

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

علی اصغر امام‌دوست<sup>۱\*</sup>، فرزانه رشیدی<sup>۲</sup>، عبدالله خلیلی<sup>۳</sup>

تاریخ دریافت: ۱۳۹۷/۰۳/۰۹

تاریخ پذیرش: ۱۳۹۷/۰۴/۱۵

### 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 بر اساس 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 مسأله 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 مسأله 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](mailto:a.imamdoost@yahoo.com)

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

۳. استادیار- دانشکده فنی و مهندسی، دانشگاه هرمزگان، [khalili@hormozgan.ac.ir](mailto:khalili@hormozgan.ac.ir)

## ۱. مقدمه

خوشه‌بندی عبارت است از گروه‌بندی عناصر با ویژگی‌های مشابه، به گونه‌ای 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 غالباً بیانگر مرکز خوشه نیز می‌باشد. میزان شباهت داده‌ها به مرکز خوشه، عموماً توسط پارامتری بنام معیار شباهت تعیین می‌گردد. معیار شباهت غالباً 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 اصطلاحاً 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 [۴]. از این‌رو، حل مسأله خوشه‌بندی در حالت کلی و مسأله خوشه‌بندی خودکار به صورت خاص، بعضاً 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 مسأله خوشه‌بندی خودکار، می‌توان به الگوریتم ژنتیک [۵]، الگوریتم تبرید فلزات [۶]، 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 معمولاً 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

می‌دهند الگوریتم مذکور توانسته است بطور میانگین حدود ۱۶ درصد 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، ۶ مجموعه داده مصنوعی و ۴ 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 ۱۳ درصد نسبت به الگوریتم MOCK و ۹ درصد نسبت به الگوریتم NSGA-II دقت خوشه‌بندی را افزایش دهد. در مرجع [۱۶] 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 MOPSO) بطور همزمان سه معیار ارزیابی خوشه با اسامی شاخص DB، شاخص Sym و شاخص Conn را مورد استفاده قرار

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 ۶ مجموعه داده واقعی، نشان می‌دهند الگوریتم مذکور توانسته است بطور میانگین در ۷۰ درصد موارد در تعیین تعداد صحیح خوشه‌ها موفق عمل نماید.

در [۱۲] یک روش خوشه‌بندی مبتنی بر 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 ۶ مجموعه داده مصنوعی و ۴ 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 [۱۴] 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

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 ۱۰ 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 ۲ بیان ریاضی مساله خوشه‌بندی فازی و توابع هدف مورد استفاده بصورت مختصر ارائه خواهد. در بخش ۳، الگوریتم بهینه‌سازی گرگ خاکستری تک هدفه توضیح داده می‌شود. در بخش ۴، ساختار روش پیشنهادی معرفی شده و بخش‌های مختلف آن با جزئیات ارائه خواهد شد. در بخش ۵، روش پیشنهادی بر روی ۱۰ مجموعه داده استاندارد پیاده سازی شده و نتایج مورد تجزیه و تحلیل قرار خواهند گرفت. بخش ۶ و ۷ نیز به ترتیب به نتیجه‌گیری و پیشنهاداتی جهت کارهای آینده اختصاص دارد.

## ۲. الگوریتم خوشه‌بندی FCM

الگوریتم خوشه‌بندی فازی C-Means که با نام FCM نیز شناخته می‌شود، در واقع معادل فازی الگوریتم خوشه‌بندی غیرفازی K-means است. همانطور که قبلاً 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 [۱۲]. این روش، جزء روش‌های خوشه‌بندی نرم محسوب می‌گردد؛ بدین معنی که در این روش هر داده می‌تواند با یک درجه عضویت، به چندین

داده است. الگوریتم مذکور بر روی ۱۴ 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، متأثر از دو فاکتور زیر است:

- ۱- الگوریتم بهینه‌سازی مورد استفاده.
  - ۲- 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:
۱. الگوهای موجود در یک خوشه باید تا حد امکان به یکدیگر شبیه باشند.

۲. 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

[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 فشردتر 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، منحصرًا 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

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

$$J_m = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m d_{ik}^2 \quad (1)$$

$$= \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m \|x_k - v_i\|^2$$

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  $k$ ام و  $v_i$  بیانگر نماینده یا مرکز خوشه  $i$ ام می‌باشد.  $u_{ik}$  میزان تعلق داده  $i$ ام در خوشه  $k$ ام 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  $i$ امین خوشه،  $u_{ik}$ ، طبق رابطه زیر محاسبه می‌شود [۱۸]:

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

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

$$v_i = \frac{\sum_{k=1}^M (u_{ik})^m x_k}{\sum_{k=1}^M (u_{ik})^m}, \quad 1 \leq i \leq C \quad (3)$$

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، غالباً معیار فاصله اقلیدسی کارایی بهتری نسبت به معیار فاصله مینکوفسکی دارد

دو بخش است: بخش اول، معیار هم‌پوشانی است که با استفاده از درجات فازی، هم‌پوشانی بین خوشه‌ای را محاسبه می‌کند و توسط رابطه (۴) تعریف می‌شود [۲۰، ۲۱]:

$$O_{\perp}(u_k(x_k), C) = \perp_{l=2,C}^1(\perp_{i=1,C}^1 u_k) \quad (4)$$

در رابطه فوق  $x_k$  دارای بردار عضویت  $u_k(x_k) = (u_{1k}, \dots, u_{ck})$  می‌باشد. همچنین عملگر  $\perp$  بیانگر  $t$ -norm فازی است که در این پژوهش از عملگر  $\min$  به جای آن استفاده شده است، یعنی:  $a \perp b = \min(a, b)$ . برای بخش دوم که همان شاخص جدایی است، از عملگر  $s$ -norm فازی استفاده می‌شود [۲۱]. 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 به صورت زیر تعریف قابل تعریف است [۲۰]:

$$OS = \frac{1}{M} \sum_{k=1}^M \frac{O_{\perp}(u_k(x_k), C)}{\max_{i=1,C} u_{ik}} \quad (5)$$

جزئیات بیشتری از تابع هدف دوم و اثبات ریاضی آن در مراجع [۲۰، ۲۱] آورده شده است.

### ۳. 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 به سزایی 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 بهینه‌سازی چندهدفه غیرخطی و نامحدب است [۱۹]، برای افزایش دقت خوشه‌بندی، ساختار الگوریتم به گونه‌ای تغییر داده شده است که امکان استفاده از آن برای حل مسائل بهینه‌سازی چندهدفه نیز وجود داشته باشد. علاوه بر این به منظور تسریع در فرآیند بهینه‌سازی و جلوگیری از گیرافتادن الگوریتم در بهینه‌های محلی، راهکارهای ابتکاری جدیدی به الگوریتم اضافه شده‌اند. در ادامه ساختار این الگوریتم به اختصار توضیح داده شده است.

الگوریتم گرگ خاکستری در سال ۲۰۱۴ بر مبنای زندگی گرگهای خاکستری ارائه شده است [۲۲]. این حیوانات به صورت گروهی زندگی می‌کنند و رهبر گروه، آلفا مسوول تصمیم‌گیری در زمینه‌هایی همچون حمله و زمان آن است. روش شکار این گونه از حیوانات که در این الگوریتم برای بهینه‌سازی مورد الهام قرار گرفته است، شامل سه مرحله زیر است:

۱: ردگیری، تعقیب و نزدیک شدن به شکار

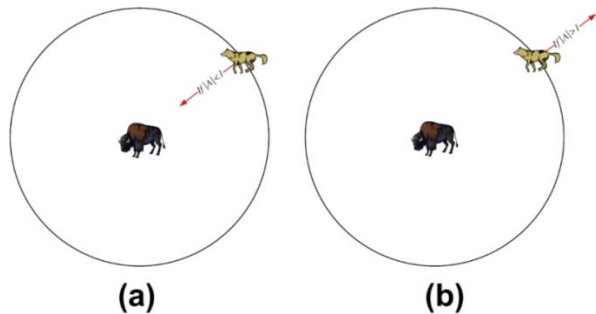
۲: دنبال کردن و محاصره کردن شکار تا زمانی که شکار از حرکت باز ایستد.

۳: حمله کردن به طرف شکار

به منظور مدل‌سازی رفتار اجتماعی گرگها، یک جمعیت تصادفی از راه‌حلها تولید، و بهترین راه‌حل بنام آلفا ( $\alpha$ )، دومین و سومین راه‌حل بهتر، نیز به ترتیب بتا ( $\beta$ ) و دلتا ( $\delta$ ) نامیده می‌شوند. همچنین سایر راه‌حل‌ها نیز به عنوان گرگهای دسته امگا ( $\omega$ ) در نظر گرفته می‌شوند. الگوریتم گرگ خاکس - تری از سه - ه جواب  $\alpha$ ،  $\beta$  و  $\delta$  جهت هدایت شکار (بهینه‌سازی) استفاده می‌کند و جوابهای  $\omega$  از این سه پیروی می‌کنند. به منظور مدل‌سازی سه - بعدی، ابتدا نقاط

همانطور که گفته شد گرگها زمانی که شکار را متوقف کردند به طرف آن حمله می کنند. به منظور مدل سازی ریاضی نزدیک شدن به طعمه، مقدار  $a$  کاهش داده می شود. باید توجه داشت که محدوده تغییرات  $A$  نیز با  $a$  کاهش خواهد یافت. به عبارت دیگر،  $A$  یک مقدار تصادفی در بازه  $[-2a, 2a]$  است که در آن  $a$  از ۲ تا ۰ در طی مراحل تکرار الگوریتم، کاهش می یابد. به ازای مقادیر تصادفی  $A$  در بازه  $[-1, 1]$ ، موقعیت بعدی عامل جستجو، می تواند در هر مکانی بین موقعیت فعلی عامل و موقعیت طعمه باشد. در الگوریتم گرگ خاکستری، موقعیت گرگها با توجه به موقعیت آلفا، بتا و دلتا بروز رسانی می شود. ممکن است بعضا لازم باشد گرگها برای پیدا کردن شکار از هم دور شوند و همیشه نزدیک شدن اتفاق نمی افتد. برای مدلسازی چنین پدیده های، از مقدار تصادفی بزرگتر از ۱ یا کمتر از -۱ برای  $A$  استفاده می شود. شکل (۱) نشان می دهد که اگر  $|A| > 1$  باشد گرگها از شکار دور می شوند، در حالیکه به ازای  $|A| < 1$ ، گرگها وادار به حمله به سمت شکار می شوند. فلوجارت الگوریتم گرگ خاکستری در شکل (۲) نشان داده شده است.

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



اطراف طعمه مشخص، سپس به سمت طعمه حرکت و در نهایت به طعمه حمله می شود. برای مشخص کردن نقاط اطراف طعمه از روابط زیر استفاده می شود [۲۲].

$$\begin{aligned} \vec{D} &= |\vec{C} \cdot \vec{X}_p(t) - \vec{X}(t)| \\ \vec{X}(t+1) &= \vec{X}_p(t) - \vec{A} \cdot \vec{D} \end{aligned} \quad (6)$$

که در آن  $t$  بیانگر تکرار فعلی الگوریتم،  $\vec{A}$  و  $\vec{C}$  بردارهای ضریب،  $\vec{X}_p$  بردار موقعیت طعمه و  $\vec{X}$  بیانگر بردار موقعیت گرگ خاکستری می باشد. بردارهای  $\vec{A}$  و  $\vec{C}$  نیز از رابطه زیر بدست می آیند [۲۲]:

$$\begin{aligned} \vec{A} &= 2\vec{a} \cdot \vec{r}_1 - \vec{a} \\ \vec{C} &= 2 \cdot \vec{r}_2 \end{aligned} \quad (7)$$

در رابطه فوق،  $\vec{r}_1$  و  $\vec{r}_2$  بردارهای تصادفی در بازه  $[0, 1]$  هستند. پارامتر  $\vec{a}$  نیز در طی فرایند اجرای الگوریتم از ۲ تا صفر کاهش می یابد. با توجه به اینکه در فضای جستجوی اولیه ایده ای در مورد موقعیت شکار وجود ندارد، فرض می شود گرگهای  $\alpha$ ،  $\beta$  و  $\delta$  به ترتیب سه تا از بهترین جوابهایی هستند که تاکنون بدست آمده اند. سایر جوابها (که گرگهای  $\omega$  نامیده می شوند) می بایست موقعیت خود را جهت همگرایی به سمت جواب بهینه، مطابق روابط زیر تغییر دهند [۲۲].

$$\begin{aligned} \vec{D}_\alpha &= |\vec{C}_1 \cdot \vec{X}_\alpha - \vec{X}| \\ \vec{D}_\beta &= |\vec{C}_2 \cdot \vec{X}_\beta - \vec{X}| \\ \vec{D}_\delta &= |\vec{C}_3 \cdot \vec{X}_\delta - \vec{X}| \\ \vec{X}_1 &= \vec{X}_\alpha - \vec{A}_1 \cdot (\vec{D}_\alpha) \\ \vec{X}_2 &= \vec{X}_\beta - \vec{A}_2 \cdot (\vec{D}_\beta) \\ \vec{X}_3 &= \vec{X}_\delta - \vec{A}_3 \cdot (\vec{D}_\delta) \\ \vec{X}(t+1) &= \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \end{aligned} \quad (8)$$

بهرتر آن نیست و غالباً 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 [۲۴]. 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.

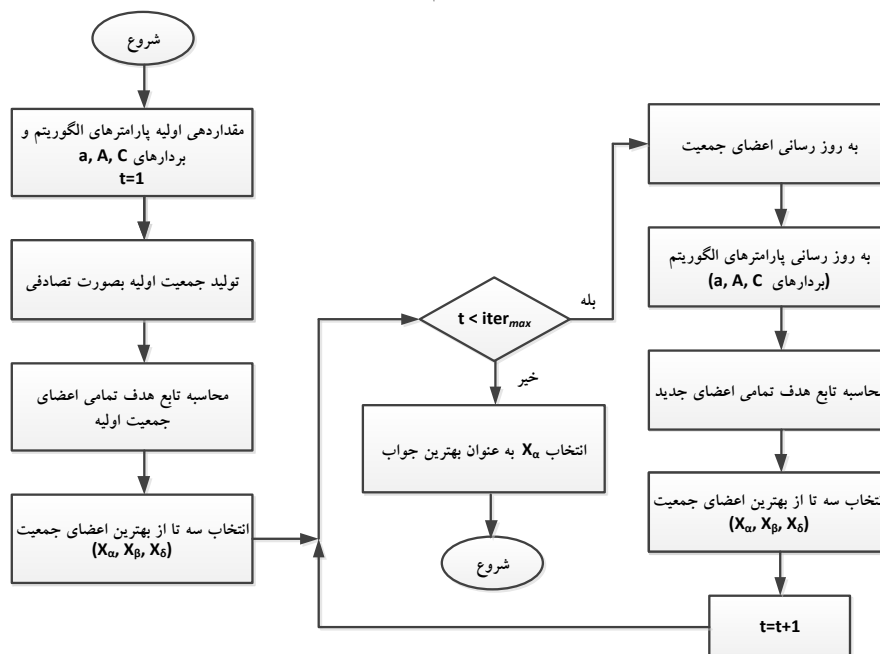
$$a(t) = \left( a_{max} - (a_{max} - a_{min}) \frac{t}{Maxgen} \right) \quad (9)$$

در رابطه فوق،  $Maxgen$  حداکثر تعداد تکرار الگوریتم،  $a_{max}$  و  $a_{min}$  نیز بیانگر محدوده تغییرات  $a$  است که مقادیر آنها مطابق با فرایندی که در بخش‌های بعدی ارائه خواهد شد، تعیین می‌شود. همانطور که مشاهده می‌شود همانند الگوریتم تبرید فلزات، پارامتر  $a$  از یک مقدار مشخص 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.

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 نسبتاً بهتری در مقایسه با سایر روش‌های بهینه‌سازی داشته باشد. اما طبق نتایج ارائه شده در مرجع [۲۴]، 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 لزوماً به معنای کارایی

شکل ۲. فلوجارت الگوریتم گرگ خاکستری [۲۲]



MOGWO اعمال می‌شوند. این تقسیم‌بندی پیش‌نیازی برای انجام روش یادگیری نیمه‌نظارتی است که در مراحل پایانی تشریح می‌گردد.

#### ۴-۲. روش کدگذاری راه‌حل

در حل مساله خوشه‌بندی با استفاده از الگوریتم‌های بهینه‌سازی هوشمند، روش‌های مختلفی جهت کدگذاری وجود دارد که به تناسب ماهیت مسئله و الگوریتمی که به کار 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 آنجایی که تشخیص تعداد مناسب خوشه‌ها غالباً از طریق راه‌حل‌های با طول متغیر انجام می‌گیرد و الگوریتم گرگ خاکستری فقط با راه‌حلی با طول ثابت کار می‌کند، روش رمزگذاری مورد استفاده نیز روش اعشاری با طول

#### ۴. ساختار روش پیشنهادی

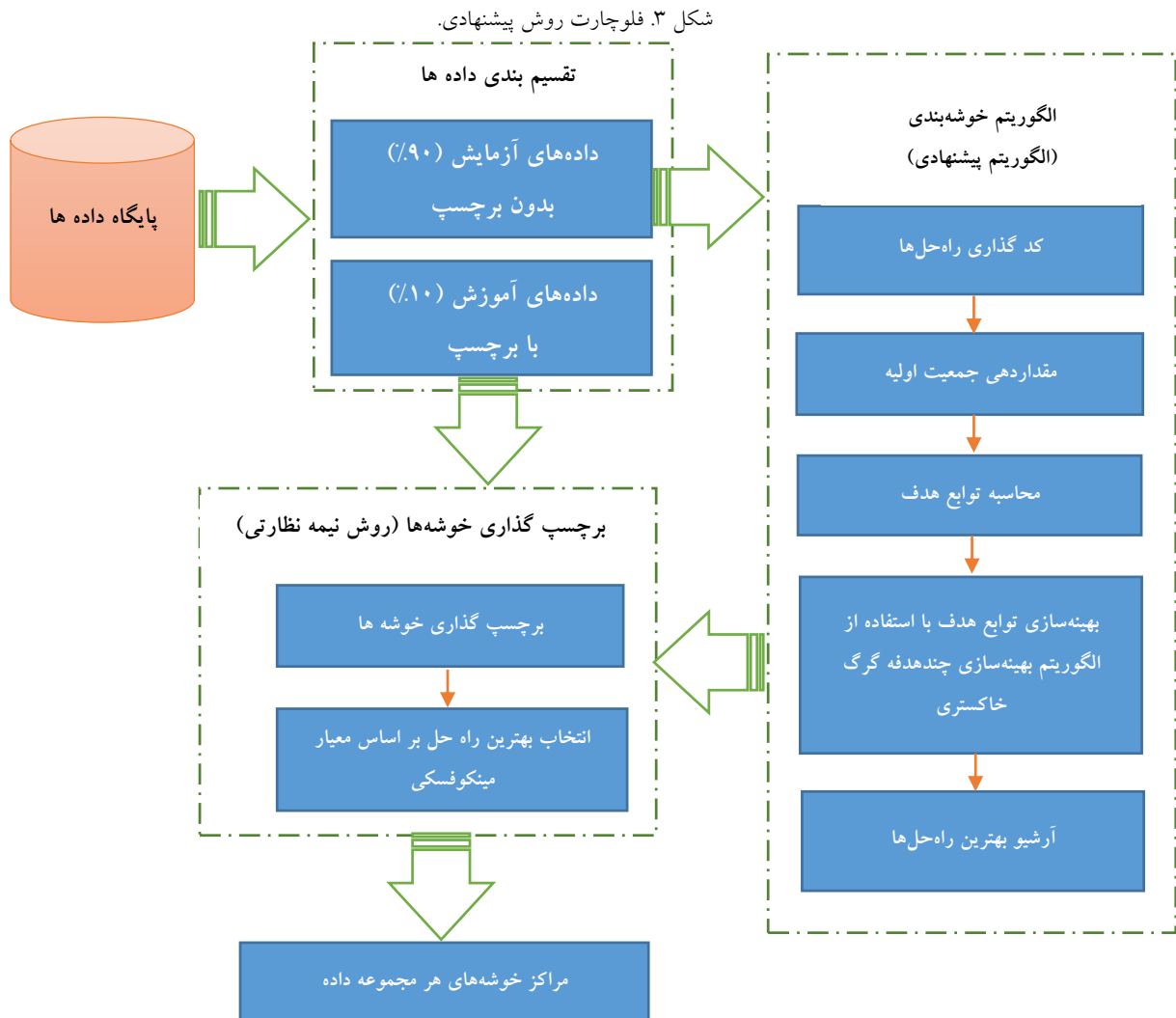
شکل (۳) فلوجارت الگوریتم خوشه‌بندی فازی مبتنی بر روش بهینه‌سازی گرگ خاکستری چندهدفه (FCM-MOGWO) را نشان می‌دهد. در ادامه بخش‌های مختلف این فلوجارت تشریح خواهد شد. پس از آن به معرفی الگوریتم FCM-MOGWO به‌عنوان هسته اصلی روش پیشنهادی پرداخته می‌شود.

#### ۴-۱. 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 ۱۰ درصد کل داده‌ها هستند دارای برچسب کلاس بوده درحالی‌که ۹۰ درصد باقیمانده (داده‌های آزمایش) فاقد برچسب کلاس هستند. داده‌های آموزش به‌عنوان ورودی به الگوریتم FCM-

دارند، معمولاً از یک بردار پوشش یا آستانه فعال‌سازی استفاده می‌شود. در صورتی که مقدار آستانه به ازای یک سرخوشه از ۰/۵ بیشتر بود، سرخوشه متناظر فعال و در غیر این صورت غیرفعال خواهد بود. شکل (۴) 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.

ثابت انتخاب شده است. در این روش، تمام راه‌حل‌های خوشه‌بندی، دارای تعداد از پیش تعیین‌شده  $K_{max}$  سرخوشه هستند. از این رو، اندازه تمام اعضای جمعیت در طول فرآیند بهینه‌سازی یکسان و به اندازه  $D \times K_{max}$  خواهد بود که  $D$  بیانگر ابعاد یا تعداد ویژگی‌های مجموعه داده است. همچنین، برای تعیین این که کدام سرخوشه‌ها در فرآیند خوشه‌بندی در هر تکرار الگوریتم شرکت



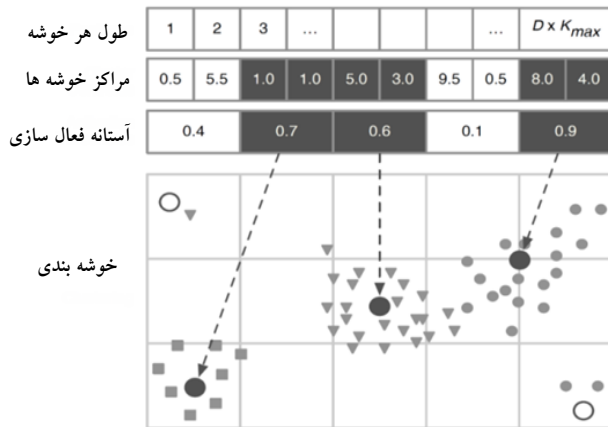
متغیرهای تصمیم‌گیری شامل ویژگی‌های سرخوشه‌ها خواهد بود. به عبارت دیگر، هر گرگ به عنوان یک راه‌حل، مختصات  $k$  سرخوشه را در خود دارد. همچنین در این روش، برای نمایش یک راه‌حل از یک بردار به طول

در الگوریتم بهینه‌سازی گرگ خاکستری، هر گرگ بیانگر یک راه‌حل برای مساله می‌باشد. هر راه‌حل نیز شامل تعدادی متغیرهای تصمیم‌گیری است که در این جا به دلیل استفاده از روش رمزگذاری مبتنی بر مرکز،

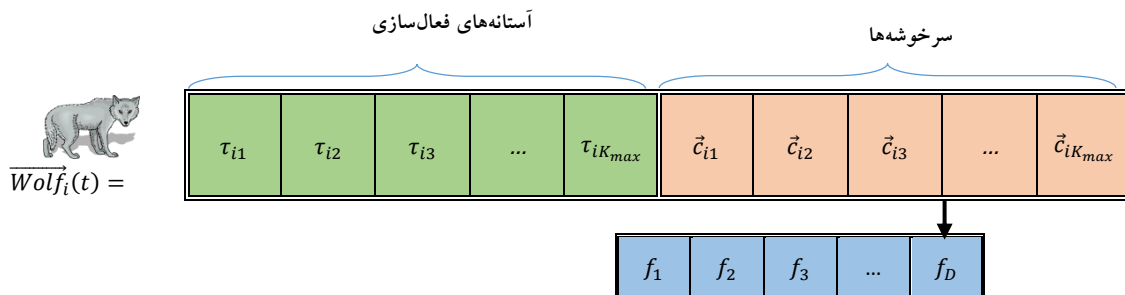
در تکرار  $t$ ام غیرفعال خواهد شد. این قانون به صورت رابطه زیر نیز قابل بیان است:

$$\begin{cases} \tau_{ij} > 0.5 & : \quad \vec{c}_{ij} = Active \\ \tau_{ij} < 0.5 & : \quad \vec{c}_{ij} = Deactive \end{cases} \quad (10)$$

شکل ۴. روش رمزگذاری اعشاری با طول ثابت



شکل ۵. یک نمونه گرگ خاکستری به عنوان راه حل خوشه بندی.



می گردد. از آنجایی که در مسئله خوشه بندی حداقل تعداد خوشه ها باید ۲ باشد ( $K_{min} = 2$ ) تعداد فعال سازهای با مقدار آستانه بیشتر از ۰/۵، نباید کمتر از ۲ باشد. به همین دلیل، در روش پیشنهادی از یک واحد کنترلی برای کنترل این موضوع و جلوگیری از تولید جمعیت

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

محتویات کروموزوم اول تا کروموزوم  $K_{max}$ ، اعداد مثبت اعشاری در بازه [۰، ۱] هستند که بیان گر فعال بودن یا غیر فعال بودن سرخوشه متناظر می باشند. این کروموزوم ها که با  $\tau$  نشان داده شده اند، آستانه فعال سازی نامیده می شوند. سایر موقعیت های بردار نیز به  $K_{max}$  مرکز سرخوشه ( $\vec{c}$ )، با ابعاد ( $f$ ) اختصاص دارد. برای درک بهتر ساختار، یک بردار راه حل (یک گرگ) در شکل (۵) آورده شده که بیان گر موقعیت گرگ  $i$  در تکرار  $t$  است. در این روش تمام راه حل های خوشه بندی دارای تعداد  $K_{max}$  سرخوشه می باشند و مقدار آستانه برابر ۰/۵ در نظر گرفته شده است. در این حالت،  $J$ امین مرکز خوشه در گرگ  $i$ ، در تکرار  $t$ ام در صورتی فعال خواهد بود که  $\tau_{ij} > 0.5$  باشد و اگر  $\tau_{ij} < 0.5$  بود،  $J$ امین مرکز خوشه از گرگ  $i$  در فرآیند خوشه بندی

### ۳-۴. تولید جمعیت اولیه

به منظور تولید جمعیت اولیه در روش پیشنهادی، ابتدا به کمک یک عملگر تصادفی قسمت آستانه فعال سازی مربوط به راه حل ایجاد می گردد. برای این منظور به تعداد  $K_{max}$  عدد تصادفی اعشاری در بازه [۰، ۱] تولید

مشکل دار استفاده شده و تا زمانی که تعداد متغیرهای فعال‌ساز با مقدار آستانه  $0/5$ ، 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 به اندازه  $K_{max}$  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.

#### ۴-۴. 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 منحصرآ جوابهای نامغلوب هستند. یعنی جوابهایی که اولاً جزو جوابهای بهینه مساله بوده و ثانیاً هیچ ارجحیتی نسبت

به هم 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 به دست آمده‌اند، با اعضای فعلی آرشیو مقایسه خواهند شد. یکی از سه حالت زیر ممکن است رخ دهد:

- اگر عضو جدید به وسیله 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.
- در صورتی که 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، به ترتیب سه تا از بهترین جواب‌های الگوریتم تا آن تکرار هستند. اما در مسائل بهینه‌سازی چندهدفه همانطور که قبلاً 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 [۲۶].

## ۶-۴. ارزیابی و انتخاب بهترین راه حل

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 DB، DI، 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.

$$MS = \frac{\sqrt{\sum_i \binom{n_i}{2} + \sum_j \binom{n_j}{2} - 2 \sum_{ij} \binom{n_{ij}}{2}}}{\sqrt{\sum_j \binom{n_j}{2}}} \quad (13)$$

در رابطه فوق،  $n_{ij}$  بیان‌گر تعداد داده‌هایی است که بین دو خوشه  $i$  و  $j$  مشترک هستند.  $n_i$  تعداد داده‌های موجود در خوشه  $i$  و  $n_j$  نیز تعداد داده‌های موجود در شاخه  $j$  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.

$$DB(C) = \frac{1}{K} \sum_{c_k \in C} \max_{c_r \in C \setminus c_k} \left\{ \frac{S(c_k) + S(c_r)}{d_e(\bar{c}_k, \bar{c}_r)} \right\} \quad (11)$$

که در آن:

$$S(c_k) = \frac{1}{n_k} \sum_{x_i \in c_k} d_e(X_i, \bar{c}_k) \quad (12)$$

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

## ۵-۴. 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 فرآیند 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].

## 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 کار گرفته 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.  $M_i$  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	۴	۱۰۰۰	۴×۲۵۰
Square4 [30]	۲	۴	۱۰۰۰	۴×۲۵۰
Sizes5 [30]	۲	۴	۱۰۰۰	۷۶۹,۷۷,۷۷,۷۷
Iris [31]	۴	۳	۱۵۰	۵۰, ۵۰, ۵۰
BreastCancer [31]	۹	۲	۶۸۳	۴۴۴, ۲۳۹
Newthyroid [31]	۵	۳	۲۱۵	۱۵۰, ۳۵, ۳۰
Wine [31]	۱۳	۳	۱۷۸	۵۹, ۷۱, ۴۸
LiverDisorders [31]	۶	۲	۳۴۱	۱۴۲, 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 شده‌اند:

$$N_p = 40, MaxGen = 100$$

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 تأثیر پارامتر  $a$  بر کارایی الگوریتم، همانند تأثیر پارامتر  $ma$  در الگوریتم بهینه‌سازی تبرید فلزات است، مقادیر  $a_{min}$  و  $a_{max}$  نیز با الهام از روابط زیر که جزو پرکاربردترین روشهای تعیین  $ma$  در الگوریتم تبرید فلزات است، محاسبه خواهند شد [۳۲].

$$a_{max} = -\frac{|\max(J_m) + \max(OS)|}{\ln p_0} \quad (14)$$

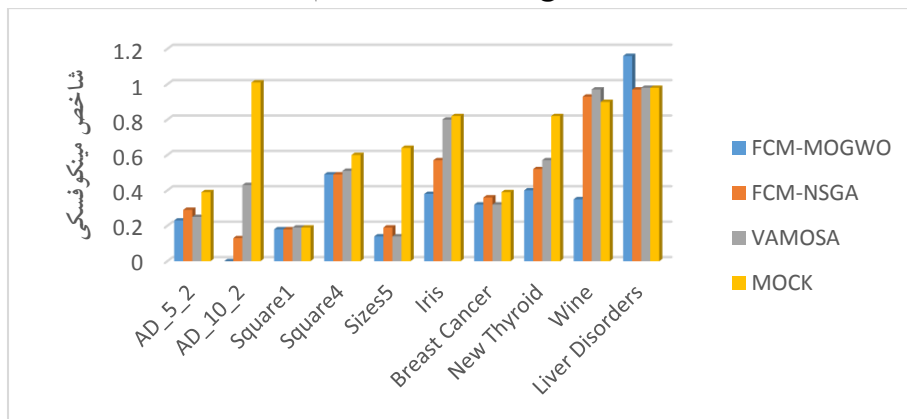
$$a_{min} = -\frac{|\min(J_m) + \min(OS)|}{\ln p_0}$$

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 قبلاً 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]		VAMOS[33]		MOCK[34]	
		K	MS	K	MS	K	MS	K	MS
AD_5_2	۵	۵	۰.۲۳	۵	۰.۲۹	۵	۰.۲۵	۶	۰.۳۹
AD_10_2	۱۰	۱۰	۰.۰۰	۱۰	۰.۱۳	۱۰	۰.۴۳	۶	۱.۰۱
Square1	۴	۴	۰.۱۸	۴	۰.۱۸	۴	۰.۱۹	۴	۰.۱۹
Square4	۴	۴	۰.۴۹	۴	۰.۴۹	۴	۰.۵۱	۴	۰.۶۰
Sizes5	۴	۴	۰.۱۴	۴	۰.۱۹	۴	۰.۱۴	۲	۰.۶۴
Iris	۳	۳	۰.۳۸	۳	۰.۵۷	۲	۰.۸۰	۲	۰.۸۲
Breast Cancer	۲	۲	۰.۳۲	۲	۰.۳۶	۲	۰.۳۲	۲	۰.۳۹
New Thyroid	۳	۳	۰.۴۰	۳	۰.۵۲	۵	۰.۵۷	۲	۰.۸۲
Wine	۳	۳	۰.۳۵	۳	۰.۹۳	۳	۰.۹۷	۳	۰.۹۰
Liver Disorders	۲	۲	۱.۱۶	۲	۰.۹۷	۲	۰.۹۸	۳	۰.۹۸

شکل ۶. نمودار میله‌ای مقایسه نتایج روش پیشنهادی با الگوریتم‌های بهینه‌سازی چندهدفه



مینکوفسکی است که کمترین مقدار آن نشان‌دهنده بهترین عملکرد می‌باشد. نمودار میله‌ای حاصل از این آزمایش در شکل (۶) 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

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]، VAMOS ارائه شده در مرجع [۳۳] و MOCK ارائه شده در مرجع [۳۴] 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

است. برای این مجموعه داده نیز، همه روش‌ها بجز روش MOCK، قادر به تخمین تعداد صحیح خوشه‌ها را به درستی تخمین زده‌اند (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 (۴ 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 ۰/۴۹ بهترین عملکرد و روش MOCK با مقدار مینکوفسکی ۰/۶۰ بدترین عملکرد را دارا بوده‌اند. در این مورد نیز هر چهار روش، تعداد خوشه‌های مجموعه داده را به درستی تخمین زده‌اند (۴ خوشه).

❖ در مجموعه داده Sizes5، روش پیشنهادی به‌همراه روش VAMOSa با مقدار مینکوفسکی مشترک ۰/۱۴ بهترین عملکرد را داشته‌اند. روش MOCK با مقدار مینکوفسکی ۰/۶۴ بدترین عملکرد را دارا بوده

❖ 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 ۰/۴۰ 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).

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 (۴ 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 ۰/۴۹ بهترین عملکرد و روش MOCK با مقدار مینکوفسکی ۰/۶۰ بدترین عملکرد را دارا بوده‌اند. در این مورد نیز هر چهار روش، تعداد خوشه‌های مجموعه داده را به درستی تخمین زده‌اند (۴ خوشه).

❖ در مجموعه داده Sizes5، روش پیشنهادی به‌همراه روش VAMOSa با مقدار مینکوفسکی مشترک ۰/۱۴ بهترین عملکرد را داشته‌اند. روش MOCK با مقدار مینکوفسکی ۰/۶۴ بدترین عملکرد را دارا بوده

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 بهتری از نظر مقدار مینکوفسکی داشته است.

الگوریتم VAMOSA که از دو معیار Sym-index و XB-index به ترتیب به منظور 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 به خاطر محدودیت دو تابع هدف استفاده شده در این الگوریتم باشد که باعث عدم کارایی آن در مواجهه با داده‌های هم‌پوشان شده است. از آنجایی که 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 به وسیله

❖ 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؛ به طوری که 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 ۹۰ درصد 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

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 این جا، 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.

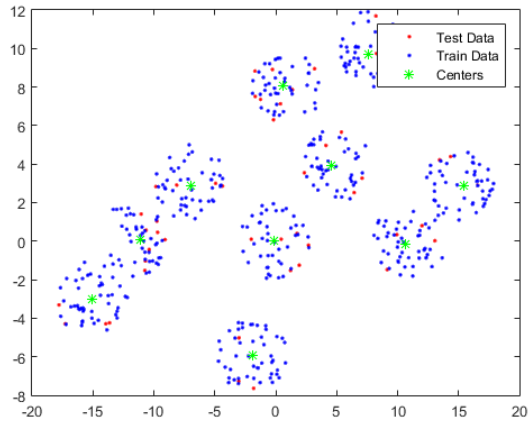
ستاره (\*) سبز رنگ مشخص شده است. همانطور که مشاهده می‌شود روش پیشنهادی علاوه بر تعیین تعداد بهینه خوشه‌ها، عمل خوشه‌بندی را برای مجموعه داده‌های غیرهم‌پوشان (شکل ۷ و ۸)، مجموعه داده‌های با خوشه‌های نابرابر (شکل ۹) و مجموعه داده‌های هم‌پوشان (شکل‌های ۱۰ و ۱۱) با دقت بسیار خوبی انجام دهد.

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

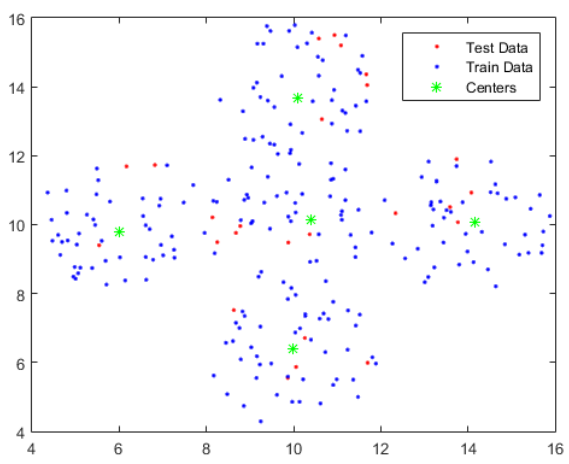
جدول ۳. نتایج روش پیشنهادی در مقایسه با رویکردهای تک‌هدفه

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	۵	۵	۰.۲۳	۵	۰.۲۳	۱۱	۱.۵۹	۵	۰.۲۵
AD_10_2	۱۰	۱۰	۰.۰۰	۱۰	۰.۱۳	۱۷	۱.۵۳	۷	۰.۸۴
Square1	۴	۴	۰.۱۸	۴	۰.۱۸	۲۱	۱.۷۷	۴	۰.۲۰
Square4	۴	۴	۰.۴۹	۴	۰.۴۸	۳۰	۱.۸۳	۵	۰.۵۲
Sizes5	۴	۴	۰.۱۴	۲	۰.۷۰	۲۲	۰.۶۲	۵	۰.۲۲
Iris	۳	۳	۰.۳۸	۲	۱.۱۹	۱۲	۱.۵۱	۳	۰.۶۲
Breast Cancer	۲	۲	۰.۳۲	۲	۰.۳۶	۲۴	۱.۳۳	۲	۰.۳۷
New Thyroid	۳	۳	۰.۴۰	۲	۰.۹۹	۱۵	۰.۸۷	۳	۰.۵۸
Wine	۳	۳	۰.۳۵	۳	۰.۵۷	۱۳	۱.۳۹	۶	۰.۹۷
Liver Disorders	۲	۲	۱.۱۶	۲	۱.۲۱	۱۸	۱.۴۱	۲	۰.۹۸

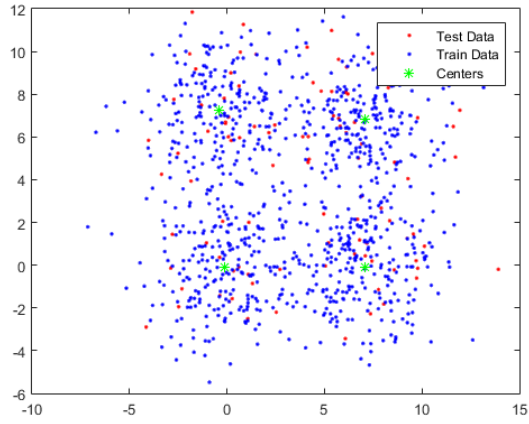
شکل ۸. مجموعه داده AD\_10\_2 با شاخص مینکوفسکی صفر



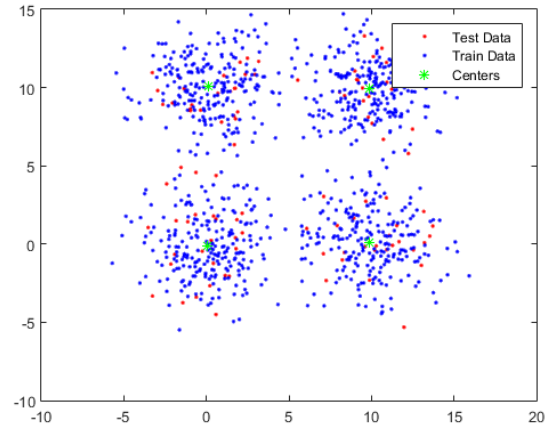
شکل ۷. مجموعه داده AD\_5\_2 با شاخص مینکوفسکی ۰,۲۳



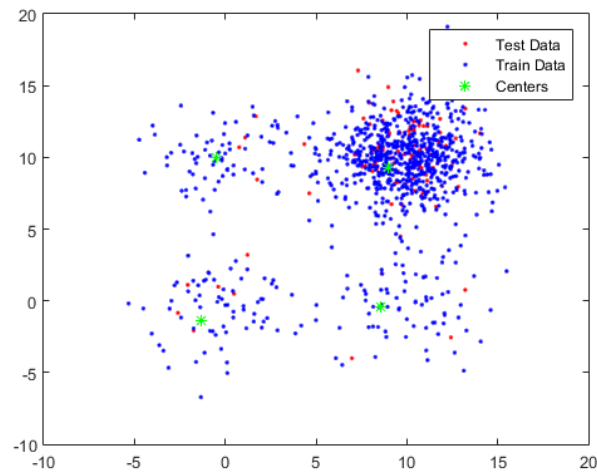
شکل ۱۰. مجموعه داده Square4 با شاخص مینکوفسکی ۰,۴۹



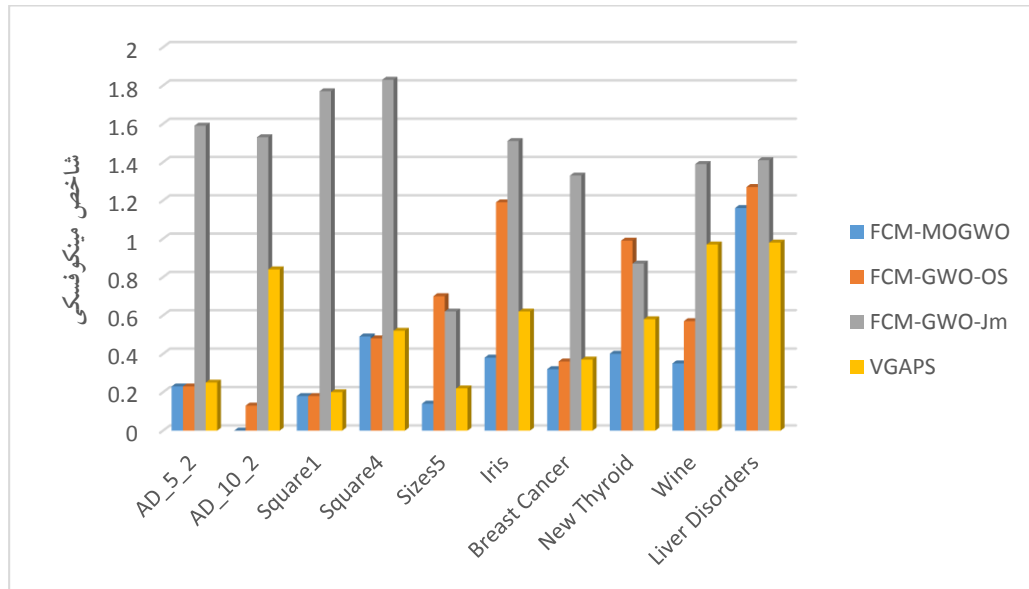
شکل ۹. مجموعه داده Square1 با شاخص مینکوفسکی ۰,۱۸



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



شکل ۱۲. نمودار میله‌ای مقایسه نتایج رویکردهای تک‌هدفه



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

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.

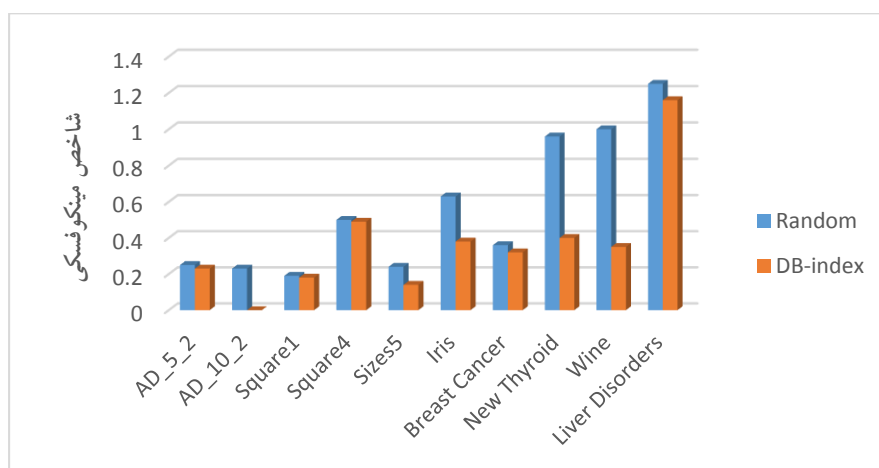
رویکرد FCM-GWO-OS تخمین‌های قابل قبولی از تعداد خوشه‌های بسیاری از مجموعه داده‌های مورد بررسی داشته است و در مورد تخمین تعداد خوشه‌ها برای مجموعه داده‌های Iris و Sizes5 و New Thyroid ناموفق عمل کرده است که نابرابری تعداد اعضای خوشه‌ها و هم‌پوشانی بالا در این مجموعه داده‌ها

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.

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

جدول ۴. نتایج تأثیر روش انتخاب گرگ رهبر در روش پیشنهادی

Datasets	Actual K	Random		DB-index	
		K	MS	K	MS
AD_5_2	۵	۵	۰.۲۵	۵	۰.۲۳
AD_10_2	۱۰	۱۰	۰.۲۳	۱۰	۰.۱۳
Square1	۴	۴	۰.۱۹	۴	۰.۱۸
Square4	۴	۴	۰.۵۰	۴	۰.۴۹
Sizes5	۴	۴	۰.۲۴	۴	۰.۱۴
Iris	۳	۳	۰.۶۳	۳	۰.۳۸
Breast Cancer	۲	۲	۰.۳۶	۲	۰.۳۲
New Thyroid	۳	۳	۰.۹۶	۳	۰.۴۰
Wine	۳	۳	۱.۰۰	۳	۰.۵۰
Liver Disorders	۲	۵	۱.۲۵	۵	۱.۱۶



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

پارامترهای قابل تنظیم جهت آزمایش‌های این بخش به دو دسته شامل روش انتخاب گرگ رهبر و روش حذف

در پایان به بررسی تأثیر تنظیمات مختلف پارامترها بر روی کارایی روش پیشنهادی پرداخته می‌شود.

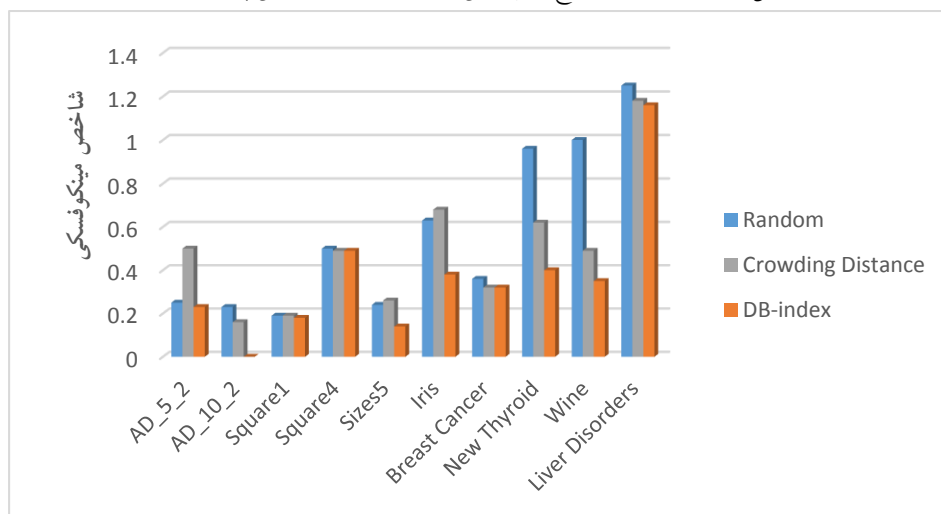
DB منجر به بهبود دقت خوشه‌بندی شده است. همانطور که قبلاً اشاره شد، در الگوریتم‌های بهینه‌سازی چندهدفه، ظرفیت مشخصی برای لیست آرشیو در نظر گرفته می‌شود. در صورتیکه تعداد اعضای موجود در لیست آرشیو از ظرفیت آن بیشتر شود، می‌بایست اعضای اضافی از لیست حذف شوند.

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 (۴) آورده شده است. نمودار میله‌ای مقایسه نتایج حاصل از این آزمایش نیز در شکل (۱۳) آمده است. همانطور که مشاهده می‌شود برای تمامی مجموعه داده‌ها انتخاب گرگ رهبر با استفاده از معیار

جدول ۵. 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	۵	۵	۰.۴۳	۵	۰.۵۰	۵	۰.۲۳
AD_10_2	۱۰	۱۰	۰.۲۹	۱۰	۰.۱۶	۱۰	۰.۰۰
Square1	۴	۴	۰.۲۱	۴	۰.۱۹	۴	۰.۱۸
Square4	۴	۴	۰.۵۰	۴	۰.۴۹	۴	۰.۴۹
Sizes5	۴	۴	۰.۲۵	۴	۰.۲۶	۴	۰.۱۴
Iris	۳	۳	۰.۶۴	۳	۰.۶۸	۳	۰.۳۸
Breast Cancer	۲	۲	۰.۳۲	۲	۰.۳۲	۲	۰.۳۲
New Thyroid	۳	۳	۰.۶۱	۳	۰.۶۲	۳	۰.۴۰
Wine	۳	۳	۰.۴۲	۳	۰.۴۹	۳	۰.۳۵
Liver Disorders	۲	۵	۱.۱۹	۵	۱.۱۸	۲	۱.۱۶

شکل ۱۴. نمودار میله‌ای نتایج تأثیر روش حذف از آرشیو در روش پیشنهادی



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 به جای استفاده از الگوریتم بهینه‌سازی گرگ خاکستری به تنهایی، 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

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 (۱۴) 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.

## ۶. نتیجه گیری

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

- [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*

شد. می‌توان با توجه به کاربرد و ماهیت داده‌ها، از سایر روش‌های خوشه‌بندی نرم و یا سخت، مانند روش‌های سلسله‌مراتبی یا روش‌های مبتنی بر 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.

- 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.
- 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