فصلنامه علمی مطالعات مدیریت صنعتی – سال هفدهم، شماره ۵۵، زمستان ۹۸
صفحات ۱۸۳-۱۶۱

توسعه روش طبقه‌بندی دیتاست‌های نامتوازن با استفاده از الگوریتم‌های تکاملی چندهدفه

امیر دانشور،* مهدی همایون‌فر،** الهام اخوان***

تاریخ دریافت: ۹۷/۱/۲۹
تاریخ پذیرش: ۹۷/۱۰/۲۲

چکیده

طبقه‌بندی داده‌ها از مباحث اساسی علم مدیریت است که از رویکردهای مختلفی مورد بررسی قرار گرفته است. روش-های هوش مصنوعی از مهمترین روش‌های طبقه‌بندی هستند که اغلب آنها تابع دقت کل را در ارزیابی عملکرد مد نظر قرار می‌دهند. از آنجاییکه در دیتاست‌های نامتوازن، این تابع، هزینه خطاهای پیش‌بینی را یکسان در نظر می‌گیرد، در این پژوهش علاوه بر تابع دقت کل، از تابع حساسیت نیز به منظور افزایش دقت در هر یک از کلاس‌های از پیش‌تعریف‌شده، استفاده شده است. به‌علاوه، بدلیل پیچیدگی فرآیند کسب اطلاعات از تصمیم‌گیرنده، از الگوریتم فرا ابتکاری NSGA II جهت استنتاج مقادیر پارامترها، (بردار وزن و سطوح برش بین کلاس‌ها) استفاده گردیده است. در هر تکرار، الگوریتم با استفاده از بردار وزن برآورد شده و دیتاست‌ها، امتیاز هر آلترناتیو را با تابع سام پروداکت[1] محاسبه نموده و در مقایسه با سطوح برش تخمینی، آن آلترناتیو را به یکی از دسته‌ها تخصیص می‌دهد. سپس با استفاده از توابع برازش، دسته تخمینی و دسته واقعی را مقایسه نموده و این فرایند تا بهینه‌سازی پارامترها ادامه می‌یابد. مقایسه نتایج الگوریتم‌های NSGA II و NRGA، نشان‌دهنده کارایی بالای الگوریتم ارائه شده است.

واژگان کلیدی: الگوریتم ژنتیک با رتبه‌بندی نامغلوب (**NSGA II**)، طبقه‌بندی چند کلاسه، دیتاست‌های نامتوازن، الگوریتم **NRGA**

---

* استادیار مدیریت صنعتی، دانشکده مدیریت، واحد الکترونیکی، دانشگاه آزاد اسلامی، تهران، ایران (نویسنده مسئول)
daneshvar.amir@gmail.com
** استادیار مدیریت صنعتی، دانشکده مدیریت و حسابداری، واحد رشت، دانشگاه آزاد اسلامی، رشت، ایران
*** دانش آموخته مدیریت صنعتی، دانشکده مدیریت، واحد الکترونیکی، دانشگاه آزاد اسلامی، تهران، ایران

[1] Sum Product

مقدمه

با افزایش روز افزون داده‌های ذخیره شده و شکل‌گیری انبارهای بزرگی از داده‌ها، بکارگیری الگوریتم‌های هوشمند برای جستجو و استخراج الگوهای موجود در آنها را اجتناب ناپذیر گشته است. در مبانی نظری، روش‌های متعددی برای طبقه‌بندی داده‌ها وجود دارند که از میان آنها می‌توان به بیز ساده و شبکه‌های بیزی، نزدیک‌ترین همسایه، شبکه‌های عصبی، درخت تصمیم، مدل‌های رگرسیونی و الگوریتم‌های تکاملی اشاره کرد. طبقه‌بندی و پیش‌بینی دو نوع عملیات برای تحلیل داده‌ها و استخراج مدل به منظور توصیف دسته‌های مهم داده‌ها و پیش-بینی رفتار آینده آنها هستند. برای تحلیل داده‌های گسسته و طبقه‌ای از مدل‌های دسته‌بندی و به منظور تحلیل داده‌های پیوسته از مدل‌های پیش‌بینی یا رگرسیون استفاده می‌شود (نیکام، ۲۰۱۵).

هدف اصلی در مدل‌های طبقه‌بندی مبتنی بر هوش مصنوعی[1] آن است که وزن معیارها و سطح برش، بوسیله الگوریتم‌های شناخته شده‌ای مانند: الگوریتم ژنتیک، بهینه‌سازی ازحام ذرات و غیره برآورد شوند (هدهمی کعبی و همکاران، ۲۰۱۵؛ ژوهانگ یو و همکاران، ۲۰۱۳). پس از تعیین وزن معیارها، امتیاز نهایی هر یک از آلترناتیوها با استفاده از قوانین تجمیع محاسبه می‌گردد. مبنای استفاده از مدل‌های طبقه‌بندی مبتنی بر هوش مصنوعی، یک دیتاست آموزشی در دسترس است که فرآیند یادگیری طبقه‌بندی بر اساس آنها انجام می‌شود (هدهمی کعبی و دیگران، ۲۰۱۵).

در سال‌های اخیر، شاهد رشد مطالعات در حوزه مسائل طبقه‌بندی نامتوازن بوده‌ایم (چن و همکاران، ۲۰۰۸) که نقش اساسی در یادگیری ماشین بازی می‌کنند (روت و همکاران، ۲۰۱۸). مساله ناشی از داده‌های نامتوازن این است که در آن یک طبقه در مقایسه با طبقه دیگر دارای تعداد زیادی نمونه است (باراندلا و همکاران، ۲۰۰۳). داده‌های نامتوازن باعث ایجاد تنگنای قابل توجهی در عملکرد روش‌های یادگیری استاندارد می‌شوند (لیو، ۲۰۰۸) که توزیع کلاس‌ها را متوازن فرض می‌کنند. این موضوع به عنوان یکی از مباحث مطرح در تحقیقات آتی یادگیری ماشینی محسوب می‌شود. هنگام یادگیری از داده‌های نامتوازن،

[1] Artificial Intelligence

روش‌های داده‌کاوی سنتی دقت پیش‌بینی بالایی را برای کلاس‌ها دارای اکثریت مشاهدات دارند، اما دقت پیش‌بینی آنها در کلاس‌های دارای مشاهدات کم، پایین است (ژو و لیو، ۲۰۰۶). دلیل این مساله آن است که کلسیفایرهای سنتی به‌دنبال عملکرد دقیق در طیف گسترده‌ای از نمونه‌ها هستند. آنها برای بررسی مسائل یادگیری نامتوازن مناسب نیستند، زیرا گرایش به دسته‌بندی همه داده‌ها در طبقه اکثریت دارند که معمولاً طبقه کم اهمیت‌تری است (چن و همکاران، ۲۰۰۸).

در به‌کارگیری معیار دقت طبقه‌بندی[1] که در طبقه‌بندی بر اساس روش‌های هوش مصنوعی اغلب به‌عنوان تابع هدف مورد استفاده قرار می‌گیرد، سه نکته حائز اهمیت است: (۱) این معیار به‌طور ضمنی هزینه‌های متفاوت هر دو نوع خطا)متقاضی بدحساب به‌عنوان خوش‌حساب طبقه‌بندی شود و بالعکس (را نادیده می‌گیرد)پرووست و فاست، ۱۹۹۷؛ مارکوییز و دیگران، ۲۰۱۲)، (۲) این معیار برای ارزیابی عملکرد کلسیفایرهایی که بر روی دیتاست‌های نامتوازن عمل می‌کنند، مناسب نیست، زیرا فرض می‌کند که توزیع مثال‌ها بین کلاس‌ها ثابت و تقریباً متوازن است)پرووست و فاست، ۱۹۹۷؛ کروز رامیرز و دیگران، ۲۰۱۴) و (۳) هنگامی‌که تعداد داده‌ها در برخی از کلاس‌ها بسیار کمتر از سایر کلاس‌ها است (دیتاست‌های نامتوازن)، استفاده از چند تابع خطا مرسوم است (مارکوییز و دیگران، ۲۰۱۲). از کاربردهای دیتاست-های نامتوازن می‌توان به تشخیص پزشکی، تشخیص نشت نفت و صنعت مالی اشاره کرد (روت و همکاران، ۲۰۱۸).

منحنی آر اُ سی[2] یک راه‌حل شناخته شده برای دسته مسائل نامتوازن است که فقط در مسائل دو کلاسه استفاده می‌شود. از آنجاییکه در مسائل چند کلاسه با افزایش تعداد کلاس‌ها، پیچیدگی محاسباتی مسائلی که از منحنی آر اُ سی استفاده کنند به‌طور نمایی افزایش می‌یابد، بکارگیری این منحنی بسیار محدود کننده و از نظر کارایی، غیر منطقی است (پدرو آنتونیو گوتیرز و دیگران ۲۰۱۲). بنابراین، در این مقاله از یک رویکرد دوهدفه برای مدل‌سازی مساله

[1] Accuracy
[2] Receiver Operating Characteristic (ROC(

طبقه‌بندی چندکلاسه با استفاده از الگوریتم تکاملی استفاده شده است.

پیشینه پژوهش

در رابطه با طبقه‌بندی داده‌ها تحقیقات متعددی بر پایه رویکردهای مختلف شکل گرفته‌اند. از آن جمله: محتشمی (۱۳۹۳) یک مدل ریاضی چندهدفه (حداکثرکردن نرخ تولید، حداقل‌کردن هزینه‌ها و حداکثرکردن کیفیت محصولات) جهت تخصیص افزونگی در سیستم‌های تولیدی نمودند. جهت حل مدل پیشنهادی از دو الگوریتم فراابتکاری تکاملی الگوریتم ژنتیک با مرتب‌سازی نامغلوب و بهینه‌سازی ازدحام ذرات چندهدفه استفاده شده‌است. نتایج حاصل از مقایسه این دو الگوریتم نشان‌دهنده کیفیت بالاتر جواب‌های الگوریتم ژنتیک با مرتب‌سازی نامغلوب در این مساله است. دانشور و همکاران (۱۳۹۴) یک روش جدید پیشنهاد کردند که در آن الگوریتم ژنتیک طی فرآیند یادگیری، به‌طور همزمان تمامی پارامترهای مدل ELECTRE TRI را از داده‌های آموزشی استنتاج نموده و در خاتمه فرآیند، پارامترهای استنتاج شده را برای طبقه‌بندی داده‌ها به‌کار می‌گیرد. زرین صدف و دانشور (۱۳۹۵) روشی ارائه نمودند که با یادگیری مقادیر پارامترها از داده‌های آموزشی با استفاده از الگوریتم بهینه‌سازی تراکم ذرات (PSO)، آنها را در طبقه‌بندی موجودی‌های جدید به‌کار می‌برد. روش پیشنهادی برخلاف مدل‌های استاندارد داده‌کاوی که طبقه‌بندی را به‌صورت اسمی انجام می‌دهند، متناسب با روش ABC اقلام موجودی را به‌صورت رتبه‌ای طبقه‌بندی می‌کند. عظیمی و همکاران (۱۳۹۴) یک مدل ترکیبی را بر اساس نگرش خوشه‌بندی و انتخاب تامین‌کنندگان ارائه داده است. به این صورت که ابتدا روش خوشه‌بندی Kهارمونیک برای خوشه‌بندی تامین‌کنندگان مورد استفاده قرار گرفته است، سپس بر اساس خروجی حاصل از خوشه‌بندی، یک مدل چند هدفه برای انتخاب مناسب‌ترین تامین‌کننده در نظر گرفته می‌شود. از آنجاییکه مساله چندهدفه مورد مطالعه به دسته مسائل NP-hard تعلق دارد، برای حل مدل پیشنهادی در یک زمان موجه از الگوریتم‌های ژنتیک مرتب‌سازی نامغلوب NSGA II و ژنتیک رتبه‌بندی نامغلوب NRGA استفاده شده است. نتایج محاسباتی

بدست آمده نشان می‌دهد که آنالیز خوشه‌بندی می‌تواند به‌عنوان یک راهکار موثر در انتخاب تامین‌کنندگان در نظر گرفته شود.

دب و همکاران (۲۰۰۲) یک الگوریتم تکاملی چند هدفه نامغلوب مبتنی بر مرتب‌سازی با نام NSGA II را برای حل مسائل چند هدفه ارائه نمودند که بسیاری از مشکلات مربوط به الگوریتم تکاملی چند هدفه را برطرف می‌سازد. نتایج شبیه‌سازی نشان می‌دهد که NSGA II در بسیاری از مسائل، قادر به پیدا کردن جواب‌های بهتر و دارای همگرایی بهتر در نزدیکی پارتو بهینه است. موکرجی و همکاران (۲۰۰۲) ابزاری جدید کاربرد الگوریتم ژنتیک چندهدفه مبتنی بر NSGA II را برای مساله مدیریت اعتباری بانک ارائه کردند. هدف از این امر برقراری تعادل مناسبی بین اهداف چندگانه بیشینه‌سازی بازدهی و کمینه‌سازی ریسک است. فناوری جدید تخمینی برای مجموعه راهکارهای بهینه پارتو ارائه می‌کند که می‌تواند انعطاف‌پذیری تصمیم‌گیری مدیریت بانک را افزایش دهد و در مقایسه با روش سنتی برنامه-ریزی چند هدفه مقید از لحاظ محاسباتی کاراتر است.

الجدان و همکاران (۲۰۰۸) یک الگوریتم ترکیبی از الگوریتم انتخاب چرخ رولت و الگوریتم رتبه‌بندی جمعیت مبتنی بر پارتو، با نام الگوریتم ژنتیک غیر مسلط ارائه نمودند. نتایج شبیه-سازی الگوریتم با استفاده از داده‌های مسائل موجود نشان می‌دهد که الگوریتم ژنتیک غیر مسلط در مقایسه با الگوریتم NSGA II در بیشتر مواقع قادر به دستیابی به جواب‌های بهتر و همگرایی سریع‌تر به پارتو بهینه است. چن و همکاران (۲۰۰۸) رویکرد داده‌کاوی مبتنی بر گرانوله اطلاعات را برای طبقه‌بندی اقلام مهم با فراوانی کم ارائه داده‌اند. روش پیشنهادی که از توانایی انسان برای پردازش اطلاعات تقلید می‌کند، دانش را از گرانول‌های اطلاعات کسب می‌کند. روش ارائه شده، یک ابزار استخراج ویژگی مبتنی بر شاخص نامتوازن را به منظور کاهش ابعاد داده‌ها ارائه نموده است. با توجه به نتایج، روش ارائه شده به طور قابل توجهی توانایی طبقه‌بندی داده‌های نامتوازن را افزایش می‌دهد. عبدو (۲۰۰۹) قابلیت برنامه‌ریزی ژنتیک در تحلیل مدل‌های اعتبارسنجی را با استفاده از داده‌های بانک‌های بخش عمومی مصر، ارزیابی نموده و این تکنیک را با آنالیز پروبیت مقایسه کردند. نتایج تجربی

نشان می‌دهند که برنامه‌ریزی ژنتیک در مقایسه با دو تکنیک دیگر، به بالاترین میزان دقت و نیز کمترین خطاهای نوع اول و دوم می‌رسد. با این وجود، رویکرد برنامه‌ریزی ژنتیک دارای کمترین هزینه طبقه‌بندی نادرست است.

کابی و همکاران (۲۰۱۵) یک روش خودکار یادگیری را برای مساله طبقه‌بندی موجودی ارائه دادند که وزن معیارها را به منظور ارائه یک طبقه‌بندی که تابع هزینه موجودی را حداقل سازد، مورد استنتاج قرار می‌دهد. روش ارائه شده آنها از تکنیک تاپسیس برای محاسبه امتیاز هر آلترناتیو و از جستجوی پیوسته همسایگی متغیر[1] (CVNS) برای تعیین وزن معیارها استفاده کرده است. روت و همکاران (۲۰۱۸) در مطالعه خود به بررسی مسائل مربوط به طبقه‌بندی داده‌های نامتوازن پرداختند. در این تحقیق آنها ضمن مرور راه‌حل‌های این دسته از مسائل در سه گروه: (۱) رویکردهای سطح داده‌ها، (۲) رویکردهای سطح الگوریتم و (۳) روش‌های ترکیبی و گروهی، معیارهای ارزیابی عملکرد داده‌های نامتوازن را مورد بررسی قرار دادند. کاربونرو- روز و همکاران (۲۰۱۷) یک معیار دو بعدی مبتنی بر دقت برای ارزیابی عملکرد طبقه-بندی ارائه دادند. در مطالعه آنها دقت به صورت میانگین موزون نرخ طبقه‌بندی هر دسته مورد بررسی قرار گرفته است و برای ارزیابی عملکرد طبقه‌بندی، یک اندازه عملکرد گرافیکی، که در فضای دو بعدی با توجه به دقت و پراکندگی تعریف شده، پیشنهاد گردیده است.

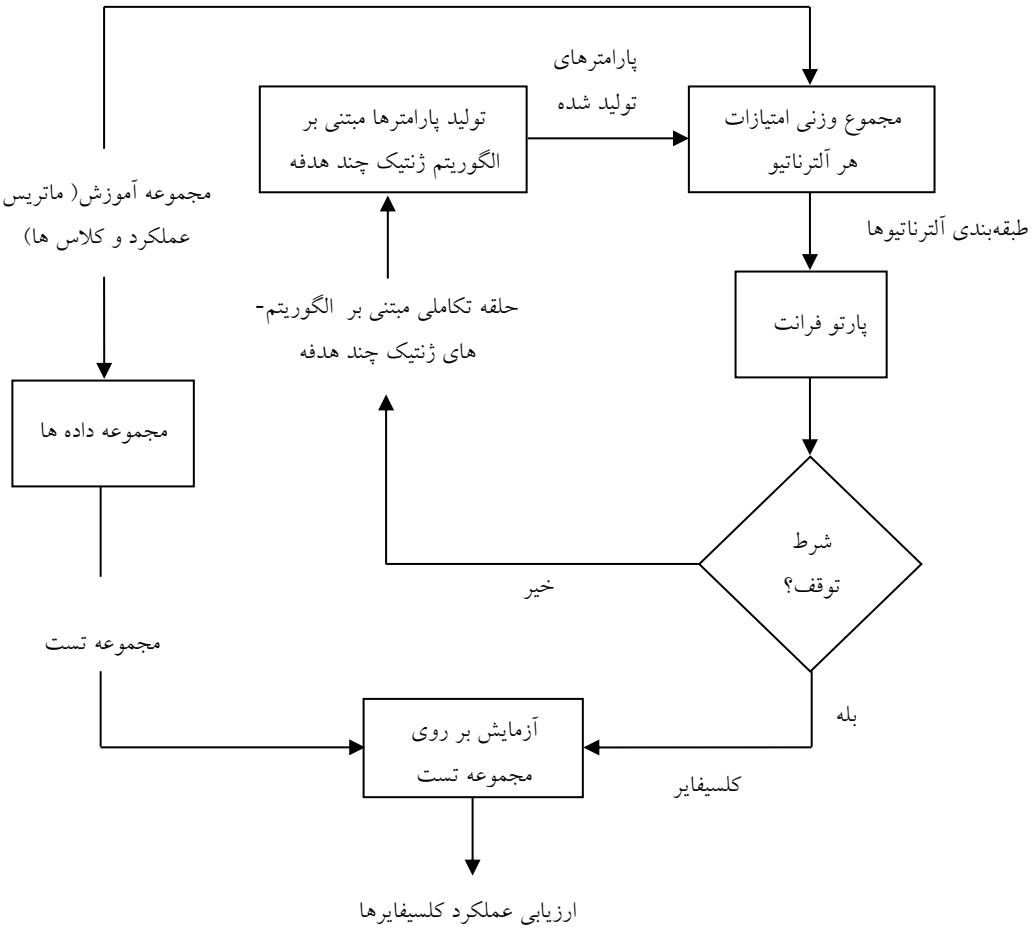
با بررسی مطالعات صورت گرفته در مبانی نظری که برخی از آنها در بخش فوق مورد اشاره قرار گرفتند، مشخص می‌شود که مدل مناسبی برای طبقه‌بندی دیتاست‌های نامتوازن چند کلاسه با توجه به اهداف چندگانه و متعارض وجود ندارد. بنابراین، تحقیق حاضر سعی بر پوشاندن خلاء اشاره شده و توسعه مدل‌های موجود دارد.

روش پیشنهادی

در روش یادگیری پیشنهادی از الگوریتم هوشمند چندهدفه NSGA II برای دستیابی به وزن معیارها و نقاط برش و از تابع "سام پروداکت" برای امتیازدهی به آلترناتیوها استفاده شده و

[1] Continuous Variable Neighborhood

براساس امتیاز هر عضو، داده‌ها طبقه‌بندی می‌گردند. با توجه به عدم وجود تحقیق مشابه در مبانی نظری پژوهش جهت ارزیابی مدل پیشنهادی، از مقایسه نتایج الگوریتم ارائه شده و الگوریتم توسعه‌یافته NRGA استفاده شده است. مراحل پیاده‌سازی روش پیشنهادی به‌صورت زیر می‌باشد:



شکل ۱. فلوچارت انجام پژوهش

بطور کلی جهت پیاده‌سازی پژوهش پس از تولید بردار وزن معیارها و سطوح برش اولیه برای هر عضو جمعیت، امتیاز آن با استفاده از تابع "سام پروداکت" محاسبه خواهد شد. سپس داده-ها براساس امتیاز هر عضو و مقادیر نقاط برش طبقه‌بندی می‌شوند و در ادامه توابع برازش محاسبه گردیده و مجموعه جواب‌های پارتو فرانت تولید می‌شوند. در نهایت جواب‌های روش پیشنهادی بر اساس دیتاست‌های واقعی با جواب‌های یک الگوریتم توسعه یافته مقایسه شده و اعتبارسنجی می‌گردد. در ادامه به الگوریتم NSGA II به عنوان مبنای شکل‌گیری الگوریتم ارائه شده خواهیم پرداخت.

الگوریتم NSGA II

الگوریتمNSGAII یکی از پرکاربردترین و قدرتمندترین الگوریتم‌های موجود برای حل مسائل بهینه‌سازی چند هدفه است که کارایی آن در حل مسائل مختلف، مورد تایید قرار گرفته است (اسرینیباس و دِب، ۲۰۰۰). با توجه به حساسیت نسبتاً زیادی که نحوه عملکرد و کیفیت جواب‌های الگوریتم NSGA به پارامترهای اشتراک برازندگی و سایر پارامترها دارد، نسخه دوم الگوریتم NSGA با نام الگوریتم فرا ابتکاری NSGA II معرفی گردید (دِب و همکاران، ۲۰۰۲). این الگوریتم یکی از اساسی‌ترین الگوریتم‌های بهینه‌سازی چندهدفه تکاملی است که می‌توان آن‌ها را نسل دوم این گونه روش‌ها دانست.

تطبیق NSGA II با مساله دو هدفه طبقه‌بندی داده‌های نامتوازن

تولید جمعیت اولیه: کارکرد الگوریتم تکاملی در این مساله یافتن پارامترهای بردار وزن و نقاط برش است. بطور کلی، در یک مساله با m معیار و n کلاس، بردار جواب شامل ۱ (m+n) – درایه خواهد بود. بنابراین، بردار جواب ایجاد شده برای مساله مورد بررسی با ۴ معیار و ۳ کلاس، شامل ۶ درایه است که ۴ درایه اول نشان‌دهنده وزن هر یک از معیارها است، بطوریکه مجموع وزن معیارها برابر با ۱ شود ($\sum_{j=1}^{4} w_j = 1$) و دو درایه آخر (X_{AB} و X_{BC}) به ترتیب نشان‌دهنده نقطه برش کلاس A و B و کلاس B و C است، بطوریکه $X_{AB} < X_{BC}$ باشد.

عملگر جهش: برای انجام عملیات جهش از عملگر جهش یکنواخت[1] استفاده شده است. به این صورت که برای هر ژن، به‌صورت تصادفی یک نقطه همسایه در فاصله $\pm a$ بدست می‌آید. مقدار a وابسته برابر ۱۰ درصد از بازه متغیرها در نظر گرفته شده است.

عملگر تقاطع: برای انجام عملیات تقاطعی از سه اپراتور: (۱) تقاطع تک نقطه‌ای، (۲) تقاطع دو نقطه‌ای و (۳) تقاطع پیوسته استفاده شده است. در تقاطع تک نقطه‌ای، یک خط برش به‌صورت تصادفی شکل گرفته و ژن‌های والدین به‌صورت تقاطعی برای ایجاد دو فرزند، با هم ترکیب می‌شوند. در تقاطع دو نقطه‌ای دو خط برش به صورت تصادفی ایجاد شده و ژن‌های والدین به‌صورت تقاطعی ترکیب می‌شوند تا دو فرزند حاصل گردد. نهایتاً در تقاطع پیوسته میانگین وزنی ژن‌های والدین محاسبه می‌شود. به‌این منظور بردار a به‌صورت تصادفی به تعداد کل ژن‌ها (در بازه صفر و یک) ایجاد می‌شود و سپس فرزندان مطابق فرمول زیر بدست می‌آیند:

$$Q_i = \alpha x_{1i} + (1-\alpha) x_{2i} \qquad (۱)$$

برای هر دو والد نیز به‌صورت تصادفی یکی از سه اپراتور فوق برای انجام عملیات تقاطعی استفاده می‌شود. دلیل استفاده از چند روش تقاطعی در الگوریتم، نتایج آزمون سعی و خطا است که نشان می‌دهد استفاده از روش‌های تقاطعی به‌صورت همزمان، نتایج بهتری را توسط الگوریتم پیشنهادی ایجاد می‌کند.

روش انتخاب والدین: برای انتخاب والدین از روش تورتمنت[2] استفاده شده است. در این روش، در هر مرحله دو عضو به‌صورت تصادفی انتخاب شده و هر کدام که از رتبه بهتری برخوردار باشد، به‌عنوان والد انتخاب می‌شود. در صورت یکسان بودن رتبه‌ها نیز، عضو دارای فاصله ازدحامی[3] بیشتر، به عنوان والد انتخاب خواهد شد.

ادغام و انتخاب: پس از اعمال عملگرهای جهش و تقاطع بر جمعیت هر نسل، همه جواب‌ها با هم ادغام می‌شوند. برای مرتب‌سازی جواب‌ها در هر جبهه، در این پژوهش با بهره‌گیری از معیارهای فاصله ازدحامی از روش ادغام و انتخاب استفاده شده است.

[1] Uniform Mutation
[2] Binary Tournament Selection
[3] Crowding Distance

موجه‌سازی جواب‌ها: از آنجاییکه عملگرهای ژنتیک کروموزوم‌ها را دستکاری می‌کنند، ممکن است فرزندان غیرموجه تولید شوند. مساله مهم در استفاده از الگوریتم ژنتیک در مسائل بهینه‌سازی محدود شده، نحوه اداره محدودیت‌ها است. در مبانی نظری روش‌های مختلفی برای اداره محدودیت‌ها ارائه شده است که از مهمترین آنها می‌توان به استراتژی‌های میخاییلویچ (۱۹۹۵) اشاره کرد. وی استراتژی‌های موجود را به: (۱) استراتژی رد کردن، (۲) استراتژی تعمیر کردن، (۳) استراتژی تغییر عملگرهای ژنتیک و (۴) استراتژی جریمه کردن، تقسیم کرده است. در این پژوهش در برخورد با جواب‌های غیر موجه از استراتژی رد کردن استفاده شده است. استراتژی رد کردن تمام کروموزوم‌های غیرموجه را که در طول فرآیند تکامل تولید می‌شوند، کنار می‌گذارد که یک انتخاب معمول در بسیاری از الگوریتم‌های ژنتیک است. این روش هنگامیکه ناحیه جستجوی موجه محدب است، ممکن است عقلانی باشد. با این وجود، چنین روشی دارای محدودیت‌های جدی است. برای مثال، برای بسیاری مسائل بهینه‌سازی محدود شده که جمعیت اولیه از کروموزوم‌های غیرموجه تشکیل شده باشد، نیاز به بهبود این کروموزوم‌ها است. به‌علاوه، اگر امکان عبور از ناحیه غیرموجه وجود داشته باشد، بسیاری از سیستم‌ها می‌توانند به نقطه بهینه برسند (مخصوصاً در فضاهای جستجوی غیرمحدب). این استراتژی احتمال کم شدن جمعیت را به همراه دارد، لذا کروموزوم‌های رد شده را با کروموزوم‌های قوی (موجه) ترکیب می‌کنند تا از ناحیه غیرموجه به فضای موجه انتقال یابد.

توابع برازش: توابع برازش بکار رفته جهت اندازه‌گیری عملکرد کلسیفایر[1]، عبارتند از:

- تابع دقت کل[2] که از نوع بیشینه‌سازی است و به عنوان تابع هدف در الگوریتم‌های تکاملی به‌منظور حل مشکلات طبقه‌بندی (به‌صورت روتین) استفاده می‌شود. این متریک به طور خاص در مسائل طبقه‌بندی اعتباری و به‌طور عام در مسائل مالی بیشترین کاربرد را دارد.

[1] Classifier
[2] Total Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (۲)$$

در رابطه فوق، خطای نوع اول (FP_{rate}) نرخ متقاضیان بد است که به عنوان خوب طبقه بندی شده‌اند. وقتی این اتفاق رخ دهد، موسسه مالی در معرض ریسک اعتباری بالایی قرار خواهد گرفت.

$$FP_{rate} = \frac{FP}{FP + TN} \quad (۳)$$

خطای نوع دوم (FN_{rate}) نرخ متقاضیان خوبی را اندازه‌گیری می‌کند که به‌عنوان متقاضی بد طبقه‌بندی شده‌اند. وقتی این اتفاق رخ بدهد؛ موسسه مالی با از دست دادن مشتریان خوب دارای هزینه فرصت خواهد بود.

$$FN_{rate} = \frac{FN}{FN + TP} \quad (۴)$$

هزینه طبقه‌بندی اشتباه مرتبط با خطای نوع اول معمولاً بسیار بیشتر از این هزینه در خطای نوع دوم است.

جدول ۱. ماتریس طبقه‌بندی‌پیش‌بینی شده

		درست (T)	اشتباه (F)
طبقه واقعی	مثبت (P)	مثبت درست (TP)	مثبت اشتباه (FP)
	منفی (N)	منفی درست (TN)	منفی اشتباه (FN)

- تابع حساسیت[1] که از نوع بیشینه‌سازی است و دقت پیش‌بینی در هر کلاس را اندازه-گیری می‌کند. بر اساس تعریف، حساسیت؛ تعداد آلترناتیوهای درست طبقه‌بندی شده تقسیم بر تعداد کل آلترناتیوهای متعلق به کلاس مورد نظر می‌باشد.

$$Sensitivity = \frac{TP}{TP + FN} \quad or \quad \frac{TN}{TN + FP} \quad (۵)$$

[1] Sensitivity function

چنانچه مشخص است، توابع حساسیت و دقت کل رابطه عکس داشته و رفتاری خلاف جهت یکدیگر دارند. بنابراین، این دو تابع با یکدیگر در تعارض بوده و زوج مناسبی برای هدایت الگوریتم چندهدفه به‌شمار می‌روند.

پارامترهای الگوریتم NSGA II

پارامترهای الگوریتم NSGA II بر اساس آزمایش و خطا به شرح جدول (۱) تعریف شده‌اند:

جدول ۲. پارامترهای الگوریتم

نرخ‌جهش	نرخ‌تقاطع	اندازه جمعیت	تعداد تکرار
۰/۱	۰/۶	۶۰	تعداد آلترناتیوها ×۳

نرخ جهش و نرخ تقاطع در جدول فوق، درصدی از اندازه جمعیت اولیه می‌باشند که برای انجام عملیات جهش یا تقاطع انتخاب گردیده‌اند. منظور از تعداد آلترناتیوها نیز تعداد رکوردهای موجود در در هر دیتاست می‌باشد. شکل ۱ شبه کد الگوریتم پیشنهادی را نشان می‌دهد.

گام ۱: پارامترهای اولیه الگوریتم را تعریف کنید.
گام ۲: جمعیت اولیه (P) را با سایز N ایجاد کنید.
گام ۳: داده‌ها را طبقه‌بندی کنید.
گام ٤: توابع هدف (حساسیت و دقت) را ارزیابی کنید.
گام ٥: بر اساس جواب‌های نامغلوب، رتبه (جبهه) را تخصیص دهید.
گام ٦: فاصله ازدحامی را برای جمعیت محاسبه کنید.
گام ۷: تا برآورده نشدن معیار پایان، گام‌های زیر را دنبال کنید:
گام ۸: تقاطع را انجام دهید و فرزند (زاد و ولد) Q_1 را ایجاد کنید.
گام ۹: جهش را انجام دهید و فرزند (زاد و ولد) Q_2 را ایجاد کنید.
گام ۱۰: طبقه بندی را انجام دهید و توابع هدف را برای Q_1 و Q_2 ارزیابی کنید.
گام ۱۱: یک جمعیت واحد تشکیل دهید: $P \cup Q_1 \cup Q_2 = \hat{P}$
گام ۱۲: تا زمانیکه جمعیت کوچکتر از N است، گامهای زیر را دنبال کنید:

گام ۱۳: جوابهای حاصل از جبهه کنونی را بر اساس فاصله ازدحامی‌شان مرتب کنید.
گام ١٤: برای هر جواب موجود در جبهه مرتب شده، موارد زیر را انجام دهید:
گام ١٥: اگر جمعیت کوچکتر از N باشد، آنگاه:
گام ١٦: جواب را در جمعیت قرار بدهید.
گام ۱۷. پایان شرط **اگر**.
گام ۱۸: پایان شرط **برای**.
گام ۱۹: به جبهه پارتو بعدی بروید.
گام ۲۰: پایان شرط **تا**.
گام ۲۱: پایان شرط **تا**.
گام ۲۲: به P بازگردید.

شکل ۱. شبه کد الگوریتم پیشنهادی

نتایج محاسباتی

از آنجاییکه در مبانی نظری پژوهش الگوی مشابهی برای مقایسه با روش پیشنهادی وجود ندارد، برای اعتبارسنجی نتایج، ابتدا الگوریتم NRGA بر اساس پارامترهای جدول (۱) اجرا شده و در ادامه نتایج حاصل از دو الگوریتم با توجه شاخص‌های آماری زیر مقایسه می‌گردند.

- فاصله‌گذاری[1]: انحراف‌معیار فاصله نقاط پارتو را نشان می‌دهد (اسکات، ۱۹۹۵).
- پوشش مجموعه[2]: نسبت جواب‌های مجموعه B که توسط جواب‌های مجموعه A غلبه می‌شود (زیتزلر، ۱۹۹۹).
- فاصله از جواب ایده‌آل[3]: میانگین فاصله جواب‌های پارتو از مبدأ مختصات است (کریمی و زندیه، ۲۰۱۰).
- بیشترین گسترش[4]: طول قطر مکعب فضائی که توسط مقادیر انتهایی اهداف برای مجموعه جواب نامغلوب به‌کار می‌رود را اندازه‌گیری می‌کند (زیتزلر، ۱۹۹۹).
- جواب‌های پارتو[5]: بیانگر تعداد جواب‌های پارتو روش بکار رفته است.

[1] Spacing
[2] Coverage
[3] MID
[4] Maximum spread
[5] NOS

قابل ذکر است که شاخص‌های فاصله‌گذاری و فاصله از جواب ایده‌آل دارای ماهیت منفی و شاخص‌های پوشش مجموعه، بیشترین گسترش و جواب‌های پارتو دارای ماهیت مثبت می‌باشند.

دیتاست‌ها

تمامی دیتاست‌های استفاده شده در پژوهش، غیرمتوازن هستند، به‌این معنا که داده‌ها به‌صورت نرمال و متوازن بین کلاس‌های مختلف توزیع نشده‌اند و برخی از کلاس‌ها در اقلیت و برخی در اکثریت هستند. مشخصات دیتاست‌های مورد استفاده و نرخ عدم توازن[1] (تعداد داده‌های کلاس اکثریت تقسیم بر تعداد داده‌های کلاس اقلیت) آنها در جدول (۲) آورده شده است.

جدول ۳. دیتاست‌های استفاده شده

نام دیتاست	مقیاس معیارها	تعداد آلترناتیوها	تعداد معیارها	تعداد کلاس‌ها	نرخ عدم توازن
Balance	رتبه‌ای	٦٢٥	٤	٣	٥/٨٨
Flare	رتبه‌ای	١٣٨٩	١٠	٧	٧/٧٦
Glass	حقیقی	٢١٤	١٠	٧	٨/٢٣
Dermatology	رتبه ای، صحیح	٣٦٦	٣٣	٦	٥/٦٦
Wine	حقیقی، صحیح	١٧٨	١٣	٣	٩/٨٢
Hayes	رتبه‌ای	١٦٠	٥	٤	٧/٥٦
Haberman	صحیح	٣٠٦	٣	٢	٢/٧٨
Page-blocks0	حقیقی، صحیح	٥٤٧٢	١٠	٢	٨/٧٩
Vehicle0	صحیح	٨٤٦	١٨	٢	٣/٢٥
Yeast1	حقیقی	١٤٨٤	٨	٢	٢/٤٦
Wisconsin	صحیح	٦٨٣	٩	٢	١/٨٦

[1] Imbalance Rate

خروجی الگوریتم‌ها

در این بخش خروجی الگوریتم‌ها بر روی شاخص‌های چندهدفه ارائه گردیده است. قابل ذکر است که الگوریتم‌های فوق بر روی هر دیتاست، ۱۰ بار اجرا شده و میانگین خروجی مبنا قرار داده شده است. برای کدنویسی الگوریتم‌های فرا ابتکاری پیشنهاد شده نیز از نرم‌افزار متلب استفاده گردیده است. خلاصه نتایج حاصل از اجرای هر یک از الگوریتم‌ها، در جدول (۳) نشان داده شده است. با توجه به میانگین شاخص‌های ارزیابی عملکرد (سطر آخر جدول) مشاهده می‌شود که الگوریتم NSGA II در معیارهای MID، NOS،Spacing ، Spread و Coverage از عملکرد بهتری نسبت به الگوریتم NRGA برخوردار است و تنها در معیار CPU time عملکرد الگوریتم NRGA بطور نسبی مناسب‌تراست.

جدول ٤. خروجی الگوریتم NRGA و NSGA II برروی شاخص‌های ارزیابی عملکرد الگوریتم‌های چندهدفه

Data Set	NSGA II						NRGA					
	MID	Spacing	CPU Time	Spread	NOS	Coverage	MID	Spacing	CPU Time	Spread	NOS	Coverage
Balance	۱/۰۸۰۹	۰/۹۶۰۳	۷۱۷/۰۶	۱/۸۹۴۱	۹	۰/۶۶	۱/۱۰۱۵	۰/۳۸۷۱	۶۹۱/۲۸	۱/۸۹۵۵	۱۰	۰
Flare	۱/۲۲۰۱	۰/۴۱۵۲	۵۰۲/۴۱	۱/۵۷۵۷	۲۳	۰/۵۲	۱/۲۲۰۷	۰/۶۲۵۵	۵۱۰/۷۷	۰/۸۶۷۱	۱۰	۰
Glass	۰/۰۶۹۲	۰/۳۴۲۹	۱۱۷/۳۴	۰/۲۱۵۲	۱۴	۰/۶۴	۰/۰۷۴۳	۰/۵۲۹۵	۱۱۶/۰۶	۰/۱۵۴۳	۹	۰/۱۱
Dermatology	۱/۱۲۲۱	۰/۷۶۳۰	۱۸۷/۹۰	۰/۸۳۶۰	۲۶	۰/۵۳	۱/۱۳۴۲	۰/۸۶۳۰	۱۹۲/۸۸	۰/۲۴۳۰	۱۶	۰
Wine	۱/۰۸۹۷	۰/۳۷۵۲	۸۸/۵۷	۰/۸۹۳۰	۱۹	۰/۴۲	۱/۱۳۱۷	۰/۶۲۳۴	۸۰/۲۵	۰/۷۶۳۰	۲۲	۰/۱۸۱
Hayes	۰/۴۶۱۸	۰/۳۲۰۷	۶۷/۹۰	۰/۷۱۶۳	۳۴	۰/۵۸	۰/۴۷۷۴	۰/۱۰۷۷	۶۳/۱۳	۰/۵۹۵۱	۳۶	۰/۱۱
Haberman	۱/۰۳۵۰	۰/۴۶۷۹	۲۸۳/۷۳	۲/۴۶۸۶	۱۴	۰/۳۴	۱/۰۳۹۵	۰/۳۱۷۲	۲۸۲/۴۷	۰/۴۸۳۸	۱۵	۰
Page-Blocks0	۰/۶۳۷۲	۰/۳۰۹۵	۲۰۴۱/۰۹	۳/۰۵۷۳	۴۵	۰/۷۶	۰/۸۶۳۹	۰/۲۷۸۱	۲۰۲۳/۵۰	۴/۸۰۶۵	۵۷	۰/۰۵
Vehicle0	۰/۷۱۸۴	۰/۶۲۴۴	۷۹۹/۷۸	۱/۱۵۶۱	۹	۰/۴۴	۰/۷۷۴۰	۰/۲۲۷۴	۷۹۸/۳۴	۱/۴۹۱۹	۱۳	۰/۰۷
Yeast1	۰/۹۲۰۷	۰/۶۹۱۱	۹۵۲/۳۷	۳/۵۳۹۶	۴۱	۰/۶۷	۰/۹۲۸۴	۱/۱۲۹۹	۹۲۶/۳۲	۴/۰۹۹۳	۴۱	۰/۱۱
Wisconsin	۱/۲۶۵۵	۰/۵۹۰۰	۷۳۵/۵۱	۲/۴۴۰۴	۱۶	۰/۴۳	۱/۳۳۵۱	۱/۳۴۴۱	۷۲۲/۹۹	۲/۴۸۰۰	۱۲	۰/۰۸
Average	۰/۸۷۴۳	۰/۵۳۲۷	۵۹۰/۳۳	۱/۷۰۶۵	۲۲/۷۲۷۲	۰/۵۴۴۵	۰/۹۱۶۴	۰/۵۸۴۸	۵۸۲/۵۴۳۸	۱/۸۰۴۵	۲۱/۹۰۹۰	۰/۰۶۴۶

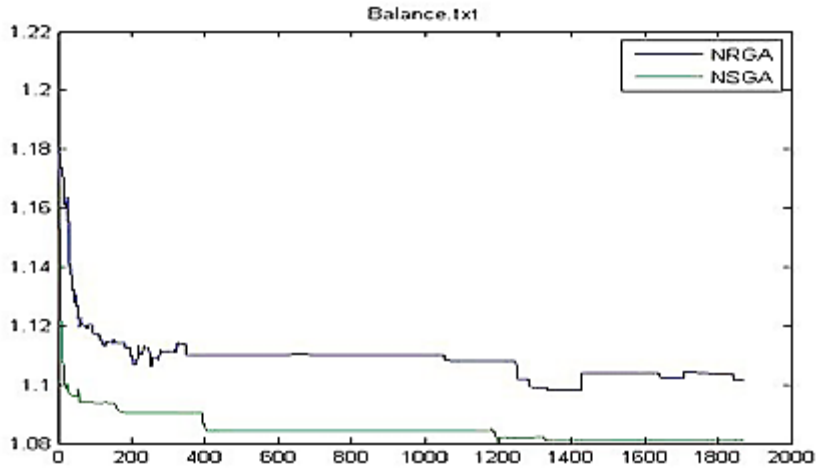
بعلاوه، به‌منظور بررسی تطبیقی الگوریتم‌های بکار رفته و تحلیل آماری نتایج حاصل، از آزمون‌های آماری من- ویتنی و ویلکاکسون استفاده شده است. نتایج آزمون‌های آماری ویلکاکسون و من-ویتنی در سطح معناداری ۵ درصد به‌ترتیب در جدول شماره (۴) آمده است:

جدول ٤. نتایج آزمون فرض آماری

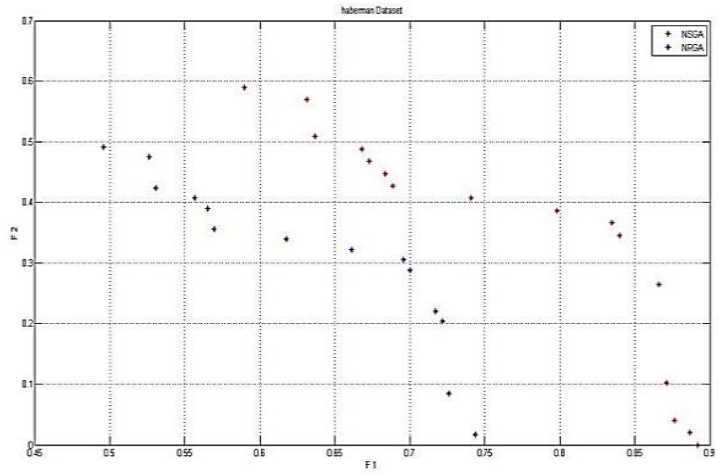
نام شاخص‌ها	ویلکاکسون		من- ویتنی	
	عدد معناداری	نتیجه آزمون	عدد معناداری	نتیجه آزمون
MID	۰/۰۰۳	رد H.	۰/۶۰۶	عدم رد H.
Spacing	۰/۶۵۷	عدم رد H.	۰/۸۹۸	عدم رد H.
CPU Time	۰/۰۶۲	عدم رد H.	۰/۸۴۷	عدم رد H.
NOS	۰/۹۲۹	عدم رد H.	۰/۹۴۹	عدم رد H.
Coverage	۰/۷۵۹	عدم رد H.	۰/۶۵۲	عدم رد H.
NOS	۰/۰۰۳	رد H.	۰/۰۰۰	رد H.

با توجه به نتایج آزمون ویلکاکسون، عملکرد الگوریتم NSGA II (با توجه به شاخص‌های ارزیابی عملکرد الگوریتم‌های چندهدفه) در معیارهای MID و NOS بهتر از الگوریتم NRGA می‌باشد. در آزمون آماری من- ویتنی نیز تنها در معیار NOS عملکرد دو الگوریتم تفاوت معناداری دارد (خروجی الگوریتم NSGA II بهتر از الگوریتم NRGA است). در مجموع، در تمام مواردی که تفاوت معناداری بین عملکرد دو الگوریتم وجود دارد، با توجه به شاخص-های ارزیابی عملکرد الگوریتم‌های چندهدفه، خروجی الگوریتم NSGA II بر الگوریتم NRGA برتری دارد.

نمودار (۱) مقدار شاخص MID حاصل از اجرای الگوریتم‌های NSGA II و NRGA بر روی دیتاست Balance را که معادل با نمودار همگرایی در الگوریتم‌های تک هدفه است، نشان می‌دهد. هر نقطه از این نمودار، با مقدار MID فرانت نسل اول متناظر است و نشان می‌دهد که با توجه به عملکرد مناسب الگوریتم؛ شیب نمودار طی نسل‌های پیاپی به‌تدریج کمتر و به عبارت بهتر فرانت اول نسل‌ها به پارتو بهینه نزدیک‌تر می‌شوند.



نمودار ۱. منحنی‌های MID الگوریتم‌ها برای دیتاست Balance



نمودار ۲. مقایسه پارتو فرانت دو الگوریتم روی دیتاست Haberman

جواب‌های پارتوی به‌دست آمده از الگوریتم‌های NSGA II و NRGA بر روی دیتاست Haberman، در نمودار (۲) نشان داده شده است. در این نمودار، محور افقی نشان‌دهنده مقدار تابع برازش حساسیت و محور عمودی مقدار تابع برازش دقت کل می‌باشد. چنانچه مشاهده

می‌شود، اغلب جواب‌های حاصل از الگوریتم NSGA II (نقاط آبی رنگ) بر الگوریتم NRGA (نقاط قرمز رنگ) غلبه دارند.

نتیجه‌گیری و پیشنهادها

در این پژوهش، حل مساله طبقه‌بندی چند کلاسه داده‌های نامتوازن با استفاده از الگوریتم‌های تکاملی مورد بررسی قرار گرفت. از آنجاییکه: (۱) استفاده صرف از تابع دقت در الگوریتم تکاملی، باعث نادیده گرفتن هزینه‌های متفاوت هر دو نوع خطا می‌شود و (۲) ثابت و متوازن بودن توزیع نمونه‌ها بین کلاس‌ها، این نوع طبقه‌بندی را برای دیتاست‌های نامتوازن نامناسب می‌سازد، در نظر گرفتن مساله به صورت تک هدفه، محدودیت‌هایی را به همراه دارد. در نتیجه، برای دستیابی به عملکرد بالا در هر کلاس، در توابع برازش از دو تابع افزایش دقت و افزایش کمترین حساسیت استفاده گردیده است.

در مجموع، با بکارگیری روش طبقه‌بندی مبتنی بر الگوریتم‌های تکاملی چندهدفه در این پژوهش، در عین برخورداری از سطوح بالای دقت، عملکرد هرکلاس نیز افزایش یافته است. کارکرد الگوریتم چند هدفه NSGA II در تعیین وزن معیارها و نقاط برش و کارکرد روش SAW در امتیازدهی به آلترناتیوها و طبقه‌بندی آنها بر اساس امتیاز هر عضو است.

با توجه به عدم مشاهده تحقیقات مشابه در مبانی نظری، به‌منظور ارزیابی نتایج مدل پیشنهاد شده، الگوریتم هوشمند NRGA توسعه داده شد و نتایج حاصل از دو الگوریتم با یکدیگر مقایسه گردید. نهایتاً، الگوریتم‌های طراحی شده در نرم‌افزار متلب بر روی دیتاست‌های متنوع دو کلاسه و چند کلاسه، به دفعات (۱۰ مرتبه) اجرا گردید و میانگین خروجی الگوریتم‌ها بر روی شاخص‌های چند هدفه MID، Spacing، Cpu time ، Spread، NOS و Coverage مبنای ارزیابی قرار گرفت. بر اساس نتایج، الگوریتم NSGAII در بسیاری از معیارها، عملکرد بهتری را در مقایسه با NRGA نشان می‌دهد. به‌منظور بررسی تطبیقی الگوریتم‌های بکار رفته و تحلیل آماری نتایج حاصل، از آزمون‌های آماری من- ویتنی و ویلکاکسون در نرم افزار SPSS استفاده گردید. نتایج آزمون‌های آماری نشان داد که الگوریتم NSGA II نسبت به الگوریتم

NSGA II در اغلب شاخص‌ها از عملکرد بهتری برخوردار است. برای انجام پژوهش‌های آتی پیشنهاد می‌شود مدل مناسب برای اهداف متعارض چندگانه توسعه داده شود و با استفاده از نسخه‌های پیشرفته‌تر الگوریتم‌های فرا ابتکاری، مانند NSGA III مورد پردازش قرار گیرد.

منابع

دانشور، ا.، زندیه، م.، ناظمی، ج. (۱۳۹۴). *یک روش تکاملی برای طبقه‌بندی اعتباری مبتنی بر رویکرد تجمیع زدایی ترجیحات*. مطالعات مدیریت صنعتی، شماره ۳۹، صفحات ۱-۳۴.

زرین صدف، م.، دانشور، ا. (۱۳۹۵). *روش کارای یادگیری ترجیحات مبتنی بر مدل ELECTRE TRI به‌منظور طبقه‌بندی چندمعیارهٔ موجودی*. مدیریت صنعتی، دوره ۸، شماره ۲، صفحات ۱۹۱-۲۱۶.

عظیمی، پ.، گلدار، ف.، مهدی‌زاده، ا. (۱۳۹۴). *ارائه مدلی ترکیبی برای انتخاب تامین کنندگان مبتنی بر رویکرد خوشه‌بندی و حل آن با استفاده از الگوریتم‌های NRGA و NSGA-II*. مطالعات مدیریت صنعتی، شماره ۳۶، صفحات ۱۱۵-۱۴۲.

محتشمی، ع. (۱۳۹۳). *یک روش تلفیقی جدید جهت تخصیص افزونگی در سیستم‌های تولیدی با استفاده از NSGA-II و MOPSO اصلاح شده*. مطالعات مدیریت صنعتی، شماره ۳، صفحات ۹۷-۱۲۴.

Abdou ,H .A .(۲۰۰۹) .*Genetic programming for credit scoring: The case of Egyptian public sector banks* .Expert Systems with Applications , ۳۶(۹), ۱۱۴۰۲–۱۱۴۱۷.

Al Jadaan, O ,.Rajamani, L., Rao, C. R. (2008 .(*Non-Dominated Ranked Genetic Algorithm for Solving Multi-Objrctive Optimisation Problems* :*NRGA* .Journal of Theoretical and Applied Information Technology. ۶۰-۶۷ ,

Barandela, R., Sanchez, J. S., Garcia, V., Rangel, E. (2003 .(*Strategies for learning in class imbalance problems* .Pattern Recognition, 36, 849–851.

Carbonero-Ruz, M., Martínez-Estudillo, F. J., Fernández-Navarro ,F., Becerra-Alonso, D., Martínez-Estudillo, A. C. (2017 .(*A two dimensional accuracy-based measure for classification performance* . Information Sciences, 382, 60-80.

Chen ,M. C., Chen, L. S., Hsu, C. C., Zeng, W. R. (2008 .(*An information granulation based data mining approach for classifying imbalanced data* .Information Sciences, 178 (16), 3214-3227.

Deb, K., Pratap, A., Agarwal ,S., Meyarivan, T. A. M. T. (2002 .(*A Fast and Elitist Multi-objective Genetic Algorithm: NSGA-II Kalyanmoy* .IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION. ۱۹۷-۱۸۲,(۲) ٦ ,

Hollander, M., Wolfe, D.A .(۱۹۷۳) .*Non-parametric Statistical Methods* .John Wiley *&* Sons.

Kaabi ,H ,.Jabeur ,K., Enneifar, L. (2015 .*(Learning criteria weights with TOPSIS method and continuous VNS for multi-criteria inventory classification* .Electronic Notes in Discrete Mathematics, 47, 197-204.

Karimi, N., Zandieh, M., Karamooz, H. R., (2010 .(*Bi-objective group scheduling in hybrid flexible flow shop: A multi-phase approach* . Expert Systems with Applications, 37, 4024-4032.

Michalewiez, Z., (1995 .(*A survey of constraint handling techniques in evolutionary computation methods* .Evolutionary programming IV, MIT Press, Cambridge, MA, 98-108.

Marqués, A. I., García, V ,.Sánchez, J. S .(۲۰۱۲) .*Exploring the behavior of base classifiers in credit scoring ensembles* .Expert Systems with Applications, 39, 10244-10250.

Mukerjee ,A., Biswas, R., Deb, K. Mathur, A. (2002 .(*Multi-objective evolutionary algorithms for the risk-return trade-off in bank loan management* .International Transactions in Operational Research ,(۵)۹ و ۵۹۷–۵۸۳.

Provost, F., Fawcett, T. (1997 .(*Analysis and visualization of classifier performance: Comparison under imprecise class and cost distributions* .Proceeding of the Third International Conference on Knowledge Discovery and Data Mining (KDD-97). Newport beach, CA, 43-48.

Gutiérrez, P. A ,.Hervás-Martínez, C ,.Martínez-Estudillo, F. J ,. Carbonerob ,M .(۲۰۱۲) .*A two-stage evolutionary algorithm based on*

sensitivity and accuracy for multi-class problems .Information Sciences, 197, 20-37.

Nikam, S. S .(۲۰۱۵) .*A Comparative Study of Classification Techniques in Data Mining Algorithms* .Computer Science and Technology, 8 (1.۱۹-۱۳,(

Rout, N., Mishra, D., Mallick, M. K .(۲۰۱۸(.*Handling Imbalanced Data: A Survey* .International Proceedings on Advances in Soft Computing, Intelligent Systems and Applications.٤٤٣-٤٣١ ,

Srinivas, N., Deb, K. (2000 .(*Multi-Objective function optimization using non-dominated sorting genetic algorithms* .Evolutionary Computation, 2 (3), 221-248.

Schott, J. R. (1995 .(*Fault tolerant design using single and multi-criteria genetic algorithms optimization* .Master thesis ,Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge.

You, Z. H., Lei, Y. K., Zhu, L., Xia, J ,.Wang, B .(۲۰۱۳) .*Prediction of protein-protein interactions from amino acid sequences with ensemble extreme learning machines and principal components analysis* .BMC Bioinformatics ,(۸) ۱۴ ,xx-xx.

Zitzler, E. (1999 .(*Evolutionary Algorithms for Multi-objective Optimization: Methods and Applications* .Ph. D Dissertation, Swiss Federal Institute of Technology (ETH.(

Zhou, Z. H ,.Liu, X. Y. (2006 .(*Training cost-sensitive neural networks with methods addressing the class imbalance problem* .IEEE Transactions on Knowledge and Data Engineering, 18, 63–77.