



DOI: 10.22084/ier.2023.27356.2114

پیاده‌سازی رویکرد ترکیبی میانگین متحرک یکپارچه خودرگرسیون-شبکه عصبی و میانگین متحرک یکپارچه خودرگرسیون-روش پیش‌بینی محلی در سیستم مدیریت ترافیک. مطالعه موردی: فرودگاه یک کلان‌شهر

شادی صدری^۱، سید محمدتقی فاطمی‌قمی^{۲*}

۱. دانشجوی دکتری گروه مهندسی صنایع، دانشکده مهندسی صنایع، دانشگاه صنعتی امیرکبیر، تهران، ایران
۲. استاد گروه مهندسی صنایع، دانشکده مهندسی صنایع، دانشگاه صنعتی امیرکبیر، تهران، ایران

خلاصه

امروزه، باتوجه به ضرورت انجام پیش‌بینی در بسیاری از مسائل دنیای واقعی، پیش‌بینی سری‌های زمانی یکی از موضوعات اصلی مورد بحث و تحقیق می‌باشد. مدیریت حمل‌ونقل کلان‌شهرها یکی از مسائل چالش‌برانگیز در حیطه پیش‌بینی است. پیش‌بینی دقیق و قابل اطمینان از جمله نیازهای یک سیستم حمل‌ونقل هوشمند می‌باشد. در این مطالعه، هدف پیش‌بینی تعداد مسافری ورودی به فرودگاه یک کلان‌شهر جهت ارائه خدمات مطلوب به مسافری است. متغیر مورد مطالعه در این مقاله مشابه مفهوم متغیر جریان در مدیریت ترافیک می‌باشد. اغلب مطالعات انجام شده در حیطه مدیریت ترافیک از روش‌های خطی و یا غیرخطی برای پیش‌بینی آینده بهره گرفته‌اند و مزایای استفاده از رویکردهای ترکیبی مورد توجه قرار نگرفته است. در مطالعات پیشین، روش‌های ترکیبی میانگین متحرک یکپارچه خودرگرسیون-شبکه عصبی مصنوعی (ARIMA-ANN) متنوعی توسعه یافته‌اند که روابط میان داده‌ها را بررسی می‌کنند. مطالعه حاضر، از یک رویکرد ARIMA-ANN برای تجزیه داده‌ها به دو قسمت کم‌نوسان و پرنوسان و انجام پیش‌بینی دقیق استفاده می‌کند. علاوه بر این، در این مطالعه محققین یک رویکرد ترکیبی جدید، میانگین متحرک یکپارچه خودرگرسیون-تکنیک پیش‌بینی محلی (ARIMA-Local) برای بررسی کارایی سایر روش‌های غیرخطی ارائه کرده‌اند. نتایج عددی حاصل از پیاده‌سازی روش‌های مذکور بر روی مطالعه موردی، دقت بالای روش ARIMA-ANN در پیش‌بینی و همچنین قابلیت بهتر روش ARIMA-Local در مقایسه با روش‌های انفرادی شبکه عصبی مصنوعی و هموارسازی نمایی را نشان می‌دهد.

اطلاعات مقاله

تاریخچه مقاله:

دریافت ۱۴۰۱/۲/۲۲

پذیرش ۱۴۰۱/۴/۸

(مقاله پژوهشی)

کلمات کلیدی:

شبکه عصبی مصنوعی (ANN)

میانگین متحرک خودهمبسته

یکپارچه (ARIMA)

رویکرد ترکیبی -ARIMA

ANN

رویکرد ترکیبی -ARIMA

Local

مدیریت ترافیک هوشمند

۱. مقدمه

آینده، پیش‌بینی سری‌های زمانی مورد توجه قرار گرفته است. سری زمانی دنباله‌ای از مشاهدات در نقاط زمانی گسسته است. برای پیش‌بینی سری‌های زمانی، یک مدل بر روی داده‌های گذشته برازش می‌شود تا روابط میان داده‌ها شناسایی شده و سپس از این رابطه برای پیش‌بینی آینده استفاده می‌گردد. شبکه‌های عصبی مصنوعی یکی از انواع مدل‌هایی هستند که می‌توانند روابط غیرخطی میان داده‌ها را شناسایی

در دنیای واقعی، عدم قطعیت درباره آینده یکی از حقایق انکارناپذیر زندگی روزمره بوده و درواقع می‌توان گفت آینده به نوعی وابسته به اتفاقات گذشته است. پیش‌بینی، نوعی تخمین از آینده مبتنی بر مشاهدات پیشین از همان رویداد است. در برخی موارد نیز آینده مستقل از گذشته است. امروزه، در راستای کاهش ریسک تصمیم‌گیری برای

* نویسنده مسئول: سید محمدتقی فاطمی‌قمی

تلفن: ۰۲۱-۶۴۵۴۵۳۸۱؛ پست الکترونیکی: fatemi@aut.ac.ir

مروری کوتاه بر ادبیات پیش‌بینی در حوزه‌های مختلف نشان می‌دهد در بیشتر مطالعات، داده‌ها خطی یا غیرخطی در نظر گرفته شده و از تکنیک‌های مرتبط برای پیش‌بینی استفاده شده است. با این حال، برخی از محققان معتقدند که ورودی‌های سیستم را می‌توان به دو بخش خطی و غیرخطی تجزیه کرد. در چنین شرایطی، نتایج اجرای ARIMA یا ANN به‌طور جداگانه دقیق نخواهند بود. مدل‌های خطی (به‌عنوان مثال ARIMA) نمی‌توانند با ویژگی‌های غیرخطی داده‌ها مطابقت داشته باشند و بالعکس. پس برای حل این مشکل، چندین تکنیک ترکیبی ARIMA-ANN برای استفاده از قابلیت هر دو روش و افزایش دقت پیش‌بینی توسعه داده شده‌اند.

یکی از مطالعات اولیه جهت توسعه روش‌های ترکیبی توسط ژانگ [۱۷] انجام شده است. وی روش پیشنهادی خود را بر مبنای دو روش کاربرد برای شناسایی روابط خطی (ARIMA) و غیرخطی (ANN) توسعه داده است. بعد از مدتی، ولنزولا و همکاران [۱۸] یک رویکرد دیگری از ترکیب ARIMA و ANN ارائه کردند. هدف آن‌ها از این مطالعه این بود که ثابت کنند هیچ اولویتی بین روش ARIMA و ANN وجود ندارد و نمی‌توان گفت کدام‌یک به‌ازای تمام مسائل عملکرد بهتری دارد. فاروک [۱۹] یک ARIMA و ANN ترکیبی را برای استفاده از نقاط قوت رویکرد سنتی سری زمانی و شبکه‌های عصبی مصنوعی در زمینه کیفیت آب پیشنهاد کرد. خاشعی و بیجاری [۱] ترکیب جدیدی از مدل‌های ANN و ARIMA را برای پیش‌بینی دقیق‌تر متغیرها پیشنهاد کردند. بسیاری از مدل‌های ترکیبی ARIMA-ANN که در ادبیات وجود دارند، یک مدل ARIMA را برای داده‌های سری زمانی داده شده اعمال می‌کنند، خطای بین داده‌های اصلی و پیش‌بینی شده توسط ARIMA را به‌عنوان یک مؤلفه غیرخطی در نظر می‌گیرند، و آن را با استفاده از ANN به روش‌های مختلف مدل‌سازی می‌کنند. بابو و ردی [۲۰] نیز نوسانات داده‌ها را با استفاده از میانگین متحرک (MA) شناسایی کرده و از مدل‌های ANN و ARIMA برای شناسایی روابط خطی و غیرخطی داده‌ها و انجام پیش‌بینی دقیق استفاده کرده‌اند.

مرور مطالعات انجام شده در حوزه مدیریت ترافیک، نشان می‌دهد که در این حوزه توجهی به قابلیت روش‌های پیش‌بینی ترکیبی نشده است. با مقایسه روش‌های توسعه یافته که در بالا اشاره شد، روش پیشنهادی بابو و ردی [۲۰] بهترین عملکرد و دقت را در پیش‌بینی سری‌های زمانی داشته است. از این‌رو، مطالعه حاضر این روش را بر روی یک مثال واقعی در حوزه مدیریت ترافیک پیاده‌سازی می‌کند. علاوه بر این، یک رویکرد ترکیبی جدید مبتنی بر روش ARIMA و یک روش پیش‌بینی محلی توسعه داده شده است تا کارایی روش‌های ترکیبی در مقایسه با روش‌های انفرادی سنجیده شود. برای بررسی کارایی هر یک از روش‌ها، از داده‌های مربوط به مسافریان یک شرکت هواپیمایی در یکی از کلان‌شهرهای ایران استفاده شده است. مدیران شرکت علاقه‌مند به ارائه خدمات فرودگاهی نوینی هستند که در این راستا بایستی تعداد مسافریان پروازهای ورودی به فرودگاه شهر پیش‌بینی گردد. طبق ادبیات، می‌توان این متغیر را معادل متغیر ترافیک دانست [۷]. از آنجایی که ارائه

کنند [۱]. علاوه بر این، رویکردهای مختلفی برای مدل‌سازی روابط خطی و برون‌یابی آن توسعه یافته است. پرکاربردترین روش‌های مدل‌سازی خطی، روش هموارسازی نمایی [۲]، میانگین متحرک [۳] و میانگین متحرک یکپارچه خودرگرسیون [۴] هستند.

سیستم حمل‌ونقل هوشمند یکی از زمینه‌های پرکاربرد روش‌های پیش‌بینی است. محققین این زمینه، اغلب بر پیش‌بینی سرعت، زمان سفر، جریان ترافیک تمرکز داشته‌اند. خومه‌ای و همکاران [۵] از ماشین بردار پشتیبانی برای پیش‌بینی اولیه زمان سفر استفاده کردند و الگوریتم فیلتر کالمن را برای پیش‌بینی زمان سفر به‌کار گرفتند. کاربرد دیگری از این تکنیک توسط جیانگ و همکاران [۶] بوده است. هوارت و همکاران [۷] معتقد بودند که پیش‌بینی دقیق متغیرهای ترافیک از جمله اهداف اصلی سیستم حمل‌ونقل هوشمند می‌باشد. بنابراین، آن‌ها یک الگوریتم یادگیری ماشین مبتنی بر کرنل برای استخراج مشاهدات از داده‌های ترافیک ارائه کردند. الهنوی و همکاران [۸] نیز با در نظر گرفتن حد پایین و بالای زمان‌های سفر دینامیک یک الگوریتم برای خوشه‌بندی داده‌ها و مدل‌سازی ژنتیک ارائه کردند. رویکرد دیگر برای پیش‌بینی تراکم ترافیک شهری، داده‌های مسیر حرکت خودروهای شناور است که توسط کنگ و همکاران [۹] پیشنهاد شده است. آن‌ها از یک روش ارزیابی جامع فازی جدید برای پیش‌بینی متغیر استفاده کردند. اخیراً، سان و سان [۱۰] یک مدل شبکه بیزی پویا از داده‌های ترافیک سری زمانی را برای توصیف همبستگی بین وقوع تصادف و داده‌های پویای وضعیت سرعت پیشنهاد کردند. جاودانی سادابی و همکاران [۱۱] مدلی مبتنی بر میانگین متحرک یکپارچه کسری خودکار رگرسیون فصلی (SARFIMA) و سری زمانی فازی (FTS) برای افزایش دقت پیش‌بینی سری‌های زمانی فصلی پیشنهاد کرد. آن‌ها از یک الگوریتم بهینه‌سازی ازدحام ذرات (PSO) برای تخمین پارامترهای مدل استفاده کردند.

از آنجایی که برنامه‌ریزی کارآمد نرخ پذیرش فرودگاه نقشی حیاتی در مدیریت ترافیک ایفا می‌کند، مورکا و هانسمن [۱۲] یک چارچوب داده‌محور را برای پیش‌بینی پیاده‌سازی کردند. آن‌ها جنبه‌های وابستگی متقابل عملیاتی سیستم‌های Metroplex را در نظر گرفتند. خو و همکاران [۱۳] شاخص‌های آماری در صنعت هوانوردی را با استفاده از مدل SARIMA-SVR پیش‌بینی کردند. آن‌ها سری‌های زمانی را توسط SARIMA تجزیه و تحلیل کرده و نویز سفید گاوسی را محاسبه نمودند. مقایسه روش پیشنهادی با چهار روش دیگر نشان‌دهنده کارایی آن در کاربرد است. لیچنگ و همکاران [۱۴] بر پیش‌بینی روزانه جریان ترافیک متمرکز شدند و از یک شبکه عصبی عمیق مبتنی بر داده‌های تاریخی جریان ترافیک و داده‌های عامل زمینه‌ای استفاده کردند. آن‌ها برای کاهش زمان یادگیری، از روش یادگیری دسته‌ای استفاده کردند. ویلارویا و همکاران [۱۵] نیز از مدل‌های شبکه عصبی برای پیش‌بینی حجم ترافیک در شهر والنسیا اسپانیا بهره گرفته‌اند. صادقی گرگری و همکاران [۱۶] نیز برای پیش‌بینی متغیر ترافیک در یکی از بنادر ایران، از روش‌های شبکه عصبی و میانگین متحرک خودهمبسته یکپارچه فصلی استفاده کرده‌اند.

باقی‌مانده بررسی شده و یک آزمون استاندارد کای-اسکوئر پیاده‌سازی می‌شود. آماره آزمون عبارت است از: $Q = (N - d) \sum_{k=1}^K r_e^2(k)$ که N برابر است با تعداد داده‌ها، k نشان‌دهنده تعداد نمونه‌ها $N/3 \leq k$ و $r_e(k)$ بیانگر خودهمبستگی نمونه‌هاست.

۲-۱-۲. شبکه عصبی مصنوعی (ANN):

شبکه‌های عصبی می‌توانند برای سطوح مختلفی از پیش‌بینی استفاده شوند [۲۲]. این روش همبستگی میان داده‌ها را با راحتی شناسایی کرده و یادگیری می‌کند و مانند روش رگرسیون نیازی به افزودن اطلاعات از قبیل نوع همبستگی ندارد. شبکه‌های عصبی بر اساس داده‌های پیشین آموخته می‌شوند و بر این اساس روابط میان داده‌ها را شناسایی کرده و از این رابطه برای پیش‌بینی مقادیر آینده استفاده می‌کنند. در واقع، داده‌های گذشته به‌عنوان ورودی شبکه عصبی تلقی شده و مقادیر پیش‌بینی شده، خروجی آن خواهند بود. در اغلب مواقع، یک شبکه عصبی سه لایه‌ای شامل لایه ورودی، لایه پنهان و لایه خروجی در پیش-بینی مقادیر یک متغیر عملکرد مناسبی دارد. ورودی‌های شبکه عصبی تابعی خطی از مقادیر داده‌های گذشته $\{x_{t-1}, x_{t-2}, \dots, x_{t-N}\}$ است.

۲-۱-۳. رویکرد ترکیبی ARIMA و ANN

سه رویکرد ترکیبی ارائه شده توسط ژانگ [۱۷]، خاشعی و بیجاری [۱] و بابو و ردی [۲۰] شناخته‌شده‌ترین و کاربردی‌ترین رویکردها در ادبیات موضوع هستند. مطالعات پیشین نشان می‌دهند که رویکردهای ترکیبی عملکرد بهتری در مقایسه با رویکردهای انفرادی ARIMA و ANN دارند. ژانگ [۱۷] هر دو جنبه خطی و غیرخطی بودن داده‌های سری زمانی را در نظر گرفته و یک مدل ARIMA بر روی این داده‌ها و یک مدل ANN بر روی دنباله مقادیر خطا برازش کرده است. نقطه ضعف این روش این است که روابط خطی و غیرخطی داده‌ها از هم جدا نشده‌اند، پس مدل خطی برازش شده ARIMA نمی‌تواند روابط غیرخطی را شناسایی کند. به‌طور مشابه، خاشعی و بیجاری [۱] نیز روابط خطی و غیرخطی را جداسازی نکرده و یک مدل ARIMA بر روی داده‌ها برازش کرده و مقادیر متغیر را پیش‌بینی کرده‌اند. سپس مقادیر واقعی گذشته، مقادیر پیش‌بینی شده و مقادیر خطای داده‌های گذشته به‌عنوان ورودی وارد شبکه عصبی شده‌اند.

باتوجه به ضعف هر یک از این روش‌ها برای استفاده از تمام قابلیت‌های روش‌های ترکیبی، بابو و ردی [۲۰] روشی ارائه کرده‌اند که ابتدا داده‌ها را به دو بخش خطی و غیرخطی تقسیم‌بندی کرده و نوسانات داده‌های سری زمانی را ارزیابی کرده‌اند. باتوجه به قابلیت بالای این روش و دقت نتایج حاصله، در این مقاله نیز این روش بر روی مطالعه موردی پیاده‌سازی شده است. در این روش فرض شده است که هر داده سری زمانی (x_t) به‌صورت مجموع یک قسمت کم‌نوسان (l_t) و یک قسمت پرنوسان (h_t) است به‌گونه‌ای که $x_t = l_t + h_t$ همان‌طور که قبلاً اشاره شد، بعد از بررسی ایستایی قسمت l_t ، یک مدل خطی ARIMA بر روی داده‌های این قسمت برازش می‌شود. سپس مدل غیرخطی ANN نیز بر روی مقادیر قسمت h_t برازش داده می‌شود. مقدار تخمینی متغیر x_t معادل حاصل جمع مقادیر پیش‌بینی شده l_t و h_t خواهد بود

خدمات فرودگاهی به مسافری مستلزم صرف هزینه گزافی است، از این‌رو، در این مطالعه از رویکردهای ترکیبی جهت انجام پیش‌بینی دقیق‌تر و کاهش هزینه‌های ریسک ناشی از عدم قطعیت استفاده شده است.

در ادامه، ساختار مقاله حاضر به این صورت است: در بخش ۲، رویکردهای ترکیبی ARIMA-ANN معرفی شده در ادبیات و رویکرد ترکیبی ARIMA-Local توسعه یافته در این مقاله شرح داده شده است. بخش ۳ نتایج حاصل از پیاده‌سازی این روش‌ها بر روی مثال موردی را نشان می‌دهد. در بخش ۴ به تجزیه و تحلیل بیشتر نتایج و روش‌ها پرداخته شده و در بخش ۵ نتیجه‌گیری و پیشنهادات برای مطالعات آتی ارائه شده است.

۲. روش حل

در این بخش، روش‌های حل پیشنهادی شامل روش ترکیبی ARIMA-ANN و روش ARIMA-Local شرح داده می‌شود.

۲-۱-۱. روش ترکیبی ARIMA-ANN

قبل از توضیح روش پیاده‌سازی شده ARIMA-ANN، توضیح مختصری در مورد هر یک از روش‌های ARIMA و ANN در زیر بیان شده است.

۲-۱-۱-۱. ARIMA (p,d,q)

در اغلب رویدادهای دنیای واقعی، مشاهدات سری زمانی بسیار همبستگی دارند و فرض استقلال این مشاهدات فرض مناسبی نیست. رویکرد هموارسازی نمایی، به‌عنوان یکی از رایج‌ترین رویکردهای پیش‌بینی، نمی‌تواند همبستگی این داده‌ها را مدل کند. از این‌رو، ARIMA یک تکنیک مدل‌سازی خطی است که به‌صورت زیر فرموله می‌شود:

$$\phi_p(B) \nabla^d x_t = \theta_q(B) \varepsilon_t \quad (1)$$

در این رابطه، ε_t خطای سفید با میانگین صفر و واریانس ثابت σ_ε^2 ، p پارامتر خود رگرسیون و q پارامتر میانگین متحرک می‌باشد [۲۱]. مدل‌سازی ARIMA در سه مرحله انجام می‌شود. ابتدا، داده‌ها از لحاظ ایستایی سنجیده می‌شوند. در صورتی که داده‌ها غیرایستا باشند، عملیات مشتق بر روی داده‌ها انجام می‌شود تا به داده‌های ایستا تبدیل شوند. در رابطه (۱)، d معادل تعداد دفعات انجام عملیات مشتق‌گیری است. پارامترهای p و q بر اساس تحلیل همبستگی و ماهیت تابع خودهمبستگی و تابع خود همبستگی جزئی محاسبه می‌شوند. در این روش، فرض شده است در هر نقطه زمانی، داده‌ها به p داده قبل $\{\varepsilon_t - \phi_p x_{t-p} + \phi_{p-1} x_{t-p+1} + \dots + \phi_1 x_{t-1}\}$ و q پارامتر خطا $\{-\theta_q \varepsilon_{t-q} + \theta_{q-1} \varepsilon_{t-q+1} + \dots + \theta_1 \varepsilon_t\}$ وابسته هستند $(\varepsilon_t \sim N(0, \sigma_\varepsilon^2))$. در مرحله دوم، بر اساس مدل برازش شده $(ARIMA(p,d,q))$ ، پارامترهای مدل $(\phi_1, \phi_2, \dots, \phi_p)$ و $(\theta_1, \theta_2, \dots, \theta_q)$ با استفاده از تکنیک مطلوب بهینه‌سازی برآورد می‌شوند. در آخر، صحت مدل ارزیابی می‌گردد. در یک مدل بهینه، مقادیر باقیمانده همان خطاهای سفید با میانگین صفر و واریانس σ_ε^2 هستند. برای سنجش این فرض، همبستگی مقادیر

هزینه متناسب با تعداد مسافری پروازهای ورودی خواهد بود. با در نظر داشتن محدودیت بودجه، مدیران و مسئولین ذی‌ربط علاقه‌مند به استفاده از یک روش پیش‌بینی سری‌های زمانی با دقت بالا هستند. همان‌طور که قبلاً اشاره شد، براساس مطالعات پیشین قابلیت و عملکرد رویکردهای ترکیبی در مقایسه با رویکردهای انفرادی اثبات شده است. در راستای پیاده‌سازی سیاست‌های جدید و افزایش چشم‌گیر تعداد پروازها و مسافری، اطلاعات گذشته قابل اطمینان نبوده و برای انجام عملیات پیش‌بینی تنها اطلاعات ۶ ماه اخیر در دسترس هستند که این اطلاعات در شکل (۱) نشان داده شده‌اند.

در ادامه جزئیات به‌کارگیری هریک از رویکردهای بحث شده در بخش ۲ نشان داده شده است.

گام ۱) تجزیه داده‌ها با استفاده از فیلتر میانگین متحرک

مطابق الگوریتم‌های مشروح در بخش ۲، در ابتدا بایستی داده‌ها با استفاده از میانگین متحرک به دو بخش کم‌نوسان و پر‌نوسان تجزیه شوند. رابطه مورد نظر برای میانگین متحرک عبارت است از:

$$x_t = \frac{1}{m} \sum_{t=T-m+1}^T x_t \quad (2)$$

در این رابطه، m بیانگر پارامتر میانگین متحرک است که در این مطالعه ۷ در نظر گرفته شده است. نتایج مربوط به قسمت‌های کم‌نوسان و پر‌نوسان هریک از داده‌ها به‌ترتیب در شکل (۲) و (۳) نشان داده شده‌اند.

گام ۲) برازش یک مدل ARIMA

مطابق توضیحات بخش قبلی، برازش یک مدل ARIMA در سه مرحله صورت می‌گیرد. بایستی توجه داشت که مدل ARIMA بر روی داده‌های قسمت کم‌نوسان (x_{tr}) برازش شود.

مرحله ۱: بررسی ایستایی داده‌ها

برای این امر، همبستگی میان K نمونه مطابق رابطه زیر محاسبه می‌شود:

$$r_k = \frac{\sum_{t=1}^{N-k} (x_{tr} - \hat{x}_{tr})(x_{tr+k} - \hat{x}_{tr})}{\sum_{t=1}^N (x_{tr} - \hat{x}_{tr})^2} \quad (3)$$

$$k = 0, 1, \dots, K$$

تابع همبستگی جزئی نیز برای سنجش ایستایی و تعیین پارامترهای ARIMA مناسب است. این تابع عبارت است از:

$$r_j = \phi_{k1}r_{j-1} + \phi_{k2}r_{j-2} + \dots + \phi_{kk}r_{j-k} \quad (4)$$

انجام محاسبات با استفاده از روابط (۳) و (۴)، ضرایب تابع همبستگی جزئی ($\hat{\phi}_{11}, \hat{\phi}_{22}, \dots, \hat{\phi}_{kk}$) را نتیجه می‌دهد. با در نظر داشتن ضرایب توابع همبستگی و همبستگی جزئی می‌توان یک مدل ARIMA مناسب برازش کرد. لازم به ذکر است در این مطالعه از نرم‌افزار Statgraphics برای برازش مدل ARIMA بر روی داده‌های کم‌نوسان استفاده شده است. طبق نتایج محاسبات خودکار نرم‌افزار، مدل ARIMA (2,2,1) به‌عنوان مدلی مناسب برازش شده است.

($\hat{x}_t = \hat{l}_t + \hat{h}_t$). در ادامه گام‌های الگوریتم شرح داده شده است:

الف) هریک از داده‌های در اختیار را با استفاده از فیلتر میانگین متحرک به دو قسمت کم‌نوسان و پر‌نوسان تقسیم کنید ($x_{tr} =$ کم‌نوسان بوده و مقادیر باقی‌مانده معادل جز پر‌نوسان هستند ($x_{res} = x_t - x_{tr}$)). مقادیر به‌دست آمده x_{tr} نشان‌دهنده جز

ب) یک مدل ARIMA مناسب بر روی مقادیر x_{tr} برازش کنید تا روابط میان متغیرها شناسایی شود.

ج) یک شبکه عصبی سه لایه‌ای بر روی مقادیر x_{res} برازش کنید.

د) نتیجه نهایی از حاصل جمع خروجی‌های ARIMA و ANN به‌دست می‌آید.

۲-۲. توسعه روش ترکیبی ARIMA-Local

تفاوت روش ترکیبی جدید پیشنهادی با روش تعریف شده توسط بابو و ردی [۲۰] در تجزیه و تحلیل بخش غیر خطی داده‌های داده شده است. در این روش، یک روش محلی برای شناسایی روابط داده‌های غیر خطی و کشف ویژگی‌های آن وجود دارد. این روش در ابتدا داده را به دو قسمت خطی و غیرخطی تقسیم می‌کند. سپس مشابه گام‌های الف و ب الگوریتم ARIMA-ANN یک مدل ARIMA بر روی قسمت‌های خطی داده‌ها برازش می‌کند. در ادامه گام‌های الگوریتم برای مدل‌سازی قسمت غیرخطی داده‌ها تشریح شده است. فرض شده است که برای پیش‌بینی مقدار متغیر h_{s+T} از T داده قبلی ($h_s, h_{s+1}, \dots, h_{s+T-1}$) استفاده شده است.

الف) در T سری داده‌های گذشته، k نقطه را بیابید که فاصله $\forall t = \|h_t - h_s\|, 1, 2, \dots, T$ کمینه شود. به عبارت دیگر، k نقطه مجاور نزدیک به h_s را بیابید. این نقاط را \hat{h}_t بنامید.

ب) یک ماتریس H تعریف کنید که سطرهای آن برابر \hat{h}_t باشد.

ج) مقادیر آتی \hat{h}_{t+T} را در یک بردار Y قرار دهید و یک مدل خطی $Y = WH$ برازش کنید. به‌طور معمول، $w = (H^T H)^{-1} H^T Y$ می‌باشد.

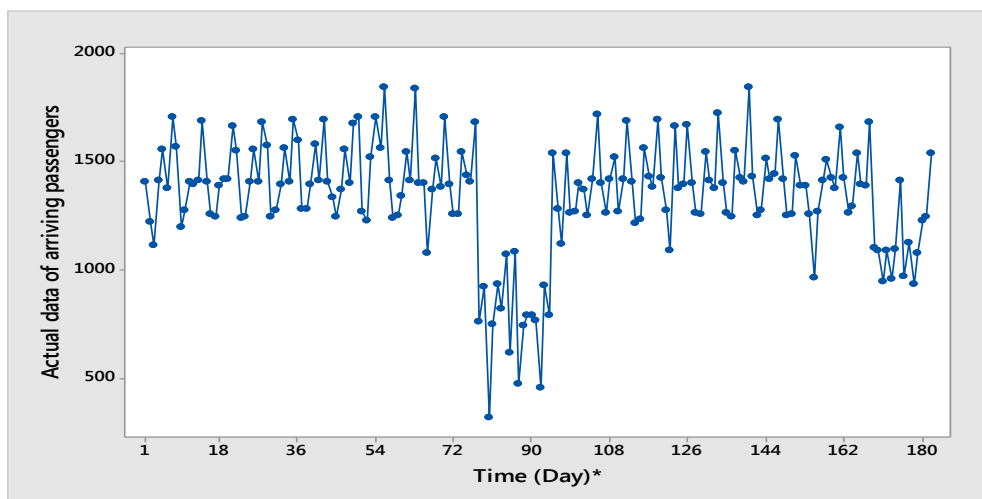
د) مقادیر آتی را براساس رابطه زیر پیش‌بینی کنید:

$$\hat{h}_{s+T} = wh_s$$

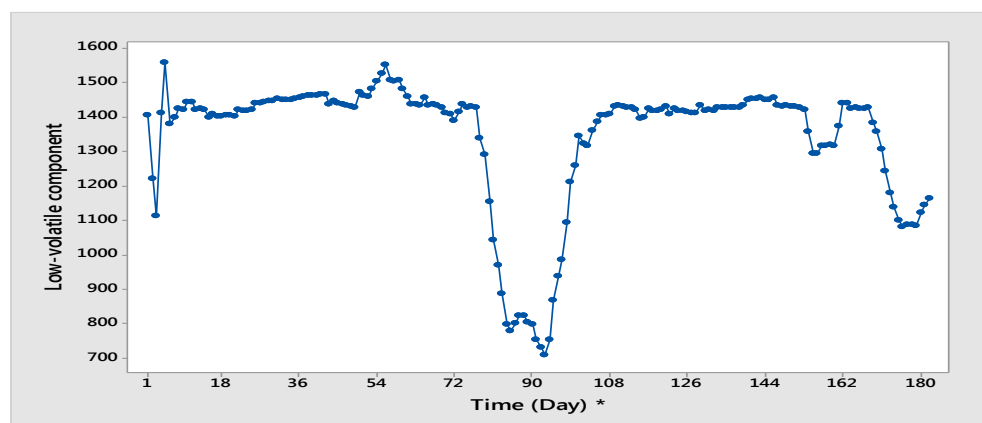
این یک روش خودرگرسیونی محلی است و هرچه k به N نزدیک می‌شود، این مدل عمومی می‌شود. در بخش بعدی جزئیات پیاده‌سازی هر دو رویکرد ترکیبی بر روی مطالعه موردی تشریح شده است.

۳. مطالعه موردی

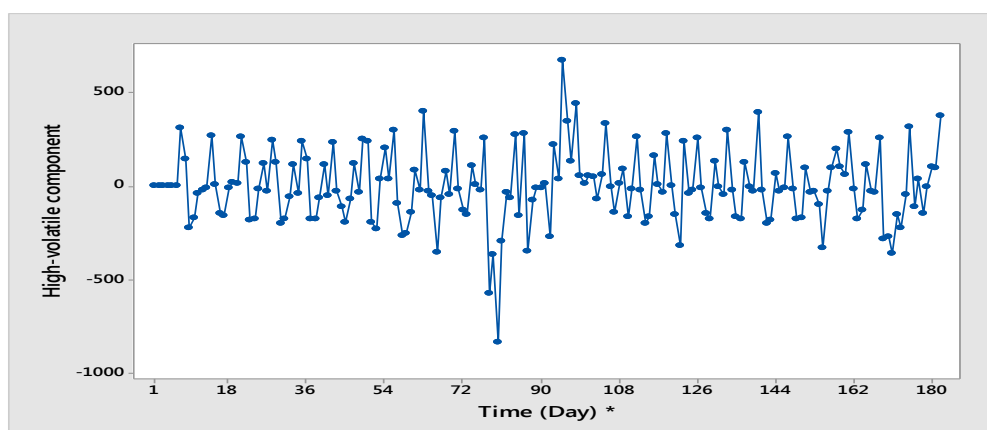
مدیران یک شرکت هواپیمایی در ایران، با افزایش میزان سرمایه در سازمان در پایان سال ۱۳۹۹ تعداد پروازها و مسافری خود را در سال ۱۴۰۰ افزایش داده‌اند. در این راستا، مدیریت تصمیم به ارائه خدمات فرودگاهی جدید برای مسافری پروازهای ورودی به فرودگاه یکی از کلان‌شهرهای ایران را دارد. ایجاد و راه‌اندازی امکانات و تسهیلات جدید نیازمند صرف هزینه گزافی است که براساس مطالعات انجام شده این



شکل (۱): تعداد مسافرین ورودی به فرودگاه یک کلان‌شهر در ۶ ماه گذشته
* این اطلاعات مربوط به بازه زمانی ۱۰ دی ۱۳۹۹ تا ۹ تیر ۱۴۰۰ هستند



شکل (۲): قسمت کم‌نوسان داده‌های ۶ ماه گذشته



شکل (۳): قسمت پرنوسان داده‌های ۶ ماه گذشته

مرحله ۲: تخمین

در این مرحله یک تخمین با خطای حداقل مربعات برای داده‌های کم‌نوسان صورت می‌گیرد. نتایج حاصل از اجرای این مرحله بر روی شکل (۴) نشان داده شده است.

مرحله ۳: تست خطا

در این مرحله بایستی صحت مدل برازش شده سنجیده شود تا در

صورت عدم تطابق با اطلاعات واقعی، اصلاحات لازم صورت گیرد. در حالت مطلوب، مقادیر باقی‌مانده یا اصطلاحاً مقادیر خطا بایستی از نوع خطای سفید (با میانگین صفر) باشند.

این بدین معنی است که همبستگی مقادیر خطا هیچ ساختاری نداشته و تفاوت معنی‌داری با صفر نداشته باشند. چنانچه در یک سری زمانی، مقادیر خطا همبستگی داشته باشند، بایستی در مدل اصلاحاتی

ورودی است. تابع فعال‌سازی فقط به ورودی‌ها و وزن‌ها بستگی دارد. تابع خروجی که معمولاً استفاده می‌شود تابع سیگموئید است که به صورت زیر نمایش داده می‌شود:

$$O_j(\bar{x}, \bar{w}) = \frac{1}{1 + e^{A_j(\bar{x}, \bar{w})}} \quad (۶)$$

شکل (۶) بخش غیرخطی داده‌های اصلی و مقادیر تخمینی توسط مدل ANN را نشان می‌دهد.

گام ۴) پیاده‌سازی رویکرد ترکیبی ARIMA-ANN

شکل (۷) نتایج حاصل از پیاده‌سازی رویکرد ترکیبی ARIMA-ANN که توسط بابو و ردی [۲۰] توسعه یافته است را نشان می‌دهد.

گام ۵) پیاده‌سازی رویکرد ترکیبی ARIMA-Local

همان‌طور که در بخش ۲ ذکر شد، این مقاله یک روش ترکیبی جدید، روش ARIMA-Local را پیشنهاد می‌کند. الگوریتم محلی پیاده‌سازی شده، بخش غیرخطی داده‌های واقعی را تجزیه و تحلیل می‌کند. شکل (۸) نتایج نهایی برای مقادیر پیش‌بینی شده با این روش را گزارش می‌کند. براساس شواهد می‌توان گفت این روش رابطه میان داده‌ها را به خوبی شناسایی کرده و با تحلیل روابط خطی و غیرخطی، تقریب مناسبی از آن‌ها را به دست آورده است.

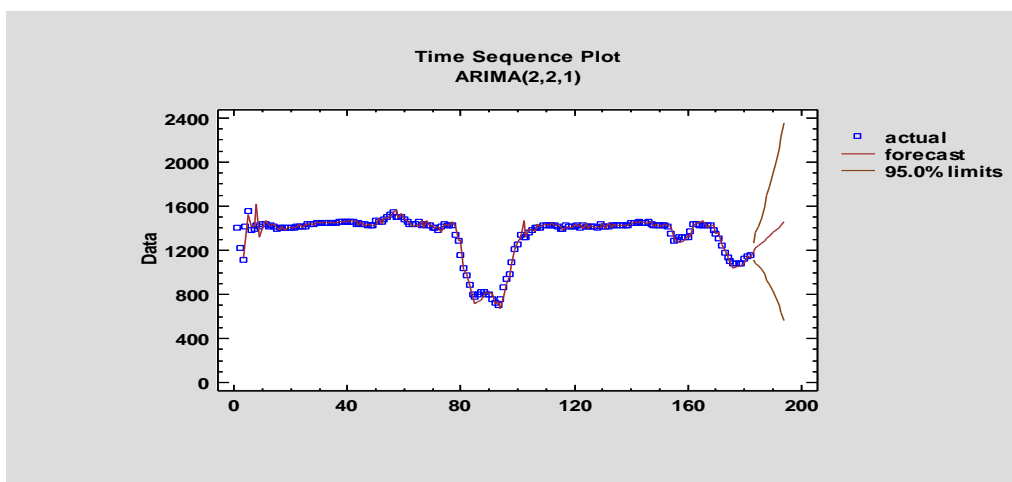
صورت بگیرد تا این همبستگی از بین برود. شکل (۵) همبستگی میان خطاهای داده‌های کم‌نوسان را نشان می‌دهد. با مشاهده شکل (۵) می‌توان نتیجه گرفت مدل برازش شده مدل مناسبی بوده و روابط میان داده‌ها را به خوبی نشان می‌دهد. جهت بررسی عدم همبستگی خطاهای داده‌های کم‌نوسان، آزمون استاندارد کای-اسکوئر پیاده‌سازی شده است. آماره آزمون عبارت است از: $Q = (N - d) \sum_{k=1}^K r_e^2(k)$ که N برابر است با تعداد داده‌ها، k نشان‌دهنده تعداد نمونه‌ها $K \leq N/3$ و $r_e(k)$ بیانگر خودهمبستگی نمونه‌هاست.

گام ۳) مدل شبکه عصبی مصنوعی (ANN)

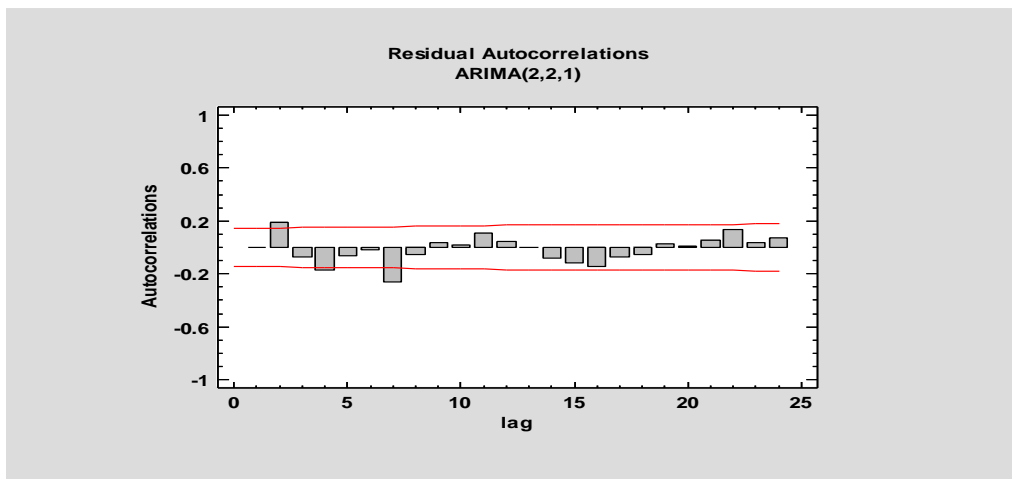
یک شبکه عصبی سه لایه‌ای برای پیش‌بینی مقادیر قسمت غیرخطی داده‌های اصلی استفاده می‌شود. براساس پارامتر میانگین متحرک، تعداد گره‌های لایه ورودی ۷ عدد است که با توجه به میانگین مربعات خطا، این تعداد در لایه پنهانی برابر ۵ است. تابع فعال‌سازی نورون‌ها به صورت مجموع وزنی بوده که می‌توان آن را به صورت زیر تعریف کرد:

$$A_j(\bar{x}, \bar{w}) = \sum_{i=0}^n x_i w_{ji} \quad (۵)$$

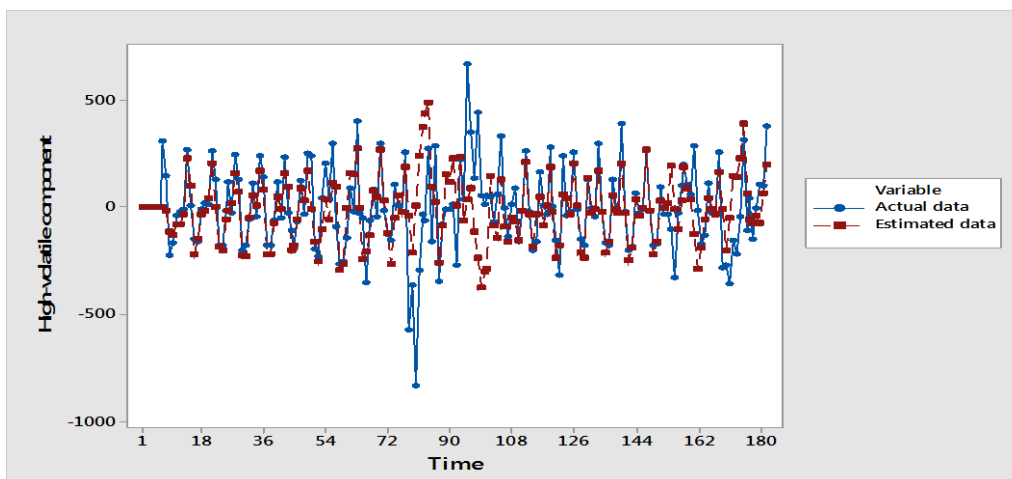
در این رابطه w_{ji} وزن اتصال یا لینک بین گره i در لایه ورودی و گره j در لایه پنهان را نشان می‌دهد و x_i نشان‌دهنده نامین گره



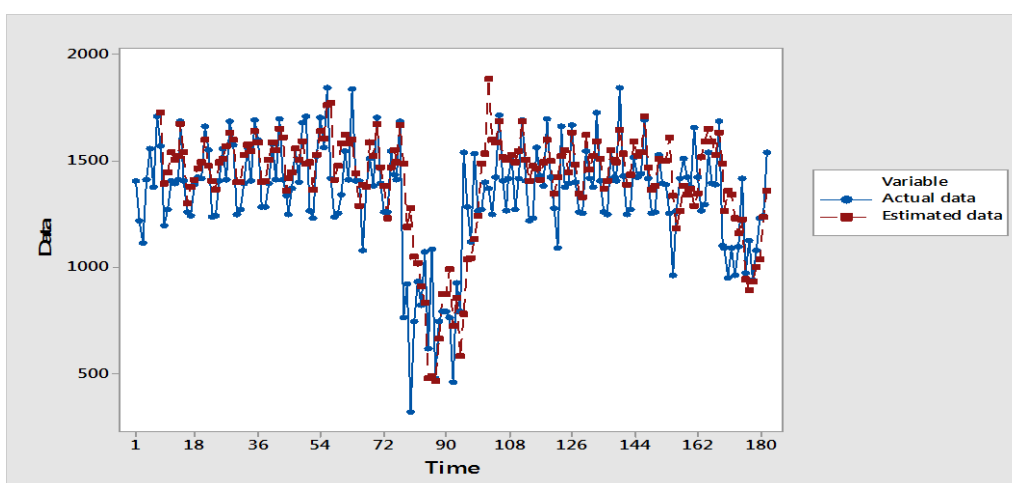
شکل (۴): مقادیر پیش‌بینی شده برای متغیر تعداد مسافرین ورودی در ۶ ماه



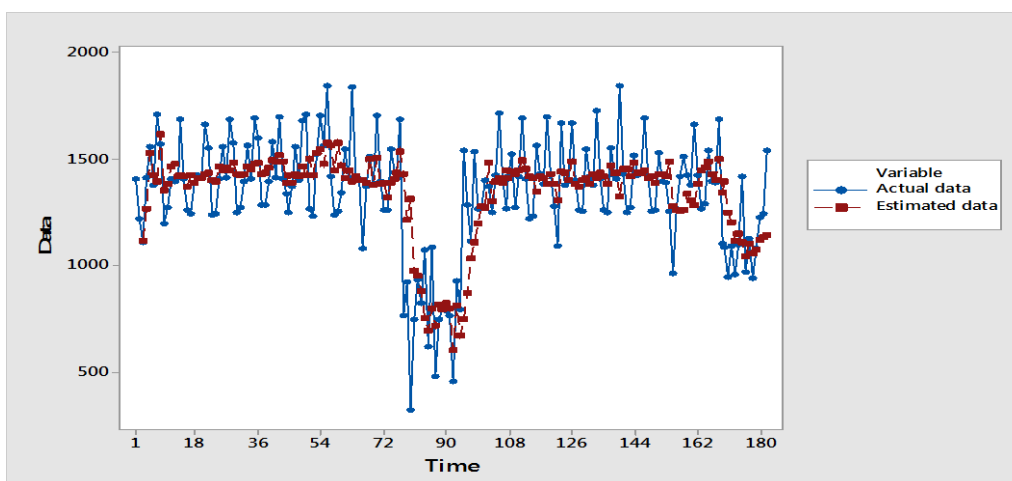
شکل (۵): همبستگی خطاهای داده‌های کم‌نوسان



شکل (۶): تخمین بخش غیر خطی داده‌های ۶ ماه



شکل (۷): نتایج مدل ARIMA-ANN



شکل (۸): نتایج مدل ARIMA-Local

۴. بحث

(MSE) برای ارزیابی دقت پیش‌بینی مدل‌های ARIMA-ANN و ARIMA-Local در مقایسه با روش هموارسازی نمایی ($\alpha = 0.3055$)، میانگین متحرک خودرگرسیون (ARMA) و شبکه عصبی مصنوعی استفاده شده‌اند.

در این بخش، قابلیت و عملکرد مدل‌های ترکیبی پیاده‌سازی در پیش‌بینی مقادیر سری‌های زمانی ارزیابی می‌شود. در این مقاله، دو معیار عملکرد، میانگین خطای مطلق (MAE) و میانگین مربعات خطا

در سال‌های گذشته قابل استناد نیستند و تنها اطلاعات ۶ ماه اخیر برای پیش‌بینی تعداد مسافری قابل استفاده است. مروری بر ادبیات موضوع مدیریت ترافیک نشان می‌دهد که در این حوزه هیچ توجهی به روش‌های ترکیبی پیش‌بینی نشده است. در سال‌های اخیر، رویکردهای ترکیبی متنوعی توسعه یافته‌اند که می‌توانند هر دو جنبه خطی و غیرخطی بودن داده‌ها را در نظر بگیرند. در این مطالعه نیز، داده‌ها به دو قسمت کم‌نوسان و پرنوسان تقسیم‌بندی شده و مدل‌های خطی (ARIMA) و غیرخطی (ANN و Local) مناسب برای پیش‌بینی مقادیر سری‌های زمانی برازش شده‌اند.

با وجود اینکه ANN یک مدل غیرخطی بسیار پرکاربرد برای سنجش روابط غیرخطی داده‌ها است، محققین یک روش پیش‌بینی محلی دیگر نیز برای سنجش روابط غیرخطی توسعه داده‌اند. نتایج حاصل از پیاده‌سازی هر دو روش ترکیبی ARIMA-ANN و ARIMA-Local عملکرد و دقت بالایی در مقایسه با روش‌های انفرادی مطرح در ادبیات دارند. جهت مقایسه روش‌های مختلف، دو معیار عملکرد میانگین خطای مطلق و میانگین مربعات خطا استفاده شده‌اند.

از آنجایی که در برخی از حوزه‌ها داده‌ها ماهیت فصلی داشته و از روند مشخصی تبعیت می‌کنند، این نوع داده‌ها نیز بسیار حائز اهمیت می‌باشند. بنابراین، با توجه به نتایج حاصله از این تحقیق، برای سایر محققین پیشنهاد می‌شود که روش‌های مطرح شده در این مقاله را برای داده‌های فصلی به کار گرفته و قابلیت رویکردهای ترکیبی در انجام پیش‌بینی برای داده‌های فصلی را ارزیابی کنند. علاوه بر این، می‌توان ترکیب‌های دیگری از روش‌های خطی و غیرخطی، مانند ترکیب روش ARIMA و GARCH، را توسعه داده و با روش‌های مشروح در این مقاله مقایسه کرد. علاوه بر دو معیار عملکرد MAE و MSE، روش شبیه‌سازی مونت کارلو نیز ابزاری مناسب برای سنجش دقت و عملکرد روش‌های پیش‌بینی سری‌های زمانی می‌باشد.

تقدیر و تشکر

نویسندگان این مقاله از شرکت هواپیمایی (X) که اطلاعات خود را در اختیار محققین قرار داده و همکاری لازم جهت جمع‌آوری اطلاعات اولیه را داشته‌اند تقدیر و تشکر می‌نمایند.

مراجع

- [1] Khashei, M., Bijari, M. (2011). A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *Applied Soft Computing*, 11: 2664-2675.
- [2] شیخ‌الاسلامی، عبدالرضا، خاکسار، حسن، احمدی، نادر. ۱۳۹۳. برنامه‌ریزی حمل‌ونقل: ارائه روش‌های متداول ادهاک و سری‌های زمانی در پیش‌بینی مؤلفه‌های ترافیکی، هشتمین کنگره ملی مهندسی عمران، ۱۳۹۳.
- [3] خسروشاهی، حسین، معطر حسینی، سید محمد، مرجانی، محمدرضا. ۱۳۹۳. اندازه‌گیری اثر شلاق چرمی در یک زنجیره تأمین خطی سه‌سطحی با استفاده از روش میانگین متحرک برای برآورد تقاضا، پژوهش‌های مهندسی صنایع در سیستم‌های تولید، شماره ۴، ۲۱-۳۷.

روابط (۷) و (۸) به ترتیب ساختار ریاضی خطای MAE و MSE را نشان می‌دهند [۲۳].

$$MAE = \frac{\sum_{t=1}^n |x_t - \hat{x}_t|}{n} \quad (7)$$

$$MSE = \frac{\sum_{t=1}^n (x_t - \hat{x}_t)^2}{n} \quad (8)$$

در این رابطه x_t مقدار پیش‌بینی شده برای متغیر در لحظه زمانی t ، \hat{x}_t مقدار واقعی متغیر در لحظه زمانی t ، و n تعداد داده‌ها را نشان می‌دهد. به طور کلی، MAE بیانگر متوسط خطای خالص در افق برنامه‌ریزی و MSE میانگین مربعات خطا را در این مدت‌زمان نشان می‌دهد [۲۰]. هر پنج روش ARIMA-Local، ARIMA-ANN، هموارسازی نمایی، ARMA و شبکه عصبی مصنوعی بر روی کلیه داده‌های گذشته که در دسترس هستند، پیاده‌سازی شده‌اند. نتایج حاصل از این پیاده‌سازی‌ها مقادیر پیش‌بینی شده (x_t) هستند که با مقادیر واقعی در دست به‌ازای هر نقطه زمانی، مقایسه می‌شوند. هرچه میزان خطای میان مقادیر واقعی و مقادیر پیش‌بینی شده کمتر باشد، آن روش دقت و عملکرد بالایی دارد. جدول (۱) مقادیر خطای MAE و MSE را برای روش‌های ARIMA-Local، ARIMA-ANN، هموارسازی نمایی، ARMA و شبکه عصبی مصنوعی جهت مقایسه دقت هر یک از روش‌ها نشان می‌دهد.

بر اساس این نتایج می‌توان گفت که روش ARIMA-ANN از عملکرد و دقت بسیار بالایی برخوردار است. روش ARIMA-Local نیز در مقایسه با روش‌های انفرادی پرکاربرد در ادبیات، عملکرد مناسبی داشته است. علت این امر این است که روش‌های ترکیبی هر دو جنبه خطی و غیرخطی داده‌ها را به‌طور هم‌زمان در نظر گرفته‌اند و مقادیر پیش‌بینی شده به مقادیر واقعی نزدیک‌تر هستند.

جدول (۱): مقایسه معیار عملکرد به‌ازای هر یک از روش‌ها

روش	MAE	MSE
ARIMA-ANN	۲۳/۰۸	۳۶/۶۴
ARIMA-Local	۱۵۳/۰۵	۱۷۰/۵۹
هموارسازی نمایی	۱۶۵/۳۳	۲۱۷/۰۶
ARMA	۱۶۸/۰۵	۲۲۱/۳۴
شبکه عصبی مصنوعی	۱۶۰/۷۶	۲۰۵/۴۳

۶. نتیجه‌گیری

امروزه، با رشد روزافزون ساختارها و تجهیزات حمل‌ونقل، مدیریت ترافیک یکی از موضوعات مورد توجه محققین قرار گرفته است. یکی از موضوعات اصلی در این حوزه، پیش‌بینی متغیر ترافیک شامل جریان و زمان سفر است. در مقاله جاری، فرودگاه یک کلان‌شهر در ایران به‌طور موردی مورد مطالعه قرار گرفته است. در این مثال، مدیران یک شرکت هواپیمایی علاقه‌مند به ارائه خدمات فرودگاهی جدید به مسافری پروازهای ورودی به فرودگاه هستند. باتوجه به سیاست‌های جدید شرکت، تغییرات چشم‌گیری در تعداد پروازها و مسافری اتفاق افتاده است که بدین سبب اطلاعات مربوط به تعداد مسافری ورودی

- [4] غفاری، فرهاد، فرهادی چشمه مرواری، عقیق. ۱۳۹۴. بررسی توان پیش‌بینی مدل‌های ARIMA-GARCH، GARCH، ARIMA و State space به کمک روش شبیه‌سازی مونت کارلو. اقتصاد کاربردی، شماره ۱۶، ۳۳-۴۲.
- [5] Xumei, C., Huibo, G., Wang, J. (2012). BRT vehicle travel time prediction based on SVM and Kalman filter. *Journal of Transportation Systems Engineering and Information Technology*, 12: 29-34.
- [6] Jiang, Z., Zhang, C., Xia, Y. (2014). Travel Time Prediction Model for Urban Road Network based on Multi-source Data. *Procedia- Social and Behavioral Sciences*, 138: 811-818.
- [7] Haworth, J., Shawe-Taylor, J., Cheng, T., Wang, J. (2014). Local online kernel ridge regression for forecasting of urban travel times. *Transportation Research Part C: Emerging Technologies*, 46: 151-178.
- [8] Elhenawy, M., Chen, H., Rakha, H. A. (2014). Dynamic travel time prediction using data clustering and genetic programming. *Transportation Research Part C: Emerging Technologies*, 42: 82-98.
- [9] Sun, J., Sun, J. (2015). A dynamic Bayesian network model for real-time crash prediction using traffic speed conditions data. *Transportation Research Part C: Emerging Technologies*, 54: 176-186.
- [10] Sadaei, H.J., Guimarães, F.G., da Silva, C.J., Lee, M.H., and Eslami, T. (2017). Short-term load forecasting method based on fuzzy time series, seasonality and long memory process. *International Journal of Approximate Reasoning*, 83: 196-217.
- [11] Murça, M.C.R., and Hansman, R.J. (2018). Predicting and planning airport acceptance rates in metroplex systems for improved traffic flow management decision support. *Transportation Research Part C: Emerging Technologies*, 97: 301-323.
- [12] Kong, X., Xu, Z., Shen, G., Wang, J., Yang, Q., Zhang, B. (2016). Urban traffic congestion estimation and prediction based on floating car trajectory data. *Future Generation Computer Systems*. 61: 97-107.
- [13] Xu, S., Chan, H.K., Zhang, T. (2019). Forecasting the demand of the aviation industry using hybrid time series SARIMA-SVR approach. *Transportation Research Part E: Logistics and Transportation Review*, 122: 169-180.
- [14] Qu, L., Li, W., Li, W., Ma, D., Wang, Y. (2019). Daily long-term traffic flow forecasting based on a deep neural network. *Expert Systems with applications*, 121: 304-312.
- [15] Villarroya, C., Calafate, C. T., Onaindia, E., Cano, J. C., Martinez, F. J. (2022). Neural Network-based Model for Traffic Prediction in the City of Valencia, 26th International Conference on Knowledge-Based and Intelligent Information & engineering Systems (KES 2022), 552- 562.
- [16] Sadeghi Gargari, N., Panahi, R., Akbari, H., Ng, A.K.Y. (2022). Long-Term Traffic Forecast Using Neural Network and Seasonal Autoregressive Integrated Moving Average: Case of a Container Port, *Transportation Research Record*, 2676: 236-252.
- [17] Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50: 159-175
- [18] Valenzuela, O., Rojas, I., Rojas, F., Pomares, H., Herrera, L. J., Guillén, A. Pasadas, M. (2008). Hybridization of intelligent techniques and ARIMA models for time series prediction. *Fuzzy Sets and Systems*, 159: 821-845.
- [19] Faruk, D. Ö. (2010). A hybrid neural network and ARIMA model for water quality time series prediction. *Engineering Applications of Artificial Intelligence*, 23: 586-594.
- [20] Babu, C. N., Reddy, B. E., 2014. A moving-average filter-based hybrid ARIMA-ANN model for forecasting time series data. *Applied Soft Computing*, 23: 27-38.
- [21] Box, G. E., Jenkins, G. M., Reinsel, G. C., Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.
- [22] Yaghini, M., Khoshraftar, M. M., Fallahi, M. (2013). A hybrid algorithm for artificial neural network training. *Engineering Applications of Artificial Intelligence*, 26: 293-301.
- [23] Toro, C. H. F., Meire, S. G., Gálvez, J. F., Fdez-Riverola, F. (2013). A hybrid artificial intelligence model for river flow forecasting. *Applied Soft Computing*, 13, 3449-3458.



DOI: 10.22084/ier.2023.27356.2114

Implementation of Hybrid ARIMA-ANN and ARIMA- Local Prediction Techniques in the Traffic Management System. Case Study: Airport of a Metropolis

S. Sadri¹, S. M. T. Fatemi Ghomi^{2*}

¹ Ph.D. student, Department of Industrial Engineering, Faculty Industrial Engineering, Amirkabir University of Technology, Tehran, Iran

² Professor, Department of Industrial Engineering, Faculty of Industrial Engineering, Amirkabir University of Technology, Tehran, Iran

ARTICLE INFO

Article history:

Received: 2022/05/15

Accepted: 2022/06/29

Keywords:

Artificial Neural Network (ANN)
Auto-Regressive Integrated
Moving Average (ARIMA)
Hybrid ARIMA-ANN
Hybrid ARIMA-Local Method
Intelligent Traffic Management

ABSTRACT

Nowadays, time series prediction is one of the fundamental research purposes, owing to the importance of prediction in various real-world applications. Transportation management is one of the main issues of each municipality that needs prediction. Accurate and reliable forecasting is one of the fundamental goals of an intelligent transportation system. In the literature, it is defined several variables of traffic management such as speed, time, and flow. In the proposed paper, a case of an Airline is considered. The associated managers are planning to propose a new service for passengers. There is a need to predict the number of passengers on arriving flights to a metropolis in Iran. This variable is inherently similar to the flow in traffic management. In the literature on traffic management, most of the studies implemented a linear or nonlinear modeling method to predict the future and ignore the advantage of hybrid methods. Several hybrid ARIMA-ANN methods have been proposed to specify the underlying relationships among the data. This paper utilizes a hybrid ARIMA-ANN model which decomposes the data into low-volatile, and high-volatile components to predict accurately. Also, the current paper develops a new hybrid method, ARIMA-Local method, to specify the efficiency of other provided nonlinear methods in a hybrid structure. The obtained results for the discussed case are reported. This study signifies the accuracy of ARIMA-ANN model in predicting, while also the ARIMA-Local method is efficient in forecasting in comparison to the individual models of ANN and Exponential smoothing.

* Corresponding author. S. M. T. Fatemi Ghomi
Tel.: 021-64545381; E-mail address: fatemi@aut.ac.ir