فصلنامه علمی - پژوهشی فرماندهی و کنترل، سال دوم، شماره ۱، بهار ۱۳۹۷

خوشه‌بندی خودکار فازی داده‌ها با استفاده از الگوریتم بهینه‌سازی چندهدفه گرگ خاکستری

علی‌اصغر امام‌دوست۱ * فرزان رشیدی۲ عبدالله خلیلی۳

تاریخ دریافت: ۱۳۹۷/۰۳/۰۹ تاریخ پذیرش: ۱۳۹۷/۰٤/۱۵

چکیده:

در این مقاله مسأله خوشه‌بندی خودکار فازی، در قالب یک مسأله بهینه‌سازی چندهدفه ارائه شده است. دو تابع هدف یکی بر پایه اتصال خوشه‌ها و دیگری براساس هم‌پوشانی-جدایی خوشه‌ها در نظر گرفته شده که جهت تعیین تعداد بهینه خوشه‌ها و افزایش کیفیت خوشه‌بندی، این دو تابع بطور همزمان بهینه می‌شوند. با توجه به اینکه مساله مورد نظر از نوع مسائل بهینه‌سازی غیر خطی، چندهدفه و نامحدب می‌باشد، برای حل آن نیز یک روش بهینه‌سازی چندهدفه مبتنی بر الگوریتم گرگ خاکستری پیشنهاد شده است. به منظور تسریع در فرآیند بهینه‌سازی و جلوگیری از گیر افتادن الگوریتم در بهینه‌های محلی، راهکارهای ابتکاری جدیدی به الگوریتم اضافه شده است. نتیجه اعمال این الگوریتم بر مساله خوشه‌بندی، منجر به مجموعه‌ای از جوابهای بهینه پارتو خواهد شد که نشان‌دهنده ناحیه مصالحه بین توابع هدف است. برای انتخاب جواب نهایی از بین چندین راه‌حل بهینه موجود، از معیار ارزیابی **DB** استفاده شده است. برای بررسی عملکرد الگوریتم پیشنهادی، شبیه‌سازی‌های متعددی بر روی مجموعه داده مصنوعی و واقعی انجام و نتایج با چند مقاله دیگر مقایسه گردیده است. نتایج آزمایشها نشان می‌دهند مدل پیشنهادی قادر به شناسایی تعداد بهینه خوشه‌ها و افراز مناسب داده‌ها در انواع مجموعه داده‌های هم‌پوشان و غیر هم‌پوشان است.

واژگان کلیدی: خوشه‌بندی خودکار فازی، الگوریتم گرگ خاکستری، بهینه‌سازی چندهدفه، شاخص ارزیابی خوشه

۱. کارشناسی ارشد- دانشکده فنی و مهندسی، دانشگاه هرمزگان، a.imamdoost@yahoo.com

۲. *استادیار- دانشکده فنی و مهندسی، دانشگاه هرمزگان (نویسنده مسئول) rashidi@hormozgan.ac.ir

۳. استادیار- دانشکده فنی و مهندسی، دانشگاه هرمزگان، khalili@hormozgan.ac.ir

۱. مقدمه

خوشه‌بندی عبارت است از گروه‌بندی عناصر با ویژگی‌های مشابه، به‌گونه‌ای که عناصر موجود در یک گروه، بیشترین شباهت را نسبت به هم و بیشترین تفاوت را با عناصر موجود در دیگر گروه‌ها داشته باشند. هدف خوشه‌بندی، دستیابی سریع و مطمئن به اطلاعات همبسته و شناسایی ارتباط منطقی میان آنها است. هر خوشه دارای نماینده‌ای است که معرف خوشه بوده و غالبا بیانگر مرکز خوشه نیز می‌باشد. میزان شباهت داده‌ها به مرکز خوشه، عموما توسط پارامتری بنام معیار شباهت تعیین می‌گردد. معیار شباهت غالبا بر اساس اصل حداکثر کردن جدایی بین خوشه‌ها و در عین حال حداقل کردن پراکندگی داخل خوشه‌ها تعیین می‌شود [۱].

خوشه‌بندی می‌تواند به دو صورت انجام شود: خوشه‌بندی غیرفازی و خوشه‌بندی فازی. در خوشه‌بندی غیرفازی هر داده تنها به یک خوشه تعلق می‌گیرد درحالیکه در خوشه‌بندی فازی هر داده می‌تواند به چند خوشه با درجه عضویت بین صفر و یک تعلق داشته باشد [۲]. می‌توان خوشه‌بندی غیرفازی را حالت خاصی از خوشه‌بندی فازی دانست که میزان تعلق هر داده به خوشه‌ای که عضو آن است یک، و به مابقی خوشه‌ها صفر است. یکی از انواع پیچیده مسائل خوشه‌بندی، زمانی طرح می‌شود که تعداد خوشه‌ها نیز نامعلوم بوده و الگوریتم خوشه‌بندی موظف به پیدا کردن تعداد خوشه‌ها باشد. این مسأله اصطلاحا خوشه‌بندی خودکار نامیده می‌شود [۳].

هر چند در سالهای اخیر، الگوریتم‌های متعددی برای خوشه‌بندی داده‌های پیچیده و حجیم ارائه شده است؛ ولی در بیشتر این الگوریتمها، تعداد خوشه‌ها می‌بایست از قبل معلوم بوده و خود الگوریتم نمی‌تواند تعداد بهینه خوشه‌ها را بدست آورد. این درحالیست که برای بسیاری از داده‌های واقعی، تعداد خوشه‌ها معلوم نبوده و حتی نمی‌توان تقریبی از تعداد آنها را نیز مشخص کرد [٤]. از این‌رو، حل مسأله خوشه‌بندی در حالت کلی و مسأله خوشه‌بندی خودکار به صورت خاص، بعضا می‌تواند خارج از توان الگوریتم‌های رایج خوشه‌بندی باشد. یکی از راهکارهای پیشنهادی برای این موضوع، تبدیل مسأله خوشه‌بندی به یک مسأله بهینه‌سازی و حل آن با استفاده از الگوریتم‌های بهینه‌سازی هوشمند است. الگوریتم‌های بهینه‌سازی هوشمند مورد استفاده در حل مسائل خوشه‌بندی، به دو دسته کلی الگوریتم‌های تک هدفه و چندهدفه تقسیم‌بندی می‌شوند. از جمله الگوریتم‌های بهینه‌سازی تک هدفه مورد استفاده در حل مساله خوشه‌بندی خودکار، می‌توان به الگوریتم ژنتیک [۵]، الگوریتم تبرید فلزات [٦]، الگوریتم ازدحام ذرات [۷]، الگوریتم تکامل تفاضلی [۸] و الگوریتم رقابت استعماری [۹] اشاره کرد.

با توجه به اینکه هیچ یک از معیارهای ارزیابی خوشه‌بندی نمی‌توانند به‌تنهایی برای همه نوع از مجموعه داده‌ها خوب عمل کنند، از طرفی اهمیت نسبی معیارهای مختلف ارزیابی خوشه‌بندی معمولا ناشناخته بوده و بهینه‌سازی یک معیار خوشه‌بندی به‌تنهایی، قادر به شناسایی خوشه‌های هم‌پوشان موجود در مجموعه داده‌ها نیست، بنابراین الگوریتم‌های بهینه‌سازی تک هدفه نمی‌توانند در مورد خوشه‌بندی داده‌های با ابعاد بالا و یا داده‌های با ساختار پیچیده، کارآیی مناسبی داشته باشند [۱۰]. از این‌رو در سالهای اخیر استفاده از الگوریتم‌های بهینه‌سازی چندهدفه در حل مسائل داده‌کاوی بطور فزاینده‌ای مورد توجه قرار گرفته است. بطور نمونه، مرجع [۱۱] سعی نموده با بهره‌گیری از

فصلنامهٔ علمی - پژوهشی فرماندهی و کنترل، سال دوم، شمارهٔ ۱، بهار ۱۳۹۷

الگوریتم تبرید فلزات چندهدفه و با لحاظ کردن دو معیار ارزیابی به صورت همزمان، عمل خوشه‌بندی خودکار داده‌ها را انجام دهد. مقایسه نتایج حاصل از الگوریتم فوق بر روی ۸ مجموعه داده مصنوعی و ٦ مجموعه داده واقعی، نشان می‌دهند الگوریتم مذکور توانسته است بطور میانگین در ۷۰درصد موارد در تعیین تعداد صحیح خوشه‌ها موفق عمل نماید.

در [۱۲] یک روش خوشه‌بندی مبتنی‌بر الگوریتم ژنتیک چندهدفه (NSGA-II) پیشنهاد شده است که به‌طور هم‌زمان دو شاخص XB و FCM را به‌عنوان توابع هدف، بهینه‌سازی می‌کند. کارایی الگوریتم فوق، با چهار الگوریتم خوشه‌بندی و با ۳ مجموعه داده مصنوعی و ۱۳ مجموعه داده واقعی مورد مقایسه قرار داده شده است. نتایج نشان می‌دهند الگوریتم مذکور در اکثر آزمایشها دقت خوشه‌بندی بهتری نسبت به چهار الگوریتم دیگر دارد.

مرجع [۱۳]، کارایی دو نوع الگوریتم تکامل تفاضلی چندهدفه یعنی MODE و DEMO، را بر روی مسئله خوشه‌بندی فازی بررسی و با چند الگوریتم دیگر مقایسه نموده است. نتایج آزمایش‌ها، براساس ٦ مجموعه داده مصنوعی و ٤ مجموعه داده واقعی با پیچیدگی‌های مختلف، نشان می‌دهند الگوریتم DEMO توانسته است در بیشتر آزمایشها نتایج بهتری نسبت به الگوریتم‌های MODE، MOCK و NSGA-II داشته باشد.

مرجع [١٤] نیز با استفاده از دو شاخص XB و FCM و بهره‌گیری از نسخه اصلاح شده الگوریتم تکامل تفاضلی چندهدفه بنام AFCMDE به حل مساله خوشه‌بندی فازی پرداخته است. نتایج خوشه‌بندی این الگوریتم بر روی داده‌های IRIS و مقایسه با الگوریتم‌های FCM، ACDE و MoDEAFC نشان می‌دهند الگوریتم مذکور توانسته است بطور میانگین حدود ١٦درصد دقت خوشه‌بندی را افزایش دهد.

مرجع [۱۵] تلاش کرده است با الهام گرفتن از دو الگوریتم تکامل تفاضلی و ژنتیک، الگوریتم جدید چندهدفه‌ای بنام GADE ارائه نماید. سپس به کمک این الگوریتم و استفاده از دو شاخص ARI و SI به عنوان شاخصهای ارزیابی، به حل مساله خوشه‌بندی پرداخته است. برای ارزیابی کارایی الگوریتم GADE، ٦ مجموعه داده مصنوعی و ٤ مجموعه داده واقعی مورد آزمایش قرار گرفته و نتایج با الگوریتم‌های MOCK و NSGA-II مقایسه شده است. نتایج نشان می‌دهند الگوریتم پیشنهادی توانسته است در مجموع ۱۳درصد نسبت به الگوریتم MOCK و ۹درصد نسبت به الگوریتم NSGA-II دقت خوشه‌بندی را افزایش دهد. در مرجع [١٦] با تلفیق الگوریتم بهینه‌سازی انتخاب کلونال و الگوریتم ازدحام ذرات، الگوریتم بهینه‌سازی چندهدفه جدیدی بنام MOIMPSO پیشنهاد و از آن برای حل مساله خوشه‌بندی خودکار استفاده شده است. مقایسه نتایج حاصل از این الگوریتم با دو الگوریتم MOCK و MOCLONAL نشان می‌دهند برای مجموعه داده‌هایی که دارای هم‌پوشانی زیادی هستند، الگوریتم پیشنهادی توانسته است دقت بهتری نسبت به دو الگوریتم دیگر داشته باشد، هرچند که در مورد مجموعه داده‌های با هم‌پوشانی کم، کارایی این الگوریتم در برخی موارد نسبت به دو الگوریتم دیگر ضعیف‌تر است. در مرجع [۱۷] نیز برای حل مساله خوشه‌بندی خودکار از ترکیب دو الگوریتم تبرید فلزات و ازدحام ذرات استفاده شده است. الگوریتم پیشنهادی (با نام MOPSOSA) بطور همزمان سه معیار ارزیابی خوشه با اسامی شاخص DB، شاخص Sym و شاخص Conn را مورد استفاده قرار

داده است. الگوریتم مذکور بر روی ١٤ مجموعه داده مصنوعی و ۵ مجموعه داده واقعی پیاده سازی و نتایج با چهار الگوریتم خوشه‌بندی خودکار دیگر شامل GenClustMOO، GenClustPESA2، MOCK، VGAPS مقایسه شده است. نتایج آزمایش‌ها حاکی از آن است که الگوریتم MOPSOSA توانسته است در بیشتر آزمایشها دقت بهتری نسبت به چهار الگوریتم دیگر داشته باشد.

در حالت کلی کارایی الگوریتم‌های خوشه‌بندی مبتنی بر روشهای بهینه‌سازی هوشمند، متاثر از دو فاکتور زیر است:

١- الگوریتم بهینه‌سازی مورد استفاده.

٢- شاخص‌های ارزیابی و توابع هدف.

الگوریتم بهینه‌سازی مورد استفاده می‌بایست علاوه بر داشتن سرعت همگرایی مناسب، از گیرافتادن الگوریتم در بهینه‌های محلی جلوگیری کند. شاخص‌های ارزیابی و توابع هدف نیز می‌بایست دو وجه خوشه‌بندی زیر را مدنظر قرار دهند:

١. الگوهای موجود در یک خوشه باید تا حد امکان به یکدیگر شبیه باشند.

٢. خوشه‌ها باید تا حد امکان از هم فاصله داشته باشند.

هدف این مقاله، ارائه یک روش کارا جهت خوشه‌بندی خودکار فازی داده‌های هم‌پوشان و غیرهمپوشان است. برای دستیابی به این هدف، از الگوریتم بهینه‌سازی چندهدفه گرگ خاکستری استفاده شده است. به منظور تسریع در فرآیند بهینه‌سازی و جلوگیری از گیر افتادن الگوریتم در بهینه‌های محلی، راهکارهای ابتکاری جدیدی نیز به الگوریتم اضافه شده‌اند. با ارائه یک ساختار مناسب از کروموزم‌های با طول ثابت و استفاده از معیارهای مناسب ارزیابی خوشه، همزمان با خوشه‌بندی بهینه داده‌ها، تعداد بهینه خوشه‌ها نیز بصورت خودکار تعیین می‌شوند. برای اینکه الگوریتم پیشنهادی قادر به حل مجموعه داده‌های همپوشان هم باشد، از یک تابع هدف فازی استفاده شده است. جهت اعتبارسنجی الگوریتم، نتایج حاصل از خوشه‌بندی خودکار بر روی ١٠ مجموعه داده استاندارد (شامل ۵ مجموعه داده مصنوعی و ۵ مجموعه داده واقعی) آزمایش و نتایج مورد تجزیه و تحلیل قرار داده شده‌اند. ساختار مقاله بصورت زیر است.

در بخش ٢ بیان ریاضی مساله خوشه‌بندی فازی و توابع هدف مورد استفاده بصورت مختصر ارائه خواهد. در بخش ٣، الگوریتم بهینه‌سازی گرگ خاکستری تک هدفه توضیح داده می‌شود. در بخش ٤، ساختار روش پیشنهادی معرفی شده و بخش‌های مختلف آن با جزئیات ارائه خواهد شد. در بخش ۵، روش پیشنهادی بر روی ١٠ مجموعه داده استاندارد پیاده سازی شده و نتایج مورد تجزیه و تحلیل قرار خواهند گرفت. بخش ٦ و ٧ نیز به‌ترتیب به نتیجه‌گیری و پیشنهاداتی جهت کارهای آینده اختصاص دارد.

٢. الگوریتم خوشه‌بندی FCM

الگوریتم خوشه‌بندی فازی C-Means که با نام FCM نیز شناخته می‌شود، درواقع معادل فازی الگوریتم خوشه‌بندی غیرفازی K-means است. همانطور که قبلا نیز اشاره شد، در روش K-means هر داده تنها به یک خوشه تعلق می‌گیرد درحالیکه در FCM هر داده می‌تواند به چند خوشه با درجه عضویت بین صفر و یک تعلق داشته باشد [١٢]. این روش، جزء روش‌های خوشه‌بندی نرم محسوب می‌گردد؛ بدین معنی که در این روش هر داده می‌تواند با یک درجه عضویت، به چندین

خوشه تعلق داشته باشد. الگوریتم FCM تابع هدف زیر را بهینه می‌کند [۱۸].

$$\mathbf{J}_m = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^m d_{ik}^2 = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^m \|x_k - v_i\|^2 \quad (۱)$$

در رابطه فوق، m یک عدد حقیقی بزرگ‌تر از یک است که در اکثر پژوهش‌ها مقدار آن دو در نظر گرفته می‌شود. d_{ik}^2 بیانگر معیار شباهت در فضای n بعدی است. x_k بیانگر داده kام و v_i بیانگر نماینده یا مرکز خوشه iام می‌باشد. u_{ik} میزان تعلق داده iام در خوشه kام را نشان می‌دهد. علامت $\|*\|$ میزان تشابه (فاصله) داده با مرکز خوشه است که می‌توان از هر تابعی که بیانگر تشابه داده و مرکز خوشه باشد استفاده کرد. پارامتر C نیز بیانگر تعداد خوشه‌ها است. میزان تعلق داده x_k به iامین خوشه، u_{ik}، طبق رابطه زیر محاسبه می‌شود [۱۸]:

$$u_{ik} = \frac{1}{\sum_{j=1}^{C}\left(\frac{d_{ik}}{d_{jk}}\right)^{\frac{2}{(m-1)}}}, \quad 1 \le k \le M, \quad 1 \le i \le C \quad (۲)$$

همچنین ماتریس مربوط به مراکز خوشه‌ها نیز از رابطه زیر قابل محاسبه است [۱۹].

$$v_i = \frac{\sum_{k=1}^{M}(u_{ik})^m x_k}{\sum_{k=1}^{M}(u_{ik})^m}, \quad 1 \le i \le C \quad (۳)$$

مهم‌ترین معیارهای شباهت مورد استفاده در حل مسائل خوشه‌بندی، معیار فاصله است. از پرکاربردترین معیارهای فاصله نیز می‌توان به معیار فاصله اقلیدسی و مینکوفسکی اشاره کرد. با توجه به اینکه در خوشه‌بندی مجموعه داده‌های با ابعاد بالا، غالبا معیار فاصله اقلیدسی کارایی بهتری نسبت به معیار فاصله مینکوفسکی دارد [۱۹]، در این مقاله از معیار فاصله اقلیدسی استفاده شده است.

بیشتر توابع هدفی که در حل مسائل خوشه‌بندی مورد استفاده قرار می‌گیرند، به‌طور معمول مبتنی بر دو شاخص فشردگی و جدایی هستند. معیار فشردگی، به پراکندگی بین داده‌های درون یک خوشه یا بین داده‌ها و سرخوشه‌ها اشاره دارد که بایستی کمینه گردد. معیار جدایی، دور بودن فاصله بین خوشه‌های مختلف را بیان می‌کند که برخلاف فشردگی بایستی بیشینه گردد [۱۹]. اگر تنها معیار فشردگی مورد استفاده قرار گیرد در آن‌صورت هر داده می‌تواند به صورت یک خوشه در نظر گرفته شود. چرا که هیچ خوشه‌ای فشرده‌تر از خوشه‌ای با یک داده نمی‌باشد. اگر تنها معیار جدایی در نظر گرفته شود در آن‌صورت بهترین خوشه‌بندی این است که کل داده‌ها را یک خوشه بگیریم.

اکثر معیارهای خوشه‌بندی که مبتنی بر شاخص جدایی و یا فشردگی هستند، منحصرا از اطلاعات مراکز استفاده می‌کنند. استفاده از اطلاعات مراکز به تنهایی، فقط ساختار هندسی خوشه‌ها را مشخص می‌کند [۱۹]. در اغلب موارد، الگوریتم‌های خوشه‌بندی مبتنی بر ساختار هندسی خوشه‌ها، قادر به تفکیک خوشه‌های همپوشان نیستند. از این‌رو نیاز به مکانیزمی است که بتوان بر اساس آن و بدون داشتن دانش قبلی، علاوه بر تعیین تعداد بهینه خوشه‌ها، خوشه‌بندی مناسب داده‌ها را نیز انجام داد.

در این مقاله برای حل مساله خوشه‌بندی خودکار از دو شاخص فشردگی و همپوشانی–جدایی استفاده شده است. تابع هدف اول که بر اساس معیار فشردگی است، همان تابع هدف ارائه شده در رابطه (۱) است. دومین تابع هدف نیز، معیار هم‌پوشانی و جدایی است که قابلیت مواجهه با مشکل هم‌پوشانی را دارد. این شاخص دارای

دو بخش است: بخش اول، معیار هم‌پوشانی است که با استفاده از درجات فازی، هم‌پوشانی بین خوشه‌ای را محاسبه می‌کند و توسط رابطه (٤) تعریف می‌شود [٢٠, ٢١]:

$$O_{\perp}(u_k(x_k), C) = \perp_{l=2,C}^{1}(\perp_{i=1,C}^{1} u_k) \qquad (٤)$$

در ابطه فوق $\boldsymbol{x_k}$ دارای بردار عضویت $\mathbf{u_k(x_k)} = (\mathbf{u_{1k}}, \dots, \mathbf{u_{ck}})$ می‌باشد. همچنین عملگر $\perp$ بیانگر t-norm فازی است که در این پژوهش از عملگر min به جای آن استفاده شده است، یعنی: $\mathbf{a} \perp \mathbf{b} = \mathbf{min(a, b)}$. برای بخش دوم که همان شاخص جدایی است، از عملگر s-norm فازی استفاده می‌شود [٢١]. در این مقاله از عملگر max، به عنوان s-norm فازی استفاده شده است. در نهایت دومین تابع هدف که بیانگر معیار همپوشانی–جدایی (OS) بین M الگو است، از تقسیم شاخص اول بر شاخص دوم محاسبه می‌شود که به‌صورت زیر تعریف قابل تعریف است [٢٠]:

$$\mathrm{OS} = \frac{1}{\mathrm{M}} \sum_{\mathrm{k}=1}^{\mathrm{M}} \frac{\mathrm{O}_{\perp}(\mathrm{u_k}(\mathrm{x_k}), \mathrm{C})}{\max\limits_{\mathrm{i}=1,\mathrm{C}} \mathrm{u_{ik}}} \qquad (٥)$$

جزئیات بیشتری از تابع هدف دوم و اثبات ریاضی آن در مراجع [٢٠, ٢١] آورده شده است.

٣. الگوریتم بهینه‌سازی گرگ خاکستری

در بخش قبل مساله خوشه‌بندی فازی در قالب یک مساله بهینه‌سازی دو هدفه، فرمول‌بندی شد. انتخاب یک الگوریتم بهینه‌سازی مناسب برای حل مساله خوشه‌بندی خودکار تاثیر به‌سزایی بر نتایج خوشه‌بندی خواهد داشت. برای نیل به این هدف، در این مقاله از الگوریتم بهینه‌سازی گرگ خاکستری استفاده شده است.

با توجه به اینکه مساله خوشه‌بندی خودکار فازی از نوع مسائل بهینه سازی چندهدفه غیرخطی و نامحدب است [١٩]، برای افزایش دقت خوشه‌بندی، ساختار الگوریتم به‌گونه‌ای تغییر داده شده است که امکان استفاده از آن برای حل مسائل بهینه سازی چندهدفه نیز وجود داشته باشد. علاوه بر این به منظور تسریع در فرآیند بهینه‌سازی و جلوگیری از گیرافتادن الگوریتم در بهینه های محلی، راهکارهای ابتکاری جدیدی به الگوریتم اضافه شده‌اند. در ادامه ساختار این الگوریتم به اختصار توضیح داده شده است.

الگوریتم گرگ خاکستری در سال ٢٠١٤ بر مبنای زندگی گرگهای خاکستری ارائه‌شده است [٢٢]. این حیوانات به‌صورت گروهی زندگی می‌کنند و رهبر گروه، آلفا مسوول تصمیم‌گیری در زمینه‌هایی همچون حمله و زمان آن است. روش شکار این‌گونه از حیوانات که در این الگوریتم برای بهینه‌سازی مورد الهام قرارگرفته است، شامل سه مرحله زیر است:

١: ردگیری، تعقیب و نزدیک شدن به شکار

٢: دنبال کردن و محاصره کردن شکار تا زمانی که شکار از حرکت باز ایستد.

٣: حمله کردن به‌طرف شکار

به‌منظور مدل سازی رفتار اجتماعی گرگها، یک جمعیت تصادفی از راه‌حلها تولید، و بهترین راه‌حل بنام آلفا (α)، دومین و سومین راه‌حل بهتر، نیز به‌ترتیب بتا (β) و دلتا (δ) نامیده می شوند. همچنین سایر راه‌حل‌ها نیز به‌عنوان گرگهای دسته امگا (ω) در نظر گرفته می‌شوند. الگوریتم گرگ خاکستری از سه جواب α، β و δ جهت هدایت شکار(بهینه‌سازی) استفاده می‌کند و جوابهای ω از این سه پیروی می‌کنند. به‌منظور مدلسازی سه بعدی، ابتدا نقاط

اطراف طعمه مشـــخص، سپس به سمت طعمه حرکت و درنهایت به طعمه حمله می‌شود. برای مشخص کردن نقاط اطراف طعمه از روابط زیر استفاده می‌شود [۲۲].

$$\vec{D} = \left|\vec{C}.\overrightarrow{X_p}(t) - \vec{X}(t)\right|$$
$$\vec{X}(t+1) = \overrightarrow{X_p}(t) - \vec{A}.\vec{D} \qquad (۶)$$

که در آن $\boldsymbol{t}$ بیانگر تکرار فعلی الگوریتم، $\vec{\boldsymbol{A}}$ و $\vec{\boldsymbol{C}}$ بردارهای ضریب، $\overrightarrow{\boldsymbol{X_p}}$ بردار موقعیت طعمه و $\overrightarrow{\boldsymbol{X}}$ بیانگر بردار موقعیت گرگ خاکستری می‌باشد. بردارهای $\vec{\boldsymbol{A}}$ و $\vec{\boldsymbol{C}}$ نیز از رابطه زیر بدست می‌آیند [۲۲]:

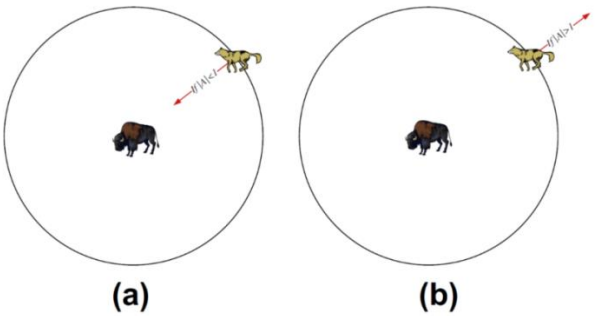
$$\vec{A} = 2\vec{a}.\overrightarrow{r_1} - \vec{a}$$
$$\vec{C} = 2.\overrightarrow{r_2} \qquad (۷)$$

در رابطه فوق، $\overrightarrow{\boldsymbol{r_1}}$ و $\overrightarrow{\boldsymbol{r_2}}$ بردارهای تصادفی در بازه [۰و۱] هستند. پارامتر $\vec{\boldsymbol{a}}$ نیز در طی فرایند اجرای الگوریتم از ۲ تا صفر کاهش می‌یابد. با توجه به اینکه در فضای جســتجوی اولیه ایده‌ای در مورد موقعیت شــکار وجود ندارد، فرض می‌شود گرگهای $\boldsymbol{\alpha}$، $\boldsymbol{\beta}$ و $\boldsymbol{\delta}$ به‌ترتیب سه تا از بهترین جوابهایی هستند که تاکنون بدست آمده‌اند. سایر جوابها (که گرگهای $\boldsymbol{\omega}$ نامیده می‌شوند) می‌بایست موقعیت خود را جهت همگرایی به سمت جواب بهینه، مطابق روابط زیر تغییر دهند [۲۲].

$$\overrightarrow{D_\alpha} = \left|\overrightarrow{C_1}.\overrightarrow{X_\alpha} - \vec{X}\right|$$
$$\overrightarrow{D_\beta} = \left|\overrightarrow{C_2}.\overrightarrow{X_\beta} - \vec{X}\right|$$
$$\overrightarrow{D_\delta} = \left|\overrightarrow{C_3}.\overrightarrow{X_\delta} - \vec{X}\right|$$
$$\overrightarrow{X_1} = \overrightarrow{X_\alpha} - \overrightarrow{A_1}.\left(\overrightarrow{D_\alpha}\right) \qquad (۸)$$
$$\overrightarrow{X_2} = \overrightarrow{X_\beta} - \overrightarrow{A_2}.\left(\overrightarrow{D_\beta}\right)$$
$$\overrightarrow{X_3} = \overrightarrow{X_\delta} - \overrightarrow{A_3}.\left(\overrightarrow{D_\delta}\right)$$
$$\overrightarrow{X}(t+1) = \frac{\overrightarrow{X_1} + \overrightarrow{X_2} + \overrightarrow{X_3}}{3}$$

همانطــور که گفته شــد گرگها زمانی که شــکار را متوقــف کردنــد به‌طرف آن حملــه می‌کنند. به‌منظور مدل‌ســازی ریاضــی نزدیک شــدن به طعمه، مقدار a کاهش داده می‌شــود. باید توجه داشــت که محدوده تغییرات A نیز با a کاهش خواهد یافت. به‌عبارت‌دیگر، A یک مقدار تصادفی در بازه $[-2a, 2a]$ است که در آن a از ۲ تا ۰ در طی مراحل تکرار الگوریتم، کاهش می‌یابد. به ازای مقادیر تصــادفی A در بازه [1 ,1-]، موقعیت بعدی عامل جســتجو، می‌تواند در هر مکانی بین موقعیت فعلی عامل و موقعیت طعمه باشــد. در الگوریتم گرگ خاکســتری، موقعیت گرگها با توجه به موقعیت آلفا، بتا و دلتا بروز رســانی می‌شــود. ممکن اســت بعضا لازم باشد گرگها برای پیداکردن شکار از هم دور شــوند و همیشه نزدیک شدن اتفاق نمی‌افتد. برای مدلســازی چنین پدیده‌ای، از مقدار تصــادفی بزرگتر از ۱ یا کمتر از ۱- برای A اســتفاده می‌شود. شکل (۱) نشان می‌دهد که اگر $|A| > 1$ باشد گرگها از شــکار دور می‌شــوند، درحالیکه به ازای $|A| < 1$، گرگ‌ها وادار به حمله به ســمت شــکار می‌شــوند. فلوچارت الگوریتم گرگ خاکســتری در شکل (۲) نشان داده شده است.

شکل ۱. (a) حمله به شکار، (b) جستجو برای شکار [۲۲]



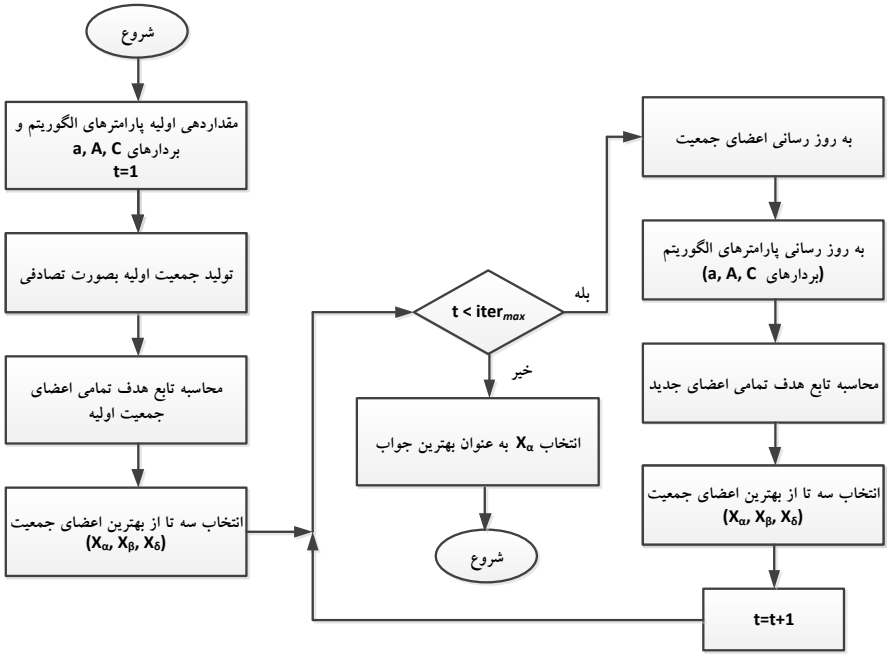
طبق مرجع [۲۲]، پیاده‌سازی این الگوریتم بر روی چند تابع ریاضی استاندارد، حاکی از آن است که الگوریتم مذکور قادر است در مورد اکثر توابع ریاضی، جواب نسبتا بهتری در مقایسه با سایر روش‌های بهینه‌سازی داشته باشد. اما طبق نتایج ارائه شده در مرجع [۲٤]، عدم وجود پارامتر کنترلی در این الگوریتم، باعث خواهد شد در مسائل بهینه‌سازی نامحدب و یا مسائلی که دارای ابعاد و پیچیدگی بالایی هستند، الگوریتم دچار همگرایی زودرس شده و در بهینه محلی گیر کند. مهم‌ترین عاملی که کارایی و دقت یک الگوریتم بهینه‌سازی هوشمند را کنترل می‌کند، ایجاد مصالحه بین اکتشاف (Exploration) و استخراج (Exploitation) است [۲۳]. مفهوم اکتشاف به الگوریتم این امکان را می‌دهد که بتواند جهت دستیابی به پاسخهای جدید، فضای جواب مساله را با بالاترین راندمان و بدون گیرافتادن در بهینه‌های محلی جستجو نماید. به عبارت دیگر مفهوم اکتشاف به معنای توانایی الگوریتم در جستجوی مناطق مختلف فضای جواب، جهت یافتن پاسخهای جدید است. در حالیکه مفهوم استخراج باعث می‌شود الگوریتم بتواند مکانهای بهینه را بصورت محلی و متمرکز برای یافتن بهترین جواب، جستجو نماید. به عبارت دیگر، استخراج به معنی قابلیت متمرکز کردن جستجو در محدوده مطلوب است تا جواب مورد نظر موشکافی شود. بنابراین برای رسیدن به جواب بهینه سراسری، می‌بایست مصالحه‌ای بین دو مفهوم اکتشاف و استخراج صورت پذیرد.

در یک الگوریتم بهینه‌سازی هوشمند، ایجاد مصالحه بین اکتشاف و استخراج توسط پارامترهای کنترلی الگوریتم صورت می‌پذیرد. هر چند زیاد بودن تعداد پارامترهای کنترلی یک الگوریتم بهینه‌سازی لزوما به معنای کارایی بهتر آن نیست و غالبا هر چه تعداد پارامترها کمتر باشد، امکان تنظیم پارامترها جهت افزایش سرعت همگرایی و جلوگیری از گیر افتادن در بهینه محلی بیشتر است، اما الگوریتم بهینه‌سازی گرگ خاکستری فاقد هرگونه پارامتر کنترلی است. تنها پارامتری که در این الگوریتم طی فرایند بهینه‌سازی تغییر می‌کند پارامتر a است که مقدار آن برای تمامی مسائل بهینه‌سازی بصورت نزولی از دو تا صفر کاهش می‌یابد. تاثیر مقدار a بر کارایی الگوریتم، همانند تاثیر پارامتر دما در الگوریتم تبرید فلزات است [۲٤]. از این‌رو در این مقاله برای ایجاد تعادل بین خاصیت اکتشاف و استخراج و جلوگیری از گیر افتادن الگوریتم در بهینه محلی، از رابطه زیر برای محاسبه پارامتر a استفاده شده است.

$$a(t) = \left(a_{max} - (a_{max} - a_{min}) \frac{t}{Maxgen}\right) \quad (۹)$$

در رابطه فوق، $Maxgen$ حداکثر تعداد تکرار الگوریتم، a_{max} و a_{min} نیز بیانگر محدوده تغییرات a ا ست که مقادیر آنها مطابق با فرایندی که در بخش های بعدی ارائه خواهد شد، تعیین می شود. همانطور که م شاهده می‌شود همانند الگوریتم تبرید فلزات، پارامتر a از یک مقدار م شخص شروع و در هر مرحله مقدار آن نسبت به مرحله قبل کمتر می‌شود. به عبارت دیگر ابتدا با انرژی بالایی به هر عضو از جمعیت، اجازه فرار از بهینه محلی داده می‌شود و طی فرایند جستجو، انرژی نیز کاهش یافته و در نهایت الگوریتم به سمت راه حل بهینه سراسری همگرا خواهد شد.

شکل ۲. فلوچارت الگوریتم گرگ خاکستری [۲۲]



٤. ساختار روش پیشنهادی

شکل (۳) فلوچارت الگوریتم خوشه‌بندی فازی مبتنی بر روش بهینه‌سازی گرگ خاکستری چندهدفه (FCM-MOGWO) را نشان می‌دهد. در ادامه بخشهای مختلف این فلوچارت تشریح خواهد شد. پس از آن به معرفی الگوریتم FCM-MOGWO به‌عنوان هسته اصلی روش پیشنهادی پرداخته می‌شود.

٤-١. تقسیم‌بندی داده‌های ورودی

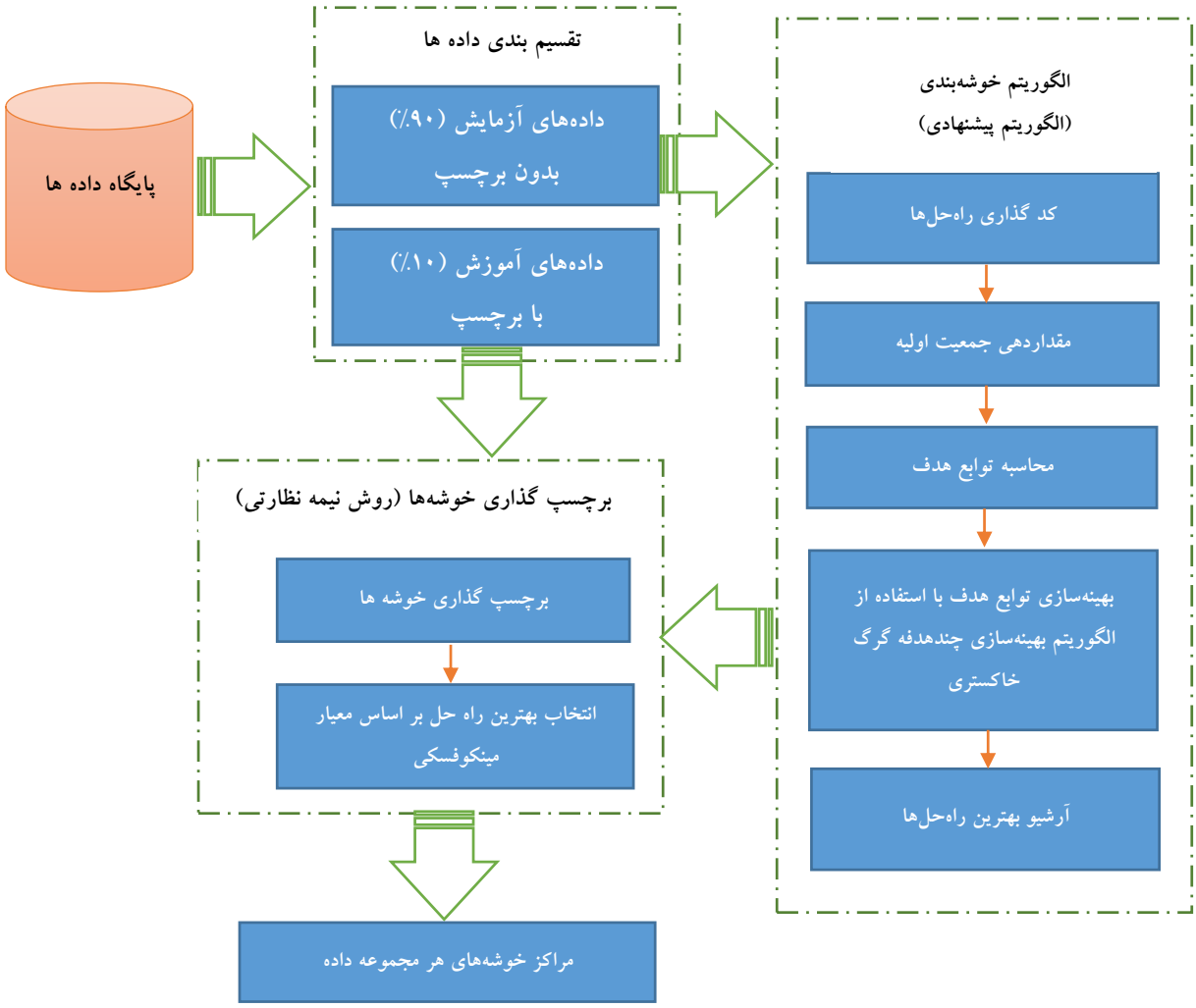
روند کار به این صورت است که داده‌های ورودی با استفاده از روش Hold Out، به‌صورت تصادفی به دو بخش داده‌های آموزش و داده‌های آزمایش تقسیم‌بندی می‌شوند. داده‌های آموزش که ۱۰درصد کل داده‌ها هستند دارای برچسب کلاس بوده درحالیکه ۹۰درصد باقیمانده (داده‌های آزمایش) فاقد برچسب کلاس هستند. داده‌های آموزش به‌عنوان ورودی به الگوریتم FCM-MOGWO اعمال می‌شوند. این تقسیم‌بندی پیش‌نیازی برای انجام روش یادگیری نیمه‌نظارتی است که در مراحل پایانی تشریح می‌گردد.

٤-٢. روش کدگذاری راه‌حل

در حل مساله خوشه‌بندی با استفاده از الگوریتم‌های بهینه‌سازی هوشمند، روش‌های مختلفی جهت کدگذاری وجود دارد که به‌تناسب ماهیت مسئله و الگوریتمی که به‌کار گرفته می‌شود، بایستی مناسب‌ترین روش انتخاب گردد. در این مقاله با توجه به پیوسته بودن فضای جستجو، از روش کدگذاری مبتنی بر مراکز با داده‌های اعشاری استفاده شده است. همچنین، از آنجایی‌که تشخیص تعداد مناسب خوشه‌ها غالبا از طریق راه‌حل‌های با طول متغیر انجام می‌گیرد و الگوریتم گرگ خاکستری فقط با راه‌حل‌هایی با طول ثابت کار می‌کند، روش رمزگذاری مورد استفاده نیز روش اعشاری با طول

ثابت انتخاب شده است. در این روش، تمام راه‌حل‌های خوشه‌بندی، دارای تعداد از پیش تعیین‌شده K_{max} سرخوشه هستند. ازاین‌رو، اندازه تمام اعضای جمعیت در طول فرآیند بهینه‌سازی یکسان و به‌اندازه $D \times K_{max}$ خواهد بود که D بیانگر ابعاد یا تعداد ویژگیهای مجموعه داده است. همچنین، برای تعیین این‌که کدام سرخوشه‌ها در فرایند خوشه‌بندی در هر تکرار الگوریتم شرکت دارند، معمولاً از یک بردار پوشش یا آستانه فعال‌سازی استفاده می‌شود. درصورتی‌که مقدار آستانه به ازای یک سرخوشه از ۰/۵ بیشتر بود، سرخوشه متناظر فعال و در غیر این صورت غیرفعال خواهد بود. شکل (٤) یک رمزگذاری با طول ثابت و با $K_{max}=5$ را نشان می‌دهد که با توجه به مقادیر بردار آستانه، فقط K=3 سرخوشه فعال خواهد بود.

شکل ۳. فلوچارت روش پیشنهادی.



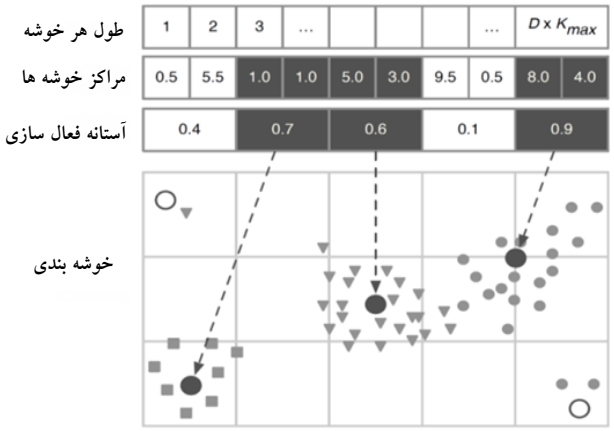
در الگوریتم بهینه‌سازی گرگ خاکستری، هر گرگ بیان‌گر یک راه‌حل برای مساله می‌باشد. هر راه‌حل نیز شامل تعدادی متغیرهای تصمیم‌گیری است که در این‌جا به‌دلیل استفاده از روش رمزگذاری مبتنی‌بر مرکز، متغیرهای تصمیم‌گیری شامل ویژگی‌های سرخوشه‌ها خواهد بود. به‌عبارت‌دیگر، هر گرگ به‌عنوان یک راه‌حل، مختصات $\boldsymbol{k}$ سرخوشه را در خود دارد. همچنین در این روش، برای نمایش یک راه‌حل از یک بردار به طول

$K_{max} + K_{max} \times D$ استفاده‌شده که ساختار آن بشرح زیر است:

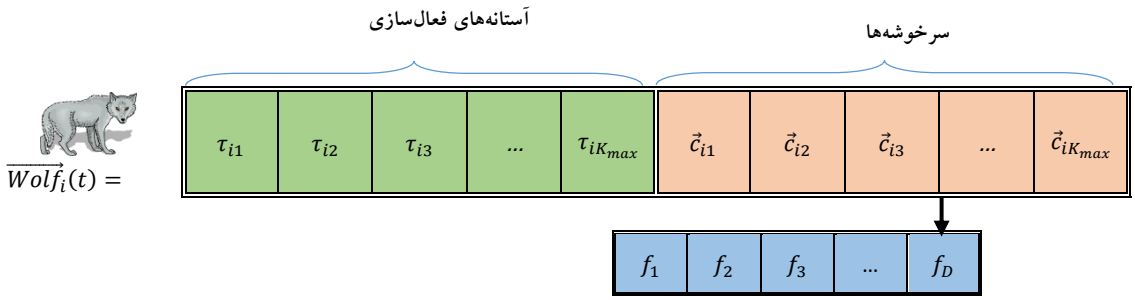
محتویات کروموزم اول تا کروموزم K_{max}، اعداد مثبت اعشاری در بازه [۱و۰] هستند که بیان‌گر فعال بودن یا غیر فعال بودن سرخوشه متناظر می‌باشند. این کروموزم‌ها که با τ نشان داده شده‌اند، آستانه فعال‌سازی نامیده می‌شوند. سایر موقعیت‌های بردار نیز به K_{max} مرکز سرخوشه ($\vec{c}$)، با ابعاد (f) اختصاص دارد. برای درک بهتر ساختار، یک بردار راه‌حل (یک گرگ) در شکل (۵) آورده شده که بیان‌گر موقعیت گرگ i در تکرار t است. در این روش تمام راه‌حل‌های خوشه‌بندی دارای تعداد K_{max} سرخوشه می‌باشند و مقدار آستانه برابر ۰/۵ در نظر گرفته شده است. در این حالت، jامین مرکز خوشه در گرگ i، در تکرار tام در صورتی فعال خواهد بود که $\tau_{ij} > 0.5$ باشد و اگر $\tau_{ij} < 0.5$ بود، jامین مرکز خوشه از گرگ i در فرآیند خوشه‌بندی در تکرار tام غیرفعال خواهد شد. این قانون به‌صورت رابطه زیر نیز قابل بیان است:

$$\begin{cases} \tau_{ij} > 0.5 \ : \ \vec{c}_{ij} = Active \\ \tau_{ij} < 0.5 \ : \ \vec{c}_{ij} = Deactive \end{cases} \quad (۱۰)$$

شکل ٤. روش رمزگذاری اعشاری با طول ثابت



شکل ۵. یک نمونه گرگ خاکستری به‌عنوان راه‌حل خوشه‌بندی.



۳-٤. تولید جمعیت اولیه

به‌منظور تولید جمعیت اولیه در روش پیشنهادی، ابتدا به کمک یک عملگر تصادفی قسمت آستانه فعال‌سازی مربوط به راه‌حل ایجاد می‌گردد. برای این منظور به تعداد K_{max} عدد تصادفی اعشاری در بازه [۱و۰] تولید می‌گردد. ازآنجایی‌که در مسئله خوشه‌بندی حداقل تعداد خوشه‌ها باید ۲ باشد ($K_{min} = 2$) تعداد فعال‌سازهای با مقدار آستانه بیشتر از ۰/۵، نباید کمتر از ۲ باشد. به‌همین دلیل، در روش پیشنهادی از یک واحد کنترلی برای کنترل این موضوع و جلوگیری از تولید جمعیت

مشکل‌دار استفاده شده و تا زمانی‌که تعداد متغیرهای فعال‌ساز با مقدار آستانه ۰/۵، کمتر از ۲ باشد، عملگر تصادفی را وادار به تولید دوباره می‌نماید. پس از ساخته شدن قسمت کنترل آستانه فعال‌سازی، به ساخت ادامه راه‌حل پرداخته می‌شود. برای این منظور، ابتدا یک جایگشت تصادفی به‌اندازه K_{max} از M نمونه موجود در مجموعه داده فراهم می‌شود. این K_{max} نمونه، به‌عنوان سرخوشه‌های اولیه از راه‌حل ایجاد شده در نظر گرفته می‌شوند. سپس، هر دو قسمت آستانه و سرخوشه‌ها در کنار هم قرار گرفته و یک راه‌حل را تشکیل می‌دهند. فرآیند گفته شده تا تولید کامل جمعیت اولیه تکرار می‌گردد. پس از تولید جمعیت اولیه، اعضای جمعیت می‌بایست توسط توابع هدف مورد ارزیابی قرار گیرند. توابع هدف مورد استفاده در این مقاله همان توابع هدف معرفی شده در روابط (۱) و (۵)، هستند.

٤-٤. الگوریتم بهینه‌سازی گرگ خاکستری چندهدفه

با توجه به اینکه مساله خوشه‌بندی فازی مطرح شده در این مقاله، یک مساله بهینه‌سازی چندهدفه است، برای حل آن می‌بایست از الگوریتم‌های بهینه‌سازی چندهدفه استفاده شود. اما الگوریتم بهینه‌سازی گرگ خاکستری که در بخش قبل معرفی شد بیشتر برای حل مسائل بهینه‌سازی تک هدفه بکار می‌رود. از این‌رو هدف این بخش تعمیم الگوریتم مذکور به یک الگوریتم بهینه‌سازی چندهدفه است.

در اولین قدم، برای تبدیل الگوریتم بهینه‌سازی گرگ خاکستری تک هدفه به الگوریتم چندهدفه، لیستی بنام لیست آرشیو در نظر می‌گیریم که اعضای آن منحصرا جوابهای نامغلوب هستند. یعنی جوابهایی که اولا جزو جوابهای بهینه مساله بوده و ثانیا هیچ ارجحیتی نسبت به‌هم ندارند. در شروع الگوریتم، این لیست خالی بوده و هیچ جوابی در آن قرار ندارد. بعد از ارزیابی جمعیت توسط توابع هدف (روابط (۱) و (۵))، تمامی جوابها با هم مقایسه شده و جوابهایی که توسط هیچ یک از اعضای جمعیت مغلوب نشده‌اند به این لیست منتقل می‌شوند. در طول فرآیند تکرار الگوریتم، راه‌حل‌های مطلوبی که تاکنون به‌دست‌آمده‌اند، با اعضای فعلی آرشیو مقایسه خواهند شد. یکی از سه حالت زیر ممکن است رخ دهد:

- اگر عضو جدید به‌وسیله حداقل یکی از اعضای فعلی آرشیو مغلوب گردد، به عضو جدید اجازه ورود به آرشیو داده نمی‌شود.
- اگر عضو جدید بر یک یا چند عضو فعلی آرشیو غلبه کند، راه‌حل(های) مغلوب از آرشیو حذف و راه‌حل جدید وارد آرشیو می‌شوند.
- درصورتی‌که نه راه‌حل جدید و نه اعضای آرشیو هیچ‌کدام بر یکدیگر غلبه نکنند، راه‌حل جدید به آرشیو اضافه می‌شود.

موضوع دیگری که می‌بایست مد نظر قرار گیرد آن است که در الگوریتم‌های بهینه‌سازی تک هدفه، گرگ‌های آلفا، بتا و دلتا در هر تکرار، به‌ترتیب سه تا از بهترین جواب‌های الگوریتم تا آن تکرار هستند. اما در مسائل بهینه‌سازی چندهدفه همانطور که قبلا ذکر شد هیچ یک از جوابهای پارتو (جوابهای موجود در لیست آرشیو) بر همدیگر برتری ندارند. در این مقاله با توجه به نوع مساله (که در اینجا خوشه‌بندی است) از شاخص DB جهت انتخاب گرگ‌های رهبر استفاده شده است. این معیار تابعی از نسبت مجموع پراکندگی درون خوشه‌ای به پراکندگی بین خوشه‌ها است. شاخص DB، با استفاده از رابطه ریاضی زیر بیان می‌شود [٢٦].

$$DB(C) = \frac{1}{K}\sum_{c_k \in C} \max_{c_r \in C \setminus c_k} \left\{ \frac{S(c_k) + S(c_r)}{d_e(\bar{c}_k, \bar{c}_r)} \right\} \qquad (۱۱)$$

که در آن:

$$S(c_k) = \frac{1}{n_k}\sum_{x_i \in c_k} d_e\,(X_i, \bar{c}_k) \qquad (۱۲)$$

خوشه‌بندی با کمترین مقدار از این شاخص، بهترین نتیجه را خواهد داشت.

٤-٥. برچسب‌زنی به داده‌های آزمایشی

پس از پایان فرآیند خوشه‌بندی فازی، داده‌ها در خوشه‌های مرتبط گروه‌بندی می‌شوند. اکنون بایستی با توجه به داده‌های آموزشی که برچسب کلاس آن‌ها مشخص می‌باشد، برچسب کلاس داده‌های آزمایشی نیز مشخص گردد. به‌طور معمول، برچسب‌زنی به داده‌های بدون برچسب آزمایشی، براساس معیار نزدیک‌ترین مرکز به نمونه انجام می‌شود؛ یعنی برچسب نزدیک‌ترین مرکز به نمونه، برچسب نمونه خواهد بود. با این حال، در روش پیشنهادی به‌منظور افزایش دقت خوشه‌بندی فازی، از ترکیب معیار نزدیک‌ترین مرکز به همراه درجه عضویت فازی استفاده شده است. فرآیند برچسب‌زنی بدین‌صورت انجام می‌گیرد که ابتدا نمونه‌های برچسب‌دار از داده آموزش به خوشه‌های متناظر با بیشترین درجه عضویت نگاشت می‌شوند. سپس فراوانی هر کلاس از داده آموزشی نسبت به هر خوشه محاسبه می‌گردد. در انتها، کلاس به خوشه‌ای اختصاص داده می‌شود که دارای بیشترین فراوانی باشد. برای جلوگیری از برچسب‌گذاری تکراری خوشه‌ها، بعد از برچسب‌گذاری هر خوشه، خوشه مورد نظر از فرایند برچسب‌گذاری حذف خواهد شد. همچنین اگر فراوانی نمونه‌ها در دو گروه برابر باشد، برچسب‌زنی براساس نمونه‌هایی با بیشترین درجه عضویت در خوشه متناظر، صورت خواهد گرفت [۱۹].

٤-٦. ارزیابی و انتخاب بهترین راه‌حل

معیارهای ارزیابی خوشه‌بندی خودکار فازی، میزان بهینه بودن نتیجه یک خوشه‌بندی را نشان می‌دهند. یک معیار ارزیابی خوشه‌بندی می‌بایست دو هدف زیر را دنبال کند:

۱- مشخص کردن تعداد بهینه خوشه‌ها

۲-بدست آوردن بهینه‌ترین حالت خوشه‌بندی با توجه به تعداد خوشه‌ها.

شاخص‌های ارزیابی مختلفی برای ارزیابی خوشه‌بندی فازی وجود دارد که از جمله آنها می‌توان به معیار DI، DB، CS، MS، XB، و WGS اشاره کرد. فرمول‌بندی ریاضی معیارهای ارزیابی فوق، به همراه مزایا و محدودیت‌های هر یک از آنها به تفصیل در مرجع [۲۷] آورده شده است. در این مقاله از معیار مینکوفسکی (MS) که طبق رابطه (۱۳) محاسبه می‌شود، استفاده شده است. مقدار بهینه این معیار صفر بوده و هر چه مقدار آن به صفر نزدیکتر باشد، نشان‌دهنده بهتر بودن کیفیت خوشه‌بندی است.

$$\boldsymbol{MS} = \frac{\sqrt{\sum_i \binom{n_{i.}}{2} + \sum_j \binom{n_{.j}}{2} - 2\sum_{ij}\binom{n_{ij}}{2}}}{\sqrt{\sum_j \binom{n_{.j}}{2}}} \qquad (۱۳)$$

در رابطه فوق، $\boldsymbol{n_{ij}}$ بیان‌گر تعداد داده‌هایی است که بین دو خوشه $\boldsymbol{i}$ و $\boldsymbol{j}$ مشترک هستند. $\boldsymbol{n_{i.}}$ تعداد داده‌های موجود در خوشه $\boldsymbol{i}$ و $\boldsymbol{n_{.j}}$ نیز تعداد داده‌های موجود در شاخه $\boldsymbol{j}$ است. پس از محاسبه این معیار برای تمامی اعضای موجود در لیست آرشیو، راه‌حلی که کمترین مقدار از لحاظ معیار مینکوفسکی را دارد، به‌عنوان بهترین راه‌حل انتخاب می‌گردد.

۵. شبیه‌سازی و تحلیل نتایج

در این بخش نتایج حاصل از روش پیشنهادی بر روی چند مجموعه از داده‌های استاندارد که در دیگر مقالات به کارگرفته شده‌اند، ارائه و با سایر کارهای انجام شده در پژوهش‌های دیگر مقایسه خواهد شد.

مجموعه داده‌های مورد استفاده در این پژوهش شامل ۵ مجموعه داده مصنوعی و ۵ مجموعه داده واقعی هستند. مجموعه داده‌های مصنوعی شامل: AD_5_2 ارائه شده در مرجع [۲۸]، AD_10_2 ارائه شده در مرجع [۲۹] و Square1 و Square4 و Sizes5 ارائه شده در مرجع [۳۰] می‌باشند. مجموعه داده‌های واقعی نیز که شامل: Iris و Breast Cancer و New thyroid و Wine و Liver Disorders می‌باشند، از داده‌های استاندارد سایت UCI دریافت شده‌اند [۳۱]. تعداد الگوها و ویژگیهای هر یک از مجموعه داده‌های فوق در جدول (۱) آورده شده است. در این جدول، **D** بیان‌گر تعداد ویژگی‌ها (ابعاد)، **K** بیان‌گر تعداد خوشه‌ها و **M** بیان‌گر تعداد نمونه‌ها می‌باشد. $\mathbf{M_i}$ نیز تعداد اعضا در هر خوشه از مجموعه داده را نشان می‌دهد.

جدول ۱. خلاصه‌ای از مجموعه داده‌ها

Dataset	D	K	M	M_i
AD_5_2 [28]	۲	۵	۲۵۰	۵۰×۵۰
AD_10_2 [29]	۲	۱۰	۵۰۰	۱۰×۵۰
Square1 [30]	۲	٤	۱۰۰۰	٤×۲۵۰
Square4 [30]	۲	٤	۱۰۰۰	٤×۲۵۰
Sizes5 [30]	۲	٤	۱۰۰۰	۷٦۹,۷۷,۷۷,۷۷
Iris [31]	٤	۳	۱۵۰	۵۰, ۵۰,۵۰
BreastCancer [31]	۹	۲	٦۸۳	٤٤٤, ۲۳۹
Newthyroid [31]	۵	۳	۲۱۵	۱۵۰, ۳۵, ۳۰
Wine [31]	۱۳	۳	۱۷۸	۵۹, ۷۱, ٤۸
LiverDisorders [31]	٦	۲	۳٤۱	۱٤۲, ۱۹۹

بررسی و تحلیل نتایج شامل مقایسه نتایج روش پیشنهادی با الگوریتم‌های بهینه‌سازی تک‌هدفه، الگوریتم‌های بهینه‌سازی چندهدفه و همچنین بررسی تأثیر تنظیم پارامترها بر کارایی الگوریتم پیشنهادی می‌باشد. شاخص ارزیابی نیز تعیین تعداد بهینه خوشه‌ها (**K**) و مقدار کمینه بدست آمده از معیار مینکوفسکی (MS) در نظر گرفته شده است. پارامترهای الگوریتم بهینه‌سازی گرگ خاکستری چندهدفه به شرح زیر انتخاب شده اند:

$$N_P = 40, MaxGen = 100$$

که N_P تعداد اعضای جمعیت و $MaxGen$ حداکثر تعداد تکرار یا همان شرط خاتمه الگوریتم است. ظرفیت لیست آرشیو نیز برابر تعداد اعضای جمعیت انتخاب شده است. با توجه به اینکه تاثیر پارامتر a بر کارایی الگوریتم، همانند تاثیر پارامتر دما در الگوریتم بهینه‌سازی تبرید فلزات است، مقادیر a_{min} و a_{max} نیز با الهام از روابط زیر که جزو پرکاربردترین روشهای تعیین دما در الگوریتم تبرید فلزات است، محاسبه خواهند شد [۳۲].

$$a_{max} = -\frac{|\max(J_m) + \max(OS)|}{lnp_0} \qquad (۱٤)$$

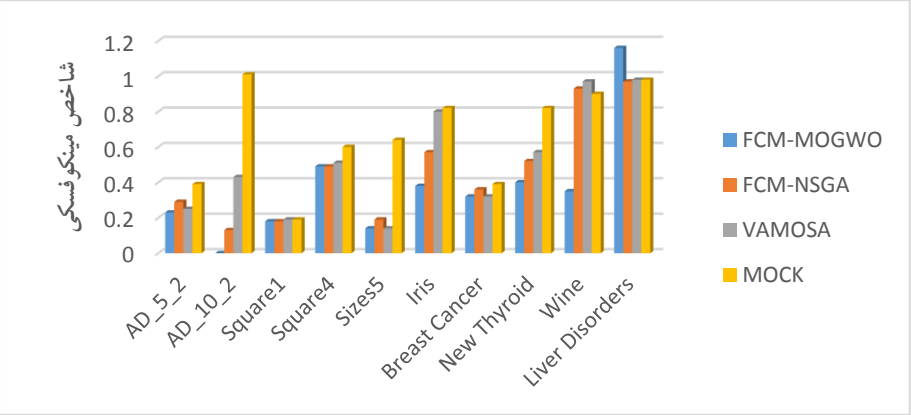
$$a_{min} = -\frac{|\min(J_m) + \min(OS)|}{lnp_0}$$

در رابطه فوق، J_m اولین تابع هدف و OS نیز دومین تابع هدف مساله است. p_0 نیز عددی در محدوده ۰/۷ تا ۰/۹ است که در این مقاله مقدار آنرا ۰/۹ در نظر گرفته‌ایم. همانطور که قبلا نیز اشاره شد مقادیر a_{min} و a_{max} وابسته به ماهیت مساله بوده و برای مجموعه داده‌های مختلف مقادیر آنها متفاوت است.

جدول ۳. نتایج خوشه‌بندی با استفاده از روش پیشنهادی و مقایسه آن با روشهای دیگر

Datasets	Actual K	FCM-MOGWO		FCM-NSGA[19]		VAMOSA[33]		MOCK[34]	
		K	MS	K	MS	K	MS	K	MS
AD_5_2	٥	**٥**	**٠.٢٣**	٥	٠.٢٩	٥	٠.٢٥	٦	٠.٣٩
AD_10_2	١٠	**١٠**	**٠.٠٠**	١٠	٠.١٣	١٠	٠.٤٣	٦	١.٠١
Square1	٤	٤	٠.١٨	٤	٠.١٨	٤	٠.١٩	٤	٠.١٩
Square4	٤	٤	**٠.٤٩**	٤	**٠.٤٩**	٤	٠.٥١	٤	٠.٦٠
Sizes5	٤	٤	٠.١٤	٤	٠.١٩	٤	٠.١٤	٢	٠.٦٤
Iris	٣	**٣**	٠.٣٨	٣	٠.٥٧	٢	٠.٨٠	٢	٠.٨٢
Breast Cancer	٢	**٢**	٠.٣٢	٢	٠.٣٦	**٢**	**٠.٣٢**	٢	٠.٣٩
New Thyroid	٣	**٣**	٠.٤٠	٣	٠.٥٢	٥	٠.٥٧	٢	٠.٨٢
Wine	٣	**٣**	**٠.٣٥**	٣	٠.٩٣	٣	٠.٩٧	٣	٠.٩٠
Liver Disorders	٢	٢	١.١٦	**٢**	**٠.٩٧**	٢	٠.٩٨	٣	٠.٩٨

شکل ٦. نمودار میله‌ای مقایسه نتایج روش پیشنهادی با الگوریتم‌های بهینه‌سازی چندهدفه



در جدول (۲)، نتایج پیاده‌سازی الگوریتم پیشنهادی بر روی ۱۰ مجموعه داده استاندارد آورده شده و با سه الگوریتم بهینه‌سازی چندهدفه FCM-NSGA ارائه شده در مرجع [۱۹]، VAMOSA ارائه‌شده در مرجع [۳۳] و MOCK ارائه‌شده در مرجع [٣٤] مقایسه شده است. در این جدول، Actual K بیان‌گر تعداد واقعی خوشه‌ها برای هر مجموعه داده، **K** بیان‌گر تعداد بهینه خوشه به‌دست‌آمده از هر الگوریتم و MS نیز مقدار معیار مینکوفسکی است که کمترین مقدار آن نشان‌دهنده بهترین عملکرد می‌باشد. نمودار میله‌ای حاصل از این آزمایش در شکل (٦) آورده شده است. با توجه به نتایج ارائه شده در جدول (۲)، موارد زیر قابل استخراج است:

- در مجموعه داده AD_5_2، روش پیشنهادی با مقدار مینکوفسکی ۰/۲۳ بهترین عملکرد و روش MOCK، با مقدار مینکوفسکی ۰/۳۹ بدترین عملکرد را به خود اختصاص داده‌اند. از نظر تخمین تعداد مناسب

خوشه برای این مجموعه داده نیز، همه روش‌ها بجز روش MOCK، قادر به تخمین تعداد صحیح خوشه‌ها بوده‌اند (۵ خوشه).

❖ در مجموعه داده AD_10_2، روش پیشنهادی به‌همراه روش با مقدار مینکوفسکی صفر که در واقع بیان‌گر دقت ۱۰۰ درصدی می‌باشد، مقتدرانه عنوان بهترین عملکرد را به خود اختصاص داده است و روش MOCK، با مقدار مینکوفسکی ۱/۰۱ بدترین عملکرد را داشته است. برای این مجموعه داده نیز، همه روش‌ها بجز روش MOCK، تعداد صحیح خوشه‌ها را به درستی تخمین زده‌اند (۱۰ خوشه).

❖ در مجموعه داده Square1، روش پیشنهادی به‌همراه روش FCM-NSGA با مقدار مینکوفسکی مشترک ۰/۱۸ بهترین عملکرد و روش‌های MOCK و VAMOSA، با مقدار مینکوفسکی مشترک ۰/۱۹ بدترین عملکرد را دارا بوده‌اند. هر چهار روش مورد بررسی، تعداد خوشه‌های این مجموعه داده را به درستی تخمین زده‌اند (٤ خوشه).

❖ در مجموعه داده Square4 که دارای فشردگی و هم‌پوشانی بیشتری نسبت به مجموعه داده Square1 می‌باشد، روش پیشنهادی به‌همراه روش FCM-NSGA با مقدار مینکوفسکی مشترک ۰/٤۹ بهترین عملکرد و روش MOCK، با مقدار مینکوفسکی ۰/٦۰ بدترین عملکرد را دارا بوده‌اند. در این مورد نیز هر چهار روش، تعداد خوشه‌های مجموعه داده را به درستی تخمین زده‌اند (٤ خوشه).

❖ در مجموعه داده Sizes5، روش پیشنهادی به‌همراه روش VAMOSA با مقدار مینکوفسکی مشترک ۰/۱٤ بهترین عملکرد را داشته‌اند. روش MOCK، با مقدار مینکوفسکی ۰/٦٤ بدترین عملکرد را دارا بوده است. برای این مجموعه داده نیز، همه روش‌ها بجز روش MOCK، تعداد صحیح خوشه‌ها را به درستی تخمین زده‌اند (٤ خوشه).

❖ برای مجموعه داده Iris، روش پیشنهادی با مقدار مینکوفسکی ۰/۳۸ با اختلاف زیاد نسبت به سایر الگوریتم‌های مورد مقایسه، بهترین عملکرد را به خود اختصاص داده است. روش FCM-NSGA، با مقدار مینکوفسکی ۰/۵۷ در رتبه دوم جای دارد و روش MOCK با مقدار مینکوفسکی ۰/۸۲ بدترین عملکرد را برای این مجموعه داده داشته است. از نظر تخمین تعداد مناسب خوشه برای این مجموعه داده نیز، به‌دلیل هم‌پوشانی بالا بین خوشه‌ها، تنها روش پیشنهادی و روش FCM-NSGA تعداد صحیح خوشه‌ها را به درستی تخمین زده‌اند (۳ خوشه).

❖ برای مجموعه داده Breast Cancer، روش پیشنهادی به‌همراه روش VAMOSA با مقدار مینکوفسکی مشترک ۰/۳۲ بهترین عملکرد را داشته‌اند. روش MOCK، با مقدار مینکوفسکی ۰/۳۹ بدترین عملکرد را دارا بوده است. برای این مجموعه داده، همه روش‌ها تعداد صحیح خوشه‌ها را به درستی تخمین زده‌اند (۲ خوشه).

❖ برای مجموعه داده New Thyroid، روش پیشنهادی با مقدار مینکوفسکی ۰/٤۰ بهترین عملکرد را نسبت به سایر روش‌ها دارا بوده است. روش MOCK نیز، با مقدار مینکوفسکی ۰/۸۲ بدترین عملکرد را برای این مجموعه داده داشته است. از نظر تخمین تعداد مناسب خوشه برای این مجموعه داده، تنها روش پیشنهادی و روش FCM-NSGA تعداد صحیح خوشه‌ها را به درستی تخمین زده‌اند (۳ خوشه).

❖ برای مجموعه داده Wine، روش پیشنهادی با مقدار مینکوفسکی ۰/۳۵ با اختلاف بسیار زیاد نسبت به سایر الگوریتم‌های مورد مقایسه، بهترین عملکرد را به خود اختصاص داده است. روش MOCK، با مقدار مینکوفسکی ۰/۹۰ در رتبه دوم جای دارد و روش VAMOSA با مقدار مینکوفسکی ۰/۹۷ بدترین عملکرد را برای این مجموعه داده داشته است. در این مورد نیز هر چهار روش، تعداد خوشه‌های مجموعه داده را به درستی تخمین زده‌اند (۳ خوشه).

❖ در مورد مجموعه داده Liver Disorders، همه روش‌ها عملکرد نامناسبی از خود نشان داده‌اند؛ به‌طوری‌که مقدار مینکوفسکی برای تمام روش‌ها عددی در حدود یک می‌باشد که نشان می‌دهد هر چهار روش در خوشه‌بندی این مجموعه داده با مشکل مواجه شده‌اند. در این میان، روش پیشنهادی با مقدار مینکوفسکی ۱/۱۶ بدترین عملکرد را برای این مجموعه داده داشته است و روش FCM-NSGA با مقدار مینکوفسکی ۰/۹۷ بهترین عملکرد را دارا بوده است. از نظر تعیین تعداد بهینه خوشه نیز فقط روش MOCK نتوانسته تعداد صحیح خوشه‌ها را به درستی تخمین بزند(۲ خوشه).

با توجه به نتایج اشاره شده، مشاهده می‌شود الگوریتم پیشنهادی توانسته است در تمامی آزمایشها تعداد بهینه خوشه‌ها را مشخص نماید. همچنین این الگوریتم در ۹۰درصد موارد، عملکرد بهتری نسبت به سایر روشها از خود نشان داده و قادر به خوشه‌بندی و تشخیص تعداد بهینه خوشه‌ها بوده است.

الگوریتم FCM-NSGA نسبت به دو الگوریتم MOCK و VAMOSA عملکرد بسیار بهتری داشته است. این الگوریتم توانسته است در همه مجموعه داده‌های مورد بررسی، تخمین درستی از تعداد خوشه‌ها داشته باشد و از این حیث موفقیت کاملی را به دست آورده است. همچنین از نظر کیفیت خوشه‌بندی نیز، مقدار مینکوفسکی قابل قبولی برای هر مجموعه داده به دست آورده است؛ با این حال در مجموع، الگوریتم FCM-MOGWO که از همان دو تابع هدف موجود در الگوریتم FCM-NSGA بهره می‌گیرد، در مجموع عملکرد بمراتب بهتری از نظر مقدار مینکوفسکی داشته است.

الگوریتم VAMOSA که از دو معیار Sym-index و XB-index به ترتیب به‌منظور فشردگی و پراکندگی استفاده می‌کند، توانسته است ساختار هم‌پوشان مجموعه داده‌های دوبعدی مصنوعی از جمله AD_5_2 و AD_10_2 را تشخیص دهد. با این حال، این الگوریتم برای داده‌هایی با هم‌پوشانی بسیار بالا (مجموعه داده‌های Iris و New Thyroid) عملکرد مناسبی از خود نشان نداده است. این موضوع می‌تواند بیانگر انتخاب نامناسب دو معیار ارزیابی مورد استفاده در این الگوریتم در مواجهه با مجموعه داده‌هایی با هم‌پوشانی بالا باشد. نتایج بررسی‌ها نشان می‌دهد که الگوریتم MOCK، نتایج ضعیفی هم از لحاظ مقدار مینکوفسکی و هم از لحاظ تخمین تعداد خوشه‌ها دارا بوده است. این موضوع می‌تواند به‌خاطر محدودیت دو تابع هدف استفاده شده در این الگوریتم باشد که باعث عدم کارایی آن در مواجهه با داده‌های هم‌پوشان شده است. از آنجایی‌که نمایش خروجی روش پیشنهادی برای مجموعه داده‌هایی با ابعاد بالاتر از سه بعد ممکن نیست، در این بخش خروجی الگوریتم پیشنهادی برای مجموعه داده‌های دو بعدی مصنوعی در شکل‌های (۷) الی (۱۱) نشان داده شده است. در این شکل‌ها، داده آموزشی با رنگ آبی، داده آزمایشی با رنگ قرمز و مراکز به‌دست آمده به‌وسیله

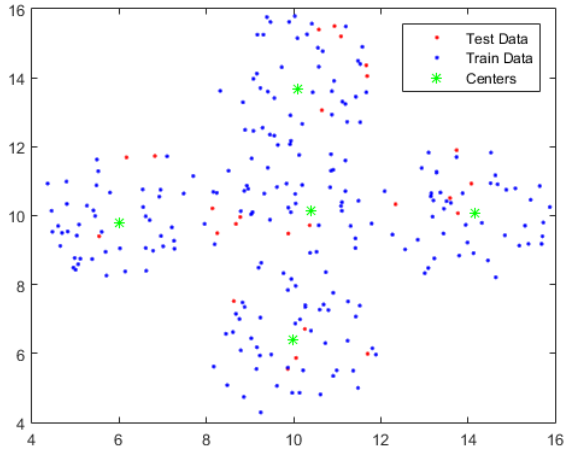
ستاره (*) سبز رنگ مشخص شده است. همانطور که مشاهده می‌شود روش پیشنهادی علاوه بر تعیین تعداد بهینه خوشه‌ها، عمل خوشه‌بندی را برای مجموعه داده‌های غیرهم‌پوشان (شکل ۷ و۸)، مجموعه داده‌های با خوشه‌های نابرابر (شکل ۹) و مجموعه داده‌های هم‌پوشان (شکل‌های ۱۰ و ۱۱) با دقت بسیار خوبی انجام دهد.

در ادامه تاثیر هر یک توابع هدف بطور مجزا (روابط (۱) و (۵))، بر روی ۱۰ مجموعه داده بررسی شده است. هدف از انجام این آزمایش آن است که نشان دهیم در مسائل خوشه‌بندی خودکار فازی، بهینه‌سازی همزمان دو تابع هدف می‌تواند دقت خوشه‌بندی را نسبت به حالتی که اهداف بصورت مجزا بهینه‌سازی می‌شوند افزایش دهد. در این‌جا، رویکردی که از تابع هدف OS (رابطه ۵) استفاده می‌کند FCM-GWO-OS و رویکردی که از تابع هدف J_m (رابطه ۱) استفاده می‌کند FCM-GWO-J_m نامیده شده است. همچنین نتایج با مرجع [۳۵] که در آن از الگوریتم تک‌هدفه VGAPS استفاده شده، مقایسه شده است.

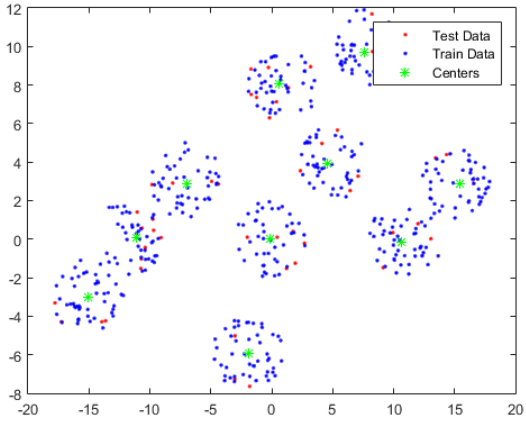
جدول ۳. نتایج روش پیشنهادی در مقایسه با رویکردهای تک‌هدفه

Datasets	Actual K	FCM-MOGWO		FCM-GWO-OS		FCM-GWO-J_m		VGAPS[35]	
		K	MS	K	MS	K	MS	K	MS
AD_5_2	٥	**٥**	**٠.٢٣**	**٥**	**٠.٢٣**	١١	١.٥٩	٥	٠.٢٥
AD_10_2	١٠	**١٠**	**٠.٠٠**	١٠	٠.١٣	١٧	١.٥٣	٧	٠.٨٤
Square1	٤	٤	**٠.١٨**	٤	**٠.١٨**	٢١	١.٧٧	٤	٠.٢٠
Square4	٤	٤	٠.٤٩	٤	**٠.٤٨**	٣٠	١.٨٣	٥	٠.٥٢
Sizes5	٤	٤	**٠.١٤**	٢	٠.٧٠	٢٢	٠.٦٢	٥	٠.٢٢
Iris	٣	**٣**	**٠.٣٨**	٢	١.١٩	١٢	١.٥١	٣	٠.٦٢
Breast Cancer	٢	**٢**	**٠.٣٢**	٢	٠.٣٦	٢٤	١.٣٣	٢	٠.٣٧
New Thyroid	٣	**٣**	**٠.٤٠**	٢	٠.٩٩	١٥	٠.٨٧	٣	٠.٥٨
Wine	٣	**٣**	**٠.٣٥**	٣	٠.٥٧	١٣	١.٣٩	٦	٠.٩٧
Liver Disorders	٢	٢	١.١٦	٢	١.٢١	١٨	١.٤١	**٢**	**٠.٩٨**

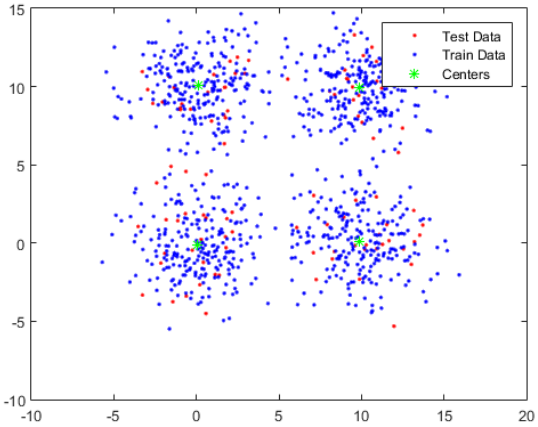
شکل ۷. مجموعه داده AD_5_2. با شاخص مینکوفسکی ۰,۲۳



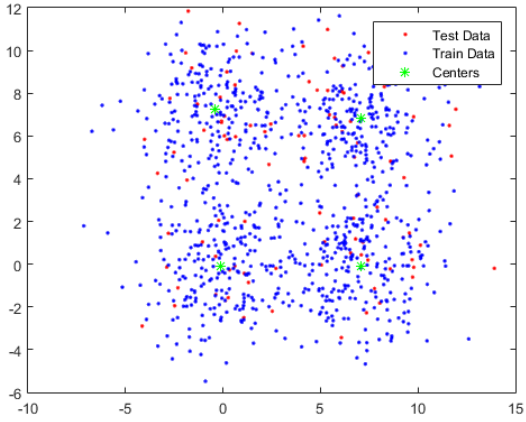
شکل ۸. مجموعه داده AD_10_2. با شاخص مینکوفسکی صفر



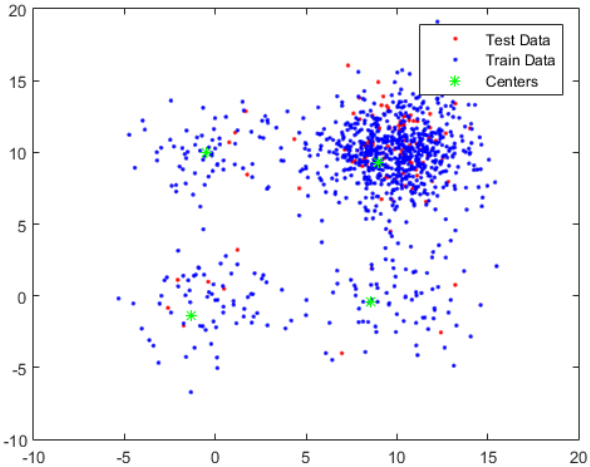
شکل ۹. مجموعه داده Square1. با شاخص مینکوفسکی ۰,۱۸



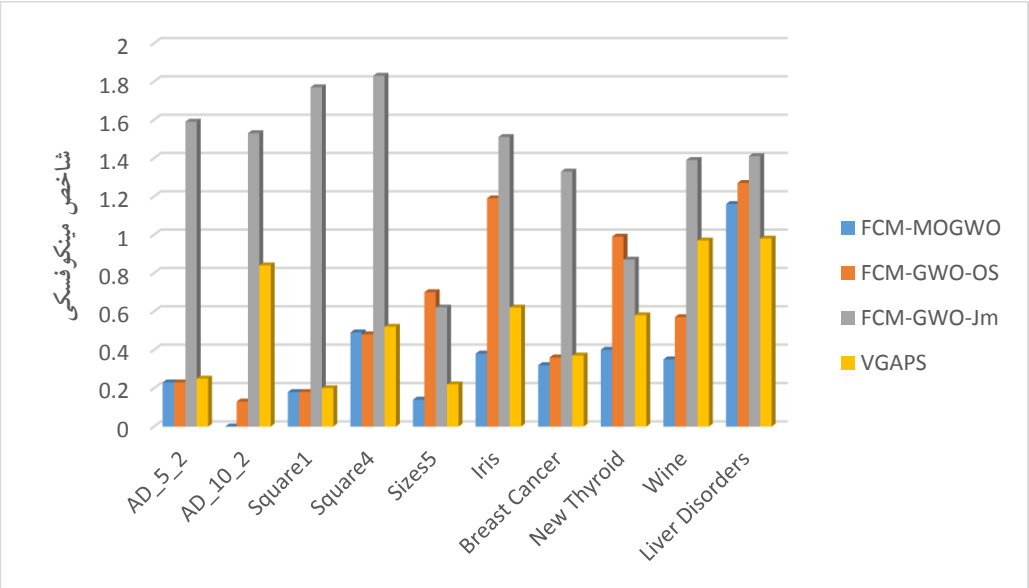
شکل ۱۰. مجموعه داده Square4. با شاخص مینکوفسکی ۰,۴۹



شکل ۱۱. مجموعه داده Sizes5. با شاخص مینکوفسکی ۰,۱۴



شکل ۱۲. نمودار میله‌ای مقایسه نتایج رویکردهای تک‌هدفه



جدول (۳) نتایج حاصل از این آزمایش را نشان می‌دهد. نمودار میله‌ای مقایسه نتایج نیز در شکل (۱۲) آورده شده است.

با بررسی نتایج ارائه شده در جدول (۳) مشاهده می‌شود الگوریتم تک‌هدفه FCM-GWO-OS، از نظر تخمین تعداد بهینه خوشه‌ها و همچنین مقدار کمینه معیار مینکوفسکی، نسبت به الگوریتم تک‌هدفه FCM-GWO-Jm و VGAPS عملکرد بهتری دارد. این موضوع نشان می‌دهد معیار OS می‌تواند شاخص ارزیابی مناسبی برای خوشه‌بندی داده‌های با هم‌پوشانی بالا باشد.

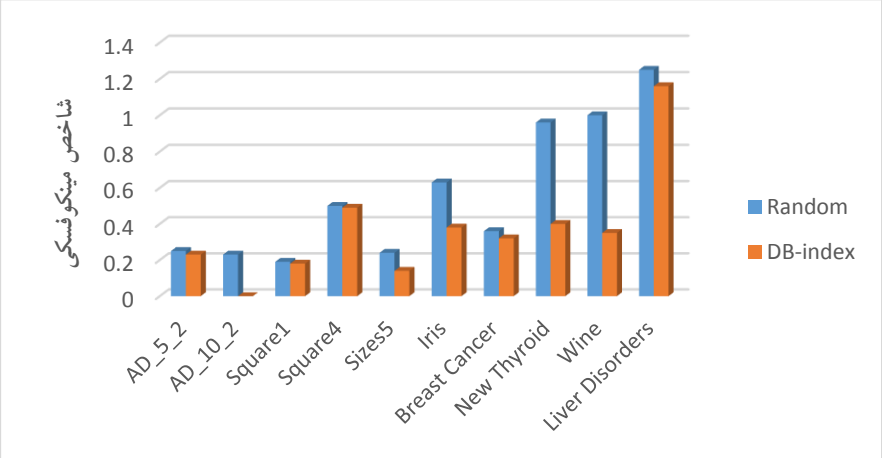
رویکرد FCM-GWO-OS تخمین‌های قابل قبولی از تعداد خوشه‌های بسیاری از مجموعه داده‌های مورد بررسی داشته است و در مورد تخمین تعداد خوشه‌ها برای مجموعه داده‌های Sizes5 و Iris و New Thyroid ناموفق عمل کرده است که نابرابری تعداد اعضای خوشه‌ها و هم‌پوشانی بالا در این مجموعه داده‌ها می‌تواند از علل عدم موفقیت این معیارها باشد. این در حالیست که رویکرد FCM-GWO-Jm تخمین بسیار بالایی برای تمام مجموعه داده‌ها داشته است و در هیچ مورد موفقیتی کسب نکرده است. از این رو می‌توان به ضعف اساسی این معیار در مواجهه با داده‌های هم‌پوشان پی برد. چرا که در این معیار با افزایش تعداد خوشه‌ها، این معیار به سمت کاهش مقدار خود متمایل می‌گردد و بنابراین در تخمین خودکار تعداد خوشه‌ها با شکست مواجه می‌گردد. همچنین نتایج رویکرد VGAPS چندان رضایت‌بخش نبوده و مشاهده می‌گردد که در چهار مجموعه داده تخمین درستی از تعداد خوشه‌ها، ارائه نداده است. این رویکرد برای مجموعه داده‌هایی با خوشه‌های فشرده و جدا از هم کارایی قابل قبولی دارد، اما در مواجهه با خوشه‌های هم‌پوشان عملکرد مناسبی ندارد. این موضوع نشان می‌دهد که معیار Sym-index

استفاده شده در این الگوریتم، به‌تنهایی قادر به شناسایی ساختار داده مختلف نمی‌باشد.

در نهایت، از مقایسه نتایج خروجی و بررسی‌های صورت گرفته، مشاهده می‌گردد که با استفاده از هر دو معیار Jm و OS در بهینه‌سازی چندهدفه روش پیشنهادی، اندازه‌گیری مقدار برازندگی اعضای جمعیت متعادل گردیده است. این موضوع باعث شده است که کارایی الگوریتم FCM-MOGWO، هم از نظر میزان معیار مینکوفسکی و هم از نظر تخمین تعداد مناسب خوشه‌ها، بهبود چشم‌گیری داشته باشد.

جدول ٤. نتایج تأثیر روش انتخاب گرگ رهبر در روش پیشنهادی

Datasets	Actual K	Random		DB-index	
		K	MS	K	MS
AD_5_2	٥	٥	٠.٢٥	٥	**٠.٢٣**
AD_10_2	١٠	١٠	٠.٢٣	١٠	**٠.١٣**
Square1	٤	٤	٠.١٩	٤	**٠.١٨**
Square4	٤	٤	٠.٥٠	٤	**٠.٤٩**
Sizes5	٤	٤	٠.٢٤	٤	**٠.١٤**
Iris	٣	٣	٠.٦٣	٣	**٠.٣٨**
Breast Cancer	٢	٢	٠.٣٦	٢	**٠.٣٢**
New Thyroid	٣	٣	٠.٩٦	٣	**٠.٤٠**
Wine	٣	٣	١.٠٠	٣	**٠.٥٠**
Liver Disorders	٢	٥	١.٢٥	٥	**١.١٦**



شکل ۱۳. نمودار میله‌ای نتایج تأثیر روش انتخاب گرگ رهبر در روش پیشنهادی

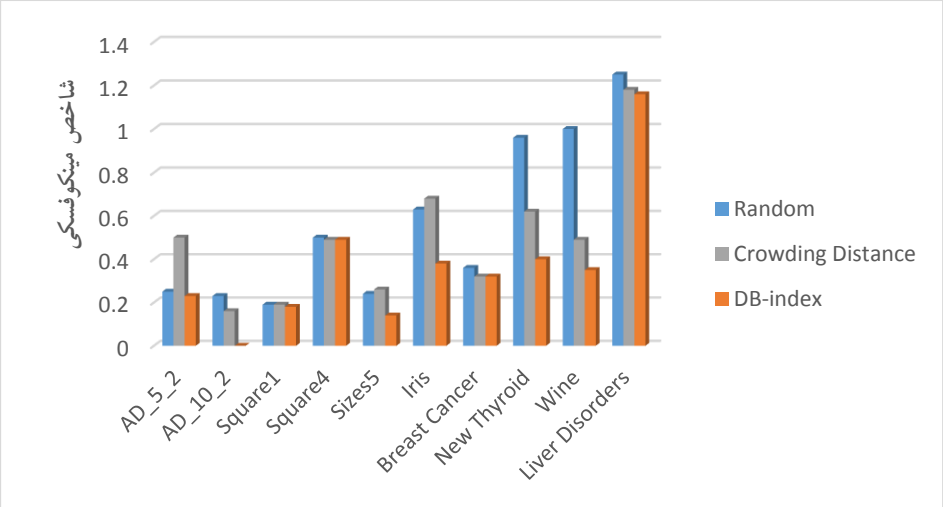
در پایان به بررسی تأثیر تنظیمات مختلف پارامترها بر روی کارایی روش پیشنهادی پرداخته می‌شود. پارامترهای قابل تنظیم جهت آزمایش‌های این بخش به دو دسته شامل روش انتخاب گرگ رهبر و روش حذف

از آرشیو می‌باشد. انتخاب گرگ رهبر از خلوت‌ترین بخش انتخاب شده، شامل دو روش انتخاب تصادفی و انتخاب بر مبنای شاخص DB می‌باشد. مقایسه نتایج این دو روش در جدول (٤) آورده شده است. نمودار میله‌ای مقایسه نتایج حاصل از این آزمایش نیز در شکل (١٣) آمده است. همانطور که مشاهده می‌شود برای تمامی مجموعه داده‌ها انتخاب گرگ رهبر با استفاده از معیار DB منجر به بهبود دقت خوشه‌بندی شده است. همانطور که قبلا اشاره شد، در الگوریتم‌های بهینه‌سازی چندهدفه، ظرفیت مشخصی برای لیست آرشیو در نظر گرفته می‌شود. در صورتیکه تعداد اعضای موجود در لیست آرشیو از ظرفیت آن بیشتر شود، می‌بایست اعضای اضافی از لیست حذف شوند.

جدول ٥. نتایج تأثیر روش حذف از آرشیو در روش پیشنهادی

Datasets	Actual K	Random		Crowding Distance		DB-index	
		K	MS	K	MS	K	MS
AD_5_2	٥	٥	٠.٤٣	٥	٠.٥٠	**٥**	**٠.٢٣**
AD_10_2	١٠	١٠	٠.٢٩	١٠	٠.١٦	**١٠**	**٠.٠٠**
Square1	٤	٤	٠.٢١	٤	٠.١٩	**٤**	**٠.١٨**
Square4	٤	٤	٠.٥٠	٤	٠.٤٩	**٤**	**٠.٤٩**
Sizes5	٤	٤	٠.٢٥	٤	٠.٢٦	**٤**	**٠.١٤**
Iris	٣	٣	٠.٦٤	٣	٠.٦٨	**٣**	**٠.٣٨**
Breast Cancer	٢	٢	٠.٣٢	٢	٠.٣٢	**٢**	**٠.٣٢**
New Thyroid	٣	٣	٠.٦١	٣	٠.٦٢	**٣**	**٠.٤٠**
Wine	٣	٣	٠.٤٢	٣	٠.٤٩	**٣**	**٠.٣٥**
Liver Disorders	٢	٥	١.١٩	٥	١.١٨	**٢**	**١.١٦**

شکل ١٤. نمودار میله‌ای نتایج تأثیر روش حذف از آرشیو در روش پیشنهادی



روشهای مختلفی برای حذف اعضای اضافی از لیست آرشیو وجود دارند که در جدول (۵) نتایج حاصل از سه روش حذف بصورت تصادفی، حذف با استفاده از روش فاصله ازدحامی و حذف از طریق شاخص DB نشان داده شده است. نمودار میله‌ای مقایسه نتایج حاصل از این آزمایش در شکل (١٤) آورده شده است. شایان ذکر است روشهای حذف بصورت تصادفی و فاصله ازدحامی در مورد تمامی مسائل بهینه‌سازی کاربرد داشته و وابسته به ماهیت مساله نیستند، اما روش حذف بر اساس معیار DB منحصر به مسائل خوشه‌بندی است. با توجه به نتایج ارائه شده در جدول (۵) مشاهده می‌شود برای تمامی مجموعه داده‌ها، مقدار مینکوفسکی برای حالتی که از شاخص DB استفاده شده نسبت به دو حالت دیگر کمتر و در نتیجه دقت خوشه‌بندی نیز بهتر است.

٦. نتیجه‌گیری

در این مقاله، مساله خوشه‌بندی خودکار فازی در قالب یک مساله بهینه‌سازی چندهدفه فرمول‌بندی گردید. توابع هدف شامل معیارهای فشردگی (J_m) و همپوشانی- جدایی (OS) بودند که بترتیب برای تعیین فشردگی داده‌های درون خوشه و تشخیص ویژگی‌های هم‌پوشانی خوشه‌ها به‌کار گرفته شدند. با توجه به اینکه مساله خوشه‌بندی از نوع مسائل بهینه‌سازی غیر خطی، چندهدفه و نامحدب می‌باشد، برای حل آن نیز یک روش بهینه‌سازی چندهدفه جدید مبتنی بر الگوریتم گرگ خاکستری پیشنهاد گردید. به منظور تسریع در فرآیند بهینه‌سازی و ایجاد مصالحه بین توابع هدف، راهکارهای ابتکاری جدیدی نیز به الگوریتم اضافه گردید. اعمال روش بهینه‌سازی پیشنهادی منجر به تولید چندین راه‌حل بهینه گردید که در نهایت با توجه به معیار مینکوفسکی، راه حل نهایی انتخاب شد. روش پیشنهادی بر روی ۱۰ مجموعه داده استاندارد مورد بررسی و آزمایش قرار گرفت و با تعدادی از روش‌های دیگر مقایسه گردید. نتایج آزمایشها نشان دادند روش پیشنهادی قادر است علاوه بر تعیین تعداد بهینه خوشه‌ها، مساله خوشه‌بندی مجموعه داده‌های پراکنده، فشرده و هم‌پوشان را نحو مطلوبی مدیریت کند. همچنین کارایی هر یک از توابع هدف نیز بطور مجزا مورد ارزیابی قرار داده شد. نتایج آزمایشها بیانگر این واقعیت بودند که هر یک از این توابع هدف به تنهایی قادر به حل مساله خوشه‌بندی خودکار نبوده و استفاده از توابع هدف مناسب می‌تواند دقت خوشه‌بندی را بطور چشم‌گیری افزایش دهد.

۷. پیشنهادات کارهای آینده

هرچند نتایج آزمایشات انجام شده بیانگر این واقعیت بود که الگوریتم بهینه‌سازی پیشنهادی در این مقاله (الگوریتم FCM-MOGWO)، دارای کارایی و عملکرد قابل قبولی است، با این حال می‌توان با اعمال تغییراتی در نسخه اصلی الگوریتم، کارایی آن‌را تا حد مطلوبی افزایش داد. استفاده از ترکیب سایر الگوریتم‌های تکاملی و فرا ابتکاری می‌تواند تا حد زیادی نقاط ضعف و مشکلات این الگوریتم‌ها را برطرف نموده و عملکرد آن‌ها را بهبود بخشد. از این رو پیشنهاد می‌گردد روش پیشنهادی به‌جای استفاده از الگوریتم بهینه‌سازی گرگ خاکستری به‌تنهایی، با سایر الگوریتم‌های مطرح در این زمینه ترکیب و جهت خوشه‌بندی فازی خودکار به‌کارگرفته شود.

در روش پیشنهادی، از الگوریتم خوشه‌بندی فازی FCM جهت خوشه‌بندی نرم داده‌های هم‌پوشان، استفاده

شد. می‌توان با توجه به کاربرد و ماهیت داده‌ها، از سایر روش‌های خوشه‌بندی نرم و یا سخت، مانند روش‌های سلسله‌مراتبی یا روش‌های مبتنی‌بر چگالی استفاده نمود. از دیگر پیشنهاداتی که می‌توان برای کارهای آینده مطرح نمود، استفاده از شاخص‌های دیگر جهت ارزیابی خوشه‌بندی است. در الگوریتم FCM-MOGWO، از دو معیار فشردگی (J_m) و همپوشانی–جدایی (OS) به‌عنوان توابع هدف استفاده شده است. در کارهای آینده می‌توان هر دو معیار یا یکی از آن‌ها را با توجه به نیاز و ماهیت مسئله با معیارهای مناسب جایگزین نمود و کارایی الگوریتم را افزایش داد.

فهرست منابع

[1] Otto C, Wang D, Jain AK. Clustering millions of faces by identity. IEEE Trans. on pattern analysis and machine intelligence, Vol. 49, No. 2, pp. 289-303, 2018.

[2] Heloulou I, Radjef MS, Kechadi MT. Automatic multi-objective clustering based on game theory. Expert Systems with Applications, Vol. 67, pp.32-48, 2017.

[3] Jose-Garcia A, Gómez-Flores W. Automatic clustering using nature-inspired metaheuristics: A survey. Applied Soft Computing, Vol. 41, pp. 192-213, 2013.

[4] Shahsamandi E P, Sadi-nezhad S, Saghaei A. Multi-objective complete fuzzy clustering approach. Intelligent Automation & Soft Computing. Vol. 23, No. 2, pp.285-94, 2017.

[5] Vo-Van T, Nguyen-Thoi T, Vo-Duy T, Ho-Huu V, Nguyen-Trang T. Modified genetic algorithm-based clustering for probability density functions. Journal of Statistical Computation and Simulation. Vol. 87, No. 10, pp. 1964-79, 2017.

[6] Yang Z, Huo H, Fang T. Automatically finding the number of clusters based on simulated annealing. Journal of Shanghai Jiaotong University (Science), Vol. 22, No. 2, pp. 139-47, 2017.

[7] Song W, Ma W, Qiao Y. Particle swarm optimization algorithm with environmental factors for clustering analysis. Soft Computing, Vol. 21, No. 2, pp. 283-93, 2017.

[8] Gupta C, Jain A, Tayal DK, Castillo O. ClusFuDE: Forecasting low dimensional numerical data using an improved method based on automatic clustering, fuzzy relationships and differential evolution, Engineering Applications of Artificial Intelligence. Vol. 31, No. 71, pp.175-89, 2018.

[9] Chaghari A, Feizi-Derakhshi MR. Automatic Clustering Using Improved Imperialist Competitive Algorithm, Journal of Signal and Data Processing, Vol. 4, No. 2, pp. 159-69, 2017.

[10] Martínez-Peñaloza MG, Mezura-Montes E, Cruz-Ramírez N, Acosta-Mesa HG, Ríos-Figueroa HV. Improved multi-objective clustering with automatic determination of the number of clusters, Neural Computing and Applications.Vol. 28, No. 8, pp.2255-75, 2017.

[11] Saha S, Bandyopadhyay S. A generalized automatic clustering algorithm in a multiobjective framework. Applied Soft Computing, Vol. 13, No. 1, pp. 89-108, 2013.

[12] Bandyopadhyay, S., U. Maulik, and A. Mukhopadhyay, Multiobjective genetic clustering for pixel classification in remote sensing imagery. IEEE transactions on Geoscience and Remote Sensing, Vol. 45, No. 5, p. 1506-1511, 2007.

[13] Suresh K, Kundu D, Ghosh S, Das S, Abraham A., Automatic clustering with multi-objective differential evolution algorithms, IEEE Congress on Evolutionary Computation,.pp. 2590-2597, 2009.

[14] Zhong, Y., S. Zhang, and L. Zhang, Automatic fuzzy clustering based on adaptive multi-objective differential evolution for remote sensing imagery. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Vol. 6, No. 5, pp. 2290-2301, 2013.

[15] Kundu D, Suresh K, Ghosh S, Das S, Abraham A, Badr Y., Automatic clustering using a synergy of genetic algorithm and multi-objective differential evolution. Hybrid Artificial Intelligence Systems, p. 177-186, 2009.

[16] Nanda, S.J. and G. Panda, Automatic clustering algorithm based on multi-objective Immunized PSO to classify actions of 3D human models. Engineering Applications of Artificial Intelligence, Vol. 26, No. 5, pp.1429-1441, 2013.

[17] Abubaker, A., A. Baharum, and M. Alrefaei, Automatic clustering using multi-objective particle swarm and simulated annealing. PloS one, Vol. 10, No. 7, pp. e0130995, 2015.

[18] Xu R, Wunsch D. Survey of clustering algorithms. IEEE Transactions on neural networks, Vol. 16, No. 3, pp. 645-78, 2005.

[19] Wikaisuksakul, S., A multi-objective genetic algorithm with fuzzy c-means for automatic data clustering. Applied Soft Computing, Vol. 24: pp. 679-691, 2014.

[20] Le Capitaine, H. and C. Frélicot, A cluster-validity index combining an overlap measure and a separation measure based on fuzzy-aggregation operators. IEEE Transactions on Fuzzy Systems, Vol. 19, No. 3, pp. 580-588, 2011.

[21] Le Capitaine, H. and C. Frélicot, A family of cluster validity indexes based on a l-order fuzzy or operator. Structural, Syntactic, and Statistical Pattern Recognition, pp. 612-621, 2008.

[22] Mirjalili, S., S.M. Mirjalili, and A. Lewis, Grey wolf optimizer. Advances in Engineering Software, Vol. 69: pp. 46-61, 2014.

[23] Rashidi F. Transmission expansion planning in a deregulated power system using multi objective differential evolution algorithm, Iranian Journal of Electrical and Computer Engineering, Vol. 15, No. 4, pp. 247-257, 2018.

[24] Kaveh A, Zakian P. Improved GWO algorithm for optimal design of truss structures. Engineering with

Computers, pp.1-23, 10.1007/doi: s00366-017-0567-1, 2017.
[25] Heloulou I, Radjef MS, Kechadi MT. Automatic multi-objective clustering based on game theory. Expert Systems with Applications, Vol. 67, pp. 32-48, 2017.
[26] Mukhopadhyay, A., U. Maulik, and S. Bandyopadhyay, A survey of multiobjective evolutionary clustering. ACM Computing Surveys (CSUR), Vol. 47, No. 4, pp. 61, 2015.
[27] Wu KL, Yang MS. A cluster validity index for fuzzy clustering. Pattern Recognition Letters, Vol. 26, No. 9, pp. 1275-91, 2005.
[28] Bandyopadhyay, S. and U. Maulik, Genetic clustering for automatic evolution of clusters and application to image classification. Pattern recognition, Vol. 35, No. 6, pp. 1197-1208, 2002.
[29] Bandyopadhyay, S. and S.K. Pal, Classification and learning using genetic algorithms: applications in bioinformatics and web intelligence, Springer Science & Business Media, 2007.
[30] Handl, J. and J. Knowles, An evolutionary approach to multiobjective clustering. IEEE transactions on Evolutionary Computation, Vol. 11, No. 1, pp. 56-76, 2007.
[31] Frank, A. and A. Asuncion, UCI Machine Learning Repository. Irvine, CA: University of California. School of information and computer science, 2010. 213.
[32] Yang XS. Engineering optimization: An Introduction with Metaheuristic Applications, John Wiley & Sons, Inc.; 2010.
[33] Saha, S. and S. Bandyopadhyay, A symmetry based multiobjective clustering technique for automatic evolution of clusters. Pattern recognition,Vol. 43, No. 3,pp. 738-751, 2010.
[34] Handl, J. and J. Knowles. Evolutionary multiobjective clustering. in PPSN. Springer, 2004
[35] Bandyopadhyay, S. and S. Saha, A point symmetry-based clustering technique for automatic evolution of clusters. IEEE Transactions on Knowledge and Data Engineering, Vol. 20, No. 11, pp. 1441-1457, 2008.