موضوعی کلیدی است که بر عملکرد SVR تأثیر می‌گذارد، و یک رویکرد مهم برای گسترش SVR از میدان خطی به میدان غیرخطی را فراهم می‌کند.

۳- مدل پیشنهادی

در این پژوهش از ده کرنل مختلف بهره برده شد که با سعی و خطا پنج کرنل از آن‌ها به‌عنوان بهترین کرنل‌ها انتخاب شده است. توابع کرنلی که در این پژوهش استفاده شده‌اند به شرح زیر است:

$$\text{kernel}_{\text{polynomial}}(u,v) = (u * v' + 1)^{\text{round}(p_1)} \quad (۷)$$

$$\text{kernel}_{\text{RBF}}(u,v) = e^{-(u-v)*(u-v)'/(2*{p_1}^2)} \quad (۸)$$

$$\text{kernel}_{\text{ERBF}}(u,v) = e^{-\sqrt{(u-v)*(u-v)'}/(2*{p_1}^2)} \quad (۹)$$

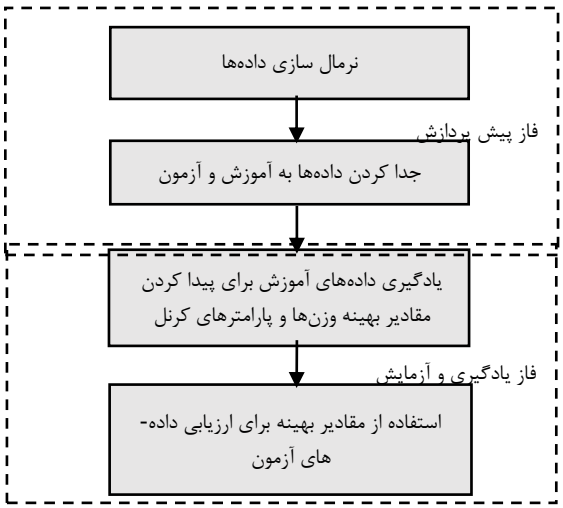
$$\text{kernel}_{\text{sigmoid}}(u,v) = \tanh\left(\frac{p_1*u*v'}{\text{length}(u)} + p_2\right) \quad (۱۰)$$

$$kernel_{gaussian}(u,v) = e^{-(u-v)^2/2*|p_1|^2} \quad (۱۱)$$

که p_1 و p_2 پارامترهای ورودی کرنل هستند که باید مقادیر بهینه برای آن‌ها توسط بهینه‌ساز پیدا شود. بهینه‌ساز با توجه به بازه مقادیر ممکن برای این دو پارامتر، در تکرارهای متوالی الگوریتم، مقادیر بهینه به‌ازای آن‌ها که منجر به بهترشدن خروجی می‌شود را پیدا می‌کند. u و v سطرهایی از ماتریس داده‌های اصلی هستند (دو بردار) که کرنل با استفاده از ضرب داخلی شباهت بین هر دوی آن‌ها را به‌دست می‌آورد. همچنین تابع round که در رابطه (۷) استفاده شده، ورودی خود را به سمت نزدیکترین عدد صحیح گرد می‌کند، نماد e نشان‌دهندهٔ عدد اویلر یا عدد نپر است، تابع tanh که در رابطه (۱۰) استفاده شده، تابع مثلثاتی تانژانت هایپربولیک[1] است و تابع length، طول بردار ورودی(تعداد کل عناصر بردار) را محاسبه می‌کند.

توابع کرنل بالا دارای صفات مربوط به خود هستند و تأثیرات متفاوتی بر عملکرد SVR دارند. دو گروه مختلف از کرنل وجود دارد: کرنل سراسری و کرنل محلی. کرنل‌های سراسری توانایی برون‌یابی قوی‌تری دارند و کرنل‌های محلی از توانایی درون‌یابی قوی‌تری برخوردار هستند. کرنل‌های استاندارد که هم‌زمان می‌توانند برون‌یابی و درون‌یابی کنند، تا حدی نادرست هستند [25]. برای مثال عملکرد کرنل RBF یک عملکرد کرنل محلی با توانایی یادگیری قوی‌تراست. با این حال ، عملکرد کرنل چندجمله‌ای یک عملکرد کرنل سراسری است که از توانایی یادگیری ضعیف‌تر برخوردار است [26]. ما با کنار هم قراردادن این کرنل‌ها از ویژگی‌های همهٔ آن‌ها استفاده می‌کنیم.

مدلی که در این مقاله ارائه شده براساس ترکیب کرنل‌ها است. هرکدام از این کرنل‌ها پارامترهایی دارند، همچنین در مدل ارائه شده به کرنل‌ها وزن هم داده شده است. پیداکردن مقادیر بهینه متغیرهای گفته شده بر عهدهٔ بهینه‌ساز است. در شکل (۱) فازهای مدل پیشنهادی ارائه شده است.



(شکل-۱): فازهای مدل پیشنهادی
(Figure-1): proposed model phases

۳-۱- فاز پیش‌پردازش

این فاز شامل مراحل نرمال‌سازی داده‌ها و جداسازی داده‌ها به آزمایش و آموزش است. در ابتدا داده‌های ورودی نرمال می‌شوند. یعنی اعداد را به بازهٔ ۱- و ۱+ منتقل

[1] Hyperbolic Tangent

[DOI: 10.52547/jsdp.19.1.39]

می‌کنیم این کار باعث می‌شود بزرگ یا کوچک‌بودن مقادیر داده‌ها، مدل را تحت تأثیر قرار ندهد، که باعث بهبود عملکرد و ثبات آموزش مدل می‌شود؛ سپس داده‌ها را به دو قسمت آموزش[1] و آزمون[2] تقسیم می‌کنیم. از مجموعه‌داده‌های آموزش برای یادگیری مقدار مناسب پارامترهای رابطه (۱۲) استفاده و از مجموعه‌داده‌های آزمون برای ارزیابی صحت پیش‌بینی استفاده می‌شود.

داده‌های آموزش را به ترکیب وزن‌داری از کرنل‌ها می‌دهیم تا داده‌ها را به فضای جدیدی منتقل کند:

$$x_{new} = \sum_{i=1}^{5} b_i * c_i * kernel_i(x, p_i) \qquad (۱۲)$$

همان‌طور که در رابطه (۱۲) مشاهده می‌کنید، ما از ترکیب ۵ کرنل که در رابطه‌های (۷) تا (۱۱) بیان شد استفاده می‌کنیم، b_i ضریبی است که مقدارش {0,1} و تعیین‌کننده این است که کرنل مورد نظر در مجموعه کرنل‌ها باشد یا نباشد. همچنین c_i وزنی است که به هر کرنل داده می‌شود و بازه [-1,1] را دربر می‌گیرد. البته c_i از لحاظ تئوری می‌تواند عملکرد b_i را با گرفتن مقدار صفر یا یک پوشش دهد. در ابتدا فقط c_i در مدل پیشنهادی قرار گرفت و آزمایش انجام شد، چون نتایج مناسب نبودند b_i نیز به‌عنوان پارامتر اضافه شد. دلیل این امر این است که نبودن یک کرنل تأثیر بسیاری در نتایج داشت و گرفتن مقدار صفر برای c_i احتمال کمی دارد.

x مجموعه‌داده‌های اصلی (نرمال شده) و p_i پارامتر ورودی کرنل است. مقادیر b_i ، c_i و p_i توسط بهینه‌ساز مشخص می‌شود. بهترین مقادیر متغیرها را که به‌ازای آنها بهترین پیش‌بینی برای داده‌های آموزش انجام شده است، به‌دست می‌آوریم و از آن‌ها برای پیش‌بینی داده‌های آزمون استفاده می‌کنیم.

در تابع شایستگی همهٔ ورودی‌های برنامه دریافت می‌شود و مراحل آماده‌سازی داده‌ها بر روی آن انجام، سپس با استفاده از رابطه (۱۲)، x_{new} ساخته می‌شود و به رگرسیون بردار پشتیبان داده می‌شود تا مدلی بسازد که مقادیری که به‌وسیلهٔ این مدل پیش‌بینی می‌شوند با دقت ε درست باشد، یعنی مقادیر پیش‌بینی‌شده با مقدار بیشتر از ε از داده‌های اصلی منحرف نشود. خروجی تابع شایستگی مقدار RMSE[3] داده‌های آموزش است. این عمل در تکرارهای متوالی انجام می‌شود تا بهترین مقادیر متغیرها برای ساخت مدل با کمترین خطا پیدا شود.

این مدل با استفاده از مدل اعتبارسنجی یک‌طرفه[4] به پیش بینی مقادیر متغیرها پیدا شد و کار بهینه‌ساز به اتمام رسید مدل مورد نظر برروی داده‌های آزمون اعمال می‌شود تا ارزیابی درستی مدل پیش‌بینی انجام شود.

۲-۳- بهینه‌سازی پارامترها

همان‌طور که گفته شد ما نیاز داریم بهترین مقادیر ممکن برای هرکدام از پارامترها و دیگر ورودی‌های مسأله را پیدا کنیم تا پیش‌بینی دقیق‌تری داشته باشیم. بهینه‌ساز این کار را برای ما انجام می‌دهد.در اینجا ما از بهینه‌ساز گرگ خاکستری[5] استفاده می‌کنیم [27]. به این دلیل از این بهینه ساز استفاده شده که توانسته در بسیاری از مسائل، به‌خصوص صورت مسأله‌های پیچیده، پاسخ مناسبی ارائه دهد. و توانایی بالایی در حل مسائل با ابعاد بالا دارد [28]. البته بهینه‌سازهای دیگری هم هستند که می‌توانستیم از آن‌ها استفاده کنیم، ولی مشاهده شد که استفاده از آن‌ها خیلی تأثیری روی نتیجه ندارد.

در این بهینه‌ساز شکار و سلسله‌مراتب اجتماعی گرگ‌های خاکستری از نظر ریاضی به‌منظور انجام بهینه‌سازی مدل می‌شوند. این بهینه‌ساز با استفاده از تقسیم سه‌سطحی **α**، **β** و **δ** که میان گرگ‌ها انجام می‌دهد می‌تواند در بیش‌تر مسائل از گیرافتادن در کمینه محلی اجتناب کند و به کمینه سراسری همگرا شود.

در شکل (۲) شبه‌کد مربوط به الگوریتم گرگ خاکستری ارائه شده است. در این شبه‌کد t نشان‌دهندهٔ تکرار فعلی، A و C بردارهای ضرایب و X بردار موقعیت گرگ خاکستری است. به عبارت دیگر A مقدار تصادفی در بازه [-2a,2a] که در آن مؤلفهٔ a از دو تا صفر در طول تکرارها برای تأکید بر شناسایی و بهره‌برداری به‌طور خطی کاهش می‌یابد. مؤلفهٔ C در رابطه با گیرافتادن نقاط در بهینه محلی به‌خصوص در تکرارهای نهایی بسیار مفید است. بردار C را می‌توان به‌صورت اثر موانع برای رسیدن به شکار در طبیعت در نظر گرفت. بسته به موقعیت گرگ، بردار C یک وزن را به شکار داده و دسترسی گرگ‌ها را به شکار سخت‌تر می‌کند یا برعکس. مناسب‌ترین راه حل در α قرار دارد.

موقعیت عامل‌های جست‌وجو براساس روابط زیر به‌روز می‌شوند:

$$\vec{D}_\alpha = \left|\vec{C}_1 * \vec{X}_\alpha - \vec{X}\right|, \vec{D}_\beta = \left|\vec{C}_2 * \vec{X}_\beta - \vec{X}\right|,$$
$$\vec{D}_\delta = \left|\vec{C}_3 * \vec{X}_\delta - \vec{X}\right| \qquad (۱۳)$$

[1] train
[2] test
[3] Root Mean Square Error
[4] Leave one out
[5] Gray Wolf Optimizer

[DOI: 10.52547/jsdp.19.1.39]

$$\vec{X}_1 = \vec{X}_\alpha - \vec{A}_1 * \vec{D}_\alpha, \vec{X}_2 = \vec{X}_\beta - \vec{A}_2 * \vec{D}_\beta, \quad \vec{X}_3 = \vec{X}_\delta - \vec{A}_3 * \vec{D}_\delta \qquad (۱۴)$$

$$\vec{X}_{(t+1)} = \frac{\vec{X}_1+\vec{X}_2+\vec{X}_3}{3} \qquad (۱۵)$$

بهینه‌ساز با اجرای برنامه در تکرارهای متوالی و بررسی مقادیر مختلف پارامترها بهترین آن‌ها را که به‌ازای آن خطای پیش‌بینی کمتر شده را یاد می‌گیرد و درنهایت بهترین مقدار آن‌هارا بر می‌گرداند.

۱. ایجاد جمعیت اولیه گرگ‌ها X_i **(i=1,2,…,n)**
۲. ایجاد متغیرهای **a**، **A** و **C**
۳. محاسبه‌ی تابع شایستگی هر عامل جست‌وجو
۴. X_α = بهترین عامل جست‌وجو
۵. X_β = دومین بهترین عامل جست‌وجو
۶. X_δ = سومین بهترین عامل جست‌وجو
۷. تا زمانی که **t** کوچکتر از حداکثر تعداد تکرار است ادامه بده
- به‌ازای هر عامل جست‌وجو
  - به روز کردن مکان عامل جست‌وجوی فعلی با استفاده از رابطهٔ (۱۵)
- به روز کردن مقادیر a، A و C
- محاسبهٔ تابع شایستگی برای تمام عامل‌های جست‌وجو
- به روز کردن مقادیر X_α، X_β و X_δ
- افزایش مقدار t به میزان یک واحد

۸. X_α را برگردان

(شکل-۲): شبه کد الگوریتم گرگ خاکستری [27]
(Figure-2): pseudo code of gray wolf algorithm [27]

بازنمایی کروموزوم (گرگ) عبارت است از مقادیر b، c و p برای هر کرنل (رابطه ۱۲). یکی از کرنل‌ها نیز دو مقدار p دارد (رابطه ۱۰). در مجموع، فضای جست‌وجو دارای شانزده بُعد است که شامل مقادیر b,c و p برای پنج تابع کرنل و همچنین یک p اضافه‌تر برای $kernel_{sigmoid}$ (رابطه ۱۰) ماست.

در شکل (۳) شبه‌کد مر بوط به تابع شایستگی ارائه شده است.

ورودی‌ها: آرایه کروموزوم(گرگ)، داده‌های آموزش سری زمانی
خروجی: مقدار RMSE داده‌های آموزش برای مقادیر این کروموزوم
۱. انتقال داده‌های ورودی (آموزش) به بازهٔ ۱- و ۱+
۲. استفاده از رابطه (۱۲) برای انتقال داده‌ها به فضای ثانویه
۳. استفاده از رگرسیون بردار پشتیبان برای ساخت مدل پیش‌بینی بر اساس داده‌های فضای ثانویه
۴. استفاده از مدل اعتبارسنجی یک طرفه بر اساس مدل ساخته شده SVR برای پیش‌بینی مقادیر
۵. بازگرداندن مقادیر اصلی و مقادیر پیش‌بینی شده به بازهٔ اصلی
۶. محاسبه مقدار RMSE مقادیر و گزارش آن به‌عنوان خروجی تابع شایستگی

(شکل-۳): شبه کد تابع شایستگی
(Figure-3): pseudo code of fitness function

با توجه به این‌که زمان اجرای الگوریتم SVR برروی n داده $O(n^3)$ است [29]، مرتبهٔ زمانی مرحلهٔ آموزش از الگوریتم برابر $O(iter*pop*(n*k+n^3))$ است که در آن iter برابر تعداد تکرار تابع شایستگی، pop برابر اندازه جمعیت بهینه‌ساز و k برابر زمان لازم برای ایجاد فضای ثانویه به‌وسیلهٔ کرنل‌ها برای یک سطر از داده است. و در مرحلهٔ آزمون با توجه به این‌که ابر صفحه مورد نظر ما در مرحلهٔ آموزش ساخته شده مرتبه زمانی O(k+1) است.

۴- ارزیابی نتایج

در این بخش به ارزیابی نتایج پژوهش‌ها می‌پردازیم و نتایج را با نتایج حاصل از پژوهش‌های گذشته مقایسه می‌کنیم.

۴-۱- سری‌های زمانی مورد استفاده

برای نشان‌دادن اثربخشی مدل پیشنهادی از پنج سری‌زمانی معیار با خصوصیات مختلف استفاده می‌کنیم. همهٔ سری‌زمانی‌های مورد استفاده در این مقاله یک بعدی و از نوع غیر ایستا[1] هستند. سری‌زمانی اول مربوط به ۱۹۲ مشاهده و ثبت CO2 (ppm) در Mauna Loa از ۱۹۶۵ تا ۱۹۸۰ است. سری‌زمانی دوم مسدودشدن ماهانه شاخص صنعتی Dow Jones را از آگوست ۱۹۶۸ تا آگوست ۱۹۸۱ ثبت می‌کند، که ۲۹۱ مشاهده درکل است. سری‌زمانی سوم شامل ۲۴۰ مشاهده است که بالاترین فرکانس رادیویی را نشان می‌دهد که می‌تواند برای پخش در واشنگتن، در بازه زمانی مه ۱۹۳۴ تا آوریل ۱۹۵۴ استفاده شود. سری‌زمانی چهارم سری‌زمانی sunspotکه تعداد سالانه لکه‌های خورشید را ثبت می‌کند، که از ۲۸۸ مشاهده از ۱۷۰۰ تا ۱۹۸۷ تشکیل شده و اغلب توسط پژوهش‌گران دیگر مورد استفاده قرار می‌گرفت و آخرین سری‌زمانی تولید ماهیانه شیر را از ژانویه ۱۹۶۲ تا دسامبر ۱۹۷۵ در پوند ثبت می‌کند، که شامل ۱۶۸ مشاهده است [10].

داده‌های سری‌زمانی نرمال شده به‌ترتیب به دو زیرمجموعه تقسیم می‌شوند: مجموعه‌داده‌های آموزش و مجموعه داده‌های آزمون. جدول (۱) اندازهٔ هر زیرمجموعه از هر سری‌زمانی که در مدل پیش‌بینی استفاده شده است را مشخص می‌کند:

[1] non stationary

[DOI: 10.52547/jsdp.19.1.39]

(جدول-۱): اندازهٔ هریک از زیرمجموعه‌ها

(Table-1): size of each subset

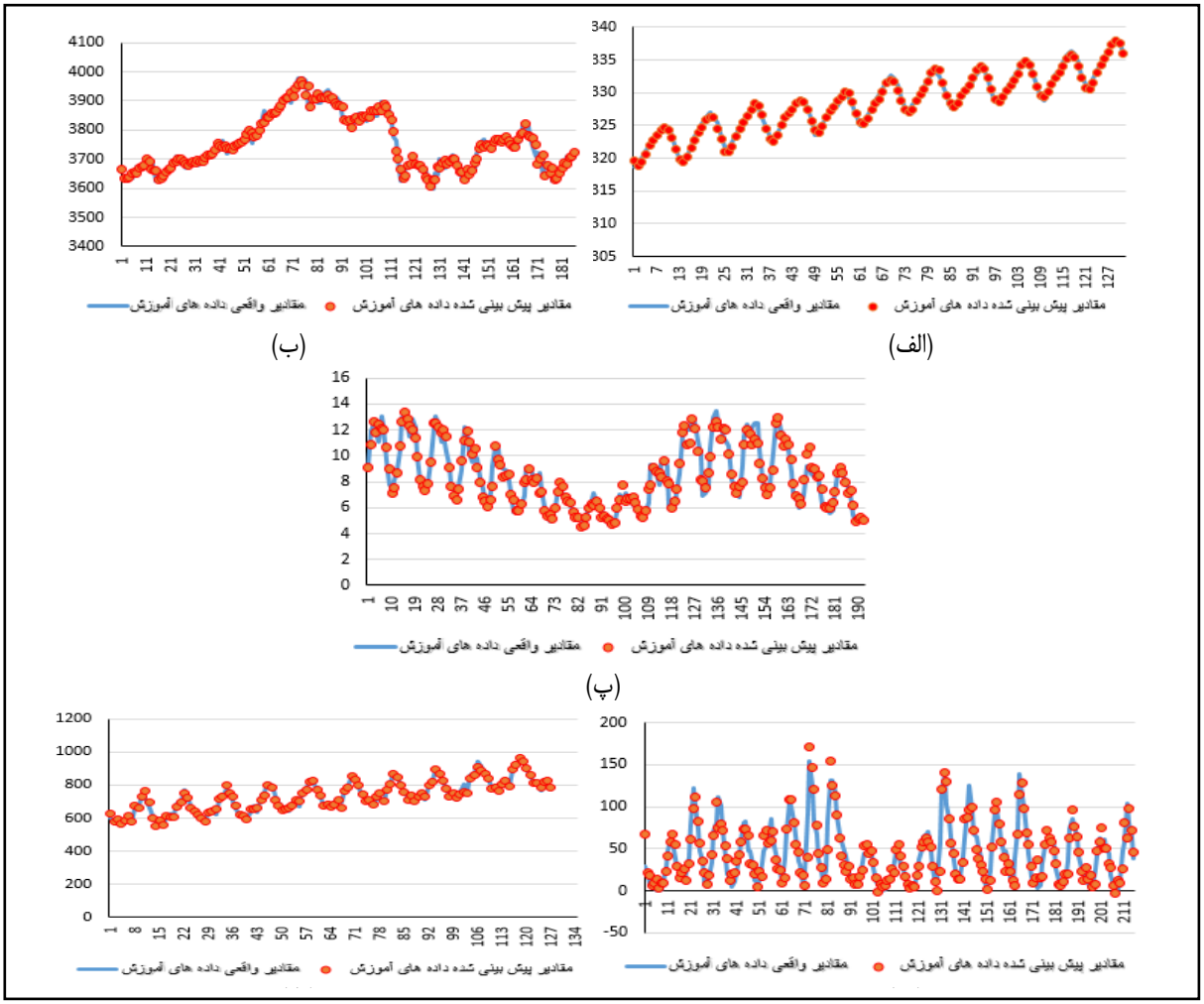
		Training	Test
Data set 1	CO2 (ppm) at Mauna Loa	163	29
Data set 2	Monthly closings of the Dow- Jones industrial index	218	73
Data set 3	Monthly critical radio frequencies	220	20
Data set 4	Sunspot time series	221	67
Data set 5	Monthly milk production per cow	134	34

برای مقایسه نتایج و محاسبه کارایی روش مورد نظر از معیار ارزیابی RMSE استفاده می‌کنیم، که بیان‌گر میزان اختلاف بین مقدار واقعی و مقدار پیش‌بینی شده است:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \mathrm{ypredict}_i)^2}{n}} \qquad (۱۶)$$

که در آن n برابر است با اندازهٔ سری‌زمانی، y_i برابر با دادهٔ واقعی و $\mathrm{ypredict}_i$ برابر با دادهٔ پیش‌بینی شده است. گفتنی است برای محاسبهٔ مقدار RMSE داده‌ها را به بازه اصلی‌شان برمی‌گردانیم.

۴-۲- معیار ارزیابی



(شکل-۴): مقایسه مقادیر واقعی و پیش‌بینی شده برای داده‌های آموزش(اعتبارسنجی) به‌ترتیب برای data set های ۱ تا ۵

(Figure-4): Comparison of actual and predicted values for training (validation) data for data sets 1 to 5, respectively

۴-۳- آزمایش‌ها و نتایج نهایی

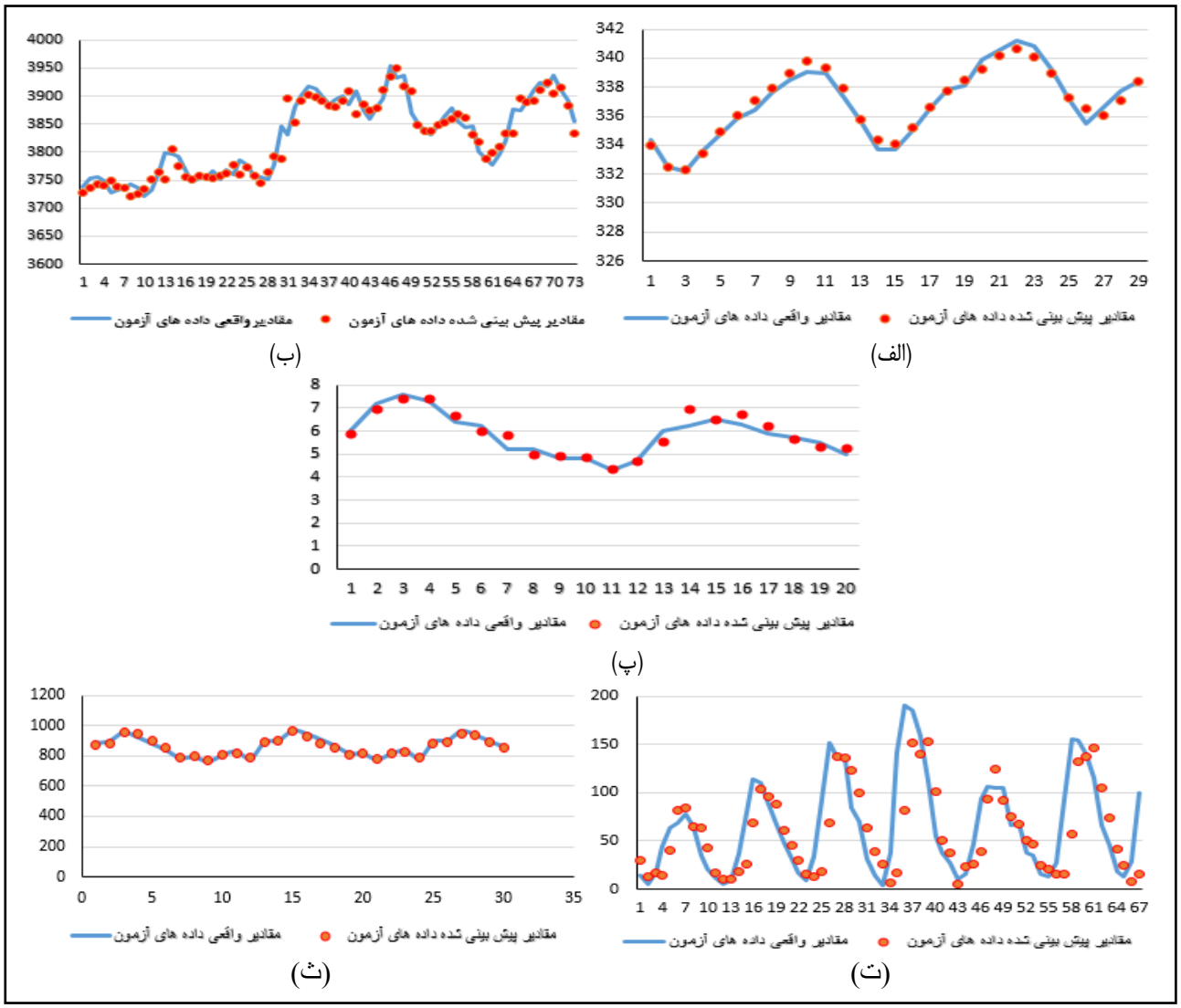
همان‌طور که گفته شد، ما داده‌ها را به دو قسمت آموزش و آزمون تقسیم کردیم که داده‌های آموزش توسط مدل اعتبارسنجی یک‌طرفه به دو قسمت آموزش واعتبارسنجی[1] تقسیم می‌شود. برای انجام آزمایش‌ها از نرم‌افزار متلب استفاده شده است.

[1] validation

[DOI: 10.52547/jsdp.19.1.39]

در شکل‌های (۴) و (۵) مقایسه مقادیر واقعی و پیش‌بینی‌شده برای داده‌های اعتبارسنجی و آزمون به‌ازای بهترین نتیجه آن در اندازه پنجره‌های مختلف برای هر سری‌زمانی نمایش داده شده است.

در جدول (۲) مقدار RMSE مربوط به هر یک از اندازه پنجره‌های داده‌های آموزش و آزمون برای هر سری‌زمانی نمایش داده شده است.



(شکل-۵): مقایسه مقادیر واقعی و پیش‌بینی شده برای داده‌های آزمون به‌ترتیب برای data set های ۱ تا ۵

(Figure-5): Comparison of actual and predicted values for testing data for data sets 1 to 5, respectively

(جدول-۲): مقدار RMSE داده‌های آموزش و آزمون در هر اندازه پنجره و رتبه آن از لحاظ دقت پیش‌بینی

(Table-2): RMSE value of training and test data in each window size and its rank in terms of predictive accuracy

	Window size	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Data set 1	RMSE داده‌های آموزش	0.788	0.724	0.655	0.632	0.625	0.610	0.528	0.445	0.429	0.419	0.413	0.413	0.425	0.427	0.417
	Rank	31	30	29	28	27	26	25	24	23	19	16	15	20	21	17
	RMSE داده‌های آزمون	1.039	0.990	0.815	0.586	0.647	0.593	0.609	0.602	0.527	0.549	0.560	0.548	0.526	0.526	0.574
	Rank	31	30	29	20	27	21	25	22	9	15	18	14	8	7	19

[DOI: 10.52547/jsdp.19.1.39]

Data set 2	RMSE داده‌های آموزش	22.415	22.553	23.173	23.033	22.980	22.916	22.491	23.437	23.601	23.712	22.915	23.543	23.534	23.758	23.441
	Rank	16	18	24	23	21	20	17	25	29	30	19	28	27	31	26
	RMSE داده‌های آزمون	23.086	23.506	23.507	24.006	25.730	25.505	24.684	25.145	24.645	25.725	27.727	29.672	25.674	24.558	24.310
	Rank	16	17	18	19	29	26	23	25	22	28	30	31	27	21	20
Data set 3	RMSE داده‌های آموزش	0.930	0.913	0.866	0.797	0.759	0.776	0.767	0.794	0.785	0.763	0.750	0.779	0.770	0.768	0.775
	Rank	31	30	29	26	14	20	16	25	23	15	12	21	18	17	19
	RMSE داده‌های آزمون	0.715	0.587	0.516	0.498	0.509	0.486	0.462	0.444	0.716	0.386	0.336	0.391	0.342	0.386	0.342
	Rank	30	29	28	26	27	25	24	23	31	18	4	20	6	17	7
Data set 4	RMSE داده‌های آموزش	13.420	12.982	13.234	12.836	12.624	12.362	12.483	12.685	12.698	13.395	13.591	13.388	13.561	13.193	13.366
	Rank	17	7	13	6	3	1	2	4	5	16	19	15	18	11	14
Data set 5	RMSE داده‌های آزمون	17.874	16.658	16.178	16.300	**15.681**	16.340	17.192	17.181	20.186	21.150	20.110	23.329	23.747	24.183	22.792
	Rank	8	5	2	3	**1**	4	7	6	10	11	9	18	21	24	17
	RMSE داده‌های آموزش	44.759	37.452	41.427	26.041	20.070	15.946	14.680	14.505	12.698	11.310	11.991	11.667	11.461	11.565	11.264
	Rank	31	29	30	28	27	26	25	24	23	2	19	10	5	7	1
	RMSE داده‌های آزمون	57.082	13.525	14.787	14.279	13.729	**12.700**	13.875	16.967	20.186	16.032	15.390	15.330	15.361	15.237	16.609
	Rank	31	7	20	17	12	**1**	13	29	30	26	24	22	23	21	27

(جدول-۲): (ادامه)

(Table-2): (continue)

	Window size	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
Dataset 1	RMSE داده های آموزش	0.428	0.418	0.411	0.410	0.409	0.404	0.404	0.397	0.393	0.383	0.390	0.392	0.372	0.384	0.384	0.377
	Rank	22	18	14	13	12	11	10	9	8	3	6	7	1	5	4	2
	RMS	0.5	0.6	0.6	0.6	0.6	0.5	0.4	0.4	0.5	0.4	0.5	0.5	0.5	0.5	0.4	**0.45**

[DOI: 10.52547/jsdp.19.1.39]

	داده E های آزمون	53	63	08	04	24	37	85	82	41	76	32	36	07	57	75	**9**
	Rank	16	28	24	23	26	12	5	4	13	3	10	11	6	17	2	**1**
Data set 2	RMS داده E های آموزش	23.010	22.275	21.610	20.829	20.687	20.209	20.060	19.933	20.041	20.180	20.523	19.973	20.180	19.754	20.035	19.996
	Rank	22	15	14	13	12	10	7	2	6	9	11	3	8	1	5	4
	RMS داده E های آزمون	24.759	22.450	22.187	20.863	20.606	21.321	21.205	21.564	21.096	21.258	21.267	20.846	22.148	20.757	20.619	**19.755**
	Rank	24	15	14	6	2	11	8	12	7	9	10	5	13	4	3	**1**
Data set 3	RMS داده E های آموزش	0.809	0.793	0.810	0.782	0.754	0.747	0.725	0.703	0.714	0.665	0.673	0.657	0.677	0.682	0.689	0.698
	Rank	27	24	28	22	13	11	10	8	9	2	3	1	4	5	6	7
	RMS داده E های آزمون	0.377	0.391	0.396	0.352	0.358	0.386	0.395	0.378	0.360	0.338	0.333	**0.312**	0.324	0.362	0.372	0.345
	Rank	14	19	22	9	10	16	21	15	11	5	3	**1**	2	12	13	8
Data set 4	RMS داده E های آموزش	13.168	13.183	13.231	13.784	13.711	13.592	13.118	14.053	14.213	13.903	14.616	14.094	14.123	14.622	14.467	13.914
	Rank	9	10	12	22	21	20	8	25	28	23	30	26	27	31	29	24
	RMS داده E های آزمون	24.129	24.148	26.423	26.529	22.732	23.630	21.655	28.375	25.720	21.549	23.616	25.621	22.693	26.405	25.290	22.2999
	Rank	22	23	29	30	16	20	13	31	27	12	19	26	15	28	25	14
Data set 5	RMS داده E های آموزش	11.892	11.851	11.763	11.910	11.563	12.290	12.098	12.577	11.773	11.679	11.580	11.836	11.775	11.440	11.617	11.440
	Rank	17	16	12	18	6	21	20	22	13	11	8	15	14	4	9	3
	RMS داده E های آزمون	13.451	13.714	14.170	12.947	15.645	13.947	13.569	13.709	13.250	16.754	12.862	14.283	14.342	13.659	13.071	14.196
	Rank	6	11	15	3	25	14	8	10	5	28	2	18	19	9	4	16

[DOI: 10.52547/jsdp.19.1.39]

در ادامه در جدول (۳) برای درک هرچه بیشتر این مسأله که استفاده از ترکیب خطی کرنل‌ها برای انتقال داده‌ها به فضای ثانویه چه تأثیری در نتایج داشته است مقادیر RMSE آزمایشاتی آورده شده که در آن‌ها برای انتقال داده‌ها به فضای ثانویه در هر آزمایش فقط یک کرنل روی داده‌ها اعمال شده و براساس این داده‌های ثانویه و با استفاده از رگرسیون بردار پشتیبان مدل پیش‌بینی ساخته شده و مقدار RMSE برای هر کدام از داده‌های آموزش و آزمون روی هر کدام از سری‌های زمانی به‌دست آورده شده است.

اندازهٔ پنجره برای هر مدل، همان اندازه پنجره‌ای انتخاب شده است که در روش ترکیب کرنل‌ها بهترین نتیجه به‌ازای آن در داده‌های آموزش و آزمون مشاهده شده است.

همان‌طور که نتایج جدول (۳) نشان می دهد روش پیشنهادی این مقاله عملکرد بهتری داشته و به نتایج بهتری دست یافته است.

(جدول-۳): مقایسه نتایج روش پیشنهادی با روش حاصل از استفاده از هر یک از کرنل‌های نشان داده شده به تنهایی برای انتقال داده‌ها به فضای ثانویه بر اساس معیار RMSE

(Table-3): Comparison of the results of the proposed method with the method of using each of the displayed kernels alone to transfer the data to the secondary space Based on RMSE criteria

		proposed model	polynomial	RBF	ERBF	sigmoid	guassian
Data set 1	آموزش	**0.372**	0.72	2.09	1.72	1	0.38
	آزمون	**0.459**	0.65	4.59	2.93	0.99	0.53
Data set 2	آموزش	19.754	**4.88**	83.23	78.35	41.57	20.32
	آزمون	**19.755**	42.44	57.55	48.96	43.24	19.74
Data set 3	آموزش	**0.657**	0.83	1.68	1.38	0.91	0.66
	آزمون	**0.312**	1.69	3.90	3.15	1.55	0.42
Data set 4	آموزش	**12.362**	21.42	32.83	32.05	24.02	12.95
	آزمون	**15.681**	28.82	54.19	42.91	31.34	16.40
Data set 5	آموزش	**11.264**	59.76	86.07	78.69	58.10	23.98
	آزمون	**12.700**	120.20	121.13	128.15	116.05	44.31

در جدول (۴) مقایسه‌ای بین نتایج روش پیشنهادی این مقاله با روش‌های مقاله‌های CNN-FCM [5]، SAE-FCM که در آن یک مدل مبتنی بر یک رمزگذار خودکار پراکنده[1] و یک FCM با مرتبه بالا برای رفع مسأله پیش‌بینی سری‌زمانی ایجاد شده است [30]، Wavelet-HFCM [10]، ANFIS که با سامانه استنتاج فازی در چارچوب شبکه‌های تطبیقی اجرا می‌شود که با استفاده از یک روش یادگیری ترکیبی، می‌تواند نقشه برداری از ورودی و خروجی را براساس جفت‌های داده ورودی و خروجی ایجاد کند [31]، AR_ model per scale که با استفاده از یک پرسپترون چندلایه برای مدل‌سازی و پیش‌بینی تجزیه سیگنال مالی، مزیت تبدیل موجک[2] را نشان داد [32]، Multiresolution AR model که یک روش multiresolution برای ترکیب فیلتر صوتی و پیش‌بینی سیگنال ایجاد کرد، که نشان می‌دهد تبدیل موجک می‌تواند وابستگی‌های کوتاه و بلندمدت را به‌صورت قدرتمند ضبط کند [33] و ANN که پیش‌بینی را با تجزیه مقیاس‌های مختلف از پنجره‌های گذشته به مقیاس‌های مختلف از موجک و پیش‌بینی ضرایب هر مقیاس از موجک‌ها با استفاده از یک پرسپترون چندلایه جداگانه شبکه عصبی انجام داده است [34]، صورت گرفته است؛ در نهایت همانطور که در جدول (۴) مشاهده می‌شود، روش پیشنهادی توانسته با بهره‌گرفتن از ترکیب وزن‌دهی‌شده کرنل‌ها و تنظیم پارامترهای آن‌ها در سه مورد از پنج سری‌زمانی بهترین پیش‌بینی را داشته و بهترین مقدار RMSE را در بین سایر روش‌ها به‌دست آورد. متغیر Rank برای یک سری‌زمانی مشخص می‌کند که هر کدام از روش‌ها روی آن سری‌زمانی در مقایسه با سایر روش‌ها چگونه پیش‌بینی کرده‌اند و در اصطلاح چه رتبه‌ای به

[1] SAE (Sparse Auto Encoder)
[2] wavelet

[DOI: 10.52547/jsdp.19.1.39]

دست آورده‌اند. متغیر Best در جدول نشان‌دهندهٔ این است که روش مورد نظر در چند سری‌زمانی از بین پنج سری‌زمانی توانسته رتبهٔ یک شود و بهترین جواب را به دست آورد و متغیر Worst به این معنا است که روش مورد نظر در چند سری‌زمانی از بین پنج سری‌زمانی بدترین جواب را پیدا کرده و رتبه هشت را در مقایسه با روش‌های موجود به‌دست آورده است.

(جدول-۴): مقایسه نتایج روش پیشنهادی با سایر روش‌ها بر اساس معیار RMSE

(Table-4): Comparison of the results of the proposed method with other methods Based on RMSE criteria

		proposed model	CNN-FCM [5]	SAE-FCM (After FT) [30]	Wavelet-HFCM [10]	ANFIS [31]	AR_ model per scale [32]	Multiresolution [33]AR model	ANN [34]
Data set 1	RMSE	0.459	0.730	**0.366**	0.560	0.910	1.350	0.812	1.695
	Rank	2	4	**1**	3	6	7	5	8
Data set 2	RMSE	**19.755**	25.189	21.335	23.159	27.526	29.822	26.733	28.532
	Rank	**1**	4	2	3	6	8	5	7
Data set 3	RMSE	**0.312**	0.566	0.490	0.547	0.651	0.902	0.662	0.652
	Rank	**1**	4	2	3	5	8	7	6
Data set 4	RMSE	**15.681**	17.948	17.390	18.916	22.753	35.262	19.186	19.901
	Rank	**1**	3	2	4	7	8	5	6
Data set 5	RMSE	12.700	30.473	**7.931**	8.258	9.578	57.717	37.838	27.113
	Rank	4	6	**1**	2	3	8	7	5
Best		3	0	2	0	0	0	0	0
Worst		0	0	0	0	0	4	0	1

۵- تحلیل نتایج

برای تجزیه و تحلیل بیشتر نتایج، در جدول (۵) از آزمون ویلکاکسون رتبه علامت‌دار[1] استفاده شده است.

آزمون ویلکاکسون رتبه علامت‌دار مشابه آزمون t زوجی[2] در روش‌های آماری غیرپارامتری است؛ بنابراین، این آزمون یک آزمون دوطرفه است که هدف آن تشخیص تفاوت‌های معنی‌دار بین دو نمونه روش، یعنی رفتار دو الگوریتم، است [35]. نتایج آزمون ویلکاکسون رتبه علامت‌دار روش پیشنهادی و هفت روش دیگر بر روی سری‌های زمانی گفته‌شده، در جدول (۵) آورده شده است. مقدار p بیان‌گر چگونگی تفاوت‌های مهم در نتایج است، هرچه مقدار p کمتر از یک باشد، روش ما قوی‌تر است. همان‌طور که در جدول مشاهده می‌شود، مقدار p برای مقایسه روش پیشنهادی و روش‌های CNN-FCM، AR_ model per scale، Multiresolution AR model **و** ANN بسیار پایین، برای روش‌های Wavelet-HFCM و ANFIS کمی پایین و برای روش SAE-FCM اندکی پایین‌تر از یک است.

همان‌طور که در جدول (۵) آمده است، تعداد سری‌زمانی‌هایی از بین پنج سری‌زمانی که روش پیشنهادی روی آن‌ها، در مقایسه با روش‌های CNN-FCM، SAE-FCM، Wavelet-HFCM، ANFIS، AR_ model per scale، Multiresolution AR model **و** ANN نتایج بهتری کسب کرده و پیش‌بینی بهتری انجام داده به‌ترتیب برابر ۵، ۳، ۴، ۴، ۵، ۵، و ۵ است.

در این قسمت از دیدگاه دیگری به تحلیل نتایج پرداخته می شود. در جدول (۶) میانگین اختلافات قله‌ها، اختلافات دره‌ها و اختلافات قله‌ها و دره‌های متوالی مقادیر واقعی داده‌های آموزش را به‌ازای شمارهٔ ترتیب آن‌ها برای سری‌های زمانی موجود به‌صورت زیر به‌دست می‌آوریم:

$$Diff_{min_i} = |min_{i+1} - min_i| \quad (۱۷)$$

$$Diff_{max_i} = |max_{i+1} - max_i| \quad (۱۸)$$

[1] Wilcoxon Signed-Rank Test
[2] paired t-test

[DOI: 10.52547/jsdp.19.1.39]

که min_i مقدار دره در هر دوره و max_i مقدار قله در هر دوره است. سپس با استفاده از روابط زیر میانگین روابط (۱۷) تا (۱۹) را به‌دست می‌آوریم.

$$Diff_{min,max_i} = \begin{cases} |max_i - min_i| & i = 2k \\ |min_i - max_i| & i = 2k+1 \end{cases} \qquad k = 1,2,\dots \qquad (۱۹)$$

(جدول-۵): آزمون ویلکاکسون رتبه علامت‌دار

(Table-5): Wilcoxon Signed-Rank Test

	مقدار P	تعداد نتیجه بهتر	تعداد نتیجه بدتر	تعداد نتیجه برابر
مدل پیشنهادی در مقایسه با CNN-FCM [5]	7/90E-03	5	0	0
مدل پیشنهادی در مقایسه با SAE-FCM (After FT) [30]	3/65E-01	3	2	0
مدل پیشنهادی در مقایسه با Wavelet-HFCM [10]	3/17E-02	4	1	0
مدل پیشنهادی در مقایسه با ANFIS [31]	3/17E-02	4	1	0
مدل پیشنهادی در مقایسه با AR_ model per scale[32]	7/90E-03	5	0	0
مدل پیشنهادی در مقایسه با Multiresolution AR model [33]	7/90E-03	5	0	0
مدل پیشنهادی در مقایسه با ANN [34]	7/90E-03	5	0	0

(جدول-۶): مقادیر میانگین، انحراف معیار و واریانس مربوط به اختلاف قله ها(max)، دره ها(min) و قله و دره متوالی به‌ازای شماره ترتیب داده برای مقادیر واقعی آموزش

(Table-6): Mean values, standard deviation and variance related to peak differences (max), valleys (min) and peaks and consecutive valleys per number arranged for actual values of training

		میانگین	انحراف معیار	واریانس
Data set 1	اختلاف max ها	12.1	0.7	0.49
	اختلاف min ها	12	1.09	1.2
	اختلاف min و max های متوالی	6.14	1.72	2.92
Data set 2	اختلاف max ها	13.33	4.10	16.88
	اختلاف min ها	14.6	4.12	17.05
	اختلاف min و max های متوالی	9.08	9.02	81.49
Data set 3	اختلاف max ها	11.8	2.28	5.22
	اختلاف min ها	11.68	1.75	3.08
	اختلاف min و max های متوالی	5.84	2.09	4.38
Data set 4	اختلاف max ها	11.11	2.15	4.65
	اختلاف min ها	10.94	1.53	2.36
	اختلاف min و max های متوالی	6.83	8.02	64.40
Data set 5	اختلاف max ها	12	0	0
	اختلاف min ها	12	0	0
	اختلاف min و max های متوالی	6	0	0

$$A = AVERAGE(Diff_{min}) \qquad (۲۰)$$

$$B = AVERAGE(Diff_{max}) \qquad (۲۱)$$

[DOI: 10.52547/jsdp.19.1.39]

$$C = AVERAGE(Diff_{min,max}) \quad (۲۲)$$

که مقادیر آن به‌ازای سری‌های زمانی مختلف در جدول (۶) نمایش داده شده است.

همان‌طور که در جدول (۶) مشاهده می‌شود، برای سری‌های زمانی که مقادیر واریانس آن‌ها کم است و با توجه به رتبه‌های مربوط به هر کدام از اندازه پنجره‌ها در جدول (۲)، برای اندازه پنجره‌هایی که در رابطه زیر صدق می‌کنند مقادیر RMSE بهتری به‌دست آمده و رتبهٔ بهتری در میان سایر اندازه پنجره‌ها داشته‌اند؛ لذا پیشنهاد می‌شود برای به‌دست‌آوردن اندازه پنجره مناسب، رابطه (۲۳) روی داده‌های آموزش اجرا شود و بهترین مقادیر i,j و k می‌تواند توسط سعی و خطا و یا الگوریتم های بهینه‌سازی گسسته جست‌وجو شود.

$$BestWindowSize = \mathrm{i} * \mathrm{A} + \mathrm{j} * \mathrm{B} + \mathrm{k} * \mathrm{C}$$
$$\mathrm{i,j,k} = 0,1,2,\ldots \quad (۲۳)$$

که A،B و C مقادیر به‌دست‌آمده از روابط (۲۰) تا (۲۲) هستند. به‌عنوان نمونه همان‌طور که در جدول (۶) مشاهده می‌شود، رابطه (۲۳) برای سری‌های زمانی ۴ و ۵ به‌ازای i,j=0 و k=1 به‌ترتیب مقادیر ۶/۸۳ و ۶ را نتیجه داده که مطابق با جدول (۲) بهترین RMSE برای این سری‌های زمانی به‌ازای اندازه پنجره‌های ۶ و ۷ حاصل شده است. برای سری‌زمانی دوم با توجه به جدول (۶) مشاهده می‌شود که مقدار واریانس آن زیاد است و رابطه (۲۳) برای آن نمی‌تواند کارا باشد.

۶- نتیجه‌گیری

در این مقاله، یک مدل جدید برای پیش‌بینی سری‌زمانی براساس رگرسیون بردار پشتیبان و ترکیب وزن‌دار کرنل‌های مختلف و بهینه‌سازی پارامترها و وزن‌های آن‌ها توسط بهینه‌ساز، ارائه شده است. بهینه‌ساز در تکرارهای متوالی تابع شایستگی، مقادیر بهینه برای ورودی تابع را، که آرایه‌ی ذرات و داده‌های آموزش هستند، یاد می‌گیرد و از آن مقادیر برای ارزیابی داده‌های آزمون استفاده می‌کند. نتایج حاصل از اعمال مدل پیشنهادی بر روی پنج سری‌زمانی استاندارد با هفت روش دیگر مقایسه شده است و در سه مورد از پنج سری‌زمانی نتایج بهتری نسبت به سایر روش‌ها کسب شده است. در ادامه برای پژوهش‌های آینده قصد داریم از روش‌های ترکیبی[1] برای مدل پیش‌بینی استفاده کنیم. با ترکیب چند مدل پایه یک مدل برای پیش‌بینی ساخته شود. همچنین با استفاده از روش‌های بهینه‌ساز گسسته به پیش‌بینی اندازه پنجره مناسب پرداخته شود.

سپاس‌گزاری

این طرح پژوهشی با استفاده از اعتبارات ویژه پژوهشی دانشگاه صنعتی نوشیروانی بابل انجام شده است؛ لذا در کمال احترام از همکاری به‌عمل‌آمدهٔ دانشگاه نهایت سپاس و قدردانی به‌عمل می‌آید.

7- References ۷- مراجع

[1] E. Kayacan, B. Ulutas and O. Kaynak, "Grey system theory-based models in time series prediction," *Expert Systems with Applications*, vol. 37, no. 2, pp. 1784-1789, 2010.

[2] B. Paaben, C. Gopfert and B. Hammer, "Time Series Prediction for Graph in Kernel andDissimilarity Spaces," *Neural Processing Letters*, no. 48, pp. 669-689, 2018.

[3] S. C. Nayak, B. B. Misra and H. S. Behera, "Efficient financial time series prediction with evolutionary virtual data position exploration," *Neural Computing and Application*, no. 31, pp. 1053-1074, 2019.

[4] M. A. Villegas, D. J. Pedregal and J. R. Trapero, "A support vector machine for model selection in demand forecasting application," *Computers & Industrial Engineering*, vol. 121, pp. 1-7, 2018.

[5] P. Liu, J. Liu and K. Wu, "CNN-FCM: System modeling promotes stability of deep learning in time series prediction," *Knowledge-Based Systems*, vol. 203, 2020.

[6] J. H. Sadaei, P. Cândido de Lima e Silva, F. Gadelha Guimarães and M. Hisyam Lee, "Short-term load forecasting by using a combined method of convolutional neural networks and fuzzy time series," *Energy*, vol. 175, pp. 365-377, 2019.

[7] J. Hu, X. Wang, Y. Zhang, D. Zhang, M. Zhang and J. Xue, "Time Series Prediction Method Based on Variant LSTM Recurrent Neural Network," *Neural Processing Letters*, 2020.

[8] K. Yuan, J. Liu, S. Yang, K. Wu and F. Shen, "Time series forecasting based on kernel mapping and high-order fuzzy cognitive maps," *Knowledge-Based Systems*, vol. 206, 2020.

[1] Ensemble

[DOI: 10.52547/jsdp.19.1.39]

[9] J. Wang, Z. Peng, X. Wang, C. Li and J. Wu, "Deep Fuzzy Cognitive Maps for Interpretable Multivariate Time Series Prediction," *IEEE Transactions on Fuzzy Systems*, pp. 1-1, 2020.

[10] S. Yang and J. Liu, "Time Series Forecasting based on High-Order Fuzzy Cognitive Maps and Wavelet Transform," *IEEE Transactions on Fuzzy System*, vol. 26, no. 6, pp. 3391-3402, 2018.

[11] Q. Xiao, "Time series prediction using bayesian filtering model and fuzzy neural networks," *Optik - International Journal for Light and Electron Optics*, vol. 140, pp. 104-113, 2017.

[12] H. Omranpour, F. Azadian, "Presenting a Fuzzy Approach to Optimize Predicting High Order Time series," Signal and Data Processing vol. 15, no. 2, pp. 3-16, 2018.

[۱۲] ح. عمرانپور, ف. آزادیان, "ارائه یک رویکرد فازی برای بهینه سازی پیش بینی سری زمانی با مرتبه ی بالا," پردازش علائم و داده‌ها, جلد ۱۵, شماره ۲, ۱۳۹۷.

[13] C. Bergmeir, R. J. Hyndman and B. Koo, "A note on the validity of cross-validation for evaluating autoregressive time series prediction," *Computational Statistics and Data Analysis*, vol. 120, pp. 70-83, 2018.

[14] W. Xu, H. Peng, X. Zeng, F. Zhou, X. Tian and X. Peng, "Deep belief network-based AR model for nonlinear time series forecasting," *Applied Soft Computing Journal*, vol. 77, pp. 605-621, 2019.

[15] L. Bianchi, M. Dorigo, L. M. Gambardella and W. J. Gutjahr, "A survey on metaheuristics for stochastic combinatorial optimization," *Natural Computing*, vol. 8, pp. 239-287, 2009.

[16] G. Cornuéjols , "Valid inequalities for mixed integer linear programs," *Mathematical Programming*, vol. 112, pp. 3-44, 2008.

[17] M. Avriel, Nonlinear Programming: Analysis and Methods, New York: Dover Publications, 2003.

[18] A. H. Land and A. G. Doig, "An automatic method for solving discrete programming problems," *50 Years of Integer Programming 1958-2008*, pp. 105-132, 2010.

[19] A. R. Simpson, G. C. Dancy and L. J. Murphy, "Genetic algorithms compared to other techniques for pipe optimization," *Journal of water resources planning and management*, vol. 120, no. 4, pp. 423-443, 1994.

[20] S. Mirjalili, "The Ant Lion Optimizer," *Advances in Engineering Software*, vol. 83, pp. 80-98, 2015.

[21] B. T. Ojemakinde, Support Vector Regression for Non-Stationary Time Series, Knoxville: University of Tennessee, 2006.

[22] S. Lin, S. zhang, J. Qiao, H. Liu and G. Yu, "A Parameter Choosing Method of SVR for Time Series Prediction," in *The 9th International Conference for Young Computer Scientists*, Liaoning, China, 2008.

[23] C. Hsin, J. M. Ho and D. T. Lee, "Travel-Time Prediction With Support Vector Regression," *IEEE Transactions on Intelligent Transportation Systems*, vol. 5, no. 4, pp. 276-281, 2004.

[24] X. Ma, Y. Zhang, H. Cao, S. Zhang and Y. Zhou, "Nonlinear Regression with High-Dimensional Space Mapping for Blood Component Spectral Quantitative Analysis," *Journal of Spectroscopy*, 2018.

[25] T. Hofmann, B. Schölkopf and A. J. Smola, "Kernel Methods in Machine Learning," *Institute of Mathematical Statistics*, vol. 36, no. 3, pp. 1171-1220, 2008.

[26] J. Xie, "Time Series Prediction Based on Recurrent LS-SVM with Mixed Kernel," in *Asia-Pacific Conference on Information Processing*, Shenzhen, China, 2009.

[27] S. Mirjalili, S. M. Mirjalili and A. Lewis, "Grey Wolf Optimizer," *Advances in Engineering Software*, vol. 69, pp. 46-61, 2014.

[28] M. R. Mosavi, M. Khishe and A. Ghamgosar, "CLASSIFICATION OF SONAR DATA SET USING NEURAL NETWORK TRAINED BY GRAY WOLF OPTIMIZATION," *Neural Network World*, vol. 4, pp. 393-415, 2016.

[29] J. Heinermann and O. Kramer, "Precise Wind Power Prediction with SVM Ensemble Regression," in *International Conference on Artificial Neural Networks*, Hamburg, Germany, 2014.

[30] K. Wu, J. Liu, P. Liu and S. Yang, "Time Series Prediction Using Sparse Autoencoder and High-order Fuzzy Cognitive Maps," *IEEE Transactions on Fuzzy Systems*, 2019.

[31] J. Shing and R. Jang, "ANFIS: adaptive-network-based fuzzy inference system," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 23, no. 3, pp. 665-685, 1993.

[32] G. Zheng, J. Starck, J. Campbell and F. Murtagh, "Multiscale transforms for filtering financial data streams," *Journal of Computational Intelligence in Finance*, vol. 7, no. 18-35, 1999.

[33] O. Renaud, J. L. Starck and F. Murtagh, "Wavelet-Based Combined Signal Filtering and Prediction," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 35, no. 6, pp. 1241-1251, 2005.

[34] A. B. Geva, "ScaleNet-multiscale neural-network architecture for time series prediction," *IEEE Transactions on Neural Networks*, vol. 9, no. 6, pp. 1471-1482, 1998.

[DOI: 10.52547/jsdp.19.1.39]

[35] J. Derrac, S. García, D. Molina and F. Herrera, "A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms," *Swarm and Evolutionary Computation*, vol. 1, no. 1, pp. 3-18, 2011.

حسام عمرانپور مدرک کارشناسی خود را در رشته مهندسی رایانه در سال ۱۳۸۵ در دانشگاه علم و صنعت و همچنین مدرک کارشناسی ارشد را در سال ۱۳۸۸ از دانشگاه امیرکبیر در گرایش هوش مصنوعی دریافت و درجه دکترای خود را از دانشگاه صنعتی امیرکبیر گرایش هوش مصنوعی در سال ۱۳۹۵ دریافت کرده است. در حال حاضر استادیار دانشگاه صنعتی نوشیروانی بابل در زمینه هوش مصنوعی و یادگیری ماشین است. زمینه پژوهشی موردعلاقه ایشان عبارتند از: هوش مصنوعی، یادگیری ماشین، شناسایی آماری الگوها و پیش‌بینی سری‌زمانی و پردازش سیگنال های حیاتی.

نشانی رایانامه ایشان عبارت است از:

h.omranpour@nit.ac.ir

حدیثه پورعلی مدرک کارشناسی خود را در رشته مهندسی رایانه در سال ۱۳۹۸ در دانشگاه صنعتی نوشیروانی بابل دریافت کرده است. در حال حاضر دانشجوی کارشناسی ارشد رشته مهندسی رایانه گرایش نرم‌افزار در دانشگاه صنعتی نوشیروانی بابل است. زمینه پژوهشی موردعلاقه ایشان عبارتند از: هوش مصنوعی، یادگیری ماشین و پیش‌بینی سری‌زمانی.

نشانی رایانامه ایشان عبارت است از:

h.poorali@nit.ac.ir

[DOI: 10.52547/jsdp.19.1.39]