

پهنه‌بندی کانسار بر اساس توزیع فضایی عیار کانسنگ با استفاده از الگوریتم خوشه‌بندی نقشه خود سازمانده در معدن چغارت

امین حسین مرشدی*

۱- استادیار دانشکده مهندسی معدن و متالورژی، دانشگاه یزد، morshedy@yazd.ac.ir

(دریافت ۲۳ تیر ۱۳۹۴، پذیرش ۲۴ آذر ۱۳۹۴)

چکیده

پهنه‌بندی کانسار، یکی از مسائل مهم در زمینه مدل‌سازی، ارزیابی و برنامه‌ریزی استخراج در امور معدنی است. در مدل‌سازی‌های معدنی، منطقه بر اساس ویژگی‌های فیزیکی مؤثر بر کانی‌زایی یا توزیع فضایی عیار به پهنه‌های مختلف تقسیم می‌شود. در این مقاله، از روش خوشه‌بندی نقشه خودسازمانده (SOM) به منظور پهنه‌بندی سه بعدی کانسار معدنی استفاده شده است و شاخص‌های اعتبارسنجی برای تعیین تعداد بهینه پهنه‌ها به کار برده شده است. به منظور اعتبارسنجی الگوریتم پیشنهادی، داده‌ها و اطلاعات معدن سنگ آهن چغارت به کار برده شده است. شاخص‌های اعتبارسنجی خوشه‌بندی روی داده‌های عیارسنجی آهن و فسفر گمانه‌های اکتشافی اجرا شد و در نتیجه تعداد دو پهنه به صورت بهینه مشخص شد. بر اساس نتایج، محدوده دو پهنه و نحوه اختصاص نمونه به هر پهنه با استفاده از الگوریتم خوشه‌بندی نقشه خودسازمانده تعیین شد. در نتیجه، خروجی الگوریتم‌های SOM و K-means بررسی شد که در حدود ۹۰٪ داده‌ها به صورت مشابه به خوشه‌های یکسان اختصاص یافته است. در روش SOM، سطح جدایش دو پهنه در راستای شمال شرقی-جنوب غربی کشیده شده که به سمت جنوب شرقی شیب دارد، در حالی که این سطح جدایش در روش K-means راستای شرقی-غربی و شیب به سمت جنوب دارد. بر اساس سطوح جدایش به دست آمده از دو روش خوشه‌بندی و مقایسه با خصوصیات بعدی-جهتی و ویژگی‌های ساختمانی مؤثر در کنترل کانی‌زایی (به ویژه گسل‌ها)، سطح جدایش SOM از نظر ساختار فضایی و جهتی، هم خوانی قابل توجهی با ویژگی‌های ساختمانی منطقه دارد.

کلمات کلیدی

پهنه‌بندی، خوشه‌بندی، شاخص اعتبارسنجی خوشه‌بندی، نقشه خودسازمانده، سنگ آهن چغارت

۱- مقدمه

در مطالعات علوم زمین، استفاده از پهنبندی می‌تواند به‌طور قابل‌توجهی موجب بهبود دقت و مدیریت عدم قطعیت مدل‌سازی شود. تغییر فضایی عمده در یک منطقه، می‌تواند موجب نقض فرض پایایی و برآورد اریب شود [۱،۲]. بر اساس میزان انحراف از فرضیه ایستایی، دو نوع رویکرد منطقه واحد و پهنبندی قابل بررسی است. رویکرد اول، برای ساده سازی مسئله، میانگینی از رفتار پهنبه‌های مختلف به‌صورت یک منطقه واحد در نظر گرفته که در مواردی باعث هموارسازی کاذب و خطای قابل‌توجه در عملیات مدل‌سازی می‌شود. در رویکرد پهنبندی، تقسیم منطقه به تعداد پهنبه‌های بهینه پیشنهاد می‌شود که ممکن است ناشی از عواملی مثل ناپیوستگی، گسل، تغییر سازند یا توزیع عیاری باشد.

روش‌های خوشه‌بندی، به‌عنوان یکی از روش طبقه‌بندی الگوی بدون ناظر، بیانگر هنر یافتن گروه‌ها یا پهنبه‌های مناسب در مجموعه داده‌ها است [۳]. مسائل مربوط به تجزیه و تحلیل خوشه‌ای در بسیاری از رشته‌ها از جمله زیست‌شناسی، پزشکی، روان‌شناسی، محیط زیست، بازاریابی، پردازش تصویر، باستان‌شناسی و غیره به‌کار می‌رود [۴]. در روش‌های خوشه‌بندی، از شاخص‌های اعتبارسنجی خوشه‌بندی برای یافتن تعداد بهینه خوشه‌ها که مبین تعداد پهنبه‌ها در یک منطقه است و همچنین به‌منظور ارزیابی کیفیت خوشه‌بندی استفاده می‌شود [۵]. الگوریتم نقشه خودسازمانده (Self-Organizing Map) ابزار مناسبی به‌منظور خوشه‌بندی داده‌ها است که بر اساس یک فرآیند رگرسیون بازگشتی ناپارامتری، مجموعه مشخصی از بردارهای مدل را به فضای بردارهای قابل مشاهده در قالب الگوریتمی تبدیل می‌کند. خوشه‌ها در یک فرآیند یادگیری رقابتی، نسبت به پارامترهای ورودی منظم می‌شوند. محل خوشه‌های تنظیم شده در شبکه به‌گونه‌ای نظم می‌یابد که برای پارامترهای ورودی، یک دستگاه مختصات معنی‌دار روی شبکه ایجاد شود [۶،۷].

در این مقاله، پهنبندی توسط الگوریتم خوشه‌بندی SOM بررسی شده و داده‌های مختلف عیاری به‌طور موردی در معدن سنگ آهن چغارت استفاده شده است. برای مدل‌سازی دقیق‌تر، تعداد بهینه پهنبه‌ها براساس شاخص‌های اعتبارسنجی خوشه‌بندی روی مجموعه داده‌ها مشخص می‌شود. براساس

نتایج خوشه‌بندی، محدوده پهنبه‌ها و سطح جدایش بین پهنبه‌ها مشخص می‌شود که بیان‌گر مرز بین دو پهنبه است. در نهایت، نتایج با روش خوشه‌بندی رایج K-means بررسی و مقایسه می‌شود.

۲- منطقه مورد مطالعه

نتایج فعالیت‌های اکتشافی از سال ۱۳۴۰ تاکنون در ایران، منجر به اکتشاف بلوک معدنی بافق با ذخیره بیش از یک سوم سنگ آهن کشور، به‌عنوان مهم‌ترین پهنبه آهن‌دار ایران شناخته شده است. در این منطقه، بیش از ۳۸ آنومالی آهن‌دار با ذخیره نزدیک به ۱/۷ میلیارد تن شناسایی شده که یکی از مهم‌ترین کناسرهای این منطقه، کناسر آهن چغارت است. معدن سنگ آهن چغارت در ۱۲ کیلومتری شمال‌شرقی شهر بافق، در ۱۲۵ کیلومتری جنوب‌شرقی شهر یزد، در ۷۵ کیلومتری جنوب غربی شهر بهاباد و در حاشیه کویر مرکزی ایران واقع شده است. این معدن در طول جغرافیایی ۵۵ درجه و ۲۸ دقیقه شرقی و عرض جغرافیایی ۳۱ درجه و ۴۲ دقیقه شمالی در حوضه بافق قرار دارد که دارای آب و هوای گرم و خشک است. این معدن از طریق جاده آسفالتی از طریق شهرستان بافق به یزد وصل می‌شود که طول این جاده حدود ۱۳۰ کیلومتر است [۸]. در شکل ۱، موقعیت و همچنین راه‌های دسترسی به آنومالی‌های آهن و منطقه مورد مطالعه در بلوک معدنی بافق نمایش داده شده است.

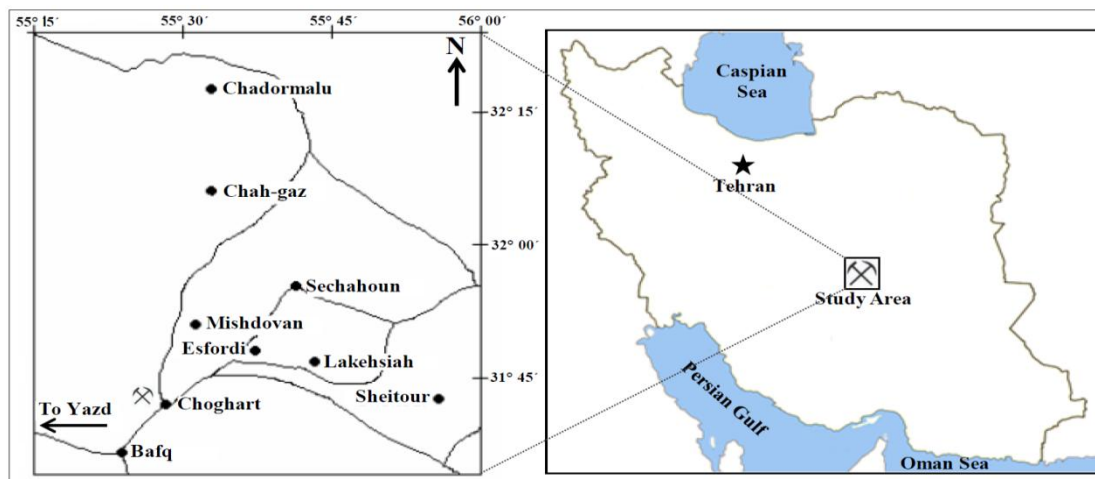
۲-۱- زمین شناسی عمومی

کناسر چغارت با ترکیب منییتی-هماتیتی، به‌صورت توده‌ای در امتداد شمال غرب به جنوب شرق با طول تقریبی ۶۰۰ متر و ضخامت ۴۰۰ تا ۷۰۰ متر در منطقه واقع شده است و تعداد زیادی دایک کناسر و سازندهای اطراف را قطع کرده است. کناسر چغارت در سازندهای پرکامبرین ایران مرکزی واقع شده که سنگ‌های در برگیرنده کناسر مجموعه‌ای از سنگ‌های رسوبی-آتشفشانی، فیلیت، شیست‌های متبلور، مرمر و گنیس است که در چینه‌شناسی از آن به سری مراد یاد می‌کنند که سری ریزو روی آن‌ها قرار گرفته است.

سری مراد از تغییراتی نظیر دگرگونی مجاورتی و ناحیه‌ای و متاسوماتیسم متأثر گشته است و با دگرشیبی خفیف، طبقات متعلق به اینفراکامبرین را می‌پوشاند و متعلق به اوایل

همچنین در کانسار پدیده‌های آلبیتیزاسیون، فلدسپاتیزاسیون، فلوگوپیتیزاسیون، آمفیبولیزاسیون و سرپانتینیزاسیون قابل مشاهده است.

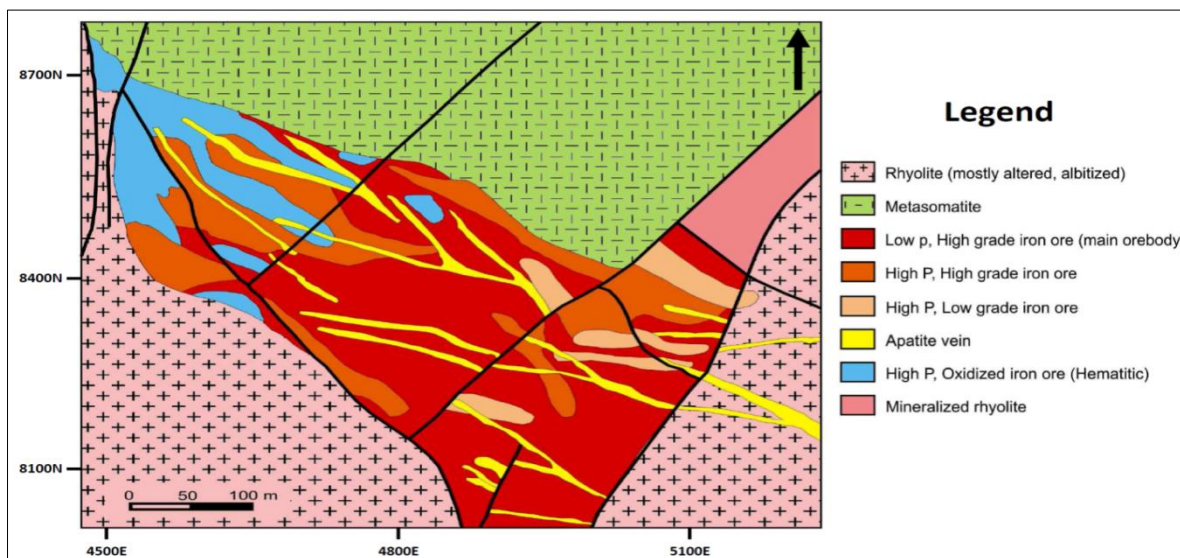
کامبرین است. سه نوع آلتراسیون سنگ میزبان در کانسار و اطراف آن قابل ذکر است که شامل آلتراسیون قبل از کانسنگ، آلتراسیون همراه کانسنگ و آلتراسیون بعد از کانسنگ است،



شکل ۱: نمایش موقعیت جغرافیایی و راه‌های دسترسی به معدن چغارت [۹]

کوارتزیت و کوارتز پورفیری که از درصد بالایی از کوارتز و فلدسپات تشکیل شده و سنگ‌هایی با درصد بالایی از آمفیبول (از نوع اکتینولیت-ترمولیت) و فلدسپات که تحت عنوان آمفیبولیت و آمفیبول پیروکسینیت در منطقه مشاهده شده است [۸،۹]. در شکل ۲، نقشه زمین‌شناسی کانسار آهن چغارت نمایش داده شده است.

ویژگی‌های ساختاری از جمله گسل‌ها، نقش اساسی در ساختار این کانسار ایفا می‌کند؛ گسل‌های موجود در منطقه منجر به دگرسانی با روند شمال غربی-جنوب شرقی، جابجایی و برشی شدن در قسمت‌هایی از کانسار شده است. کانی‌های شناخته شده در کانسار چغارت شامل منیتیت، هماتیت-مارتیت، آپاتیت، اسفن، کلسیت، ترمولیت، آلبیت، پیریت، ایلمینیت و به صورت نادر کالکوپیریت است. سنگ‌های درونگیر

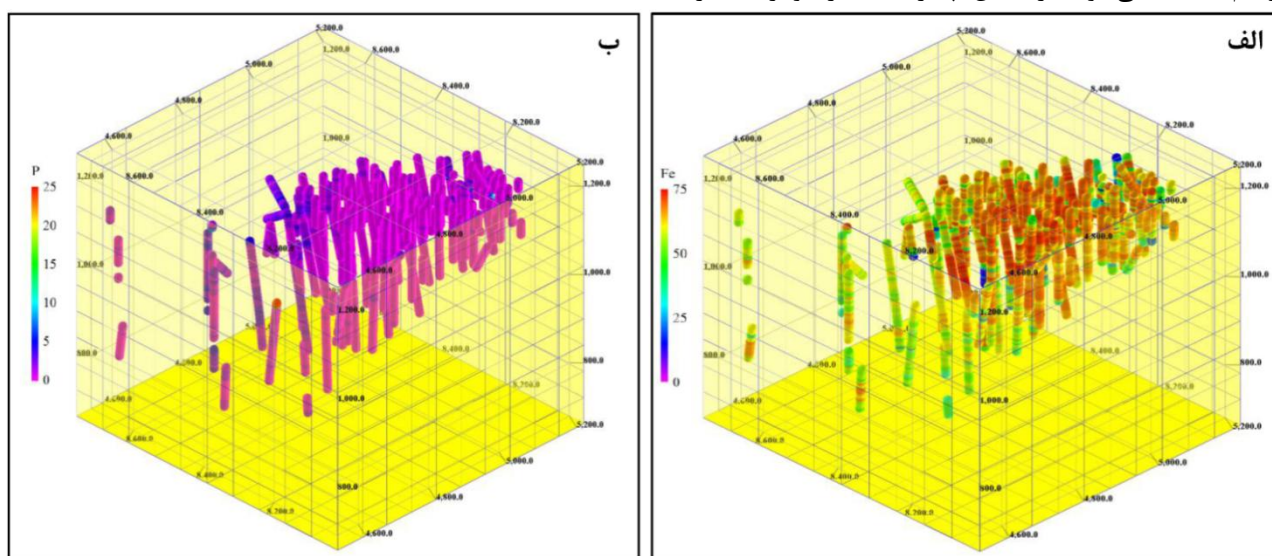


شکل ۲: نقشه زمین‌شناسی کانسار آهن چغارت [۱۰]

۲-۲- داده‌های مورد بررسی

تعداد ۱۳۷ گمانه قائم و جهتی در یک شبکه حفاری نامنظم در کانسار چغارت انجام شده و داده‌های مربوط به عیار آهن و فسفر در اعماق مختلف استخراج شد. در چنین معادنی، آهن به‌عنوان ماده با ارزش معدنی و فسفر و گوگرد به‌عنوان عناصر مزاحم مطالعه می‌شود. در معدن چغارت، مقدار گوگرد بسیار

کم است و نادیده گرفته می‌شود، اما مقدار فسفر قابل توجه است و می‌تواند باعث کاهش ارزش ماده معدنی و ایجاد مشکل در فرآیند ذوب آهن شود. در شکل ۳ توزیع سه بعدی آهن و فسفر در گمانه‌ها را نمایش می‌دهد [۹].



شکل ۳: نمایش آرایش نامنظم گمانه‌ها و توزیع مقادیر (الف) عیار آهن؛ (ب) عیار فسفر

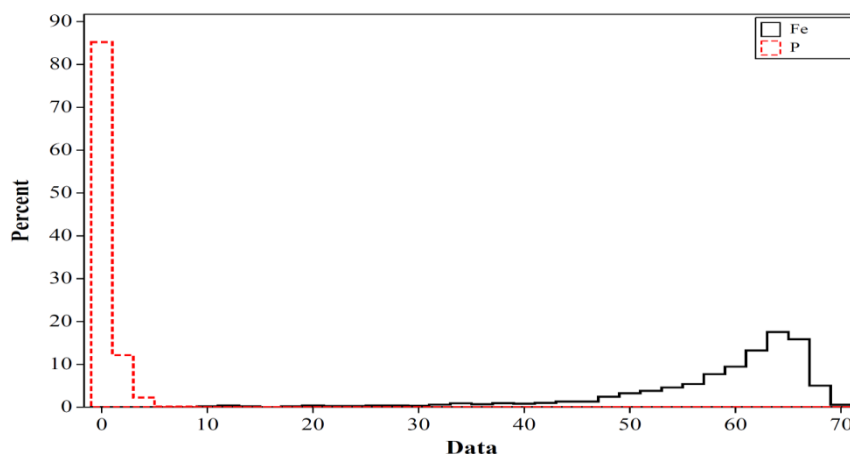
۳- ضرورت پهنه بندی در مدل‌سازی کانسارها

در فضای ناهمگن، بر اساس میزان انحراف از فرضیه ایستایی، منطقه مورد مطالعه می‌بایست به‌صورت پهنه بندی به‌تعدادی پهنه مجزا تقسیم شود تا آمار فضایی در هر پهنه مدل‌سازی شود؛ سپس به‌منظور ساخت یک مدل واحد، خروجی پهنه‌ها با توجه به‌نوع مرز بین پهنه‌ها با هم ترکیب می‌شوند. در صورتی که برای ارزیابی منطقه، یک رفتار میانگینی از پهنه‌ها در نظر گرفته شود؛ یک مدل تخمینی کاذب همراه با خطا در پی خواهد داشت. در بررسی یک کانسار ناهمگن، تعداد پهنه‌ها، محدوده هر پهنه و نوع مرز بین دو پهنه (ناگهانی، تدریجی یا فازی) براساس پایگاه اطلاعات زمین‌شناسی و اطلاعات عیاری قابل بررسی است. در مطالعات علوم زمین، پهنه بندی را می‌توان براساس شواهد فیزیکی زمین مانند ویژگی‌های ساختاری و زمین‌شناسی یا با استفاده از خوشه بندی داده‌ها انجام داد [۱۲،۱۱].

اولین قدم در انجام مطالعات عددی، مطالعات آماری پایه و استخراج خصوصیات آماری مربوط به داده‌ها است. در این مطالعه خصوصیات آمار توصیفی متغیرهای آهن و فسفر محاسبه شد که عیار آهن و فسفر به‌ترتیب دارای میانگین ۵۸٪ و ۵٪ است (جدول ۱). برای نمایش توزیع فراوانی داده‌های مربوط به آهن و فسفر، از هیستوگرام استفاده شد. داده‌های آهن و فسفر بیانگر بیانگر توزیع آماری غیر نرمال این دو متغیر است که در دامنه وسیعی از عیار در نوسان است. همان‌طور که در شکل ۴ قابل مشاهده است، آهن دارای چولگی منفی و فسفر دارای چولگی مثبت است که با توزیع آماری آهن و فسفر در کانسنگ‌های آهن هم خوانی دارد.

جدول ۱: بررسی خصوصیات آماری برای متغیرهای آهن و فسفر

متغیر	بیشینه	کمینه	میانگین	نما(مد)	میانه	انحراف معیار	چولگی	کشدگی	%۲۵	%۷۵
آهن	۷۰/۳۰	۵/۵۹	۵۸/۰۱۴	۶۵/۴	۶۱/۴۲	۱۰/۱۵۶	-۲/۱	۵/۱۲	۵۴/۹۸	۶۴/۶
فسفر	۲۳/۶۸	۰/۰۰	۰/۴۶۳۳	۰/۰۰۱	۰/۰۸۱	۱/۰۰۰۸	۷/۲۶	۱۱۱,۸۶	۰/۰۳	۰/۴۱



شکل ۴: توزیع فراوانی داده‌های خام پارامترهای آهن و فسفر

امروزه، نقشه خودسازمانده (SOM) یکی از الگوریتم‌های پر کاربرد خوشه‌بندی شبکه عصبی است که در زمینه‌های مختلفی چون علوم مهندسی، بیولوژی، اقتصاد به کار گرفته شده است [۱۳، ۱۴]. الگوریتم نقشه خودسازمانده ابزار مناسبی برای خوشه‌بندی داده‌ها است و قادر است روابط آماری غیرخطی بین داده‌های ورودی را به روابط هندسی ساده تبدیل کند. این امر بیانگر یک رابطه رگرسیونی بازگشتی ناپارامتری است، به طوری که با ارایه هر نمونه، رگرسیون به صورت بازگشتی اجرا می‌شود. بر این اساس می‌توان ادعا کرد که SOM می‌تواند میزان خطاها را بررسی و تصحیح کند. خوشه-بندی، به معنای افزایش مجموعه‌ای از داده‌ها به تعداد معینی از خوشه‌ها است و خوشه‌بندی بهینه، افزایش است که فواصل بین نمونه‌های داخلی را کمینه و فواصل بین خوشه‌ها را بیشینه کند [۶، ۱۵].

الگوریتم نقشه خودسازمانده شامل روابط دو بعدی، شبکه‌ای از واحدهای نقشه‌ای است که واحدها توسط یک رابطه همسایگی، به مجاورت آن‌ها متصل می‌شوند. تعداد واحدهای نقشه که اغلب از چند تا چندین هزار تغییر می‌کند، دقت و قابلیت تعمیم SOM را مشخص می‌کنند. هر نورون توسط یک بردار وزنی پیش‌نمونه نمایش داده می‌شود که

۴- پهنه بندی با استفاده از الگوریتم‌های خوشه بندی

خوشه‌بندی به فرآیندی اطلاق می‌شود که یک خوشه را به-عنوان مجموعه داده همگن با انسجام داخلی و جدایش خارجی معرفی می‌کند و به یافتن نظم در داده‌ها منتج می‌شود. درحقیقت، هیچ تعریف واحدی برای خوشه و خوشه‌بندی که واجد همه شرایط باشد، وجود ندارد. این تصور ممکن است موجب ایجاد مفاهیم دقیق همگن‌سازی و جدایش ریاضیاتی به صورت شاخص‌های عددی صریح باشد که منجر به ایجاد معیارهای متعدد و متنوع شده است. به طور واضح نمی‌توان، پراکندگی یک خوشه را وقتی روی یک صفحه به-نمایش در می‌آید، تعیین کرد، در حالی که یکی از ویژگی‌های معمول در فرآیند شناسایی، محاسبه فواصل نسبی بین نقاط است [۴]. محاسبه خوشه‌بندی برای دستیابی به برخی از بهینه‌سازی‌ها، از جمله فاصله کمینه درون خوشه‌ای یا فاصله بیشینه بین خوشه‌ای صورت می‌گیرد. در روش‌های خوشه‌بندی، خوشه‌ها را می‌توان بر اساس هر یک از معیارهای ارتباط بین داده‌ها، مرکز ثقل، توزیع و چگالی تعیین کرد. در ادامه روش خوشه‌بندی نقشه خودسازمانده بیان شده است.

۴-۱- الگوریتم خوشه‌بندی نقشه خود سازمانده

$$h_{Ci}(t) = \exp\left(-\frac{\|r_C - r_i\|^2}{2\sigma^2(t)}\right) \quad (3)$$

که $\sigma^2(t)$ برابر با شعاع گسترش تابع همسایگی و $r_i \in R^2$ و $r_C \in R^2$ موقعیت بردارهای نورون‌های i و C واحد بهترین میزان تطبیق روی شبکه خودسازمانده هستند که در این حالت نیز همراه با گذشت زمان و مراحل رگرسیون، شعاع گسترش کاهش می‌یابد. برای تعیین تعداد خوشه‌ها به‌طور معمول از قاعده‌ای کلی، ۲ تا \sqrt{N} برای تعیین تعداد خوشه‌ها ارائه شده است که N تعداد نمونه‌ها در مجموعه داده‌ها است. الگوریتم خودسازمانده، تابع خطا در رابطه ۴ را کمینه می‌کند [۱۶، ۱۷].

$$E = \sum_{k=1}^N \sum_{i=1}^K h_{Ci} \|x_k - m_i\|^2 \quad (4)$$

که K تعداد خوشه‌ها است، هسته همسایگی $h_{Ci}(t)$ در واحد C تمرکز یافته که بیان‌گر بهترین میزان تطبیق بردار x_i است. با توجه به رابطه ۲، SOM برای خطاهای بزرگ (فواصل بیشتر) امتیاز منفی بیشتری را در نظر می‌گیرد. داده‌های ورودی SOM از بردارهایی با k عنصر تشکیل شده‌اند. نرخ یادگیری از یک مقدار اولیه آغاز و کاهش می‌یابد و فاصله همسایگی ابتدا از بیشینه فاصله نورونی (تقریباً همه واحدها در همسایگی باشند) سپس به‌تدریج کوچک می‌شود تا در پایان فاز مرتب شدن به واحد برنده محدود و همگرا می‌شود. در این الگوریتم در هر مرتبه آموزش، همه بردارهای آموزشی هرکدام یک بار به‌صورت تصادفی به شبکه اعمال شده و ضرائب وزنی و بایاس بعد از ارائه هر بردار به‌روز رسانی می‌شوند [۱۲، ۱۵].

۴-۲- تعیین تعداد پهنه بهینه در حالت مرز ناگهانی

در اغلب الگوریتم‌های خوشه بندی، تعداد خوشه‌ها به‌عنوان یکی از پارامترهای از پیش تعیین شده و اساسی محسوب می‌شود. روش‌ها و شاخص‌های متنوعی برای تعیین تعداد بهینه خوشه‌ها ارائه شده است که اغلب بر اساس بیشینه تفاوت بین خوشه‌ای و بیشینه تشابه درون خوشه‌ای تعیین می‌شود. درستی اعتبار خوشه‌بندی به‌صورت نسبی بر اساس پارامترهای مختلف خوشه‌بندی و تعداد خوشه‌ها صورت می‌گیرد. اعتبارسنجی خوشه‌بندی، یکی از مسائل بسیار مهم در تحلیل خوشه‌بندی است، چون نتایج خوشه‌بندی را تحت تأثیر قرار می‌دهد.

شامل بردار ورودی است. در مرحله آموزش، SOM، یک شبکه شبه فیزی به‌وجود آورده که توده تشکیل شده از داده‌های ورودی را با هم در می‌آمیزد. داده‌هایی که نزدیک به هم در فضای ورودی قرار گرفته‌اند، به‌صورت واحدهای نقشه نگاشت می‌شوند. بنابراین SOM مانند یک نقشه توپولوژی است که امکان نمایش، تفسیر و ترتیب خوشه‌بندی را فراهم می‌کند و قادر است از درجه فضای داده‌های ورودی به‌شبکه دو بعدی واحدهای نقشه نگاشت شود (شکل ۵).

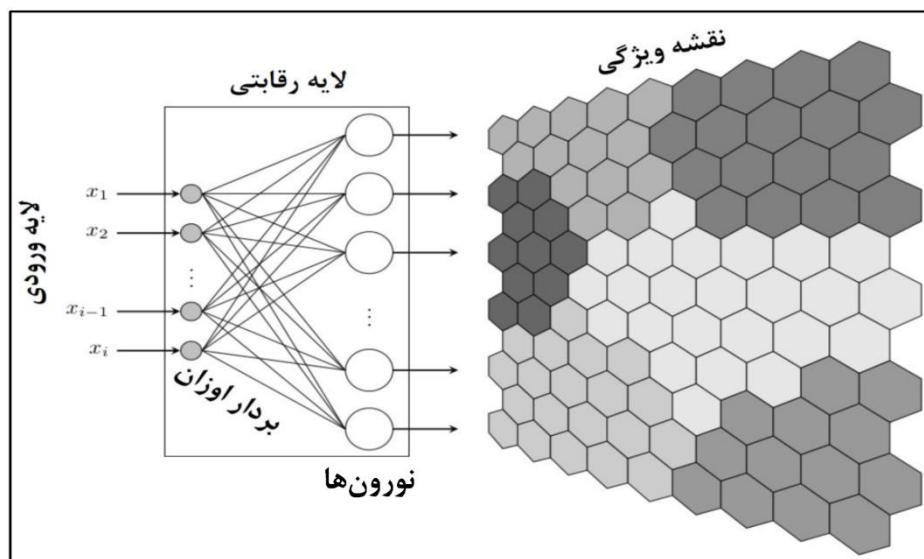
برای اجرای الگوریتم، ابتدا مقادیر تصادفی کوچک برای بردار وزن ارتباطی m_i ($i=1,2,\dots,K$) انتخاب می‌شود که k تعداد نورون‌های نقشه است. در هر مرحله از آموزش، یک بردار نمونه‌های ورودی $x=[x_1, x_2, \dots, x_N]$ از مجموعه n داده‌های ورودی به‌صورت تصادفی انتخاب شده و فواصل اقلیدسی بین بردار ورودی و همه بردارهای وزن ارتباطی تمام نورون‌ها محاسبه می‌شوند. براساس کمینه‌سازی فاصله یک نمونه با بردارهای وزن ارتباطی دیگر، بهترین میزان تطبیق توسط رابطه ۱ محاسبه می‌شود که C نمایانگر نورون برنده است [۶، ۱۶، ۱۵].

$$\|x - m_i\| = \min_i (\|x - m_i\|) \quad (1)$$

بعد از یافتن بهترین میزان تطبیق، بردارهای وزنی نقشه خودسازمانده به‌روز رسانی می‌شوند که واحد دارای بهترین میزان تطبیق و همسایه‌های توپولوژیکی آن به نزدیکی بردار ورودی در فضای ورودی منتقل می‌شوند. فرآیند تطبیق، واحد برنده را گسترش و همسایه‌های توپولوژیکی آن به سمت بردار ورودی نزدیک می‌شود. برای به‌روزرسانی بردار پیش نمونه واحد i از رابطه ۲ استفاده می‌شود.

$$m_i(t+1) = m_i(t) + \alpha(t)h_{Ci}(t)[x - m_i(t)] \quad (2)$$

که t نمایان‌گر زمان است که شبکه خودسازمانده به صورت یک فرآیند بازگشتی آموزش داده می‌شود، $\alpha(t)$ نرخ یادگیری و آموزش که بیان‌گر میزان تطبیق است و به‌صورت یکنواخت با مراحل رگرسیون (زمان) کاهش می‌یابد و $h_{Ci}(t)$ هسته همسایگی در اطراف واحد برنده C است که یک تابع کاهشی از فاصله بین واحد i و واحد برنده C روی شبکه نقشه است که روی واحد برنده متمرکز می‌شود. تابع همسایگی به‌صورت رابطه ۳ در نظر گرفته می‌شود.



شکل ۵: ساختار روش نقشه خودسازمانده

نمایش و مقایسه شاخص‌های متفاوت در کنار هم می‌توان مقادیر شاخص‌ها را نرمال‌سازی یا رتبه بندی کرد. در نتیجه، با توجه به تغییرات مقادیر شاخص‌های مختلف در برابر تعداد خوشه‌ها، تعداد بهینه خوشه‌ها یا به عبارت دیگر پهنه‌ها مشخص می‌شود.

برای اجرای اعتبارسنجی و تعیین تعداد بهینه خوشه‌ها، مقادیر عیاری آهن و فسفر مغزه‌ها به همراه موقعیت فضایی آن‌ها به عنوان پارامترهای ورودی مشخص شدند. ابتدا، رابطه مستقیم یا عکس بین شاخص‌های اعتبارسنجی خوشه بندی و تعداد خوشه‌ها بررسی می‌شود. مقادیر بیشینه شاخص‌های کالینسکی-هاراباسز، کرزانوسکی-لای، دان و سیلهوته و در مقابل مقادیر کمینه شاخص‌های دیویس-بولدین، هارتیگان و C، تعداد بهینه خوشه‌ها را مشخص می‌کند. برای نمایش شاخص‌های متفاوت در کنار هم، ابتدا می‌بایست مقدار شاخص‌ها را نرمال‌سازی کرده و در مقابل تعداد خوشه‌ها بررسی کرد که در شکل ۶ نمایش داده شده است.

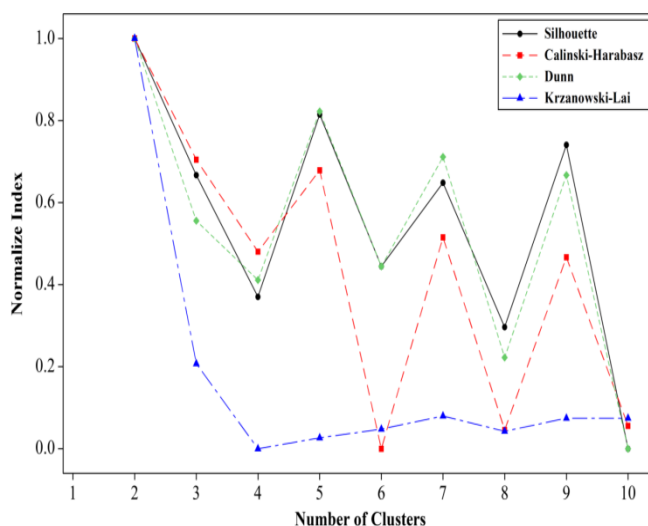
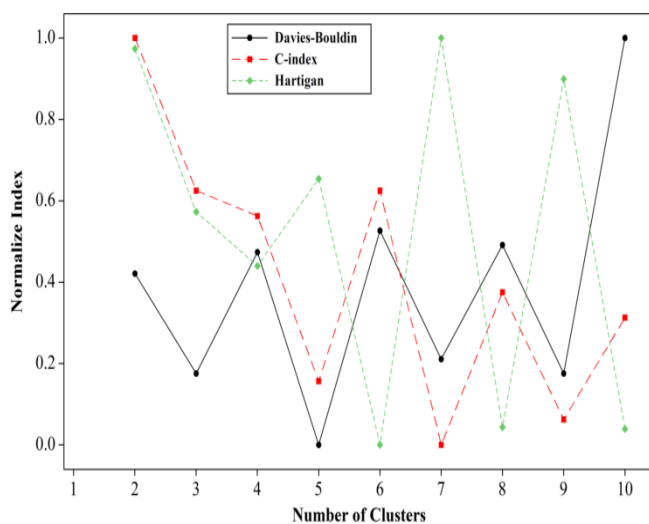
هفت شاخص، اعتبارسنجی خوشه بندی براساس رابطه مستقیم یا عکس با تعداد خوشه‌ها رتبه بندی شدند. در جدول ۳، براساس رتبه بندی ارائه شده، چهار شاخص (Krzanowski- Silhouette, Dunn, Lai) از هفت شاخص، تعداد بهینه را دو خوشه، سه شاخص دیگر، تعداد بهینه خوشه را پنج، شش و هفت شاخص نمایش می‌دهد.

از این رو، هنگامی که تعداد خوشه‌ها از قبل مشخص نیست، شاخص‌های اعتبارسنجی خوشه بندی را می‌توان برای یافتن تعداد بهینه خوشه‌ها استفاده کرد [۲۰-۱۸]. شاخص اعتبارسنجی خوشه بندی، روشی برای تعیین کیفیت افراز به دست آمده است و پاسخ به این که آیا افراز بهتری وجود دارد یا خیر. هیچ یک از شاخص‌های اعتبارسنجی به تنهایی کامل نیست، بنابراین، استفاده همزمان از چند شاخص برای تعیین تعداد بهینه خوشه‌ها پیشنهاد می‌شود [۲۱]. شاخص‌های اعتبارسنجی متفاوتی از جمله دان، دیویس-بولدین، سیلهوته، C، کالینسکی-هاراباسز، هارتیگان و کرزانوسکی-لای در جدول ۲ توضیح داده شده است.

برای یافتن تعداد خوشه‌ها که مبین تعداد پهنه‌ها در یک منطقه است، بدین منظور از مجموعه داده‌های عیاری آهن و فسفر در گمانه به همراه مختصات فضایی گمانه‌ها استفاده شده و این ویژگی‌ها به عنوان ورودی شاخص‌های اعتبارسنجی خوشه بندی تعیین می‌شوند. مقدار شاخص اعتبارسنجی خوشه بندی برای افرازهای داده‌ها برای هر تعداد خوشه محاسبه می‌شود. ابتدا، رابطه مستقیم و عکس بین شاخص‌های اعتبارسنجی خوشه بندی و تعداد خوشه‌ها بررسی می‌شود. در هر دو حالت، مقادیر بیشینه و کمینه شاخص‌های مختلف در برابر تعداد بهینه خوشه‌ها بررسی می‌شود. برای

جدول ۲: شاخص‌های مختلف اعتبارسنجی خوشه بندی به منظور تعیین تعداد بهینه خوشه

شاخص	توضیح
دان	این شاخص بر پایه این ایده که فاصله درون خوشه‌ها به هم نزدیک‌تر و خوشه‌ها به خوبی از هم جدا شده باشند بنیان گذاری شده است. هدف اصلی این شاخص، بیشینه‌سازی فاصله بین خوشه‌ها و کمینه‌سازی فاصله درون خوشه‌ای است. بنابراین، تعداد بهینه خوشه‌ها بر اساس بیشینه‌سازی شاخص اعتبارسنجی دان انجام می‌شود [۲۲].
دیویس-بولدین	شاخص اعتبارسنجی دیویس-بولدین، تابعی از نسبت مجموع پراش درون خوشه به فاصله بین خوشه‌ها است. بنابراین، هنگامی که درون خوشه به هم نزدیک و خوشه‌ها از یکدیگر دور باشد، این نسبت کوچک است. مقدار کوچک شاخص اعتبارسنجی دیویس-بولدین، نماینده خوشه‌بندی معتبر است [۲۳].
سیلهوته	روش اعتبارسنجی سیلهوته، بر اساس میانگین فاصله (یا هر معیار عدم تشابه) هر یک از نمونه‌های یک خوشه با تمام نمونه‌های موجود در همان خوشه، و میانگین فاصله کل نمونه موجود در خوشه‌های دیگر با یک خوشه مشخص تعریف می‌شود. مقدار شاخص اعتبارسنجی سیلهوته بین ۱- تا ۱ متغیر است. در صورتی که این شاخص برابر ۱ باشد، نمایانگر این است که خوشه‌بندی به درستی صورت پذیرفته است. اگر مقدار شاخص نزدیک به صفر باشد، این بدین معنی است که نمونه را می‌توان به یک خوشه نزدیک‌تر نسبت داد، در صورتی که این شاخص برابر ۱- باشد، این بدین معنی است که خوشه‌بندی به درستی صورت نپذیرفته است [۲۴].
C	شاخص C بر اساس مقایسه فواصل جفت نمونه‌های موجود در یک خوشه نسبت به هم و تمام خوشه‌ها نسبت به هم سنجیده می‌شود و به صورت یک رابطه نرمال‌سازی تعریف می‌شود. مقادیر کوچک شاخص C نمایانگر خوشه‌بندی صحیح است [۲۵].
کالینسکی-هاراباسز	این سه شاخص اعتبارسنجی، از ماتریس مجموع مربعات پراکندگی بین خوشه‌ها و درون خوشه‌ای استفاده می‌کنند؛ که شاخص اعتبارسنجی کالینسکی-هاراباسز از نسبت مجموع قطر ماتریس مجموع مربعات پراکندگی بین خوشه‌ها به مجموع قطر ماتریس مجموع مربعات پراکندگی درون خوشه‌ای استفاده می‌شود. شاخص‌های هارتینگان و کرزانوسکی-لای از مقایسه مجموع قطر ماتریس مجموع مربعات پراکندگی درون خوشه‌ای، خوشه‌ها نسبت به هم استفاده می‌شود. مقادیر کوچک شاخص هارتینگان و مقادیر بزرگ شاخص‌های کالینسکی-هاراباسز و کرزانوسکی-لای نمایانگر خوشه‌بندی مناسب است [۲۸-۲۶].
کالینسکی-هاراباسز	
هارتینگان	
کرزانوسکی-لای	



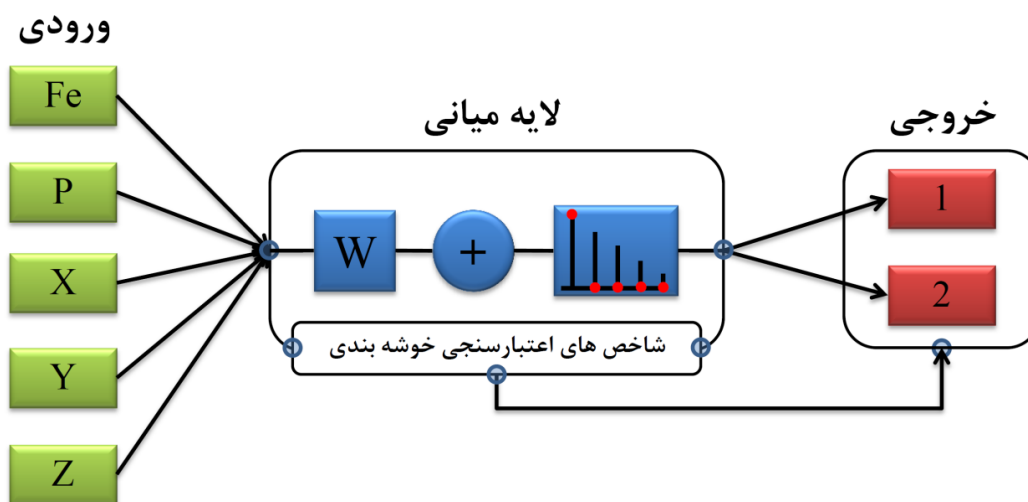
شکل ۶: تغییر مقادیر نرمال‌سازی شده شاخص‌های اعتبارسنجی خوشه‌بندی در مقابل تعداد خوشه‌ها

جدول ۳: بهینه سازی تعداد خوشه ها براساس اولویت بندی شاخص های مختلف

Silhouette	Davies-Bouldin	Calinski-Harabasz	Dunn	C-index	Krzanowski-Lai	Hartigan	تعداد بهینه خوشه
۱	۵	۱	۱	۹	۱	۸	۲
۴	۳	۲	۵	۷	۲	۵	۳
۷	۶	۵	۷	۶	۹	۴	۴
۲	۱	۳	۲	۳	۸	۶	۵
۶	۸	۹	۶	۷	۶	۱	۶
۵	۴	۴	۳	۱	۳	۹	۷
۸	۷	۸	۸	۵	۷	۳	۸
۳	۳	۶	۴	۲	۴	۷	۹
۹	۹	۷	۹	۴	۴	۲	۱۰

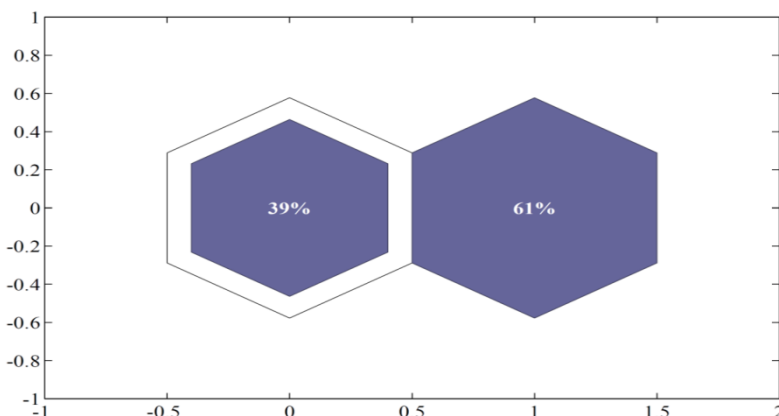
شبکه شده و خروجی شبکه براساس شاخص های اعتبارسنجی خوشه ها، داده ها به دو پهنه (خوشه) مجزا تقسیم می شوند و هر نمونه از گمانه به یک پهنه اختصاص داده می شود. بنابراین فضای ورودی یک فضای پنج بعدی است و خروجی یک فضای دو بعدی است که در شکل ۷ نمایش داده شده است.

۳-۴- استفاده از SOM در پهنه بندی کانسار
در بررسی حاضر، داده های آهن و فسفر معدن آهن چغارت برای خوشه بندی SOM استفاده شده است. داده های ورودی SOM، پارامترهای عیاری آهن و فسفر و سه پارامتر مختصاتی (X,Y,Z) متناظر آن است. در مرحله بعد، پنج پارامتر ورودی به صورت نرمال سازی شده (در دامنه ۰ تا ۱) برای آموزش وارد



شکل ۷: نمایش الگوریتم خوشه بندی خود سازمانده در منطقه مورد مطالعه

توپولوژی مسئله به صورت شش ضلعی است. یکی دیگر از قابلیت‌های این روش، تعیین تعداد نمونه‌های مرتبط با هر نورون است. نمونه‌ها با پنج ویژگی براساس یک شبکه رقابتی است و درصد اختصاص نمونه‌ها، به دو خوشه به ترتیب ۶۱٪ و ۳۹٪ نمونه تقسیم شده است (شکل ۸).



شکل ۸: تعیین تعداد نمونه‌های مربوط به هر خوشه

پهنه اول است، در حالی که در پهنه اول، عیار آهن پراکندگی بیشتری دارد. در مورد عیار فسفر، بیشترین فراوانی در دو پهنه مربوط به مقادیر نزدیک به صفر است که فراوانی این مقادیر در پهنه دوم شدیدتر است. توزیع عنصر فسفر در دو پهنه، چولگی شدید مثبتی را از خود نشان می‌دهند، این در حالی است که آهن در پهنه دوم، دارای چولگی منفی است.

۵- بحث

در این مقاله، نتایج خروجی پهنه‌بندی SOM با الگوریتم رایج خوشه‌بندی K-means مقایسه شده است. این الگوریتم یک روش خوشه‌بندی مبتنی بر مرکز ثقل است که براساس معیار کمینه‌سازی مجموع مربعات خطاهای خوشه‌ها اجرا می‌شود. محبوبیت این الگوریتم به دلیل سادگی در اجرا، کارایی و تجربیات موفق این روش در مطالعات مختلف است. مراحل مختلف الگوریتم خوشه‌بندی K-means شامل انتخاب تعداد خوشه‌ها، اختصاص هر نمونه به نزدیک‌ترین مرکز خوشه، محاسبه مراکز خوشه‌های جدید و در نهایت تکرار این مراحل

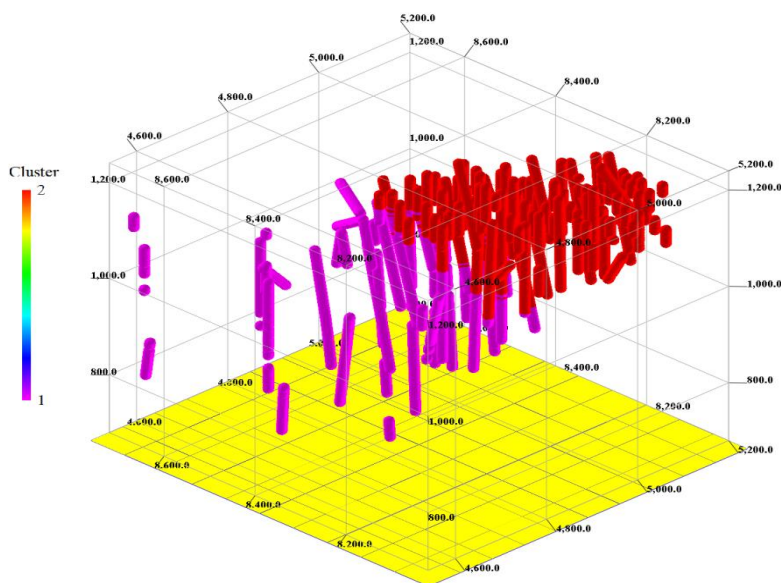
برای آموزش SOM، بردار وزن مربوط به هر نورون، برای نزدیک شدن به مرکز خوشه بردارهای ورودی، تغییر کرده و مکان خود را تغییر می‌دهد. همچنین، نورون‌هایی که براساس توپولوژی در مجاورت یکدیگر قرار دارند، برای نزدیک شدن به هم، مکان خود را تغییر می‌دهند. در این بررسی هندسه

خروجی مدل شبکه عصبی SOM، هر نمونه را به یکی از دو پهنه نسبت و اختصاص داده و دو پهنه مجزا را تشکیل می‌دهد که تقریباً مرزبندی پهنه‌ها به صورت ناگهانی است. چون در هر پهنه تغییرات عمق وجود دارد و یک مدل سه‌بعدی واقعی است، احتمال دارد قسمت بالایی گمانه، مربوط به یک پهنه و قسمت پایینی گمانه، مربوط به پهنه دیگر باشد که در شکل ۹، توزیع فضایی نمونه‌ها در دو پهنه نمایش داده شده است.

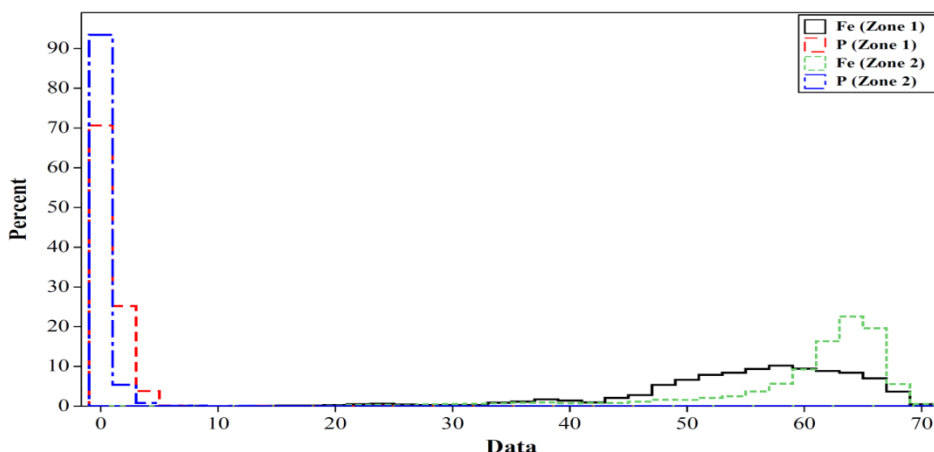
در حالت پهنه‌بندی با استفاده از روش SOM، توزیع آماری مقادیر آهن و فسفر در هریک از پهنه‌های اول و دوم بررسی شد که در شکل ۱۰ نشان داده شده است. احتمال اختصاص یک نمونه جدید تنها براساس آمار پیشین برای پهنه‌های اول و دوم به ترتیب تقریباً برابر ۰/۴ و ۰/۶ است. توزیع مقادیر آهن در پهنه دوم نشان می‌دهد که فراوانی نمونه‌های دارای مقادیر با عیار بالای آهن، بیشتر و تابع توزیع آهن در پهنه دوم، از مقدار ۴۵٪ تا ۷۰٪ تقریباً دارای یک توزیع تقریباً نرمال است. بر اساس نشان دسته میانگین، عیار آهن در پهنه دوم بیشتر از

همچنین تعداد پهنه موجود اجرا شد.

تا دستیابی به معیار همگرایی است [۲۹، ۳۰]. الگوریتم خوشه بندی K-means با توجه به توزیع فضایی عیاری آهن و فسفر و



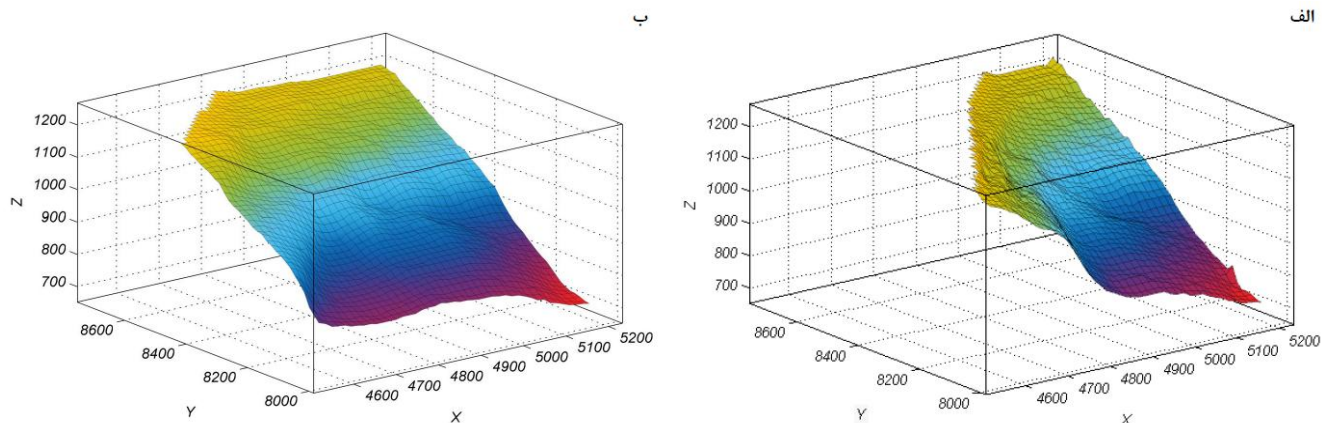
شکل ۹: اختصاص هر نمونه به پهنه مربوطه براساس منطقه بندی



شکل ۱۰: توزیع فراوانی داده های خام پارامترهای آهن و فسفر در دو پهنه

این در حالی است که سطح جدایش در روش K-means دارای راستای شرقی-غربی و شیب به سمت جنوب است (شکل ۱۱). بر اساس مقایسه خصوصیات بعدی و جهتی، سطح جدایش دو پهنه و ویژگی های ساختمانی مؤثر در کنترل کانی زایی بالاخص گسل ها، سطح جدایش به دست آمده از روش SOM از نظر ساختار فضایی و جهتی همخوانی قابل-توجهی با ویژگی های ساختمانی منطقه دارد.

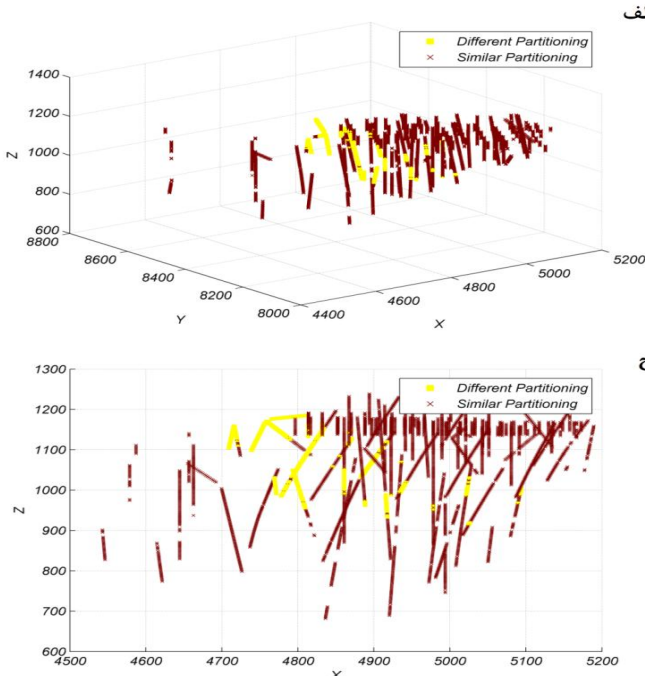
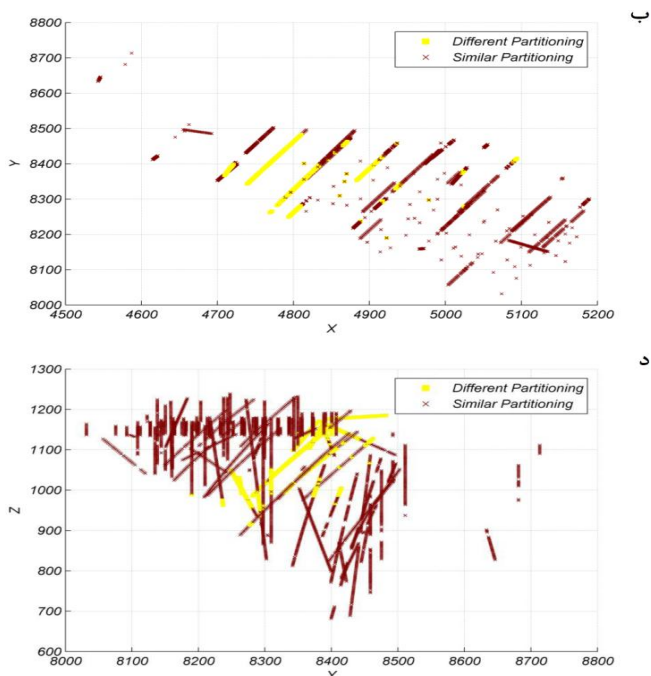
از نظر توزیع آماری آهن و فسفر در دو پهنه، خروجی های الگوریتم همخوانی بالایی با روش SOM دارد، با این تفاوت که روش SOM، دو پهنه با تمایز شاخص تری را انتخاب کرده است. برای مقایسه نتایج روش های خوشه بندی SOM و K-means، سطح جدایش دو خوشه محاسبه و ترسیم شد. سطح جدایش دو پهنه در روش SOM در راستای شمال شرقی- جنوب غربی کشیده شده که به سمت جنوب شرقی شیب دارد،



شکل ۱۱: نمایش سطح جدایش توسط الگوریتم‌های: الف) SOM، ب) K-means

خوشه آن در دو روش متفاوت است. در شکل ۱۲، تقسیم-بندی نمونه‌ها براساس افراز یکسان و متفاوت نمایش داده شده است.

افراز نمونه‌ها براساس اختصاص به هر یک از پهنه‌ها توسط الگوریتم‌های SOM و K-means بررسی شد، اغلب نمونه‌ها توسط دو روش به پهنه‌های یکسان اختصاص داده شده است. در مرز میانی دو پهنه که بیانگر سطح جدایش است که شماره



شکل ۱۲: نمایش افراز نمونه‌های اختصاص یافته مشابه و متفاوت در دو رویکرد خوشه‌بندی از نماهای مختلف

۶- نتیجه گیری

- در این مقاله، از روش خوشه بندی نقشه خودسازمانده به منظور پهنه بندی فضایی کانسار براساس توزیع فضایی عیاری گمانه ها استفاده شد. این رویکرد تعیین محدوده پهنه ها براساس مرز بندی ناگهانی است. در راستای ارزیابی الگوریتم، از داده های حفاری اکتشافی عیار آهن و فسفر معدن سنگ آهن چغارت استفاده شد.
- به منظور تعیین تعداد بهینه پهنه ها از شاخص های اعتبارسنجی مختلف استفاده شده، به دلیل اینکه هیچ یک از شاخص ها به تنهایی کامل نیستند. بر این اساس تعداد دو پهنه به عنوان پاسخ بهینه مشخص شد.
- بر اساس خروجی های شاخص های اعتبارسنجی و داده های عیاری اکتشافی، الگوریتم SOM اجرا شد و داده ها به دو پهنه اختصاص داده شد. بر اساس مطالعات آماری، متوسط عیار آهن در پهنه دوم بیشتر از پهنه اول بوده، در حالی که در پهنه اول، عیار آهن پراکندگی بیشتری دارد. در مورد عیار فسفر، بیشترین فراوانی در دو پهنه، مربوط به مقادیر نزدیک به صفر است و توزیع عنصر فسفر در دو پهنه، چولگی شدید مثبتی را از خود نشان می دهند.
- سطح جدایش دو پهنه در روش SOM، در راستای شمال شرقی-جنوب غربی امتداد داشته که به سمت جنوب شرقی شیب دارد که با ویژگی های ساختمانی مؤثر در کنترل کانی زایی (به ویژه گسل ها) هم خوانی دارد، در حالی که این سطح جدایش در روش K-means (علی رغم شباهت ۹۰٪ در اختصاص نمونه ها با روش SOM) با جهت یافتگی ساختار کانه زایی در معدن چغارت مطابقت ندارد.

مراجع

- [2] Journel, A.G. Huijbregts, C.J. 1978. Mining Geostatistics, Academic Press, London.
- [3] Kaufman, L., Rousseeuw, P.J., Finding Groups in Data: an Introduction to Cluster Analysis, Wiley, New Jersey, 342 pp, 2009.
- [4] Everitt, B., Landau, S., Leese, M., Stahl, D. Cluster Analysis (5th ed.), Hoboken, NJ: Wiley Publishing, 330 pp, 2011.
- [5] Rezaee, M.R., Lelieveldt, B.P.F., Reiber, J.H.C. A new cluster validity index for the fuzzy c-mean, Pattern Recognition Letters 19, 237-246, 1998.
- [6] Vesanto, J., Alhoniemi, E., 2000. Clustering of the self-organizing map. IEEE Transactions on Neural Networks 11, 586-600.
- [7] Kohonen, T., 2001. Self-Organizing Maps. Springer, New York.
- [8] Moor, F., Modabberi, S., 2003. Origin of Choghart iron oxide deposit, Bafq mining district, Central Iran: new isotopic and geochemical evidence. Journal of science, Islamic Republic of Iran 14, 259-269.
- [9] Hossein Morshedy, A., Torabi, S. A., Memarian, H., 2015. A new method for 3D designing of complementary exploration drilling layout based on ore value and objective functions. Arabian Journal of Geosciences, 1-21, DOI: 10.1007/s12517-014-1754-7.
- [10] Mohammad Torab, F.; "Geochemistry and metallogeny of magnetite-apatite deposits of the Bafq Mining District, Central Iran", PhD Thesis, Clausthal University of Technology, Germany, 131pp
- [11] Wingle, W.L., 1997. Evaluating Subsurface Uncertainty Using Modified Geostatistical Techniques. PhD Thesis, Colorado School of Mines, 180 pp.
- [۱۲] حسین مرشدی، ا.، معماریان، ح.، ۱۳۹۱. پهنه بندی شاخص کیفی سنگ در ساختگاه سد سمیلان، بر اساس گسل ها و شبکه عصبی خود سازمانده. فصلنامه علوم زمین، سال ۲۱، شماره ۸۴، ص. ۹۹ تا ۱۱۲.
- [13] Mwasiagi, J.I., 2011. Self Organising Maps - Applications and Novel Algorithm Design. INTECH, 702 pp.
- [14] Oja, M., Kaski, S., Kohonen, T., 2002. Bibliography of self-organizing map (SOM) papers: 1998-2001 Addendum. Neural Computing Surveys 3, 1-156.
- [15] Kohonen, T., 1998. The self-organizing map. Neurocomputing 21, 1-6.

- [1] Dagdelen, K., Turner, A.K. Importance of stationarity for geostatistical assessment of environmental contamination. ASTM Special Technical Publication 1283, 117-132, 1996.

evolutionary k-means. *Applied Soft Computing* 11, 1938-1952

- [16] Amin-Naseri, M.R., Soroush, A.R., 2008. Combined Use of Unsupervised and Supervised Learning for Daily Peak Load Forecasting. *Energy Conversion and Management*. 49, 1302-1308.
- [17] Cheng, G., Cheng, Y., Shen, L., Qiu, J., Zhang, S., 2013. Gear Fault Identification Based on Hilbert-Huang Transform and SOM Neural Network. *Measurement* 46, 1137-1146.
- [18] Kim, D.W., Lee, K.H., Lee, D., 2004, On cluster validity index for estimation of the optimal number of fuzzy clusters, *Pattern Recognition* 37, 2009–2025.
- [19] Topchy, A., Punch, J.W., 2003. Combining multiple weak clusterings Proc. Third IEEE International Conference on Data Mining (ICDM'03), pp. 331-338.
- [20] Wang, K., Wang, B., Peng, L., 2009. CVAP: Validation for cluster analyses. *Data Science Journal* 8, 88-93.
- [21] Balasko, B., Abonyi, J., Feil, B. *Fuzzy Clustering and Data Analysis Toolbox.*, University of Veszprem, Hungary, 74 pp, 2002.
- [22] Dunn J.C., 1974. Well separated clusters and optimal fuzzy partitions, *J.Cybern.* 4. 95-104.
- [23] Davies, D.L., Bouldin, D.W., 1979. A cluster separation measure, *IEEE Trans. Pattern Anal. Machine Intell.* 1 (4). 224-227.
- [24] Rousseeuw P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 20, 53-65.
- [25] Hubert, L., Schultz, J., 1976. Quadratic assignment as a general data-analysis strategy. *British Journal of Mathematical and Statistical Psychology* 29. 190-241.
- [26] Calinski, T., Harabasz, J., 1974. A dendrite method for cluster analysis. *Communications in Statistics* 3, 1-27.
- [27] Hartigan, J.A., 1975. *Clustering Algorithms.* Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Inc. New York.
- [28] Krzanowski W., Lai Y., 1985. A criterion for determining the number of groups in a dataset using sum of squares clustering. *Biometrics* 44, 23–34.
- [29] Jain, A.K. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters* 31, 651-666, 2010.
- [30] Naldi, M.C., Campello, R.J., Hruschka, E.R., Carvalho, A.C.P.L.F., 2011. Efficiency issues of