

استخراج الگوهای لرزه‌ای با استفاده از خوشه‌بندی فازی بهینه‌شده با الگوریتم بهینه‌سازی توده ذرات

رضوان قاسمی‌نژاد^۱، رحیم علی‌عباسپور^{۲*}، مسعود مجرب^۳

^۱ کارشناس ارشد سیستم‌های اطلاعات مکانی - دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی - پردیس

دانشکده‌های فنی - دانشگاه تهران

r.ghasemi@ut.ac.ir

^۲ استادیار دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی - پردیس دانشکده‌های فنی - دانشگاه تهران

abaspour@ut.ac.ir

^۳ محقق آزمایشگاه مهندسی زمین - دانشکده مهندسی معدن - پردیس دانشکده‌های فنی - دانشگاه تهران

mmojarab@alumni.ut.ac.ir

(تاریخ دریافت خرداد ۱۳۹۵، تاریخ تصویب آذر ۱۳۹۵)

چکیده

شناسایی الگوها در داده‌های لرزه‌ای از طریق خوشه‌بندی، به‌عنوان یکی از رایج‌ترین روش‌های داده‌کاوی، منجر به استخراج اطلاعات بسیار مهمی از یک حجم زیاد داده می‌شود. به دلیل ماهیت داده‌های لرزه‌ای، الگوریتم‌های خوشه‌بندی فازی نتایج واقع‌بینانه‌تری را ارائه می‌کنند. اگرچه الگوریتم‌های بسیاری بدین منظور ارائه شده است اما حساس بودن به مقادیر اولیه و به تله افتادن در جواب‌های بهینه محلی از جمله مشکلاتی است که در رابطه با روش‌های ارائه‌شده برای خوشه‌بندی وجود دارد. از این رو، در این مقاله الگوریتم‌های فرا ابتکاری به‌عنوان روش‌های بهینه‌سازی کارآمد به‌منظور رفع مشکلات روش‌های خوشه‌بندی پیشنهاد شده‌اند. در این مقاله سعی شد تا با استفاده از ترکیب الگوریتم بهینه‌سازی توده ذرات و دو الگوریتم خوشه‌بندی فازی Gustafson Kessel و Fuzzy c-means دو رویکرد برای خوشه‌بندی داده‌های لرزه‌ای ارائه شود. هر یک از این دو الگوریتم که به ترتیب PSO-GK و PSO-FCM نامیده می‌شوند بر روی داده‌های لرزه‌ای ساختگی و داده‌های لرزه‌ای ایران اعمال شدند. به‌منظور ارزیابی نتایج حاصل از دو الگوریتم، سه معیار ارزیابی خوشه‌بندی فازی یعنی FHV، متوسط چگالی بخش‌بندی و چگالی بخش‌بندی مورد استفاده قرار گرفتند. مقدار FHV در الگوریتم PSO-GK به میزان ۰/۴۲۷۲ برای داده‌های ساختگی و به میزان ۰/۰۹۴۱ برای داده‌های لرزه‌ای ایران کمتر (بهتر) از مقدار این معیار در الگوریتم PSO-FCM می‌باشد. همچنین مقادیر دو معیار ارزیابی دیگر هم برای داده‌های ساختگی و هم برای داده‌های لرزه‌ای ایران در الگوریتم PSO-GK دارای مقادیر بهتری می‌باشند که نشان از کارایی بهتر الگوریتمی است که بر مبنای Gustafson Kessel ارائه شده نسبت به الگوریتمی که بر مبنای Fuzzy c-means ارائه شده برای تحلیل داده‌های لرزه‌ای دارد.

واژگان کلیدی: بهینه‌سازی توده ذرات، Fuzzy c-means، Gustafson Kessel، تحلیل‌های لرزه‌ای

* نویسنده رابط

۱- مقدمه

تحلیل حجم زیادی از داده‌های لرزه‌ای به صورت خودکار به منظور استخراج الگوهای لرزه‌ای یکی از مسائل مطرح در زلزله‌شناسی مهندسی است [۱]. از سوی دیگر، امروزه داده‌کاو مکانی-زمانی به عنوان شناخته‌شده‌ترین رویکرد برای کشف الگو از داده‌های حجیم محسوب می‌گردد [۲]. در میان روش‌های مختلف داده‌کاو، استفاده از روش‌های خوشه‌بندی^۱ در استخراج الگوهای خوشه‌ای و گروهی کارایی خوبی از خود نشان داده‌اند. هدف این گروه از روش‌ها تعیین کلاس یا خوشه‌هایی از داده‌ها است که هر خوشه از داده‌های مشابه تشکیل و در خوشه‌های مختلف داده‌های غیرمشابه قرار می‌گیرند. مزیت عمده روش‌های خوشه‌بندی این است که به صورت غیر نظارت‌شده و بدون نیاز به داده‌های آموزشی عمل می‌کنند و به اطلاعات قبلی نیز نیاز ندارند [۲].

استفاده از خوشه‌بندی با معرفی الگوریتم k-means توسط MacQueen شروع شد [۳] که به دلیل سادگی در فهم و اجرا به سرعت مورد توجه بسیاری از محققین در حوزه‌های کاربردی مختلف قرار گرفت. با استقبال از ایده خوشه‌بندی انواع دیگری از الگوریتم‌های خوشه‌بندی ارائه شد که از نظر روش عملکردی می‌توان آن‌ها را در سه دسته روش‌های خوشه‌بندی افرازی، چگالی‌مبنا و سلسله-مراتبی دسته‌بندی کرد [۴]. کاربرد هر یک از این روش‌ها به نوع مسأله بستگی دارد. از دیدگاه نوع تخصیص داده‌ها به خوشه‌ها، روش‌های خوشه‌بندی به دو دسته قطعی و فازی تقسیم می‌شوند. هر داده در خوشه‌بندی قطعی تنها به یک خوشه تعلق می‌گیرد، در حالی که هر داده در روش‌های فازی می‌تواند به طور هم‌زمان به همه خوشه‌ها با درجه عضویت‌های مختلف نسبت داده شود [۵].

در تحلیل‌های لرزه‌ای ضروری است که یک منطقه به ناحیه‌های مختلف با بیشترین و کمترین شباهت در ویژگی‌های لرزه‌ای تقسیم شود، از آنجایی که الگوهای پنهان در داده‌های لرزه‌ای کاملاً مشخص نیست، از روش‌های خوشه‌بندی بدین منظور استفاده می‌شود. نکاتی در رابطه با داده‌های لرزه‌ای وجود دارد از جمله اینکه چگالی و تعداد رویدادهای لرزه‌ای در مناطق مختلف یکسان نیست،

رویدادهای لرزه‌ای عمدتاً در بخش‌های مختلف گسل اصلی توزیع شده‌اند و اینکه ویژگی‌های لرزه‌ای به تدریج تغییر می‌کنند به طوری که تغییرات ناگهانی در این داده‌ها وجود ندارد [۶]، بنابراین استفاده از روش‌های فازی مناسب‌تر بوده و نتایج واقع‌بینانه‌تری حاصل می‌شود.

محققین مختلفی با استفاده از خوشه‌بندی به تعیین الگوهای لرزه‌ای پرداخته‌اند که می‌توان به Aminzadeh و Chatterjee [۷] اشاره کرد که پس از توجیه استفاده از روش‌های خوشه‌بندی برای تحلیل داده‌های لرزه‌ای، از روش ISODATA بدین منظور استفاده کرد. Ouillon و همکاران [۸] با استفاده از روش k-means بهبود یافته به شناسایی گسل‌های موجود در منطقه مورد مطالعه خود پرداخته‌اند. Weatherill و Burton [۹] چشمه‌های لرزه‌ای را از طریق خوشه‌بندی رویدادهای لرزه‌ای توسط الگوریتم k-means تعیین کرده‌اند. Rehman و همکاران [۱۰] نیز از الگوریتم k-means جهت بخش‌بندی لرزه‌ای پاکستان بهره برده‌اند. Ansari و همکاران [۶] از روش خوشه‌بندی فازی Gath-Giva برای تشخیص الگوهای پنهان میان داده‌های لرزه‌ای استفاده کرده‌اند که هدف آن‌ها استخراج مدل رفتار لرزه‌ای بخش‌های مختلف ایران از طریق بررسی نتایج خوشه‌بندی می‌باشد. Panda و Nanda [۱۱] با استفاده از الگوریتم خوشه‌بندی چگالی مبنای DBSCAN پهنه‌های خطر لرزه‌ای موجود در ژاپن را تعیین کرده‌اند. همچنین Hashemi و Mehdi-zadeh [۱۲] با استفاده از کاربردهای خوشه‌بندی سلسله‌مراتبی به تعیین مناطق تک‌تونیکی ناحیه زاگرس در ایران پرداخته‌اند.

از جمله مشکلاتی که اغلب روش‌های خوشه‌بندی با آن مواجه‌اند این است که به مقادیر اولیه حساس بوده و در اغلب موارد در جواب‌های بهینه محلی گیر می‌کنند و نمی‌توانند خوشه‌های واقعی را در فضای مسئله بیابند [۱۳]؛ بنابراین مسئله خوشه‌بندی می‌تواند به عنوان یک مسئله بهینه‌سازی مطرح شود و الگوریتم‌های جستجوی سراسری بهینه برای انجام این فرایند مورد نیاز است [۱۴]. الگوریتم‌های فرا ابتکاری که جزئی از الگوریتم‌های بهینه‌سازی هوشمند می‌باشند [۴]، به طور گسترده‌ای به عنوان روش‌های کارآمد برای حل بسیاری از مسائل بهینه‌سازی از جمله خوشه‌بندی مورد استفاده قرار می‌گیرند [۱۵]. این الگوریتم‌ها به منظور یافتن جواب‌های بهینه یا نزدیک به بهینه، از استراتژی جستجوی تکراری احتمالی

^۱ clustering

الگوریتم‌ها یک بخش‌بندی فازی را ایجاد می‌کنند که به صورت ماتریس بخش‌بندی فازی $U=[u_{ij}]_{c \times n}$ نشان داده می‌شود.

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{ik}} \right)^{\frac{2}{M-1}}} \quad (2)$$

متداول‌ترین الگوریتمی که در خوشه‌بندی فازی مورد استفاده قرار می‌گیرد، الگوریتم Fuzzy c-means (FCM) است که شکل اولیه آن توسط Dunn [۱۹] ارائه و سپس توسط Bezdek [۲۰] تکمیل شد. این الگوریتم مجموعه داده را در C خوشه شبیه، که تعداد مختلفی عضو دارند، تقسیم می‌کند. هدف FCM به دست آوردن C خوشه فازی از طریق مینیمم کردن تابع ارزیابی رابطه (۱) است به طوری که d_{ij} فاصله اقلیدسی بین مرکز خوشه i ام و داده j ام می‌باشد [۲۱].

الگوریتم خوشه‌بندی فازی Gustafson Kessel (GK) [۲۲] در بیشتر موارد به عنوان روش قدرتمند خوشه‌بندی مورد استفاده قرار می‌گیرد. در واقع الگوریتم GK توسعه یافته الگوریتم FCM است. برخلاف الگوریتم FCM که از فاصله اقلیدسی استفاده می‌کند، ویژگی اصلی الگوریتم GK در نظر گرفتن فاصله Mahalanobis جهت خوشه‌بندی است که علاوه بر مرکز خوشه، ماتریس کوواریانس فازی خوشه نیز در محاسبات نقش دارند [۲۳]. نتیجه این الگوریتم به تعداد و مقادیر اولیه مراکز و ماتریس کوواریانس خوشه‌ها حساس است [۲۳]. رابطه فاصله Mahalanobis داده x_j در (۳) ذکر شده است [۲۲]:

$$d_{ij}^2 = (x_j - v_i)^T A_i (x_j - v_i) \quad (3)$$

که x_j همان داده j ام و v_i مرکز خوشه i ام است. ماتریس A_i از ماتریس کوواریانس فازی F_i طبق رابطه (۴) و (۵) به دست می‌آید:

$$F_i = \frac{\sum u_{ij}^M (x_j - v_i)(x_j - v_i)^T}{\sum u_{ij}^M} \quad (4)$$

$$A_i = [\rho_i \det(F_i)]^{1/\rho_i} F_i^{-1} \quad \forall \rho_i > 0 \quad (5)$$

بهره گرفته به طوری که پیچیدگی محاسباتی کمی داشته و در زمان محدودی به جواب می‌رسند [۱۶].

در این مقاله سعی شده است که با استفاده از کاربرد روش بهینه‌سازی توده ذرات و روش‌های خوشه‌بندی فازی الگوریتم‌هایی ارائه گردد به طوری که با تحلیل داده‌های لرزه‌ای خوشه‌های بهینه سراسری را شناسایی نمایند. هر یک از این الگوریتم‌ها بر روی داده‌های لرزه‌ای ساختگی و داده‌های لرزه‌ای ایران اعمال می‌شود. به منظور تعیین مناسب‌ترین الگوریتم جهت تحلیل داده‌های لرزه‌ای، دو روش با استفاده از معیارهای ارزیابی مقایسه می‌شوند و نتایج حاصل از آن‌ها از طریق گسل‌های منطقه مورد ارزیابی قرار می‌گیرند. در ادامه و در بخش دوم این مقاله، مفاهیم اساسی در رابطه با خوشه‌بندی و بهینه‌سازی مطرح و سپس روش‌های پیشنهادی در بخش سوم ارائه می‌شوند. در بخش چهارم پس از معرفی داده‌های ورودی و منطقه مورد مطالعه، تعداد خوشه‌های بهینه تعیین و روش‌های پیشنهادی بر روی داده‌های لرزه‌ای اعمال و نتایج حاصل در بخش پنجم نمایش داده می‌شوند. روش‌های پیشنهادی در بخش ششم مورد ارزیابی و مقایسه قرار می‌گیرند.

۲- مفاهیم نظری

۲-۱- خوشه‌بندی فازی

در خوشه‌بندی مجموعه‌ای از اشیاء به گروه‌های مجزایی به نام خوشه تقسیم می‌شوند؛ به طوری که اشیاء موجود در یک خوشه تا حد امکان به یکدیگر شبیه و با اشیاء سایر خوشه‌ها متفاوت باشد [۲]. این کار با کمینه (بیشینه) کردن برخی از معیارهای عددی انجام می‌پذیرد [۱۷]. یکی از انواع روش‌های خوشه‌بندی، الگوریتم‌های خوشه‌بندی فازی هستند که اغلب این الگوریتم‌ها از طریق کمینه کردن تابع هدف رابطه (۱) مجموعه اشیاء داده را به طور هم‌زمان به C خوشه مختلف نسبت می‌دهند [۱۸].

$$J_m = \sum_{j=1}^n \sum_{i=1}^c u_{ij}^M d_{ij}^2 \quad (1)$$

M میزان فازی بودن را کنترل و d_{ij} رابطه میان خوشه i ام و داده j ام است. u_{ij} درجه عضویت داده j ام به خوشه i ام است که از رابطه (۲) به دست می‌آید. این

در تکرار بعد تعیین می‌کند. در هر تکرار جستجو برای موقعیت بهینه از طریق به‌روزرسانی سرعت و موقعیت ذرات طبق روابط (۶) و (۷) انجام می‌شود. همچنین در هر تکرار مقدار تابع هدف موقعیت هر ذره نیز تعیین می‌شود. فرایند تکرار تا زمانی ادامه می‌یابد که تابع هدف به مقدار بهینه برسد و بهبود قابل‌ملاحظه‌ای در آن صورت نگیرد [۱۴]. سرعت هر ذره با استفاده از دو مقدار $pbest$ و $gbest$ به‌روز می‌شود که $pbest$ هر ذره بهترین موقعیتی است که آن ذره در آن قرار گرفته است و $gbest$ بهترین موقعیت در کل ذرات است.

$$V(t+1) = w \times V(t) + c_1 \times r_1(pbest - X(t)) + c_2 \times r_2(gbest - X(t)) \quad (۶)$$

$$X(t+1) = X(t) + V(t+1) \quad (۷)$$

که X و V به ترتیب موقعیت و سرعت ذره هستند. w وزن اینرسی است که میزان تأثیر سرعت مرحله قبل را در سرعت جدید کنترل می‌کند. c_1 و c_2 مقادیر ثابت مثبتی هستند که ضرایب شتاب نامیده می‌شوند و تأثیر $pbest$ و $gbest$ را در فرایند جستجو کنترل می‌کنند. r_1 و r_2 مقادیر تصادفی بین ۰ و ۱ می‌باشند.

برای حل مسئله خوشه‌بندی با استفاده از PSO، هر ذره به شکل رابطه (۸) مدل‌سازی می‌شود [۲۸]:

$$X_k = (a_{k1}, a_{k2}, \dots, a_{kc}) \quad (۸)$$

a_{kc} مرکز خوشه k ام در ذره k ام است. هدف PSO در خوشه‌بندی یافتن بهترین ذره X_k یا بهترین بخش‌بندی، از طریق بهینه (کمینه/بیشینه) کردن تابع هدف می‌باشد.

۳- روش‌های به‌کار گرفته شده

تأثیر روش خوشه‌بندی عمدتاً وابسته به تابع هدف بوده و کارایی آن، وابسته به الگوریتم جستجو می‌باشد [۱۵]. الگوریتم PSO به‌عنوان یکی از الگوریتم جستجو، دارای کارایی بالا در حل مسائل بهینه‌سازی غیرخطی بوده و به‌سادگی قابل پیاده‌سازی می‌باشد. علاوه بر این نسبت به سایر الگوریتم‌های بهینه‌سازی به پارامترهای ورودی کمتری نیاز دارد و از آنجایی که فرایند به‌روزرسانی بر اساس معادلات ساده انجام می‌پذیرد، این الگوریتم می‌تواند برای مجموعه داده‌های بزرگ مورداستفاده قرار گیرد [۱۴]. همان‌طور که در قسمت قبل ذکر شد الگوریتم PSO به یک تابع هدف نیاز

ρ_i ضریب ثابتی است و در صورتی که برابر با ۱ در نظر گرفته شود، خوشه‌ها دارای حجم یکسان خواهند شد. هدف GK نیز کمینه کردن تابع هدف f_m است.

اگرچه الگوریتم‌های سنتی خوشه‌بندی (مانند k-means و fuzzy c-means) به‌سادگی قابل فهم و پیاده‌سازی هستند، اما این الگوریتم‌ها از مشکل افتادن در دام بهینه محلی رنج می‌برند و اغلب قادر به شناسایی خوشه‌های سراسری نیستند [۲۴] همچنین اینکه نتایج نهایی این الگوریتم‌ها شدیداً به مقادیر و شرایط اولیه مسئله حساس می‌باشند [۱۶]. از آنجایی که مسائل خوشه‌بندی می‌توانند به‌عنوان مسائل بهینه‌سازی مطرح شوند، برای غلبه بر مشکلات روش‌های خوشه‌بندی، استفاده از الگوریتم‌های بهینه‌سازی پیشنهاد شده‌اند [۴].

الگوریتم‌های بهینه‌سازی دقیق، غالباً برای مسائل پیچیده زمان‌بر هستند. اما الگوریتم‌های بهینه‌سازی فرا ابتکاری که اغلب از طبیعت الهام گرفته شده‌اند به‌منظور حل مسائل پیچیده بهینه‌سازی طراحی شده‌اند [۲۵]. ویژگی اصلی این الگوریتم‌ها، توانایی آن‌ها در یافتن یک جواب تقریباً بهینه برای یک مسئله بهینه‌سازی در یک زمان کم است [۲۶]. بنابراین الگوریتم‌های بهینه‌سازی هوشمند می‌توانند دارای نقش برجسته‌ای در حل مسئله خوشه‌بندی باشند. الگوریتم‌های هوشمند متنوعی جهت انجام فرایند بهینه‌سازی ارائه شده که در ادامه یکی از کارآمدترین و رایج‌ترین این روش‌ها شرح داده می‌شود.

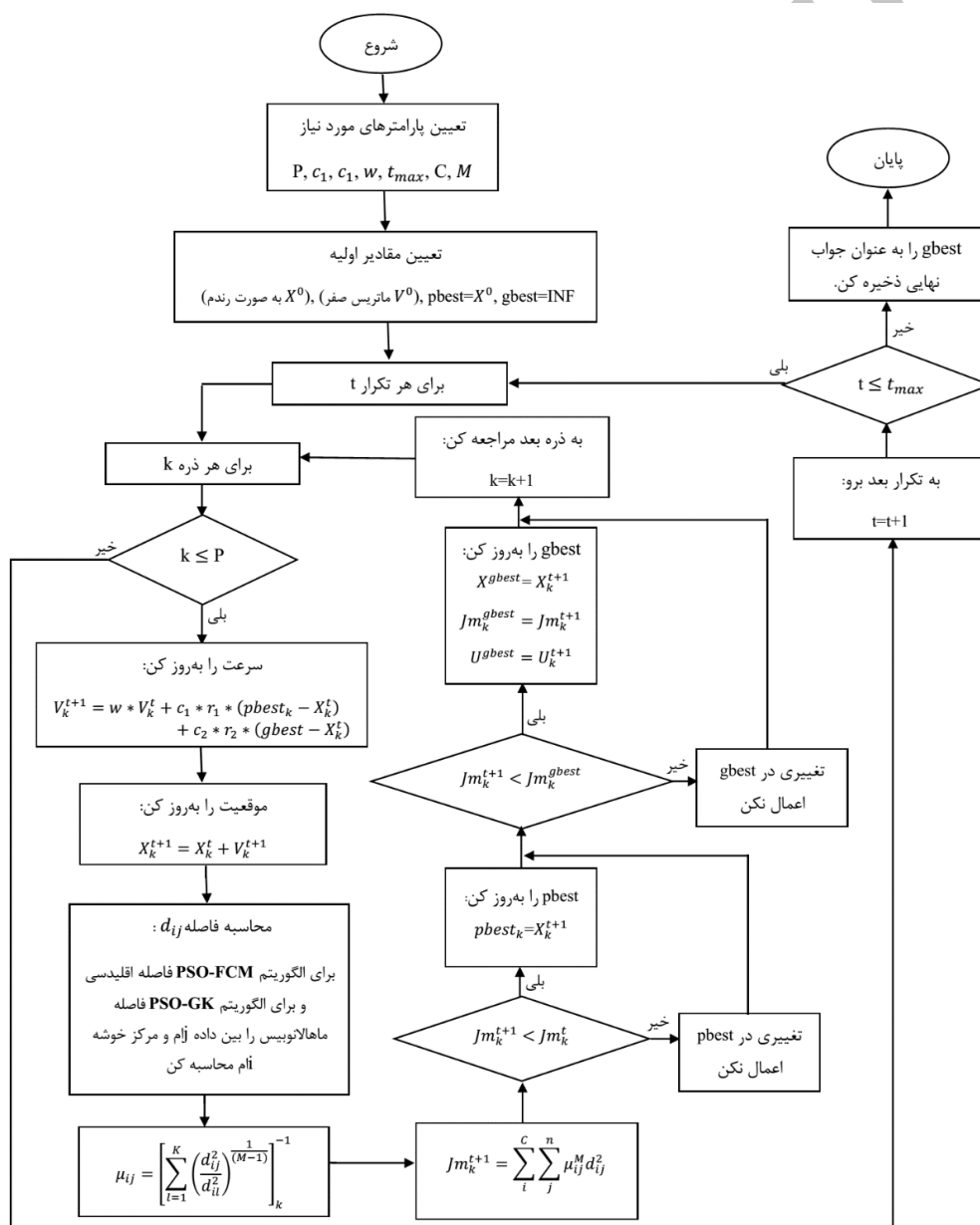
۲-۲- الگوریتم بهینه‌سازی توده ذرات^۱ (PSO)

الگوریتم PSO یکی از الگوریتم‌های بهینه‌سازی جمعیت مینا است که برای حل مسائل مختلف بهینه‌سازی یا مسائلی که دارای قابلیت تبدیل شدن به مسائل بهینه‌سازی می‌باشند، به‌سادگی قابل پیاده‌سازی است. این الگوریتم که در ابتدا توسط Kennedy و Eberhart [۲۷] ارائه شده، از حرکت دسته‌جمعی پرندگان و ماهی‌ها الهام گرفته شده است [۲۸] که از جمعیتی از ذرات اولیه که هریک دارای موقعیت‌های تصادفی می‌باشند، شروع می‌شود به‌طوری که موقعیت هر ذره می‌تواند جواب مسئله باشد. در این الگوریتم حداقل یک تابع هدف بهینه می‌شود. سرعت ذرات که در ابتدا به‌صورت تصادفی مقداردهی اولیه می‌شود، میزان جابه‌جایی آن ذره را

^۱ Particle Swarm Optimization

GK استفاده می‌شود و هم با استفاده از روش بهینه‌سازی هوشمند PSO، بهترین جواب حاصل خواهد شد. در نهایت هر داده به خوشه‌ای تعلق می‌گیرد که دارای بیشترین درجه عضویت باشد. تفاوت این دو الگوریتم در نوع فاصله مورد استفاده و در نتیجه در شکل و اندازه خوشه‌ها است. الگوریتم FCM به دلیل استفاده از فاصله اقلیدسی اشیاء داده تا مرکز خوشه‌ها، خوشه‌هایی را که ایجاد می‌کند کره‌ای شکل (دایره‌ای) ولی الگوریتم GK قادر است خوشه‌های بیضی‌شکل در جهات مختلف را شناسایی کند [۲۱]. در شکل ۱ فلوجارت مربوط به الگوریتم‌های PSO-FCM و PSO-GK آورده شده است.

دارد که در حین بهینه کردن آن به بهترین جواب برسد، از طرفی الگوریتم‌های خوشه‌بندی معرفی شده به دنبال کمینه کردن تابع J_m هستند که در رابطه (۱) آورده شده است. بنابراین می‌توان با معرفی تابع J_m مربوط به هریک از روش‌های FCM و GK به‌عنوان تابع هدف الگوریتم خوشه‌بندی PSO، آن را بهینه کرد و درعین حال خوشه‌بندی فازی را برای داده‌ها انجام داد. الگوریتم‌های بهبود یافته از این طریق به ترتیب PSO-FCM و PSO-GK نامیده می‌شوند. مزیت این کار این است که علاوه بر انجام خوشه‌بندی فازی به‌صورت بهینه، الگوریتم در جواب‌های بهینه محلی نیز گیر نمی‌کند. یعنی اینکه هم از مزیت هریک از روش‌های FCM و

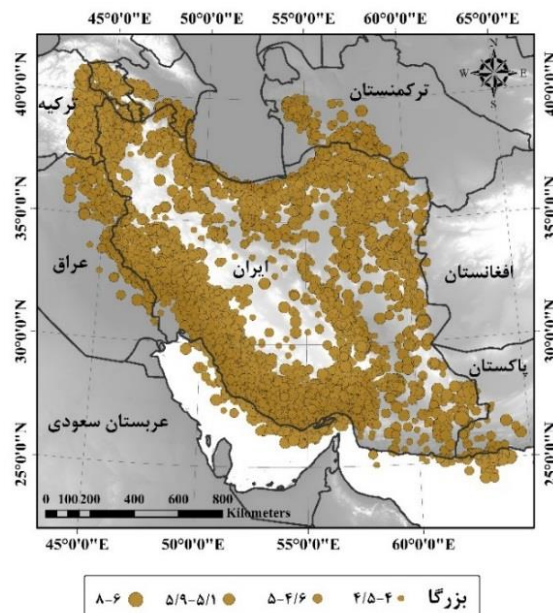


شکل ۱- فلوجارت الگوریتم‌های PSO-GK و PSO-FCM

۴- پیاده‌سازی

۴-۱- منطقه مورد مطالعه و داده‌های ورودی

فلات ایران از جمله مناطق لرزه‌خیز جهان است که توزیع مکانی و بزرگای رویدادها در همه‌جای آن یکسان نیست، بنابراین برای برآورد خطر لرزه‌ای، تقسیم آن به بخش‌های مختلف که هر یک دارای ویژگی‌های لرزه‌ای مشابهی باشند، ضروری است [۶]. هر رویداد لرزه‌ای نتیجه فعالیت پارامترهای مختلفی است؛ بنابراین اگر موقعیت داده‌های لرزه‌ای برای انجام فرایند خوشه‌بندی به کار روند، تأثیر همه پارامترهای مؤثر به‌طور صریح لحاظ می‌شود [۶]. در این مقاله موقعیت دوبعدی (طول و عرض جغرافیایی) رویدادهای لرزه‌ای سراسر ایران ورودی مسئله در نظر گرفته‌شده‌اند. کاتالوگ رویدادهای لرزه‌ای مورد استفاده توسط Shahvar و همکاران [۲۹] تهیه‌شده است که حاوی اطلاعات مربوط به رویدادهای دستگاهی دارای بزرگای بیش از ۴ در بازه زمانی سال‌های ۱۹۰۰ تا ۲۰۱۱ است که برای داده‌های لرزه‌ای تا ۲۰۱۶ به‌روز شده است. شکل ۲ توزیع رویدادها بر اساس بزرگا را نشان می‌دهد.



شکل ۲- توزیع رویدادهای لرزه‌ای ایران بر اساس بزرگا

۴-۲- پارامترهای ورودی الگوریتم

پارامترهای مورد نیاز الگوریتم عبارت‌اند از: وزن اینرسی w ، ضرایب شتاب c_1, c_2 ، تعداد ذرات (P) ، حداکثر تعداد

تکرارها (t_{max}) ، ضریب فازی M و تعداد خوشه‌ها (C) . Eberhart و Shi [۳۰] پیشنهاد کردند که ضرایب w, c_1 و c_2 را می‌توان به ترتیب با استفاده از روابط (۹)، (۱۰) و (۱۱) تعیین کرد:

$$w = \chi \quad (۹)$$

$$c_1 = \chi \varphi_1 \quad (۱۰)$$

$$c_2 = \chi \varphi_2 \quad (۱۱)$$

که φ_1 و φ_2 دو مقدار ثابت مثبت هستند به طوری که $\varphi = \varphi_1 + \varphi_2 \geq 4$ و χ با استفاده از رابطه (۱۲) به دست می‌آید:

$$\chi = \frac{2}{\varphi - 2 + \sqrt{\varphi^2 - 4\varphi}} \quad (۱۲)$$

اگر $\varphi_1 = \varphi_2 = 2.05$ در نظر گرفته شود، آنگاه ضریب χ برابر خواهد بود با 0.7298 با احتساب این مقادیر اغلب جواب‌های حاصل از الگوریتم بهترین مقدار را خواهند داشت [۳۱]. بخش‌بندی از طریق افزایش مقدار ضریب M ماهیت فازی تری را به خود می‌گیرد، و اخیراً هیچ تئوری اساسی برای تعیین مقدار بهینه M وجود ندارد [۳۲]. در این مقاله مانند بسیاری از مسائل عملی دیگر مقدار M برابر با ۲ در نظر گرفته می‌شود [۳۳]. هرگاه تعداد ذرات زیاد باشد و ذرات در فضای جستجو به‌طور یکنواخت توزیع شده باشند، تنوع در بین ذرات زیاد شده و الگوریتم راندمان بالاتری می‌یابد. البته باید توجه شود که تعداد زیاد ذرات در پیچیدگی الگوریتم ارتباط مستقیم دارند. هرچند که نسبت زمانی که تعداد ذرات کم است، تعداد تکرارهای الگوریتم کمتر است و زمان رسیدن به پاسخ بهینه کمتر است. تعداد ذرات و حداکثر تعداد تکرارهای الگوریتم پس از چندین بار اجرای الگوریتم با تعداد مختلف این مقادیر حاصل می‌شود. تعداد ذرات را برابر با ۷۰ ذره و حداکثر تعداد تکرارها ۳۰۰ تکرار جایی که تابع هدف همگرا می‌شود، تعیین شد.

۴-۳- تعیین تعداد خوشه‌های بهینه

یکی از مشکلاتی که مسئله خوشه‌بندی با آن مواجه است، تعیین تعداد بهینه خوشه‌ها است. یک راه‌حل برای این مشکل استفاده از یک شاخص اعتبار مناسب

$$XB = \frac{\sum_{j=1}^n \sum_{i=1}^c u_{ij}^M d_{ij}^2}{n \times d_{min}} = \frac{J_m}{n \times d_{min}} \quad (15)$$

d_{min} حداقل مقدار فاصله بین مراکز خوشه‌ها است که معیار جدایی آن‌ها است. تعداد خوشه‌هایی که به ازای آن شاخص XB مقدار کمتری را داشته باشند، به‌عنوان تعداد خوشه‌های مناسب برای آن مسأله استفاده می‌شود.

در این مقاله برای تعیین تعداد خوشه‌ها از ترکیبی از این سه شاخص استفاده می‌شود. از آنجایی که باید شاخص PC و XB کمینه و شاخص PE بیشینه شوند، رابطه شاخص ترکیبی به‌صورت رابطه (۱۶) می‌شود که کمترین مقدار آن نشان‌دهنده تعداد خوشه بهینه‌تری می‌باشد:

$$V = \frac{XB \times PC}{PE} \quad (16)$$

هر یک از الگوریتم‌ها به‌صورت تکراری برای تعداد ۲ تا ۳۰ خوشه اجرا و شاخص ترکیبی برای هر تعداد خوشه محاسبه گردید که نتایج حاصل در جدول ۱ آورده شده‌اند. برای الگوریتم PSO-FCM حداقل مقدار شاخص V برای ۲۲ تعداد خوشه و برای الگوریتم PSO-GK حداقل مقدار شاخص V برای ۲۳ تعداد خوشه حاصل شد.

۴-۴- هندسه خوشه‌ها

در روش‌های خوشه‌بندی که از فاصله اقلیدسی استفاده می‌کنند، خوشه‌های حاصل دارای اشکال متمایل به شکل هندسی دایره‌ای یا کروی مانند خواهند بود. از این‌رو الگوریتم FCM به دلیل استفاده از فاصله اقلیدسی بین داده‌ها تا مرکز خوشه‌ها، خوشه‌هایی را که ایجاد می‌کند کره‌ای شکل (دایره‌ای) هستند، ولی الگوریتم GK قادر است خوشه‌های بیضی‌شکل در جهات مختلف را نیز شناسایی کند [۲۱]. از طریق مقادیر و بردارهای ویژه ماتریس کوواریانس هر خوشه می‌توان اطلاعاتی را در مورد شکل و جهت خوشه به دست آورد. در واقع بزرگ‌ترین مقدار ویژه اطلاعاتی را در مورد طول خوشه و بردار ویژه متناظر با آن راستای آن را در اختیار می‌گذارد. همچنین دومین مقدار ویژه و بردار ویژه متناظر با آن به ترتیب نماینده عرض و راستای آن می‌باشند [۸]. از آنجاکه خوشه‌های حاصل از الگوریتم GK بیضی‌شکل و خوشه‌های حاصل از FCM دایره‌ای شکل هستند با

خوشه‌بندی است [۶]. به‌طور کلی شاخص اعتبار خوشه‌بندی دو هدف را دنبال می‌کند، اول اینکه، می‌تواند برای تعیین تعداد خوشه‌ها مورد استفاده قرار گیرد و دوم اینکه، می‌تواند بهترین بخش‌بندی را شناسایی کند. با اجرای الگوریتم به‌صورت تکراری برای تعداد مختلف خوشه و سپس انتخاب بخش‌بندی دارای بهترین مقدار برای شاخص اعتبار می‌توان تعداد بهینه خوشه‌ها را تعیین کرد [۳۴].

تعداد خوشه‌ای مناسب است که اولاً نمونه‌های موجود در یک خوشه تا حد امکان شبیه به یکدیگر باشند و ثانیاً نمونه‌های متعلق به خوشه‌های متفاوت تا حد امکان با یکدیگر نامشابه باشند. به عبارات دیگر، خوشه‌ها باید دارای بیشینه فشردگی^۱ باشند و تا حد امکان جدایی^۲ آن‌ها نیز زیاد باشد. شاخص‌های بسیاری برای اعتبار سنجی بخش‌بندی‌ها ارائه شده‌اند که بیشتر این شاخص‌ها در روش‌های اعتبار سنجی اغلب ویژگی‌های فشردگی و جدایی خوشه‌ها را در نظر می‌گیرند و تنها وابسته به داده‌ها هستند [۳۴]. در رابطه مربوط به شاخص‌های ارائه شده برای الگوریتم‌های خوشه‌بندی فازی، علاوه بر پارامترهای فشردگی و جدایی، میزان درجه عضویت داده‌ها به هر یک از خوشه‌ها نیز در نظر گرفته می‌شود. ضریب بخش‌بندی (PC)^۳ در رابطه (۱۳)، آنتروپی بخش‌بندی (PE)^۴ در رابطه (۱۴) از جمله پرکاربردترین این شاخص‌ها هستند [۳۵]. بیشترین مقدار برای PC و کمترین مقدار برای PE بخش‌بندی خوب را نشان می‌دهد.

$$PC = \frac{1}{n} \sum_{i=1}^c \sum_{j=1}^n u_{ij}^2 \quad (13)$$

$$PE = -\frac{1}{n} \sum_{i=1}^c \sum_{j=1}^n u_{ij} \log(u_{ij}) \quad (14)$$

شاخص Xie-Beni (XB) [۳۶] نسبت وسعت در خوشه، J_m به حداقل میزان جدایی بین خوشه‌ها را تعیین می‌کند که با این کار علاوه بر میزان درجه عضویت، فشردگی درون هر خوشه و جدایی میان خوشه‌ها را نیز در نظر می‌گیرد (رابطه ۱۵).

^۱ Compactness

^۲ Separation

^۳ Partition Coefficient

^۴ Partition Entropy

خوشه مشخص نمود زیرا دایره را می‌توان نوعی بیضی محسوب کرد.

استفاده از مقادیر و بردارهای ویژه یک خوشه می‌توان جهت‌گیری اعضای آن را از طریق برازش یک بیضی به

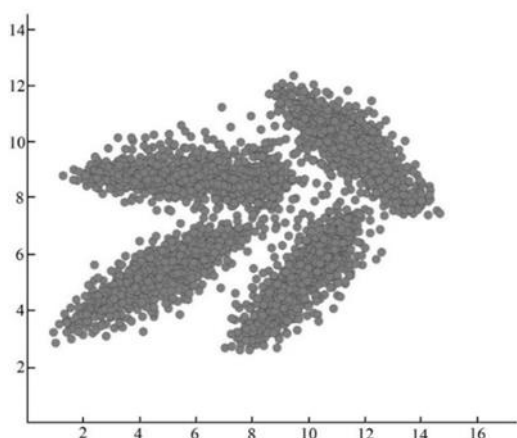
جدول ۱- مقادیر شاخص ترکیبی حاصل از دو الگوریتم PSO-GK و PSO-FCM برای تعداد مختلف ۲ تا ۳۰ خوشه

تعداد	PSO-FCM	PSO-GK	تعداد	PSO-FCM	PSO-GK	تعداد	PSO-FCM	PSO-GK
۲	۰/۳۱۰۷	۰/۹۷۴۶	۱۲	۰/۰۸۳۵	۰/۱۵۲۹	۲۲	۰/۰۷۷۰	۰/۱۴۰۵
۳	۰/۱۸۵۹	۰/۳۹۸۶	۱۳	۰/۱۱۰۹	۰/۷۹۵۳	۲۳	۰/۰۷۸۰	۰/۱۰۷۸
۴	۰/۱۱۶۷	۰/۱۴۸۴	۱۴	۰/۱۱۴۶	۰/۴۴۳۰	۲۴	۰/۱۱۱۷	۰/۲۴۴۹
۵	۰/۱۳۸۶	۰/۳۸۸۴	۱۵	۰/۱۰۲۴	۰/۲۹۶۳	۲۵	۰/۱۳۰۹	۰/۱۴۵۹
۶	۰/۱۷۳۷	۰/۱۲۳۸	۱۶	۰/۰۹۳۱	۰/۲۶۶۸	۲۶	۰/۰۹۹۵	۰/۲۰۹۵
۷	۰/۱۶۲۴	۰/۲۵۲۱	۱۷	۰/۱۱۲۹	۰/۲۰۰۱	۲۷	۰/۰۸۱۵	۰/۲۶۶۸
۸	۰/۱۳۳۲	۰/۱۸۰۷	۱۸	۰/۱۰۹۶	۰/۲۲۸۶	۲۸	۰/۱۱۵۰	۰/۲۹۶۹
۹	۰/۱۱۴۸	۰/۱۴۳۰	۱۹	۰/۰۹۳۰	۰/۲۷۸۴	۲۹	۰/۱۴۹۱	۰/۱۳۷۵
۱۰	۰/۱۱۵۰	۰/۵۸۷۵	۲۰	۰/۰۹۵۵	۰/۲۲۴۱	۳۰	۰/۱۱۶۹	۰/۱۵۳۴
۱۱	۰/۰۹۲۳	۰/۲۴۴۲	۲۱	۰/۰۸۶۳	۰/۳۷۳۶			

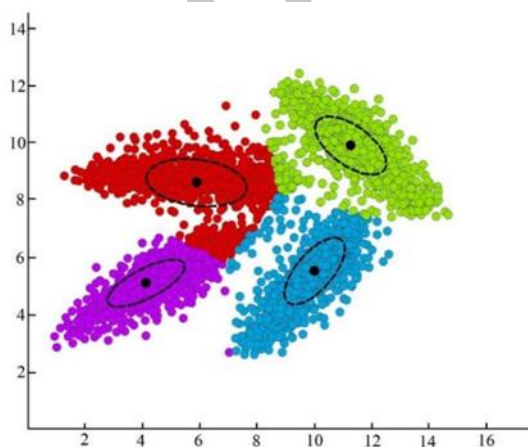
۵- نتایج حاصل از پیاده‌سازی

۵-۱- نتایج برای داده‌های ساختگی

به‌منظور اینکه قابلیت و کارایی هر الگوریتم برای انجام خوشه‌بندی داده‌های لرزه‌ای با وضوح بیشتری نشان داده شود، از داده‌های ساختگی استفاده می‌شود. از آنجایی که رویدادهای لرزه‌ای به‌طور عمده در بخش‌های مختلف گسل‌هایی که خطی شکل‌اند توزیع شده‌اند، داده‌های ساختگی به‌صورت تصادفی در راستای چند خط فرضی توزیع می‌شوند که در شکل ۳ نشان داده شده‌اند. الگوریتم‌های ارائه شده بر روی این داده‌ها اعمال می‌شود و پس از انجام خوشه‌بندی با استفاده از مقادیر و بردارهای ویژه، به هریک از خوشه‌های حاصل یک بیضی برازش داده خواهد شد. همان‌طور که در شکل ۵ نشان داده شده است، الگوریتم PSO-GK به‌خوبی توانسته است داده‌های مربوط به هر خوشه را شناسایی کرده و به هر خوشه یک بیضی در راستای صحیح برازش دهد. این در صورتی است از اجرای الگوریتم PSO-FCM خوشه‌های صحیحی حاصل نشده است (شکل ۴).

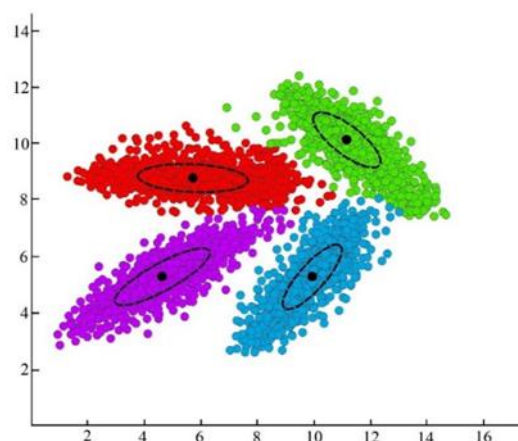


شکل ۳- داده‌های ساختگی توزیع شده تصادفی اطراف خطوط فرضی



شکل ۴- نتیجه الگوریتم PSO-FCM برای داده‌های لرزه‌ای ساختگی

u_{ij} میزان درجه عضویت نهایی داده زام به خوشه i ام و C تعداد خوشه‌ها می‌باشد. F_i ماتریس کوواریانس فازی خوشه‌های حاصل است که از رابطه (۷) به دست می‌آید. در جدول ۲ مقدار هر یک از این معیارها را برای هر الگوریتم PSO-GK و PSO-FCM مربوط به داده‌های ساختگی و داده‌های واقعی آورده شده است. همان‌طور که در جدول ۲ مشاهده می‌شود، برای داده‌های ساختگی معیار ارزیابی FHV برای الگوریتم PSO-GK کمتر از مقدار این معیار برای الگوریتم PSO-FCM است همچنین دو معیار APD و PD نیز برای الگوریتم PSO-GK بیشتر از مقدار این دو معیار برای الگوریتم PSO-FCM هستند که نشان از بهتر بودن الگوریتم PSO-GK برای داده‌های ساختگی است. برای داده‌های لرزه‌ای ایران نیز معیار ارزیابی FHV برای الگوریتم PSO-GK کمتر از مقدار این معیار برای الگوریتم PSO-FCM شده است و دو معیار APD و PD نیز برای الگوریتم PSO-GK بیشتر از مقدار این دو معیار برای الگوریتم PSO-FCM هستند که نشان از بهتر بودن الگوریتم PSO-GK برای داده‌های لرزه‌ای واقعی است. بنابراین از آنجایی که مقادیر معیارهای ارزیابی الگوریتم PSO-GK هم برای داده‌های ساختگی و هم برای داده‌های واقعی بهتر از الگوریتم PSO-FCM می‌باشند، لذا الگوریتم PSO-GK بخش‌بندی بهتری را ایجاد کرده است. از سوی دیگر، اگر گسل‌های منطقه در نظر گرفته شود، انتظار می‌رود که خوشه‌ها در موقعیت گسل‌ها مخصوصاً مناطق با چگالی بیشتر گسل‌ها حضور داشته باشند به طوری که جهت‌گیری آن‌ها نیز با راستای گسل‌ها متناسب باشد. از مقایسه موقعیت و جهت‌گیری بیضی‌های برازش داده‌شده به هر خوشه با نقشه گسل‌های منطقه در شکل ۶ و ۷ مشاهده می‌شود که خوشه‌های الگوریتم PSO-GK علاوه بر اینکه در مناطق با چگالی بالای گسل‌ها قرار دارند، جهت‌گیری مناسبی را نیز اتخاذ کرده‌اند، در حالی که خوشه‌های الگوریتم PSO-FCM در موقعیت‌ها و راستاهای کاملاً صحیحی جایابی نشده‌اند که به دلیل تفاوت در نوع فاصله و به دنبال آن شکل خوشه‌های حاصل از دو الگوریتم است.



شکل ۵- نتیجه الگوریتم PSO-GK برای داده‌های لرزه‌ای ساختگی

۵-۲- نتایج برای داده‌های لرزه‌ای ایران

هر یک از الگوریتم‌های PSO-GK و PSO-FCM بر روی رویدادهای لرزه‌ای ایران اعمال گردیده و نقشه گسل‌های منطقه به همراه نتیجه الگوریتم PSO-FCM در شکل ۶ و نتیجه حاصل از PSO-GK در شکل ۷ نشان داده است.

۶- ارزیابی

به منظور ارزیابی خوشه‌های به دست آمده از سه معیار ارزیابی فازی fuzzy hyper volume (FHV)، متوسط چگالی بخش‌بندی^۱ (APD) و چگالی بخش‌بندی^۲ (PD) [۳۷] استفاده گردید. این معیارها میزان جدایی خوشه‌ها، فشردگی آن‌ها و تراکم داده‌ها را نزدیک مراکز خوشه نشان می‌دهند به طوری که کم‌ترین مقدار برای FHV و بیشترین مقدار برای APD و PD بخش‌بندی خوب را نشان می‌دهد [۳۸]. هر یک از این سه معیار در روابط (۱۷)، (۱۸) و (۱۹) ذکر شده‌اند.

$$FHV = \sum_{i=1}^C [\det(F_i)]^{0.5} \quad (17)$$

$$APD = \frac{1}{C} \sum_{i=1}^C \frac{\sum_{j=1}^n u_{ij}}{[\det(F_i)]^{0.5}} \quad (18)$$

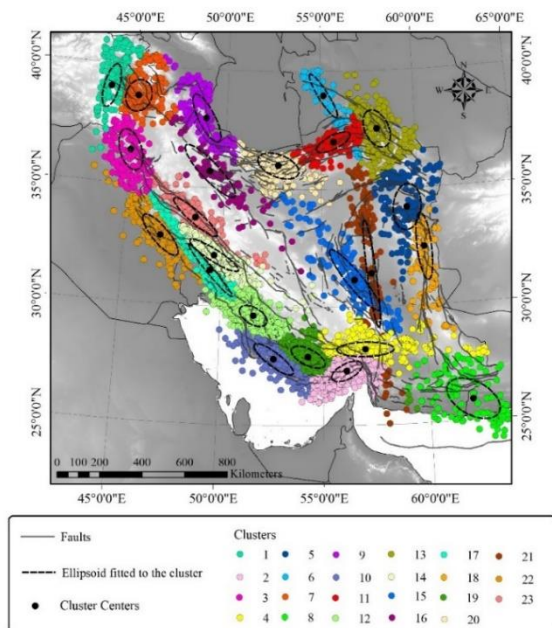
$$PD = \frac{\sum_{i=1}^C \sum_{j=1}^n u_{ij}}{\sum_{i=1}^C [\det(F_i)]^{0.5}} \quad (19)$$

^۱ Average Partition Density

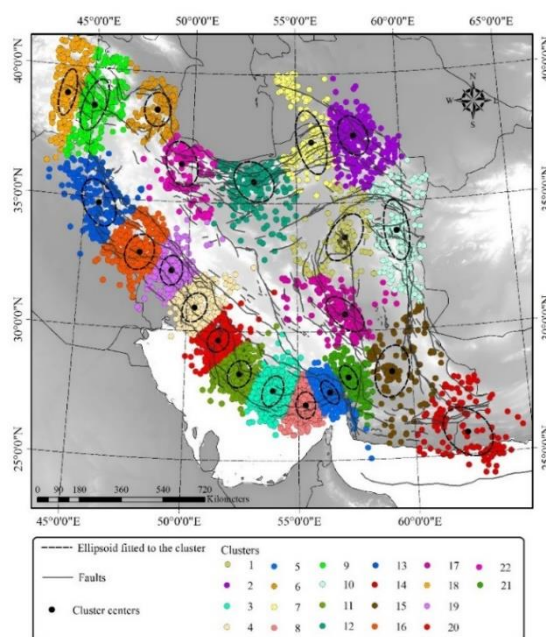
^۲ Partition Density

جدول ۲- مقدار هریک از معیارهای ارزیابی خوشه‌بندی برای هر الگوریتم PSO-FCM و PSO-GK مربوط به داده‌های ساختگی و داده‌های واقعی

	داده‌های ساختگی		داده‌های واقعی ایران	
	PSO-FCM	PSO-GK	PSO-FCM	PSO-GK
FHV	۵/۸۳۷۷	۵/۴۱۰۵	۲۱/۴۱۸۲	۲۱/۳۲۴۱
APD	۵۵۷/۱۱۵۹	۵۹۹/۱۴۳۶	۳۴۵/۰۹۶۷	۳۶۹/۴۶۶۳
PD	۵۵۲/۹۵۹۴	۵۹۶/۶۱۹۵	۳۳۵/۱۸۲۶	۳۶۴/۷۳۵۱



شکل ۷- نقشه گسل‌های منطقه به همراه نتیجه حاصل از الگوریتم PSO-GK



شکل ۶- نقشه گسل‌های منطقه به همراه نتیجه حاصل از الگوریتم PSO-FCM

۷- نتیجه‌گیری

بخش‌بندی بهتری را ایجاد کرده است. از طرف دیگر وقتی که نتایج حاصل از دو الگوریتم با گسل‌های منطقه مقایسه گردید، مشاهده شد که بیشتر خوشه‌های حاصل از الگوریتم PSO-GK موقعیت و راستای صحیحی را اتخاذ کرده‌اند. بنابراین عملکرد الگوریتم فازی PSO-GK عملکرد بهتری را در فرایند تحلیل داده‌های لرزه‌ای از خود نشان داده است که دلیل این امر می‌تواند ناشی از نحوه توزیع داده‌های لرزه‌ای در پیرامون گسل‌ها باشد که شکل بیضی بهتر می‌تواند این توزیع را به نمایش بگذارد. نهایتاً می‌توان نتیجه گرفت که در صورتی با استفاده از الگوریتم‌های بهینه‌سازی هوشمند جواب‌های درستی حاصل خواهد شد که به ماهیت داده‌ها و شکل خوشه‌های حاصل از الگوریتم مورد استفاده به عبارت دیگر به تابع هدف انتخابی نیز توجه شود.

اصولاً تحلیل حجم بالای داده‌های لرزه‌ای کاری دشوار و همراه با عدم قطعیت‌های بسیار می‌باشد. از این رو استفاده از روش‌های هوشمند جهت خوشه‌بندی، به عنوان یک روش داده‌کاوی، ابزاری مناسب برای دستیابی به تصمیمات بهتر است. نکته قابل توجه در کاربرد روش‌های هوشمند خوشه‌بندی در تمایز داده‌های لرزه‌ای، نزدیک شدن نتایج مطالعات گوناگون در یک محدوده یکسان می‌باشد. از این رو در مطالعه حاضر، با استفاده از الگوریتم بهینه‌سازی توده ذرات دو الگوریتم فازی PSO-FCM و PSO-GK برای خوشه‌بندی از طریق کمینه کردن تابع هدف الگوریتم‌های FCM و GK، پیشنهاد و مورداستفاده قرار گرفت. این دو الگوریتم بر روی داده‌های ساختگی و داده‌های سراسر ایران اعمال شد. مقادیر معیارهای ارزیابی برای دو الگوریتم نشان داد که الگوریتم PSO-GK

- [1] M. Deighton and M. Petrou, (2009) "Data mining for large scale 3D seismic data analysis," (in English), *Machine Vision and Applications*, vol. 20, no. 1, pp. 11-22.
- [2] J. Han, M. Kamber, and J. Pei, (2011) *Data mining: concepts and techniques: concepts and techniques*. Elsevier.
- [3] J. MacQueen, (1967) "Some methods for classification and analysis of multivariate observations," *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, no. 14, pp. 281-297.
- [4] S. Das ,A. Abraham, and A. Konar, (2009) "Metaheuristic Pattern Clustering – An Overview," in *Metaheuristic Clustering*, vol. 178, pp. 1-62.
- [5] W. Wang and Y. Zhang, (2007) "On fuzzy cluster validity indices," *Fuzzy Sets and Systems*, vol. 158, no. 19, pp. 2095-2117.
- [6] A. Ansari, A. Noorzad, and H. Zafarani, (2009) "Clustering analysis of the seismic catalog of Iran," *Computers & Geosciences*, vol. 35, no. 3, pp. 475-486.
- [7] F. Aminzadeh and S. Chatterjee, (1984) " Applications of clustering in exploration seismology," *Geoexploration*, vol. 23, no. 1, pp. 147-159.
- [8] G. Ouillon, C. Ducorbier, and D. Sornette, (2008) "Automatic reconstruction of fault networks from seismicity catalogs: Three-dimensional optimal anisotropic dynamic clustering," *Journal of Geophysical Research: Solid Earth*, vol. 113, no. B1.
- [9] G. Weatherill and P. W. Burton, (2009) "Delineation of shallow seismic source zones using K-means cluster analysis, with application to the Aegean region," *Geophysical Journal International*, vol. 176, no. 2, pp. 565-588.
- [10] K. Rehman, P. Burton, and G. Weatherill, (2014) "K-means cluster analysis and seismicity partitioning for Pakistan," *Journal of Seismology*, vol. 18, no. 3 ,pp. 401-419.
- [11] S. J. Nanda and G. Panda, (2015) "Design of computationally efficient density-based clustering algorithms," *Data & Knowledge Engineering*, vol. 95, pp. 23-38.
- [12] S. N. Hashemi and R. Mehdizadeh, (2015) "Application of hierarchical clustering technique for numerical tectonic regionalization of the Zagros region (Iran)," *Earth Science Informatics*, journal article vol. 8, no. 2, pp. 367-380.
- [13] O. Zhongbin, H. Jing, and Z. Keliang, (2008) "Data Mining for Seismic Exploration," in *Web Intelligence and Intelligent Agent Technology. WI-IAT IEEE/WIC/ACM International Conference on*, vol. 3, pp. 424-427.
- [14] M. Zhao, H. Tang, J. Guo, and Y. Sun, (2014) "Data Clustering Using Particle Swarm Optimization," in *Future Information Technology*, vol. 309, pp. 607-612.
- [15] D. Binu, (2015) "Cluster analysis using optimization algorithms with newly designed objective functions," *Expert Systems with Applications*, vol. 42, no. 14, pp. 5848-5859.
- [16] C.-W. Tsai, W.-C. Huang, and M.-C. Chiang, (2014) "Recent Development of Metaheuristics for Clustering," in *Mobile, Ubiquitous, and Intelligent Computing*, vol. 274(Lecture Notes in Electrical Engineering: Springer Berlin Heidelberg, pp. 629-636.
- [17] B. S. Everitt, S. Landau, M. Leese, and D. Stahl, (2011) "Optimization Clustering Techniques," in *Cluster Analysis: John Wiley & Sons, Ltd*, pp. 111-142.
- [18] H. Izakian, A. Abraham, and V. Snasel, (2009) "Fuzzy clustering using hybrid fuzzy c-means and fuzzy particle swarm optimization," in *Nature & Biologically Inspired Computing. NaBIC. World Congress on*, pp. 1690-1694.
- [19] J. C. Dunn, (1973) "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters," *Journal of Cybernetics*, vol. 3, no. 3, pp. 32-57.
- [20] J. C. Bezdek, (2013) *Pattern Recognition with Fuzzy Objective Function Algorithms*. Springer US.
- [21] G. Grekousis and H. Thomas, (2012) "Comparison of two fuzzy algorithms in geodemographic segmentation analysis: The Fuzzy C-Means and Gustafson–Kessel methods," *Applied Geography*, vol. 34, pp. 125-136.
- [22] D. Dovžan and I. Škrjanc, (2011) "Recursive clustering based on a Gustafson–Kessel algorithm," *Evolving Systems*, journal article vol. 2, no. 1, pp. 15-24.
- [23] P. Angelov, D. P. Filev, and N. Kasabov, (2010) *Evolving Intelligent Systems: Methodology and Applications*. Wiley.

- [24] P. Manikandan and S. Selvarajan, (2014) "Data Clustering Using Cuckoo Search Algorithm (CSA)," in Proceedings of the Second International Conference on Soft Computing for Problem Solving (SocProS), : Springer, pp. 1275-1283.
- [25] L. Bianchi, M. Dorigo, L. M. Gambardella, and W. J. Gutjahr, (2009) "A survey on metaheuristics for stochastic combinatorial optimization," vol. 8, no. 2, pp. 239-287.
- [26] C. Blum and A. Roli, (2003) "Metaheuristics in combinatorial optimization: Overview and conceptual comparison," ACM Comput. Surv., vol. 35, no. 3, pp. 268-308.
- [27] R. C. Eberhart and J. Kennedy, (1995) "A new optimizer using particle swarm theory," Proceedings of the sixth international symposium on micro machine and human science, vol. 1, pp. 39-43.
- [28] K. Manda, A. S. Hanuman, S. Satapathy, V. Chaganti, and A. V. Babu, (2010) "A Software Tool for Data Clustering Using Particle Swarm Optimization," in Swarm, Evolutionary, and Memetic Computing, vol. 6466, pp. 278-285.
- [29] M. P. Shahvar, M. Zare, and S. Castellaro, (2013) "A Unified Seismic Catalog for the Iranian Plateau (1900–2011) ", Seismological Research Letters, vol. 84, no. 2, pp. 233-249.
- [30] R. C. Eberhart and Y. Shi, (2000) "Comparing inertia weights and constriction factors in particle swarm optimization," in Evolutionary Computation, vol. 1, pp. 84-88 vol.1.
- [31] A. Khare and S. Rangnekar, (2013) "A review of particle swarm optimization and its applications in Solar Photovoltaic system," Applied Soft Computing, vol. 13, no. 5, pp. 2997-3006.
- [32] P. Gopala Krishna and D. Lalitha Bhaskari, (2016) "Fuzzy C-Means and Fuzzy TLBO for Fuzzy Clustering," in Proceedings of the Second International Conference on Computer and Communication Technologies: IC3T, Volume 1 New Delhi: Springer India, pp. 479-486.
- [33] H. Gu, S. Ren, F. Si, and Z. Xu, (2016) "Evolved FCM framework for working condition classification in furnace system," Soft Computing, journal article pp. 1-13.
- [34] N. Bharill and A. Tiwari, (2014) "Enhanced cluster validity index for the evaluation of optimal number of clusters for Fuzzy C-Means algorithm," in IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1526-1533.
- [35] M. K. Goyal and V. Gupta, (2014) "Identification of Homogeneous Rainfall Regimes in Northeast Region of India using Fuzzy Cluster Analysis," Water Resources Management, journal article vol. 28 ,no. 13, pp. 4491-4511.
- [36] M. Muranishi, K. Honda, and A. Notsu, (2014) "Application of xie-beni-type validity index to fuzzy co-clustering models based on cluster aggregation and pseudo-cluster-center estimation," in 14th International Conference on Intelligent Systems Design and Applications, pp. 34-38.
- [37] M. Halkidi, Y. Batistakis, and M. Vazirgiannis, (2002) "Clustering validity checking methods: part II," SIGMOD Rec., vol. 31, no. 3, pp. 19-27.
- [38] L. Vendramin, M. C. Naldi, and R. J. G. B. Campello, (2015) "Fuzzy Clustering Algorithms and Validity Indices for Distributed Data," in Partitional Clustering Algorithms Cham: Springer International Publishing, pp. 147-192.