مدیریت فناوری اطلاعات
Journal of Information Technology Management
دانشکدهٔ مدیریت دانشگاه تهران
دورهٔ ۸، شمارهٔ ۱
بهـار ۱۳۹۵
ص. ۲۶- ۱

کاربرد قواعد کشفی و الگوریتم ژنتیک در ساخت مدل ARMA برای پیش‌بینی سری زمانی

محمدرضا اصغری اسکوئی[۱]، محمد قاسم‌زاده[۲]

چکیده: برای پیش‌بینی سری زمانی ابتدا باید مدل مناسبی از آن ساخته شـود. تعیـین ابعـاد و تخمین پارامترهای مناسب برای مـدل ARMA سـری زمـانی، چالشـی اسـت کـه عـلاوه بـر روش‌های متداول آماری، از طریق محاسبات هوشمند نیز به آن توجه شده است. در ایـن مقالـه استفاده از الگوریتم ژنتیک برای تخمین پارامترهای مدل ARMA و قواعد کشفی برای تعیـین ابعاد مدل ارائه می‌شود. قواعد کشفی براساس ویژگی‌هـای سـری زمـانی اسـتخراج مـی‌شـوند. داده‌ها به‌روش پنجرهٔ لغزان در پیش‌بینی به‌کار می‌روند. مدل بر اساس معیار اطلاعاتی بیـزین و پیش‌بینی بر اساس دو معیار مجذور متوسط مربعات خطـا و متوسـط قـدر مطلـق درصـد خطـا ارزیابی می‌شود. روش ارائه‌شده روی هشت سری زمانی با ویژگی‌های مختلـف بـه‌کـار رفتـه و نتایج آن با نتایج روش آماری مقایسه شده است. نتایج به‌دسـت‌آمـده نشـان مـی‌دهـد در تمـام موارد، روش ارائه‌شده همسان یا بهتر از روش کلاسیک عمل می‌کند.

واژه‌های کلیدی: *الگوریتم ژنتیک، پنجرهٔ لغزان، پیش‌بینی سری زمانی، تخمـین مـدل ARMA، قواعد کشفی.*

۱. استادیار رایانه، دانشکدهٔ علوم ریاضی و رایانه، دانشگاه علامه طباطبائی، تهران، ایران
۲. دانشجوی کارشناسی ارشد علوم رایانه ـ گرایش سیستم‌های هوشمند، دانشکدهٔ علـوم ریاضـی و رایانـه، دانشـگاه علامه طباطبائی، تهران، ایران

تاریخ دریافت مقاله: ۱۳۹٤/۰٦/۰۹
تاریخ پذیرش نهایی مقاله: ۱۳۹٤/۱۰/۱٤
نویسندهٔ مسئول مقاله: محمد قاسم‌زاده
E-mail: Ghasemzade.mohammad@aut.ac.ir

مقدمه

شناخت ماهیت تغییرات در سری‌های زمانی متنوعی چون بازارهای مالی، اقتصاد، تولید، کنترل و منابع طبیعی برای صاحبان بنگاه‌های اقتصادی، سرمایه‌گذاران، مشتریان، مسئولان و حتی مردم عادی، اهمیت ویژه‌ای دارد و توجه بسیاری از پژوهشگران را به خود جلب کرده است. تحلیل درست از روند تغییرات و امکان پیش‌بینی آن در برنامه‌ریزی و تصمیم‌سازی اهمیت بسیار زیادی دارد (اسکویی، ۱۳۹۴). ضرورت‌های پژوهشی این مقاله به سازمان‌ها، سرمایه‌گذاران و تمام کسانی اختصاص دارد که درصدد برنامه‌ریزی برای آینده‌اند. درحقیقت پیش‌بینی دقیق، کلید حل مشکلات مردم و سازمان‌های تصمیم‌گیرنده است. امروزه در بسیاری از رشته‌ها، حوزه‌های علمی، مدیریتی و اقتصادی، علوم تجربی و غیره، استفاده از داده‌های به‌دست‌آمده و تحلیل آنها در تصمیم‌گیری‌ها اهمیت بسیاری دارد؛ زیرا در صورت موفقیت پیش‌بینی و برنامه‌ریزی، می‌توان سبب کاهش هزینه، افزایش دقت و آمادگی بیشتر در تصمیم‌گیری شد.

تغییرات و پیش‌بینی سری‌های زمانی از طریق دو رویکرد متفاوت بنیادین (توصیفی) و فنی (تکنیکال)، مطالعه و مدل‌سازی می‌شود. رویکرد بنیادین مبتنی بر مفاهیم و نظریه‌های مطرح در علم آن سری زمانی است. در این رویکرد پس از مطالعهٔ رابطهٔ متغیرهای مختلف و استخراج مدل ریاضی، سری زمانی پیش‌بینی می‌شود. رویکرد فنی بدون توجه عمیق به ماهیت معیارهای سری زمانی و فقط با مطالعهٔ روند تغییرات در دوره‌های گذشته، به پیش‌بینی تغییرات آتی مبادرت می‌کند (اسکویی، ۱۳۹۴).

در این مقاله به‌منظور مدل‌سازی و پیش‌بینی سری‌های زمانی، رویکرد فنی انتخاب شده است و درواقع هدف، پیش‌بینی رفتار سیستم است، نه روش کار آن. الگوریتم ژنتیک با رمزگذاری حقیقی برای تخمین ضرایب ARMA به‌کار می‌رود و مدل‌ها با بهره‌مندی از قواعد کشفی ساخته‌شده و با به‌کارگیری معیار اطلاعات بیزی کنترل می‌شوند و به‌کمک جریمهٔ پیچیدگی مدل، از بیش‌برازش جلوگیری می‌شود. اهداف این پژوهش ابتدا مدل‌سازی برای سری‌های زمانی و سپس بهینه‌سازی پارامترهای مدل و درنهایت پیش‌بینی با بهره‌مندی از بهترین مدل ساخته شده است. در نهایت، خطای پیش‌بینی روش مطرح‌شده در مقاله با روش‌های کلاسیک مقایسه خواهد شد.

پیشینهٔ پژوهش

به‌طورکلی، روش‌های پیش‌بینی سری‌های زمانی را می‌توان به دو دسته تقسیم کرد: روش‌های کلاسیک و روش‌های هوشمند. مدل‌های کلاسیک به مدل‌های آماری‌ای گفته می‌شود که از

ابتدای تحلیل‌های سری زمانی به آنها پرداخته شده است. روش‌های هوشمند نیز به‌کمک سامانه‌های محاسباتی و محاسبات هوشمند در دهه‌های اخیر تحقیق‌شده و توسعه یافته‌اند. مدل‌های کلاسیک سری زمانی خود به دو بخش تقسیم می‌شوند: مدل‌های ایستا و مدل‌های غیرایستا. مدل‌های ایستا نیز به انواع مدل‌های خودگردان[1] یا AR، میانگین متحرک[2] یا MA و میانگین متحرک خودگردان[3] یا ARMA دسته‌بندی می‌شوند و مدل غیرایستا با عنوان میانگین متحرک خودگردان یکپارچه[4] یا ARIMA شناخته می‌شود.

مدل‌های تصادفی ARMA توصیف‌کنندهٔ رفتار سیستم خطی تحت تأثیر اغتشاش پنهان گاوسی قرار دارند. این مدل‌ها قابلیت مدل‌سازی شهودی و پیش‌بینی مؤثری را فراهم می‌کنند، اما متأسفانه الگوریتمی با کارایی زیاد برای محاسبهٔ پارامترهای ماکزیمم‌کنندهٔ احتمال مرزی داده‌های ARMA وجود ندارد. درنتیجه تخمین‌گرهای محلی اکتشافی[5] به‌صورت تئوری ارائه و بررسی شده‌اند. البته هیچ‌یک از این تخمین‌گرها تضمینی برای یافتن پارامترهای بهینهٔ سراسری ایجاد نمی‌کنند (وایت و ون، ۲۰۱۵). مینیمم‌های چندگانه معمولاً در رویهٔ تابع خطای هدف مدل‌های سری‌های زمانی وجود دارند. جواب‌های روش‌های کلاسیک در نقاط بهینهٔ محلی به دام می‌افتند و مدل ضعیفی تولید می‌کنند (چاو مینگ و هانگ تیزر، ۱۹۹۵).

مباحث محاسبات هوشمندِ پژوهش در عملیات سبب ارائهٔ روش‌های کمّی‌تر عددی در پیش‌بینی سری‌های زمانی شده است که جایگزین روش‌های کلاسیک می‌شوند. در دو دههٔ اخیر پیش‌بینی سری‌های زمانی غیرخطی ظهور پیدا کردند که در این میان می‌توان به شبکه‌های عصبی مصنوعی به‌عنوان یکی از روش‌های مشهور اشاره کرد. شبکه‌های عصبی نیز در موارد متعددی چون طبقه‌بندی و خوشه‌بندی و تشخیص الگو به‌کار گرفته شده است. قاسمی و اصغری‌زاده در سال ۱۳۹۳ در مطالعه‌ای به‌کمک شبکهٔ عصبی، خوشه‌بندی شرکت‌های مختلف را با توجه به شاخص‌های مالی بررسی کردند. همچنین وثوق، تقوی‌فرد و البرزی (۱۳۹۳) در مقاله‌ای با توجه به قابلیت زیاد شبکهٔ عصبی در بازشناسی الگو، توانسته‌اند با دقت نسبتاً زیادی تراکنش‌های سالم و متقلبانهٔ بانکی را از هم تشخیص دهند. شبکه‌های عصبی مصنوعی نیز برای پیش‌بینی‌هایی همچون جریان رودخانه به‌کار گرفته شده‌اند. از این روش برای تخمین پارامترهای ARMA استفاده می‌شود. چنووت و همکارانش در سال ۲۰۰۰ نشان دادند

---

1. Autoregressive
2. Moving average
3. Autoregressive - Moving average
4. Autoregressive integrated Moving average
5. Heuristic

درصورتی‌که تعداد نقاط داده‌ها کمتر از ۱۰۰ باشد، شبکه‌های عصبی مصنوعی نمی‌توانند مرتبه‌های مدل ARMA را به‌صورت دقیق تخمین بزنند (محمدی و اسلامی، ۲۰۰۶).

دربارهٔ استفاده از هوش محاسباتی در کنار روش‌های کلاسیک، برای مدل‌سازی و پیش‌بینی سری‌های زمانی مطالعات فراوانی شده است. ازجمله می‌توان به استفاده از شبکهٔ عصبی پیشرو (اسکویی، ۱۳۸۳)، شبکهٔ عصبی بازگشتی (کانر، مارتین و اطلس، ۱۹۹۴)، سیستم استنتاج فازی (گریوز و پدریکز، ۲۰۰۹)، سیستم خبره (حمزاپیی، آکای و کوتای، ۲۰۰۹)، سیستم ترکیبی فازی‌ـ عصبی (ژانگ، ۲۰۰۸؛ اسکار و پاتریکا، ۲۰۰۲)، بردار حائل (گستل و همکاران، ۲۰۰۱؛ ساپانکویچ و سانکار، ۲۰۰۹)، سیستم آشوب‌گونه (ژانگ، چونگ و لو، ۲۰۰۸) و الگوریتم‌های تکاملی (کورتز، روچا و نوس، ۲۰۰۴) اشاره کرد. همچنین شعبان‌الهی، رشیدی و صادقی در سال ۱۳۹۴ با استفاده از الگوریتم ژنتیک و الگوریتم جانسون در سیستم خبره، رفتار مصرف انرژی کارکنان را تحلیل کردند. کریمی (۱۳۹۴) نیز در مطالعه‌ای، سیستم خبرهٔ فازی را برای حوزهٔ تبادلات الکترونیکی دولت و کسب‌وکارها به‌کار برد. البته در مواردی که پیش‌بینی خطی نیاز باشد، همچنان می‌توان از روش‌های قدیمی بهره گرفت (ماکریداکیس و ویلورایت، ۱۹۹۸). روش‌های کلاسیک پیش‌بینی، محدودیت‌های محاسباتی بسیاری دارند و پارامترهای آن به‌کمک روش‌های عددی، مانند روش حداقل مربعات، بهینه‌سازی می‌شوند که نتایج آن به مینیمم‌های محلی ختم می‌شود (کورتز و همکاران، ۲۰۰۴).

الگوریتم‌های هوش مصنوعی شامل روش‌هایی چون شبکه‌های عصبی مصنوعی، روش فازی، الگوریتم ژنتیک و تحلیل پوششی داده[1] می‌شوند. در حالت کلی، مطالعهٔ الگوریتم‌های تکاملی ژنتیک، پتانسیل‌های هوش مصنوعی را تقویت کرده است. هولند در سال ۱۹۷۵ الگوریتم ژنتیک را پایه‌گذاری کرد. مفهوم کلی این روش بر پایهٔ شبیه‌سازی تکامل طبیعی یک جمعیت استوار است. جواب مسئله باید برای بقا تلاش کند و تناسب آن در هر نسل بهبود می‌یابد (اسماعیل و ینگ، ۲۰۱۱). الگوریتم ژنتیک و ترکیب این الگوریتم با سایر روش‌های هوشمند در سال‌های اخیر کاربردهای متعدد و متفاوتی داشته است. از این الگوریتم می‌توان برای انتخاب ویژگی‌های خاصی (رتبه‌بندی اعتباری) از پیش‌پردازش داده‌ها یا حل مسائل بهینه‌سازی ترکیبی استفاده کرد (تقوی‌فرد، حسینی و خان‌بابایی، ۱۳۹۳؛ بهشتی‌نیا و فرازمند، ۱۳۹۴).

سال‌های اخیر الگوریتم ژنتیک برای تخمین ساختار و پارامترهای مدل ARMA ارائه شد و در سال ۲۰۰۹ ماجهی، روت، میشرا و پاندا مدل PSO-ARMA را برای پیش‌بینی فروش

1. Data envelopment analysis

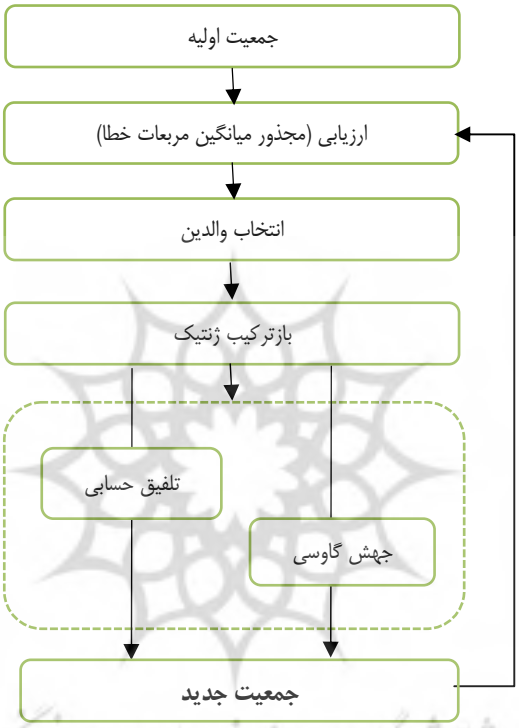
پیشنهاد دادند. به‌تازگی چند روش تکاملی مانند بهینه‌سازی ازدحام ذرات، تکامل تفاضلی[1]، بهینه‌سازی باکتریایی[2] و بهینه‌سازی گروهی گربه‌ای[3] به‌خوبی در زمینه‌های مختلفی به‌کار گرفته شده‌اند (روت و ماجهی، ۲۰۱۳). الگوریتم‌های تکاملی، روش‌های انطباق و بهینه‌سازی برمبنای نظریهٔ تکامل طبیعی داروین هستند. درواقع این الگوریتم‌ها به شکل دسته‌ای از روش‌ها بر پایهٔ اصول مشابهی هستند که اغلب برای بهینه‌سازی پارامترها کاربرد دارند. در این روش‌ها نقاط موجود در فضای جست‌وجو به‌صورت اعضای منفردی در محیطی مصنوعی مدل می‌شوند. برخلاف بیشتر روش‌های بهینه‌سازی، دسته‌ای از نقاط راه‌حل‌ها به‌عنوان جمعیت شناخته می‌شوند و در هر تکرار تغییر می‌کنند. شانس انتخاب‌شدن به‌عنوان والدین نسل بعد، به تناسب یا کیفیت اعضا در معیار ارزش‌گذاری که معمولاً تبدیلی ساده از تابع هدف است، بستگی دارد. نقاط جدید راه‌حل‌های جدید جست‌وجو، با استفاده از انتخاب و دو عملگر پایهٔ ژنتیک یعنی تلفیق[4] و جهش[5] مقداریابی می‌شوند. مزیت عمدهٔ الگوریتم‌های ژنتیک این است که به دانستن ویژگی‌های مسئلهٔ بهینه‌سازی مد نظر نیاز ندارند. آنها به ویژگی مشخصی از تابع هدف مانند تحدب[6]، قابلیت تمایز یا هموار، وابسته نیستند، البته به‌جز الگوریتم‌هایی که توانایی جست‌وجوی کل فضای پارامتری و دنبال‌کردن هر تغییر در سیستم بهینه‌سازی را دارند. الگوریتم‌های تکاملی در بیشتر موارد پاسخ‌های مناسب در زمان قابل قبولی ارائه می‌دهند. درحالی‌که روش‌های قطعی[7] معمولاً در مینیمم‌های محلی به دام می‌افتند (بلجیانیس، لیکوتاناسیس و دمیریس، ۲۰۰۱).

الگوریتم‌های تکاملی گزینهٔ مناسبی برای تخمین پارامترند؛ زیرا این الگوریتم‌ها با جست‌وجوی چندنقطه‌ای سراسری به‌سرعت مناطق باکیفیت را مشخص می‌کنند. استفاده از الگوریتم‌های تکاملی در پیش‌بینی سری‌های زمانی مزیت‌هایی چون ارائهٔ مدل صریح و جست‌وجوی انطباق‌یافتهٔ تکاملی دارند که از به دام افتادن در مینیمم‌های محلی نامطلوب جلوگیری می‌کنند. در گذشته الگوریتم‌های تکاملی به‌صورت رمزهای باینری برای بهینه‌سازی پارامترهای روش‌های پیش‌بینی کلاسیک مانند روش‌های هالت- وینترز[8] (آگاپای و آگاپای، ۱۹۹۷) یا ARMA (هانگ و یانگ، ۱۹۹۵) استفاده می‌شدند، اما بعدها بر استفاده از ژن‌هایی با

---

1. Differential Evolution (DE)
2. Bacterial Foraging Optimization (BFO)
3. Cat Swarm Optimization (CSO)
4. Crossover
5. Mutation
6. Convexity
7. Deterministic
8. Holt-Winters

مقادیر حقیقی متمرکز شدند (میچالهویسز، ۱۹۹۶؛ رولف و اسپریو، ۱۹۹۶؛ کورتز و همکاران، ۲۰۰۱)؛ زیرا این نمایش مستقیم، قابلیت مقیاس‌بندی بیشتری دارد و امکان تعریف عملگرهای ژنتیکی قوی‌تری را نیز فراهم می‌کند (کورتز و همکاران، ۲۰۰۴). شکل ۱ نشان‌دهندهٔ نمونه‌ای از رویهٔ الگوریتم تکاملی است.



شکل ۱. شمای الگوریتم تکاملی

با شبیه‌سازی فرایند تکامل طبیعی، الگوریتم‌های تکاملی قابلیت همگرایی به اکسترمم سراسری در رویه‌ای پیچیده را نشان داده‌اند. آزمایش‌های عددی نشان داده‌اند الگوریتم‌های تکاملی استفاده‌شده در تعیین مدل ARMA قابلیت تخمین هم‌زمان مرتبه‌ها و مقادیر پارامترها را به‌ازای ورودی‌های مختلف دارند (هانگ و یانگ، ۱۹۹۵).

کورتز و همکارانش در سال ۲۰۰۴ با بهره‌مندی از الگوریتم ژنتیک و قواعد کشفی، پارامترهای مدل ARMA را تعیین کردند. قواعد کشفی مجموعه‌ای از ویژگی‌های سری‌های زمانی است که میزان همبستگی‌های سری زمانی با خودش در تأخیرهای متفاوت را اندازه‌گیری می‌کند. این قواعد با انتخاب تأخیرها به‌صورت کشفی ابعاد ورودی را مشخص می‌کنند. ورودی

نیز به‌صورت متغیر از طریق روش پنجرهٔ لغزان تعیین می‌شود. این الگوریتم از انتخاب چرخهٔ رولت، تلفیق حسابی و جهش گاوسی استفاده می‌کند. اندازهٔ کروموزوم‌ها نیز با توجه به قواعد تعیین می‌شوند. عملگرهای الگوریتم ژنتیک و روش انتخاب در بخش‌های بعدی به‌طور کامل شرح داده خواهد شد.

اسکویی در سال ۱۳۹۴ به‌کمک روش پنجرهٔ لغزان و شبکهٔ عصبی، سری‌های زمانی مالی را پیش‌بینی کرد. در این روش پیش‌بینی به‌صورت مکرر روی زیرمجموعه‌های متوالی از سری زمانی صورت می‌گیرد. زیرمجموعهٔ متناهی و متوالی از سری زمانی پنجره نامیده می‌شود. اندازهٔ پنجره تعداد داده‌های درون آن است که درعمل یک دورهٔ زمانی را نشان می‌دهد. پنجره با فاصله یا قدم ثابت روی کل داده‌های سری زمانی حرکت داده می‌شود. پنجره‌های متوالی قسمت‌های مشترک دارند که به‌اصطلاح همپوشانی پنجره‌ها نامیده می‌شود. فرایند پیش‌بینی سری زمانی به‌صورت مکرر و مستقل از هم روی هر پنجره انجام می‌شود. اندازه و قدم پنجره، دو پارامتر تعیین‌کننده در مدیریت داده‌های سری زمانی و تعداد تکرار انجام پیش‌بینی است.

برای استفاده از داده‌ها به‌منظور پیش‌بینی دو رویکرد وجود دارد؛ در رویکرد اول برای پیش‌بینی فقط از داده‌های موجود استفاده می‌شود و در رویکرد دوم برای پیش‌بینی از مقادیر پیش‌بینی‌شده نیز استفاده می‌شود. بدیهی است که رویکرد دوم پیچیدگی بیشتری دارد و ممکن است گاهی به‌واسطهٔ پیش‌بینی‌های نامطلوب، پیش‌بینی‌های بعدی با خطای بیشتری ادامه پیدا کند و نتایج نامطلوبی به‌دست آید. قسمتی از داده‌ها که برای ساخت مدل، تخمین ضرایب و پارامترها استفاده می‌شود، داده‌های آموزش نام دارند. پس از مدل‌سازی قسمتی از داده‌ها نیز به‌عنوان داده‌های آزمایش مدل، استفاده می‌شود که داده‌های آزمون نام دارند. چگونگی انتخاب و تقسیم‌بندی داده‌ها خود چالشی برای مدل‌سازی و پیش‌بینی است.

از نظر ساختاری، سری‌های زمانی می‌توانند متشکل از مؤلفه‌های روند کلی، دوره‌ای و تصادفی باشند. مؤلفهٔ روند کلی، تغییرات سری زمانی را در دورهٔ طولانی نشان می‌دهد. درحالی‌که مؤلفهٔ دوره‌ای، تغییرات نوسانی در دورهٔ مشخصی را نشان می‌دهد. هر دو مؤلفه، الگوی معین ریاضی دارند و می‌توانند توصیف و مدل‌سازی شوند. مؤلفه‌ای که مقادیر آن قابل پیش‌بینی نباشد، تصادفی تلقی می‌شود و معمولاً توزیع احتمالی این مؤلفه در مدل ریاضی در نظر گرفته می‌شود. منظور از مؤلفهٔ تصادفی، نداشتن اطلاعات کامل و استنباط رابطهٔ دقیق وابستگی داده‌ها در گذشته و آینده است (اسکویی، ۱۳۹۴).

به‌منظور رفع محدودیت‌های روش‌های آماری در پیش‌بینی درصد تبادل، روش‌های تکاملی در منابع مختلف معرفی شده‌اند. روت و ماجهی (۲۰۱۳) به‌کمک پنجرهٔ لغزان داده‌های گذشته،

مشخصه‌های آماری ساده‌ای برای درصد تبادل استخراج کرده‌اند و برای آموزش ضرایب داخلی آن با استفاده از بهینه‌سازی تکامل تفاضلی[1] به‌عنوان ورودی به مدل پیش‌بینی داده‌اند.

در این مقاله از مدل تکاملی کورتز و همکاران در سال ۲۰۰۴ استفاده می‌شود و در بخش قواعد کشفی توسعه‌یافته، برای هشت سری زمانی با ویژگی‌های متفاوت آزمایش خواهد شد. در این پژوهش مدل‌های ARMA و AR انتخاب شدند که با استفاده از قواعد کشفی ابعاد آن تعیین می‌شود و داده‌ها با به‌کارگیری روش پنجرهٔ لغزان به‌عنوان ورودی به مدل‌ها داده می‌شود. سپس پارامترها، ضرایب و ثابت مدل انتخابی، با استفاده از الگوریتم ژنتیک به‌صورت بهینه مقداردهی می‌شوند. ۹۰ درصد ابتدایی داده‌ها، به‌عنوان داده‌های آموزش و باقی داده‌ها یعنی ۱۰ درصد انتهایی به‌عنوان داده‌های آزمون استفاده می‌شود. خطا در هر مرحله از آموزش محاسبه می‌شود و روند آموزش براساس مجذور میانگین مربعات خطاست. درنهایت پس از رسیدن به آستانهٔ توقف که براساس تعداد نسل‌هاست، مدل ساخته و بهینه شده است. سپس با استفاده از معیار اطلاعاتی بیزین، بهترین مدل انتخاب می‌شود. از مدل ساخته‌شده برای پیش‌بینی سری زمانی استفاده می‌شود. برای پیش‌بینی، رویکرد دوم، یعنی استفاده از پیش‌بینی‌های قبلی برای پیش‌بینی‌های جدید، انتخاب شده است.

مدل مفهومی

سری زمانی مجموعه‌ای از داده‌های متوالی است که در فواصل زمانی مساوی اندازه‌گیری و ثبت شده‌اند. برای مثال، قیمت یک نوع سهام در روزهای متوالی، ارتفاع سطح آب دریاچه در روزهای مختلف، فروش هفتگی یکی از بنگاه‌های اقتصادی و جمعیت یکی از کشورها، نمونه‌هایی از سری زمانی هستند. منظور از پیش‌بینی متغیر، تعیین مقادیر آتی آن برای دوره‌ای مشخص است. پیش‌بینی یک سری زمانی عبارت است از ساخت مدلی از داده‌های موجود برای پیش‌بینی مقادیر آتی. اساس مدل پیش‌بینی سری‌های زمانی بر این باور است که از مقادیر پیشین متغیری می‌توان برای پیش‌گویی مقادیر آتی استفاده کرد. بنابراین هدف، شناسایی الگوی حاکم بر مقادیر پیشین و تعمیم آن به مقادیر پسین است (اسکویی، ۱۳۹۴ و پاناجیوتوپولوس، ۲۰۱۲).

در میان روش‌های آماری، مدل‌های بسیاری برای افزایش دقت پیش‌بینی به‌کار گرفته شده‌اند. روش رگرسیون، روش حداقل مربعات، تحلیل سری‌های زمانی، تحلیل موجک و الگوریتم‌های دیگر به‌صورت گسترده استفاده شده‌اند. مدل ARIMA برای همبستگی‌های کم‌دامنه مناسب بوده و در پیش‌بینی‌های مختلفی به‌کار رفته است. باکس- جنکینز شخصی بود

1. Differential evolution

که مدل ARIMA را معرفی کرد. مدل یادشده در پیش‌بینی‌های اقتصادی، بازار و مسائل اجتماعی کارایی مناسبی از خود نشان داده است (سو و وانگ، ۲۰۱۴). روش تصادفی باکس‌ـ جنکینز می‌تواند پیش‌بینی‌های دقیق با اتکا بر طراحی پارامتر ثابت تولید کند. براساس دسته‌ای از داده‌های وابسته به زمان، ساختار مدل و پارامترهای آن تعیین می‌شوند. سپس مدل تعیین‌شده برای کاربردهای پیش‌بینی به‌کار می‌رود. یکی از نیازهای مهم مدل ARMA، خطی و ایستابودن سری‌های زمانی است (روت و همکاران، ۲۰۱۳).

مدل ARMA برای تحلیل داده‌های سری زمانی تک‌متغیرهٔ ایستا مناسب است. چهار مرحلهٔ اساسی شامل بررسی ایستایی، شناسایی مدل، تخمین مدل و بررسی مدل در ایجاد مدل ARMA اهمیت دارند. شناسایی مدل، مهم‌ترین مرحله در شکل‌گیری مدل‌های ARMA است؛ چراکه شناسایی نادرست سبب تخمین نادرست مدل می‌شود و هزینهٔ شناسایی دوباره را تحمیل می‌کند. بااین‌حال روش مناسبی برای غلبه بر مشکل مینیمم‌های محلی در هر مدل ARMA وجود ندارد. پیش‌بینی‌های کاربردی پیچیدگی‌هایی دارند که استفاده از روش‌های کلاسیک را امکان‌ناپذیر می‌کنند. یول در سال ۱۹۲۷ مدل خودگردان را برای پیش‌بینی سری‌های زمانی پیشنهاد کرد. پس از چند سال، واکر در سال ۱۹۳۱ مدل میانگین متحرک و ترکیب آن با مدل خودگردان را ابداع کرد (ARMA). ایجاد مدل ARMA در حالتی که مرتبه‌های p (AR) و q (MA) مناسب برای مدل در دسترس باشد، آسان است، اما در پژوهش‌های متعددی اشاره شده است که دستیابی به این دستورها دشوارتر از تخمین ضرایب است (وانگ، ۲۰۰۸).

در حالت کلی، روش تحلیلی برای تخمین پارامترهای ARMA به‌خصوص موارد غیرخطی که با ماکزیمم‌کردن تابع احتمال شکل می‌گیرد، وجود ندارد. به همین دلیل پژوهشگران از روش‌های بهینه‌سازی تکرارشونده یا جست‌وجوی مستقیم استفاده می‌کنند. بااین‌حال، روش‌های تکرارشونده ممکن است به‌دلیل وجود تابع احتمال به‌شدت غیرخطی، در پیداکردن ماکزیمم ناموفق باشند. روش‌های جست‌وجوی مستقیم نیز به تعداد زیادی ارزیابی از تابع نیاز دارند؛ زیرا اطلاعاتی دربارهٔ شیب و انحنای آن ندارند (بولاروک و دیجدور، ۲۰۱۵). الگوریتم‌های میانگین حداقل مربعات پیشرو و معکوس[1] و حداقل مربعات بازگشتی[2] (ویدرو و استرینس، ۱۹۸۵) به‌منظور دستیابی به مدل ARMA با روش‌های تکرارشونده استفاده شده‌اند، اما ماهیت چنین روش‌هایی مبتنی بر مشتق توابع است و احتمال قرارگرفتن پارامترهای آنها در مینیمم‌های محلی بیشتر است (روت و همکاران، ۲۰۱۳).

---

1. Forward and backward least mean square (FBLMS)
2. Recursive least square (RLS)

معادلات روش باکس و جنکینز به تعیین مدل، تخمین پارامتر و اعتبارسنجی مدل می‌پردازند. مزیت اصلی این روش، دقت مناسب آن در محدودهٔ گسترده‌تری از سری‌هاست. مدل کلی در این روش در قالب ترکیب خطی از مقادیر قبلی (مؤلفه‌های AR) و خطاها (مؤلفه‌های MA) با عنوان میانگین متحرک خودگردان یکپارچه (ARIMA) بیان می‌شود. رابطهٔ ۱ این مدل را نشان می‌دهد. ضرایب و ثابت‌های مدل با استفاده از روش‌های آماری مانند روش حداقل مربعات تخمین زده می‌شوند.

$$\hat{x}_t = \mu + \sum_{i=1}^{p} A_i x_{t-i} + \sum_{j=1}^{q} M_j e_{t-j}$$ رابطهٔ ۱)

در این رابطه p و q نشان‌دهندهٔ مرتبه‌های AR و MA هستند. A_i و M_j نیز ضرایب AR و MA هستند. μ مقدار ثابت و e خطای پیش‌بینی است. خطای e_t به‌ازای مقادیر واقعی Y_t و پیش‌بینی F_t در دورهٔ t عبارت است از:

$$e_t = Y_t - F_t$$ رابطهٔ ۲)

خطای پیش‌بینی سری‌های زمانی

خطای پیش‌بینی مطابق رابطهٔ ۲، از تفاضل خروجی مدل و مقدار واقعی سری زمانی محاسبه می‌شود. معیار خطا در داده‌های آموزش و داده‌های آزمون هرچه کوچک‌تر باشد، به‌ترتیب بیان‌کنندهٔ میزان مناسب برازش و تعمیم‌پذیری مدل است. خطای برازش معیاری برای سنجش کیفیت مدل، پایان‌دادن به یادگیری مدل و رفتن به مرحلهٔ آزمون مدل است. هایدمن و کوهلر در سال ۲۰۰۶ معیارهای آماریِ اندازه‌گیری خطای پیش‌بینی را در پنج گروه دسته‌بندی کردند: شاخص وابسته به مقیاس؛ شاخص درصد خطا؛ شاخص متقارن خطا؛ شاخص نسبی خطا و شاخص مقیاس خطا.

در این مقاله از دو معیار برای اندازه‌گیری عملکرد پیش‌بینی استفاده می‌شود. معیار اول مجذور مربعات خطاست که از شاخص‌های وابسته به مقیاس محسوب می‌شود. این معیار برای مقایسهٔ روش‌های مختلف پیش‌بینی روی مجموعه‌ای از داده‌ها مناسب است، اما برای مقایسهٔ نتیجهٔ یک روش بر مجموعهٔ داده‌های مختلف مناسب نیست. معیار دوم، متوسط قدر مطلق درصد خطاست که از نوع شاخص درصد خطا به‌شمار می‌رود. این معیار مستقل از مقیاس داده‌هاست و می‌توان آن را برای مقایسهٔ نتایج مجموعهٔ داده‌های مختلف به‌کار برد. این نوع

معیارها ضمن متداول‌بودن معایبی نیز دارند، از جمله اینکه به‌ازای مقادیر داده‌های صفر ($Y=0_t$) خطا به بی‌نهایت می‌رسد و به‌ازای مقادیر نزدیک به صفر خطا توزیع اریبی شدیدی پیدا می‌کند (اسکویی، ۱۳۹۴). خلاصهٔ معیارهای خطای استفاده‌شده در این مقاله و چگونگی محاسبهٔ آنها در جدول ۱ ارائه شده است.

جدول ۱. معیارهای اندازه‌گیری خطای پیش‌بینی استفاده‌شده در این مقاله

مجذور متوسط مربعات خطا[1]	$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2}$
متوسط قدر مطلق درصد خطا[2]	$MAPE = \frac{1}{n}\sum_{t=1}^{n} \left\lvert \frac{e_t}{Y_t} \times 100 \right\rvert$

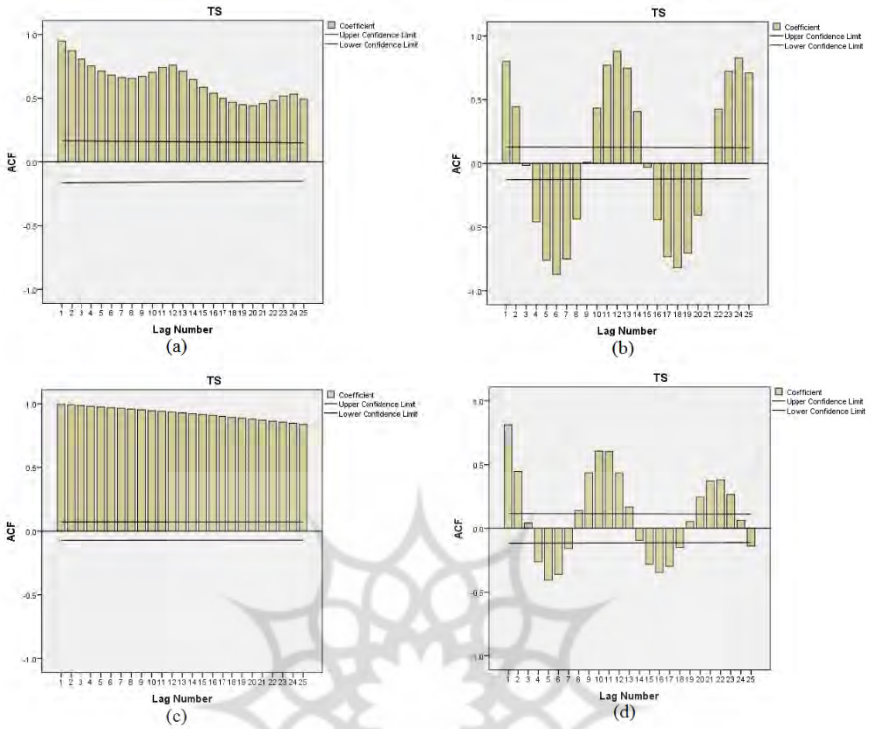
که e_t خطای پیش‌بینی و n تعداد نقاط پیش‌بینی است.

یکی از ابزارهای متداول در تحلیل سری‌های زمانی، ضریب خودهمبستگی است که اندازه‌ای از همبستگی داده‌های سری زمانی ارائه می‌دهد. خودهمبستگی، همسانی بین مشاهدات به‌عنوان تابعی از زمان جدایی بین آنهاست. این مفهوم، ابزاری ریاضی برای یافتن الگوهای تکراری است. چگونگی محاسبهٔ تابع همبستگی به‌صورت زیر است (باکس و جنکینز، ۱۹۷۶).

$$r_k = \frac{\sum_{t=1}^{s-k}(x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{s-k}(x_t - \bar{x})^2}$$ رابطهٔ ۳)

s بیانگر اندازهٔ سری زمانی است. خودهمبستگی می‌تواند برای استخراج ویژگی‌های اصلی یک سری زمانی مانند تأثیرات فصلی و روند مفید باشد. شکل ۲ نشان‌دهندهٔ خودهمبستگی چهار دسته سری زمانی مد نظر در این مقاله است. نمودار خودهمبستگی عادی سری‌های زمانی فصلی و روندی، فصلی، روندی و غیرخطی در شکل ۲ نشان داده شده است.

1. Root mean squared error
2. Mean absolute percentage error



شکل ۲. نمودار خودهمبستگی چهار دستهٔ سری زمانی
a) فصلی روندی؛ b) فصلی؛ c) روندی و d) غیرخطی

روش‌شناسی پژوهش

الگوریتم‌های تکاملی (باک، ۱۹۹۶) شامل خانواده‌ای از روش‌ها مانند الگوریتم ژنتیک یا استراتژی تکاملی‌اند که در آن جمعیتی متشکل از اعضای منفرد در حال تکامل، به دنبال راه‌حلی مناسب برای مسئلهٔ خاصی هستند. هر عضو راه‌حلی را در رشته‌ای از ژن‌ها به‌عنوان کروموزوم رمزگذاری[1] می‌کند و به هریک از آنها مقدار عددی یا تناسبی اختصاص می‌دهد. این عدد کیفیت راه‌حل را تعیین می‌کند. استفاده از روش محاسبات تکاملی، مانند الگوریتم ژنتیک، برای مدل‌سازی و پیش‌بینی یک سری زمانی، زمینهٔ نوظهوری است که با شناسایی زیرمجموعه‌ای از متغیرها در پهنای پنجره‌ای با بهترین پیش‌بینی، می‌تواند مدل‌های مناسبی ارائه کند. متأسفانه اغلب روش‌ها به بهینه‌سازی تعدادی پارامتر یا انتخاب متغیرهای توصیفی نیاز دارند. به‌طور مثال،

1. Encoding

برای معرفی‌کردن یک مدل ARIMA (p, d, q) باید متغیرهای p، d و q تعیین شوند (فلورس و گراف، ۲۰۱۲).

ARIMA مدلی گسترده‌تر از میانگین متحرک خودگردان یا ARMA است. این مدل‌ها زمانی که داده‌ها غیرایستا باشند به‌کار می‌روند. در این حالت با یک بار دیفرانسیل‌گیری، غیر ایستابودن داده‌ها از بین می‌رود و برآورد ARMA در داده‌های جدید امکان‌پذیر می‌شود. p، d، q در مدل ARIMA اعداد حقیقی غیرمنفی هستند که درجهٔ خودگردانی، یکپارچگی و میانگین متحرک را معلوم می‌کنند.

مدل پیش‌بینی تکاملی

با به‌کارگیری عملگرهای ژنتیکی شامل تلفیق و جهش، روش‌های حل جدیدی به‌وجود می‌آید و کل فرایند به‌کمک فرایندهای تصادفی و الهام از نظریهٔ انتخاب طبیعی و با انتخاب اعضای متناسب‌تر، اجرا می‌شود. الگوریتم تکاملی در این مطالعه در شکل ۱ نشان داده شده است. دو مدل برای پیش‌بینی با کمک الگوریتم ژنتیک در نظر گرفته می‌شود. در نوع اول، مدل پیش‌بینی ترکیبی خطی از مقادیر قبلی است. در این روش ژن‌های داخل کروموزوم وزن‌هایی را رمزگذاری می‌کنند که مقادیر قبلی به آنها ضرب شوند. در نوع دوم، هم مقادیر و هم خطاهای قبلی مد نظر قرار می‌گیرند، سپس با تکیه بر استراتژی مدل‌های ARMA کدهای ژن‌ها، ضرایب را تشکیل می‌دهند. هر دو مدل از یک پنجرهٔ زمانی لغزان[1] استفاده می‌کنند که مجموعه‌ای از تأخیرهای زمانی لازم برای تولید یک پیش‌بینی و تعداد ورودی‌های مدل را تعیین می‌کند. هر پنجرهٔ زمانی با نماد $< k_1, k_2, \ldots, k_n >$ برای n ورودی و k_i تأخیر زمانی در معادله‌های ۴ و ۵ نشان داده شده است که در آن g_i بیان‌کنندهٔ ژن iام از نوعی کروموزوم است.

رابطهٔ ۴)
$$AR: \hat{X} = g_0 + \sum_{i \in \{1,\ldots,n\}} g_i \times X_{t-k_i}$$

رابطهٔ ۵)
$$ARMA: \hat{X} = g_0 + \sum_{i \in \{1,\ldots,n\}} (g_i \times X_{t-k_i} + g_{i+n} e_{t-k_i})$$

دو عملگر الگوریتم ژنتیک در این مقاله استفاده شده‌اند:

1. Sliding Time Window

تلفیق حسابی: هر ژن در فرزندان ترکیب خطی‌ای از مقادیر موجود در کروموزوم والدین است. اگر a_i و b_i ژن‌های فرزندان و z_i و w_i ژن‌های اجدادی باشند، با داشتن عدد تصادفی λ روی بازه [۱ و ۰] در موقعیت i روابط زیر برقرارند.

رابطهٔ ۶) $$a_i = \lambda z_i + (1 - \lambda) \times w_i$$

رابطهٔ ۷) $$b_i = \lambda \times w_i + (1 - \lambda) \times z_i$$

جهش گاوسی: یک عملگر جهشی که مقداری گرفته‌شده از اغتشاش گاوسی با میانگین صفر را به ژن خاصی اضافه می‌کند. اغتشاشات کوچک‌تر به مقادیر بزرگ‌تر ترجیح داده می‌شوند.

بخش دیگری از الگوریتم ژنتیک، انتخاب برای گزینش جمعیت است. انتخاب، رتبه‌بندی کیفیت اعضای جمعیت و سپس روشی برای گزینش جمعیت منتخب است. در این مقاله از روش چرخهٔ رولت استفاده شده است. این روش نمونه‌برداری‌ای تصادفی با جایگزینی است. این روش مبتنی بر شانس، به این صورت اجرا می‌شود که کلیهٔ افراد برمبنای میزان شایستگی خود روی نواحی هم‌جوار یک خط نگاشته می‌شوند. اندازهٔ ناحیهٔ هر فرد با توجه به اندازهٔ شایستگی آن تعیین می‌شود. سپس یک عدد تصادفی تولیدشده و با توجه به اندازهٔ این عدد، فرد انتخاب می‌شود. این فرایند تا تأمین تعداد مد نظر والدین تکرار می‌شود.

ژن‌های جمعیت اولیه در بازهٔ [۱ و ۱−] مقادیر تصادفی را اختیار می‌کنند. اندازهٔ جمعیت اولیه ۵۰ است. کیفیت هر عضو جمعیت براساس دقت مدل نظیر آن، روی داده‌های آموزش اندازه‌گیری می‌شود. در این مطالعه در هر نسل ۴۰ درصد از اعضای نسل قبلی نگه داشته می‌شوند و ۶۰ درصد بقیهٔ اعضا با استفاده از عملگرهای ژنتیکی بیان‌شده به‌وجود می‌آیند. عملگر تلفیق مسئول تولید دو سوم از فرزندان و عملگر جهش مسئول تولید مابقی است. درنهایت الگوریتم ژنتیک پس از هزار تکرار نسل متوقف می‌شود. اندازهٔ کروموزوم متغیر با توجه به قواعد کشفی تعیین می‌شود.

روش اکتشافی برای انتخاب مدل

الگوریتم ژنتیک بیان‌شده می‌تواند برای تخمین پارامترها استفاده شود، اما همچنان انتخاب بهترین مدل برای سری زمانی به‌عنوان مسئله باقی می‌ماند. این موضوع به‌شدت وابسته به انتخاب قواعد کشفی مناسب است. به‌طور مثال در مدل‌های فصلی معمولاً از تأخیرهای <۱,۱۲,۱۳> برای سری‌های ماهانه استفاده می‌شود. هر مدل خوب باید توانایی یادگیری از داده‌های آموزشی را داشته باشد و درعین‌حال آن را به مشاهدات جدید تعمیم دهد و از

بیش‌برازش جلوگیری کند. روش قابل قبول در این زمینه، در نظر گرفتن قید تابع پیچیدگی مدل است، معیار اطلاعاتی بیزین گزینهٔ مناسبی حتی برای سری‌های غیرخطی است که در رابطهٔ ۸ بیان شده است. در این رابطه N تعداد نمونه‌های آموزشی و p تعداد پارامترهاست. در اینجا $P_{AR}=1+n$ و $P_{ARMA}=1+2\times n$ در نظر گرفته می‌شود.

$$BIC = N \times \ln\left(\frac{SSE}{N}\right) + p \times \ln N \qquad \text{رابطهٔ ۸)}$$

استفاده از قوانین کشفی، مبتنی بر نوع سری زمانی و ویژگی‌های آن است. ابعاد مدل با توجه به نوع سری زمانی و چگونگی تعیین ضریب خودهمبستگی، مشخص می‌شود. جدول ۲ نشان‌دهندهٔ هفت قاعدهٔ کشفی، توضیحات و مدل پنجرهٔ لغزان زمانیِ استفاده‌شده در این پژوهش است. قواعد کشفی، ابعاد مدل ARMA و مؤلفه‌های تأخیری را مشخص می‌کنند. قاعدهٔ ۲ نسبت به قاعدهٔ ۱ پیچیدگی محاسباتی کمتری دارد. انتظار می‌رود که قاعدهٔ ۵ برای سری‌های فصلی و روندی، قاعدهٔ ۶ برای سری‌های فصلی و قاعدهٔ ۷ نیز برای سری‌های روندی مدل مناسب‌تری ارائه دهند.

جدول ۲. قواعد کشفی و مدل‌های پنجرهٔ لغزان زمانی استفاده‌شده در این پژوهش

قواعد کشفی	توضیحات	مدل پنجرهٔ لغزان زمانی
قاعدهٔ ۱	پنجرهٔ لغزان شامل تمام تأخیرهای زمانی	<m و... و۲ و۱>
قاعدهٔ ۲	پنجرهٔ لغزان شامل تأخیرهای زوج زمانی	<m و... و۴ و۲>
قاعدهٔ ۳	پنجرهٔ لغزان شامل تأخیرهای k_i که ضریب خودهمبستگی آنها بیشتر از حد آستانه (میانگین به‌علاوهٔ واریانس تأخیرها) است.	<... و c و b a>
قاعدهٔ ۴	پنجرهٔ لغزان شامل ۴ تأخیر با بیشترین مقدار خودهمبستگی	<a و b و c d >
قاعدهٔ ۵	مناسب برای سری‌های فصلی و روندی	<۱و k و k+۱>
قاعدهٔ ۶	مناسب برای سری‌های فصلی	<۱و k>
قاعدهٔ ۷	مناسب برای سری‌های روندی	<۲و ۱> و <۱>

در جدول ۲، k دورهٔ فصلی است و در این مقاله به‌علت ماهانه‌بودن داده‌های سری‌های زمانی فصلی مقدار k، برابر با ۱۳ فرض شده است. الگوریتم ژنتیک مکاشفه‌ای نیز به‌صورت شبه‌کد در شکل ۳ نشان داده شده است.

شکل ۳. مدل‌سازی به روش الگوریتم ژنتیک و قواعد کشفی و چگونگی بررسی و مقایسهٔ نتایج آن با روش‌های کلاسیک

تجزیه و تحلیل داده‌ها (مجموعهٔ داده‌ها)

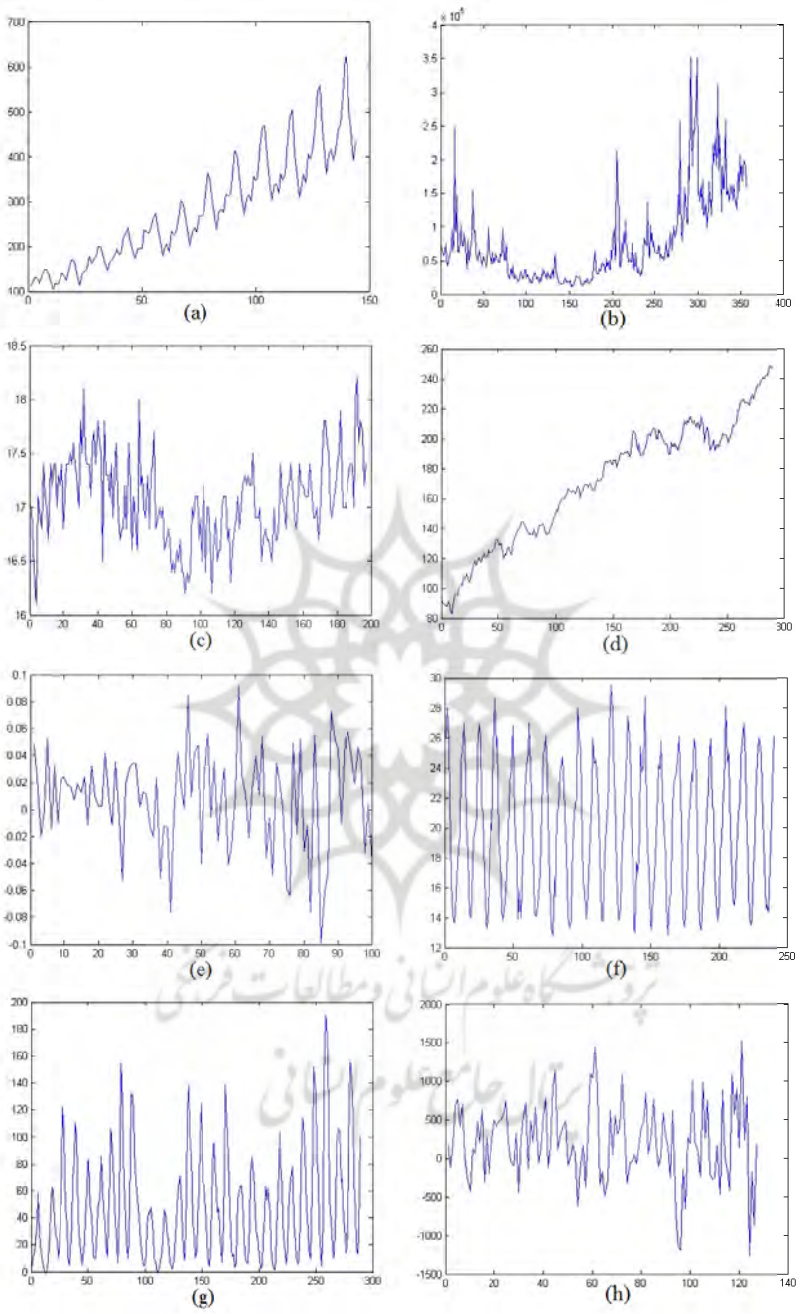
در این مقاله مجموعه‌ای شامل ۸ سری زمانی انتخاب شده است که در جدول ۳ و شکل ۴ آمده است و بنگاه‌های اقتصادی، فرایندهای طبیعی و بهابازار (بورس) را شامل می‌شود. چهار سری زمانی از مقالهٔ کورتز و همکاران (۲۰۰۴)، یک سری زمانی از مقالهٔ اسکویی (۱۳۹۴) و سه سری زمانی دیگر از سایت معتبر دیتا مارکت[1] بخش مالی گرفته شده است.

تمام سری‌های زمانی در چهار دسته تقسیم‌بندی شده‌اند:

۱. سری‌های زمانی مسافران و نفت از نوع فصلی روندی؛
۲. سری‌های زمانی شیمیایی و اپل از نوع فصلی؛
۳. سری‌های زمانی شاخص و بیشترین دما از نوع روندی؛
۴. سری‌های زمانی لکه‌های خورشیدی و استرالیا از نوع غیرخطی.

همچنین توضیحات مختصری دربارهٔ سری‌های زمانی انتخابی در جدول ۳ آمده است.

1. https://datamarket.com



شکل ٤. سری‌های زمانی a) مسافران؛ b) نفت؛ c) شیمیایی؛ d) اپل؛ e) شاخص؛ f) بیشترین دما؛ g) لکه‌های خورشیدی و h) استرالیا

جدول ۳. سری‌های زمانی استفاده‌شده در این پژوهش

سری زمانی	نوع	اندازه	محدوده	توضیحات
مسافران	فصلی و روندی	۱۴۴	[۱۰۴; ۶۲۲]	مسافران هوایی ماهانه (کورتز، ۲۰۰۴)
نفت	فصلی و روندی	۳۵۶	[۱۱۲۶۸; ۳۵۱۸۴۵]	معاملات نفت (دیتا مارکت)
شیمیایی	روندی	۱۹۸	[۱۶/۱; ۱۸/۲]	مطالعه مراقبت‌های شیمیایی (کورتز، ۲۰۰۴)
اپل	روندی	۲۸۹	[۸۳/۱۱; ۲۴۸/۹۲]	قیمت پایانی روزانه در بهابازار (اسکویی، ۱۳۹۴)
شاخص	فصلی	۱۰۰	[۰/۰۹۹۱; ۰/۰۹۱۵]	شاخص اسکولز در بازار بهابازار نیویورک (دیتا مارکت)
حداکثر دما	فصلی	۲۴۰	[۱۲/۸; ۲۹/۵]	بیشترین دما در ملبورن (کورتز، ۲۰۰۴)
لکه‌های خورشیدی	غیرخطی	۲۸۹	[۰/۰; ۱۹۰/۲]	تعداد لکه‌های خورشیدی (کورتز، ۲۰۰۴)
استرالیا	غیرخطی	۱۲۷	[-۱۲۵۶; ۱۵۲۱]	موجودی غیرکشاورزی (دیتا مارکت)

محیط پیاده‌سازی

تمام آزمایش‌ها و پیاده‌سازی الگوریتم ژنتیک و قواعد کشفی در محیط نرم‌افزاری MATLAB 2014a اجرا شد و برای مقایسهٔ این روش با روش‌های کلاسیک از نرم‌افزار قدرتمند و معروف IBM SPSS Statistics 22 استفاده شد. روش کار SPSS برای انتخاب مدل به‌صورت خبره و خودکار است، یعنی بهترین مدل با کمترین خطا انتخاب می‌شود (شکل ۳). دو روش کلی برای مدل‌سازی و پیش‌بینی سری‌های زمانی وجود دارد؛ روش اول مدل‌های ARIMA و روش دوم مدل‌های نمو هموار[1] است. درواقع در این نرم‌افزار هر یک از این روش‌ها خود شامل چندین مدل می‌شود و بسیاری از روش‌های کلاسیک زیرمجموعه این دو مجموعه‌اند (IBM SPSS Forecasting 22[2] و راهنمای نرم‌افزار IBM SPSS[3]).

1. Exponential Smoothing
2. http://www.cc.uoa.gr/fileadmin/cc.uoa.gr/uploads/files/manuals/SPSS22/IBM_SPSS_Forecasting.pdf
3. Tutorial/Algorithms/TSMODEL Algorithms

یافته‌های پژوهش

برای نمونه، عملکرد الگوریتم ژنتیک و قواعد کشفی روی دو مدل (رابطه‌های ۴ و ۵) و نتـایج دو روش کلاسیک روی سری زمانی مسافران در جدول ۴ درج شده است.

جدول ٤. نتایج الگوریتم ژنتیک و قواعد کشفی و روش‌های کلاسیک نرم‌افزار SPSS روی سری زمانی مسافران

مدل	انتخاب قاعده	تأخیرهای زمانی	تعداد پارامترها	خطای آموزش		خطای پیش‌بینی
			P	$RMSE_t$	BIC	$RMSE_f$
GA-AR	قاعدهٔ ۱	<۱۳ و... و۲ و۱ >	۱۴	۱۹/۴۷۶	۸۳۴	۳۳/۶۴۴
	قاعدهٔ ۲	<۱۲ و... و۲ >	۷	۱۵/۷۱۵	۷۳۰	۱۶/۶۷۹
	قاعدهٔ ۳	< ۱۲ و۱۱ و۳ و۲ و۱ >	۶	۱۴/۱۴۸	۷۰۷	۱۷/۵۲۸
	قاعدهٔ ۴	<۱۲ و۳ و۲ و۱>	۵	۱۴/۲۶۰	۷۰۹	۱۹/۸۴۳
	قاعدهٔ ۵	**< ۱۳ و۱۲ و۱ >**	۴	۹/۴۴۲	۵۹۸	۱۴/۰۲۴
GA-ARMA	قاعدهٔ ۱	<۱۳ و... و۲ و۱ >	۲۷	۲۱/۶۵۹	۹۲۴	۴۲/۹۶۷
	قاعدهٔ ۲	<۱۲ و... و۲ >	۱۵	۴۲/۴۵۶	۱۰۵۳	۲۰/۴۷۳
	قاعدهٔ ۳	< ۱۲ و۱۱ و۳ و۲ و۱ >	۱۳	۱۵/۳۰۴	۷۵۶	۲۱/۶۱۸
	قاعدهٔ ۴	< ۱۲ و۳ و۲ و۱ >	۹	۱۱/۰۳۵	۶۶۳	۱۷/۱۰۴
	قاعدهٔ ۵	< ۱۳ و۱۲ و۱ >	۷	۱۱/۴۸۵	۶۶۰	۲۱/۶۷۶
	SPSS-ARIMA			۷۹/۸۲۶	۷۷۷	۹/۵۸۲
	SPSS-Exponential Smoothing			۷۵/۳۹۰	۸۸۹	۱۳/۷۵۹

همان‌گونه که در جدول ۴ مشخص است، مـدل GA-AR بـا اسـتفاده از قاعـدهٔ ۵ کمتـرین پیچیدگی را دارد و بهترین مدل انتخاب می‌شود. ممکن است این مدل خطای آموزش یا خطـای پیش‌بینی بیشتری نسبت به سایر مدل‌ها داشته باشد، اما ملاک انتخاب، معیار اطلاعـاتی بیـزین است. سری زمانی مسافران از نوع فصلی و روندی است. بنابراین طبـق انتظـار، قاعـدهٔ ۵ نتـایج مطلوب‌تری نسبت به قاعدهٔ ۶ و ۷ دارد.

در جدول ۵ تمام سری‌های زمانی موجود با هر دو روش AR و ARMA آزمایش شده‌اند. نتایج سه ستون آخر در قالب میانگین ۱۵ اجرای متفاوت آورده شده است. در ۲ ستون آخر خطای پیش‌بینی در گروه آزمون $RMSE_f$ و MAPE آمده است. قید BIC زمانی که مدلی با پارامترهای کمتر انتخاب شود، بهتر عمل می‌کند.

جدول ۵. بهترین نتایج پیش‌بینی توسط الگوریتم ژنتیک براساس معیار BIC

سری زمانی	انتخاب قاعده	مدل	تأخیرهای زمانی	تعداد پارامترها	خطای آموزش	خطای پیش‌بینی	
				p	BIC	$RMSE_f$	MAPE
مسافران	قاعدهٔ ۵	AR	< ۱۳ و۱۲ و۱ >	۴	۵۹۸	۱۴/۰۲۴	۲/۳۹۴
نفت	قاعدهٔ ۵	AR	< ۱۳ و۱۲ و۱ >	۴	۶۵۷۴	۳۶۹۰	۳۰/۳۰۵
شیمیایی	قاعدهٔ ۷	ARMA	< ۱ >	۳	-۳۹۷	۰/۳۹۲۴	۱/۶۷۳
اپل	قاعدهٔ ۳	AR	<۴و۳و۲و۱>	۳	۵۷۸	۲/۴۰۳	۰/۸۰۹
شاخص	قاعدهٔ ۲	AR	< ۱۲و... و۴ و۲ >	۷	-۵۷۹	۰/۰۲۶	۹۴/۰۱۳
بیشترین دما	قاعدهٔ ۶	AR	< ۱۳ و ۱ >	۵	۱۵۲	۱/۳۲۳	۶/۲۰۲
لکه‌های خورشیدی	قاعدهٔ ۴	AR	< ۱۱ و۱۰ و۲ و۱ >	۵	۱۴۵۰	۲۵/۳۱۴	۳۲/۹۴۶
استرالیا	قاعدهٔ ۳	AR	< ۱۰ و۸ و۲ و۱ >	۵	۱۳۸۸	۷۳۱/۴۴۶	۱۱۹/۲۰۷

قواعد ۵، ۶ و ۷ برای انواع خاصی از سری‌های زمانی مناسب است و نتایج مطلوب‌تری دارند. این قواعد برای تشخیص، از اطلاعات تجزیه‌ای آنها استفاده می‌کنند. قاعدهٔ ۵ برای سری‌های زمانی فصلی و روندی بهترین نتیجه را می‌دهد. همچنین قاعدهٔ ۶ برای سری زمانی فصلی و قاعدهٔ ۷ برای سری زمانی روندی بهترین مدل‌ها را با کمترین پیچیدگی ساخته‌اند.

جدول ۶ نتایج روش‌های کلاسیک روی تمام سری‌های زمانی با استفاده از نرم‌افزار SPSS را نشان می‌دهد. همان‌طور که اشاره شد، دو روش ARIMA و نمو هموار هر یک شامل چندین مدل است. همان‌گونه که در جدول ۶ مشخص است در چهار سری زمانی روش ARIMA و در چهار سری زمانی دیگر روش نمو هموار نتیجهٔ بهتری دارد.

جدول ٦. نتایج پیش‌بینی با روش‌های کلاسیک توسط نرم‌افزار SPSS برای تمام سری‌های زمانی

سری زمانی	مدل	خطای آموزش	خطای پیش‌بینی	
		BIC	$RMSE_f$	MAPE
مسافران	Winter's Multiplicative	۶۰۵	۲۱/۴۰۷	۱/۷۱۹
نفت	MA	۶۵۸۳	۱۶۶۰۱	۲۰/۳۸۶
شیمیایی	Simple	-۴۰۶	۰/۴۰۵	۰/۶۲۹
اپل	Holt	۵۸۰	۸/۶۸۴	۱/۹۴۶
شاخص	ARMA	-۵۹۰	۰/۰۳۴	۱۶۳/۲۳۱
بیشترین دما	Simple Seasonal	۸۲	۰/۷۳۱	۱/۸۰۵
لکه‌های خورشیدی	ARMA	۱۴۶۸	۶۱/۷۷۹	۲۵/۰۸۴
استرالیا	ARMA	۱۳۹۰	۶۵۵/۹۷۹	۱۳۴/۲۶۵

مقایسه

در این بخش نتایج به‌دست‌آمده از پیش‌بینی‌های دو روش مطرح‌شدهٔ الگوریتم ژنتیک در محیط MATLAB و روش‌های کلاسیک در نرم‌افزار SPSS مقایسه شده است. جدول ۷ نشان‌دهندهٔ این نتایج است. با مقایسهٔ روش الگوریتم ژنتیک مبتنی بر قواعد کشفی و روش‌های کلاسیک مشاهده می‌شود که روش پیشنهادی نتایج همسان یا بهتری نسبت به روش‌های کلاسیک دارد. همان‌گونه که پیش‌بینی شده بود، در سری‌های نامنظم و غیرخطی این روش نتایج بهتری دارد. توانایی پیش‌بینی در سری‌های نامنظم و غیرخطی بسیار مهم است؛ زیرا بیشتر سری‌های زمانی ماهیتی نامنظم و غیرخطی دارند. تفاوت نتایج را می‌توان به‌کمک جست‌وجوی سراسری الگوریتم‌های تکاملی توضیح داد، به این‌ترتیب که تخمین پارامترها با استفاده از این الگوریتم می‌تواند بهتر از روش حداقل مربعات باشد.

انتخاب مدل بهتر براساس معیار اطلاعاتی بیزین در فضای وسیع جست‌وجو صورت گرفته است. این موضوع ویژگی اصلی این روش به‌شمار می‌رود.

جدول ۷. مقایسهٔ نتایج بهترین پیش‌بینی برای تمام سری‌های زمانی

سری زمانی	Exponential Smoothing			ARIMA			GA		
	BIC	$RMSE_f$	MAPE	BIC	$RMSE_f$	MAPE	BIC	$RMSE_f$	MAPE
مسافران	۸۸۹	۷۵/۳۹۰	۱۳/۷۵۹	۷۷۷	۷۹/۸۲۶	۹/۵۸۲	۵۹۸	۱۴/۰۲۴	۲/۳۹۴
نفت	۶۵۸۷	۵۴۵۴۹	۱۲/۶۵۶	۶۵۸۳	۱۶۶۰۱	۲۰/۳۸۶	۶۵۷۴	۳۶۹۰۴	۳۰/۳۰۵
شیمیایی	-۴۰۶	۰/۴۰۴	۰/۶۲۹	-۴۰۳	۰/۴۰۵	۰/۶۱۴	-۳۹۷	۰/۳۹۲	۱/۶۷۳
اپل	۵۸۰	۸/۶۸۴	۱/۵۵۱	۵۹۰	۸/۶۳۲	۱/۹۴۶	۵۷۸	۲/۴۰۳	۰/۸۰۹
شاخص	-۵۸۸	۰/۰۳۵	۱۳۸/۰۶۸	-۵۹۰	۰/۰۳۴	۱۶۳/۲۳۱	-۵۷۹	۰/۰۲۶	۹۴/۰۱۳
بیشترین دما	۴۱۹	۸/۴۵۵	۰/۲۶۸	۱۸۲	۱/۰۲۱	۱/۴۵۵	۱۵۲	۱/۳۲۳	۶/۲۰۲
لکه‌های خورشیدی	۱۵۵۶	۸۱/۳۵۱	۳۰/۰۹۸	۱۴۶۸	۶۱/۷۷۹	۲۵/۰۸۴	۱۴۵۰	۲۵/۳۱۴	۳۲/۶۴۶
استرالیا	۱۴۰۶	۷۵۶/۵۷۱	۲۲۷/۲۷۲	۱۳۹۳	۶۶۱/۷۲۰	۴۱۸/۱۸۱	۱۳۸۸	۷۳۱/۴۴۶	۱۱۹/۲۰۷

نتیجه‌گیری و پیشنهادها

روش‌های بهینه‌سازی مبتنی بر الگوریتم‌های تکاملی امکانات جدیدی را در زمینهٔ پیش‌بینی سری‌های زمانی به‌وجود آورده‌اند. با ادامهٔ این روند، این الگوریتم با این فرض که هیچ اطلاعاتی از رفتار اولیهٔ سری در دسترس نباشد (تبدیل‌های ویژه برای سری‌ها به‌کار گرفته شود)، به‌عنوان روشی کارآمد برای تولید مدل‌های پیش‌بینی سری‌های زمانی معرفی شده است. همچنین ساختارهای به‌وجودآمده به‌طور مستقل عمل می‌کند و به تحلیل‌های آماری برای داده‌ها نیاز ندارد. مشکل اصلی این روش پیچیدگی محاسباتی است. مشکلات زمان اجرا با یافتن مدل‌های بهتر به‌عنوان جمعیت اولیهٔ الگوریتم تکاملی ژنتیک حل می‌شود. هرچند این امر ممکن است نیازمند دانستن اطلاعات اولیه باشد. ازآنجاکه بیشتر سری‌های زمانی واقعی از داده‌های روزانه یا ماهانه استفاده می‌کنند، این مسئله مشکل‌ساز نخواهد بود.

روش پیشنهادی در این مطالعه با توجه به هوشمندی در تشخیص سری زمانی و استفادهٔ مناسب از قواعد بیان‌شده می‌تواند روشی قدرتمند و دقیق برای پیش‌بینی داده‌های آتی باشد.

در کارهای آتی، پیش‌بینی به‌کمک الگوریتم ژنتیک با مجتمع‌سازی توابع غیرخطی (لگاریتمی یا مثلثاتی) می‌تواند در نظر گرفته شود. موضوع دیگر، کاربرد روش‌های مشابه برای پیش‌بینی‌های درازمدت و چندمتغیره است. به‌محض آنکه یک روش تکاملی نتایج مناسبی برای بهینه‌سازی پارامترها و انتخاب مدل به‌دست آورد، باقی روش‌های فرامکاشفه‌ای مانند تبرید شبیه‌سازی‌شده، بهینه‌سازی ازدحام ذرات، بهینه‌سازی کلونی مورچگان و حتی ترکیب آنها می‌توانند استفاده شوند. همچنین از برنامه‌سازی ژنتیک نیز می‌توان برای ساخت مدل استفاده کرده و سپس این مدل جایگزین مدل‌های کلاسیک شود.

در بخش مدل‌ها می‌توان از مدل‌های کلاسیک دیگری جز مدل‌های میانگین متحرک خودگردان استفاده کرد یا از روش‌های هوشمند دیگری مانند برنامه‌سازی ژنتیک برای ساخت مدل بهره برد. در بخش قواعد کشفی نیز می‌توان قواعد هوشمند دیگری به آنها افزود و پایگاه قواعد قوی‌تر و غنی‌تری به وجود آورد. ترکیب این قوانین نیز می‌توانند استفاده شوند.

همچنین می‌توان با تبدیل سری‌های زمانی به سری‌های زمانی با ویژگی‌های دیگری، احتمال موفقیت در مدل‌سازی و پیش‌بینی را افزایش داد. به‌طور مثال می‌توان یک سری زمانی را به مؤلفه‌های اصلی سازندهٔ آن تقسیم کرد یا حتی می‌توان تجزیه را به‌گونه‌ای هوشمند انجام داد که کاربرد قواعد کشفی در مدل‌سازی آنها موفق‌تر عمل کند.

References

Bäck, T. (1996). *Evolutionary Algorithms in Theory and Practice*. Oxford University Press.

Beheshti-Nia, M.A. & Farazmand, N. (2015). A Novel Decision Support System for Discrete Cost-CO2 Emission Trade-off in Construction Projects: The Usage of Imitate Genetic Algorithm. *Journal of information technology management*, 7(1): 23-48. (*in Persian*)

Beligiannis, G.N., Likothanassis, S.D. & Demiris, E.N. (2001) A robust hybrid evolutionary method for ARMA model identification. in Image and Signal Processing and Analysis. *ISPA 2001. Proceedings of the 2nd International Symposium on*.

Boularouk, Y. & Djeddour, K. (2015). New approximation for ARMA parameters estimate. *Mathematics and Computers in Simulation*, 118: 116–122.

Box, G. & G. Jenkins. (1976). Time Series Analysis: Forecasting and Control. Holden Day, San Francisco, USA.

Chao-Ming, H. & Hong-Tzer, Y. (1995). A time series approach to short term load forecasting through evolutionary programming structures. in *Energy Management and Power Delivery. Proceedings of EMPD '95, 1995 International Conference on*.

Conner, J.T., Martin, R.D & Atlas, L.E. (1994). Recurrent neural networks and robust time series prediction. *Neural Networks, IEEE Transactions*, 5(2): 240- 254.

Cortez, P., Rocha, M. & Neves, J. (2004). Evolving Time Series Forecasting ARMA Models. *Journal of Heuristics*, 10(4): 415-429.

Cortez, P., M. Rocha, and J. Neves. (2001). Genetic and Evolutionary Algorithms for Time Series Forecasting. *In Monostori, L., V´ancza, J., and Alis, M. (eds.), Engineering of Intelligent Systems: Proc. of IEA/AIE 2001, LNAI* 2070: 393-402.

De Gooijer, G.J. & Hyndman, R. J. (2006). 25 years of time series forecasting. *International Journal of Forecasting*, 22(3): 443-473.

Elahi, S., Rashidi, M. & Sadeghi, M. (2015). Designing fuzzy expert system for chief privacy officer in government and businesses E-transactions. *Journal of information technology management.* 7(3): 511-530. (*in Persian*)

Flores, J. & Graff, M. (2012). Evaluative design of ARMA and ANN models for time series forecasting. *Renewable Energy*, 44: 225-230.

Ghasemi, A.R. & Asgharizadeh, E. (2014). Presenting a Hybrid ANN-MADM Method to Define Excellence Level of Iranian Petrochemical Companies. *Journal of information technology management*, 6(2): 267-284. (*in Persian*)

Graves, D. & Pedrycz, W. (2009). Fuzzy prediction architecture using recurrent neural networks. *Neurocomputing*, 72(9): 1668-1678.

Hamzaçebi, C., Akay, D. & Kutay, F. (2009). Comparison of direct and iterative artificial neural networks forecast approaches in multi-periodic time series forecasting, *Expert Systems with Applications*, 36(2): 3839-3844.

Huang, Ch. & Yang, H. (1995). A time series approach to short term load forecasting through evolutionary programming structures. *Energy Management and Power Delivery. Proceedings of EMPD 95. International Conference*, 21-23 Nov. DOI: 10.1109/EMPD.1995. 500792.

Hyndman, R.J. & Koehler, A.B. (2006). Another look at measurs of forecast accuracy. *International Journal of Forecasting.* 22(4): 679-688.

Ismail, Z. & Fong Yeng, F. (2011). Genetic Algorithm for Parameter Estimation in Double Exponential Smoothing. *Australian Journal of Basic and Applied Sciences*, 5(7): 1174-1180.

Karim, T. (2014). Desining an Expert System for Analyzing the Energy Consumption Behavior of Employees in Organizations Using Rough Set Theory. *Journal of information technology management*, 7(2): 363-384. (*in Persian*)

Majhi, R., Majhi, B., Rout, M., Mishra, S. & Panda, G. (2009). Efficient sales forecasting using ARMA-PSO model. In: *Proc. of IEEE International Conference on Nature and Biologically Inspired, Computing*, pp. 1333–1337.

Makridakis, S., Weelwright, S. & Hyndman, R. (1998). *Forecasting: Methods and Applications*, 3rd edn. New York, USA: John Wiley & Sons.

Michalewicz, Z. (1996). *Genetic Algorithms + Data Structures = Evolution Programs*, 3rd edn. USA: Springer-Verlag.

Mohammadi, K. & Eslami, H.R. (2006). Parameter estimation of an ARMA model for river flow forecasting using goal programming. *Journal of Hydrology*, 331(1-2): 293– 299.

Oscar, C. & Patricia, M. (2002). Hybrid intelligent systems for time series prediction using neural networks, fuzzy logic, and fractal theory. *Neural Networks, IEEE Transaction*, 13(6): 1395-1408.

Oskoei, M.A. (2004). Time Series Prediction Using Neural Networks, *Journal of Economic Research*, 16(47): 163-183. (*in Persian*)

Oskoei, M.A. (2015). Application of Sliding Window for Financial Time Series Prediction using Time-Delay Neural Networks. *Journal of Economic Research*, ISSN: 1735-210X. (*in Persian*)

Panagiotopoulos, A. (2012). *Optimizing Time Series Forecast Through Linear Programming*. PhD Thesis, Nottingham University.

Rolf, S. & Sprave, J. (1997). Model identification and parameter estimation Of ARMA models by means of evolutionary algorithms. *Computational Intelligence for Financial Engineering*, 237(243): 23-25.

Rout, M. & Majhi, B. (2013). Forecasting of currency exchange rates using an adaptive ARMA model with differential evolution based training. *Journal of King Saud University–Computer and Information Sciences*, 26(1): 7–18.

Sapankevych, N.I. & Ravi, S. (2009). Time Series Prediction Using Support Vector Machines: A Survey. *Computational Intelligence Magazine, IEEE*, 4(2): 24-38.

Su, Z. & Wang, J. (2014). A new hybrid model optimized by an intelligent optimization algorithm for wind speed forecasting. *Energy Conversion and Management*, 85: 443-452.

Taghavi Fard, M.T., Hosseini, F. & Khanbabaei, M. (2014). Hybrid credit scoring model using genetic algorithms and fuzzy expert systems Case study: Ghavvamin financial and credit institution. *Journal of information technology management,*. 6(1): 31-46. (*in Persian*)

Ursem, R.K. & Vadstrup, P. (2003). Parameter Identification of Induction Motors Using Differential Evolution. *Evolutionary Computation*, CEC '03. The 2003 Congress 2: 8-12.

Van Gestel, T., Suykens, J.A.K., Baestaens, D.E., Lamberechts, A., Lanckriet, G., Vandaeie, B., DeMoor, B. & Vandewalle, J. (2001). Financial time series prediction using least squares support vector machines within the evidence framework. *Neural Network, IEEE Transaction*, 12(4): 809-821.

Vosough, M., Taghavi Fard, M.T. & Alborzi, M. (2015). Bank card fraud detection using artificial neural network. *Journal of information technology management*, 6(4): 721-746. (*in Persian*)

Wang, J. & Liang, J. (2008). ARMA Model identification using Particle Swarm Optimization Algorithm. *ICCSIT, Computer Science and Information Technology, International Conference*, 223-227.

White, M. & Wen, J. (2015). Optimal Estimation of Multivariate ARMA Models. *AAAI Conference on Artificial Intelligence*, North America.

Zhang, J., Chung, H.S.H. & Lo, W.L. (2008). Chaotic Time Series Prediction Using a Neuro-Fuzzy System with Time-Delay Coordinates. *Knowledge and Data Engineering, IEEE Transactions*, 20(7): 956-964.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی