

معرفی یک روش بهینه جهت خوشه‌بندی جزءبندی داده‌های فراطیفی با استفاده از الگوریتم بهینه‌سازی توده ذرات

امین علیزاده نائینی^۱، محمد سعادت سرشت^۲، سعید همایونی^۳، ابوالفضل جمشیدزاده^{۴*}

^۱دکتری فتوگرامتری - گروه مهندسی نقشه‌برداری - پردیس دانشکده‌های فنی - دانشگاه تهران
a.alizadeh@ut.ac.ir

^۲استادیار گروه مهندسی نقشه‌برداری - پردیس دانشکده‌های فنی - دانشگاه تهران
msaadat@ut.ac.ir

^۳استادیار گروه جغرافیا - دانشگاه اتاوا - کانادا
saeid.homayouni@uottawa.ca

^۴عضو هیأت علمی گروه مهندسی نقشه‌برداری - دانشکده مهندسی - دانشگاه بجنورد
a.jamshidzadeh@ub.ac.ir

(تاریخ دریافت بهمن ۱۳۹۳، تاریخ تصویب اسفند ۱۳۹۳)

چکیده

یکی از موفقیت‌آمیزترین کاربردهای داده‌های فراطیفی، طبقه‌بندی آنها و تولید نقشه‌های پوشش زمین بدون نیاز به داده‌های واقعیت زمینی است. در میان روش‌های مختلف طبقه‌بندی غیرنظارت شده، روش‌های جزءبندی به دلیل سرعت و عملکرد بالا در خوشه‌بندی داده‌های فراطیفی مورد توجه قرار گرفته‌اند. عملکرد خوشه‌بندی مبتنی بر جزءبندی این داده‌ها تابع پنج پارامتر تعداد و موقعیت خوشه‌ها، تعداد و موقعیت باندها و معیار شباهت است. بنابراین می‌توان به مسئله‌ی خوشه‌بندی جزءبندی به صورت یک مسئله‌ی بهینه‌سازی نگاه کرد که هدف در آن پیدا کردن موقعیت بهینه‌ی پارامترهای مذکور است. بسته به اینکه کدامیک از این پنج پارامتر در فرآیند بهینه‌سازی توده ذرات وارد شوند چهار حالت معنی‌دار در این تحقیق در نظر گرفته شده و هدف پیدا کردن حالتی است که به بالاترین دقت در تهیه نقشه‌های موضوعی منجر شود. لازم به ذکر است که از میان پنج پارامتر مختلف خوشه‌بندی، معیار شباهت و تعداد خوشه‌ها به منظور جلوگیری از پدیده‌ی افزونی پارامترها وارد فرآیند بهینه‌سازی نشده و ثابت در نظر گرفته شدند. بررسی‌ها بر روی یک مجموعه‌ی داده‌ی شبیه‌سازی و دو داده‌ی واقعی نشان دادند که از میان چهار حالت مورد بررسی در این تحقیق، حالتی که تعداد باندها در یک مرحله‌ی پیش‌پردازشی توسط خوشه‌بندی باندها در فضای داده‌ها یا با استفاده از تبدیل PCA در فضای ویژگی کاهش داده می‌شود دارای دقت بالاتری در تهیه‌ی نقشه‌های موضوعی هستند.

واژگان کلیدی: داده‌های فراطیفی، طبقه‌بندی نظارت نشده، خوشه‌بندی باندها، الگوریتم بهینه‌سازی توده ذرات

* نویسنده رابط

۱- مقدمه

توانمندی منحصربه‌فرد تصاویر فراطیفی در بیان ویژگی‌های عوارض و اجسام مختلف سطح زمین، تحقیقات این شاخه را به سمت توسعه‌ی روش‌هایی هدایت می‌کند که تا حد ممکن نیاز به دخالت عامل انسانی در پردازش داده‌ها کاهش یابد [۱]. در این میان، خوشه‌بندی یکی از پرکاربردترین روش‌ها در بسیاری از پردازش‌های مطرح در تصاویر فراطیفی از قبیل طبقه‌بندی و تشخیص خودکار اشیاء است. روش‌های خوشه‌بندی موجود و متداول در سنجش از دور را می‌توان به چهار گروه تقسیم کرد: (۱) خوشه‌بندی سلسله‌مراتبی - این روش‌ها برای داده‌های سنجش از دوری بخصوص داده‌های فراطیفی با حجم بالا توصیه نمی‌شوند [۱]، (۲) خوشه‌بندی مبتنی بر نقطه اوج هیستوگرام - این روش‌ها برای داده‌های فراطیفی توصیه نمی‌شوند [۱]، (۳) خوشه‌بندی آماری - این روش‌ها به دلیل زمانبر بودنشان از یک طرف و وابستگی شدید آنها به تابع توزیع مورد استفاده برای داده‌های فراطیفی توصیه نمی‌شوند [۱]، (۴) خوشه‌بندی جزءبندی^۱ - در میان روش‌های مذکور، خوشه‌بندی جزءبندی یکی از پرکاربردترین روش‌های خوشه‌بندی در این تصاویر است. این روش‌ها طی فرآیندی خودکار یک مجموعه داده معین را به یک مجموعه کلاس‌ها یا خوشه‌ها تقسیم می‌کنند. هدف از بخش‌بندی داده‌ها در چنین فرآیندی جداسازی آنها به قسمی است که دو داده در یک خوشه تا حد امکان به هم شبیه و دو داده در دو خوشه‌ی متفاوت تا حد امکان از یکدیگر متمایز باشند [۲]. پارامترهای تأثیرگذار در این خوشه‌بندی عبارتند از: (۱) تعداد خوشه‌ها، (۲) تعداد باندهای تصویر فراطیفی، (۳) موقعیت مراکز خوشه‌ها، (۴) موقعیت باندهای تصویر فراطیفی و (۵) معیار شباهت. از میان این پنج پارامتر دو پارامتر تعداد خوشه‌ها و معیار شباهت در این تحقیق ثابت در نظر گرفته شدند. به بیان دیگر مقدار این پارامترها بایستی قبل از انجام الگوریتم بهینه‌سازی توده ذرات معلوم باشد. در ارتباط با معیار شباهت، در تحقیقات مختلف از جمله [۲] نشان داده شده است که فاصله SAM^۲ یکی از مناسب‌ترین معیارها برای داده‌ها با ابعاد بالا می‌باشند. از

سوی دیگر در [۳] نشان داده شد که از میان معیارهای شباهت SAM، ED^۳، SSV^۴، JMD^۵ و CC^۶ در حالتی که همسان‌سازی داده‌ها صورت پذیرفته باشد، معیار ED و SSV به نتایج بهتری نسبت به سایر معیارها منتهی خواهند شد. لازم به ذکر است که در حالتی که داده‌ها همسان‌سازی شوند فاصله‌ی SAM و ED به نتایج تقریباً یکسانی منجر خواهند شد [۴]. بعلاوه در [۳] نشان داده شد که در بعضی موارد ED در حالت همسان‌سازی به نتایج نسبتاً بهتری نسبت به SAM منتهی خواهد شد. بنابراین تصمیم گرفته شد که فاصله ED که دارای سرعت به نسبت بهتری نسبت به SAM و بالخصوص SSV می‌باشد، به عنوان معیار شباهت نهایی در نظر گرفته شود. معلوم در نظر گرفتن تعداد خوشه‌ها به دو دلیل می‌باشد. دلیل اول، وابستگی سایر پارامترها به تعداد خوشه‌ها است. دلیل دوم، وابستگی معیار ارزیابی خوشه‌بندی - تابع سازگاری الگوریتم بهینه‌سازی توده ذرات در حالات مورد بررسی در این تحقیق - به این پارامتر می‌باشد. اگرچه روش‌هایی وجود دارند که تعداد خوشه‌ها را قبل از انجام خوشه‌بندی محاسبه می‌کنند [۵]، اما در این تحقیق برای بررسی هر چه بهتر حالت‌های مختلف تعداد خوشه‌ها را معلوم در نظر می‌گیریم. بنابراین، در این تحقیق تأثیر هر یک از پارامترهای تعداد و موقعیت باندها و موقعیت خوشه‌ها در قالب یک سری حالات بررسی و در نهایت حالتی که منجر به بالاترین دقت در داده‌های فراطیفی می‌شود، مشخص خواهد شد. در ادامه به برخی از تحقیقات جدید انجام شده در این راستا اشاره خواهد شد.

در سال ۲۰۱۳، Villa و همکاران یک روش جدید برای خوشه‌بندی داده‌های فراطیفی با توان تفکیک مکانی پایین پیشنهاد نمودند [۶]. یکی از مشکلات تصاویر فراطیفی آن است که توان تفکیک مکانی آنها از حدود یک متر تا چند ده متر تغییر می‌کند. عوامل متعددی همچون محدودیت سیستم تصویربرداری و همچنین تلاش در جهت جبران کاهش کیفیت تصویربرداری ناشی از نویز سنجنده، پراکنش اتمسفری و غیره باعث پایین آمدن توان تفکیک مکانی تصاویر حاصل می‌گردد. بنابه دلایل مذکور در تصاویر فراطیفی تعداد زیادی پیکسل مختلط وجود

^۳ Euclidean Distance

^۴ Spectral Similarity Value

^۵ Jeffries-matusita Distance

^۶ Cross Correlation

^۱ Partitional Clustering

^۲ Spectral Angle Mapper

همکاران الگوریتمی با نام ACMD^۸ برای خوشه‌بندی اتوماتیک داده‌های فراطیفی به گروه‌های همگن ارائه کردند [۱۱]. این الگوریتم از یک دیدگاه بالا به پایین استفاده می‌کند. در این روش پیکسل‌های خوشه‌بندی شده به صورت تکراری به دو زیرخوشه تبدیل می‌شوند تا اینکه تعداد کلاس‌های بهینه مشخص شود. این عمل نصف شدن خوشه‌ها معمولاً زمانی متوقف خواهد شد که نسبت دو معیار تست مجاورت و تست واریانس، بزرگتر از یک حدآستانه‌ی از پیش تعیین شده گردد.

Paoli و همکاران با استفاده از الگوریتم بهینه‌سازی توده ذرات دوهدفه به‌طورهمزمان مولفه‌های آماری هر کلاس، تعداد و موقعیت باندهای بهینه و تعداد خوشه‌ها را به‌دست آورده‌اند [۱۲]. در این تحقیق با فرض توزیع گوسین برای خوشه‌ها، از log likelihood و فاصله bahatacharya به عنوان توابع سازگاری استفاده می‌شود و مولفه‌های آماری هر کلاس و تعداد باندهای بهینه به‌دست می‌آید. برای برآورد تعداد کلاس‌ها، مراحل بالا با تعداد کلاس‌های مختلف انجام شده و تعداد بهینه کلاس‌ها براساس معیار MDL^۹ مشخص می‌گردد [۱۲]. همین کار با تغییرات بسیار اندک توسط Melgani و Pasolli انجام شده است [۱۳].

یک روش دو مرحله‌ای برای خوشه‌بندی داده‌های فراطیفی توسط Senthilnath و همکاران پیشنهاد شده است [۱۴]. در این روش، در مرحله‌ی اول ابعاد داده‌ها با تکیه بر تبدیل PCA^{۱۰} کاهش داده می‌شود. سپس در مرحله‌ی دوم تعداد بهینه‌ی خوشه‌ها با تکیه بر روش NHAIS^{۱۱} به‌دست می‌آید و در نهایت داده‌ها براساس رای-گیری حداکثری^{۱۲} گروه‌بندی می‌شوند. براساس نتایج تجربی این تحقیق، روش NHAIS دارای عملکرد بهتری نسبت به روش ISODATA و AIS^{۱۳} است. Bilgin و همکاران یک روش بخش‌بندی برای تصاویر فراطیفی با تکیه بر (۱) یک معیار شباهت جدید و (۲) یک روش ارزیابی جدید مبتنی بر ماشین‌های بردار پشتیبان تک کلاسه ارائه نمودند [۱۵]. این روش با تکیه بر معیار ارزیابی‌اش

دارد. بر این اساس، Villa و همکاران دو روش براساس دو مفهوم جداسازی طیفی^۱ و خوشه‌بندی پیشنهاد کردند تا به این ترتیب نقشه‌های موضوعی را از داده‌های فراطیفی به‌دست آورند. Sun و همکاران در مقاله خود، یک روش خوشه‌بندی جزءبندی مبتنی بر الگوریتم کلونی زنبورهای مصنوعی ارائه داده‌اند [۷]. در این الگوریتم بعد از آنکه مراکز بهینه‌ی خوشه‌ها به‌دست آمد هر داده به نزدیکترین مرکز خوشه اختصاص داده می‌شود.

عزت آبادی و همکاران یک ایده‌ی جدید برای استفاده از سایر معیارهای شباهت به جای فاصله‌ی اقلیدسی در مدل‌های مختلف خوشه‌بندی فازی با نام‌های FCM^۲، PCM^۳ و FPCM^۴ پیشنهاد کرده‌اند [۸]. این اندازه‌گیری‌های شباهت شامل زاویه طیفی SAM، همبستگی طیفی SC^۵ و فاصله‌ی JMD می‌باشد. مدل‌های خوشه‌بندی فازی به همراه معیارهای شباهت مختلف بر روی داده‌های هایپریون اخذ شده از دلتای Okavango در شمال غربی Botswana به کار برده شده که نتایج، حاکی از بهبود عملکرد روش‌های فازی بر مبنای معیارهای شباهت جدید به‌ویژه JMD و SAM هستند.

Namin و همکاران خوشه‌بندی داده‌های فراطیفی مبتنی بر الگوریتم بهینه‌سازی توده ذرات را در دو فضای طیفی و MNF^۶ انجام دادند که نتایج نشان دهنده‌ی عملکرد بهتر الگوریتم خوشه‌بندی در فضای MNF هستند [۹]. دو الگوریتم توسط Niazmardi و همکاران در سال ۲۰۱۲ توسعه داده شد تا الگوریتم خوشه‌بندی KFCM^۷ را نسبت به پارامترهایش یعنی پارامترهای کرنل، موقعیت مراکز خوشه‌ها و باندهای مورد استفاده بهینه کنند [۱۰]. شایان ذکر است که در این الگوریتم‌ها، عملیات بهینه‌سازی توسط الگوریتم بهینه‌سازی توده ذرات صورت گرفت. اولین الگوریتم با هدف بهینه کردن KFCM نسبت به پارامتر کرنل و موقعیت مراکز خوشه‌ها توسعه داده شده است، در حالی که الگوریتم دوم علاوه بر بهینه کردن موقعیت مراکز خوشه‌ها، پارامتر بهینه‌ی کرنل را به همراه مجموعه‌ی بهینه باندها پیدا می‌کند. Shulman و

۸ Automatic Clustering of Multidimensional Data
 ۹ Minimum Distance Length
 ۱۰ Principle Component Analysis
 ۱۱ Niche Hierarchical Artificial Immune System
 ۱۲ Majority Voting
 ۱۳ Artificial Immune System

۱ Spectral Unmixing
 ۲ Fuzzy C-Means
 ۳ Probabilistic C-Means
 ۴ Fuzzy Probabilistic C-Means
 ۵ Spectral Correlation
 ۶ Minimum Noise Fraction
 ۷ Kernelized Fuzzy C-Means

خوشه‌بندی را به سمتی هدایت می‌کند تا علاوه بر انجام خوشه‌بندی تعداد بهینه خوشه‌ها نیز مشخص شود.

در هیچ یک از تحقیقات صورت پذیرفته تا به حال، تمامی پارامترهای موثر خوشه‌بندی جزءبندی در داده‌های فراطیفی به صورت یکجا و به این شکل مورد بحث و بررسی قرار نگرفته است. در این تحقیق برای اولین بار تمامی پارامترهای موثر خوشه‌بندی داده‌های فراطیفی که در ابتدای مقدمه بیان شدند به صورت یکجا در نظر شدند و به دنبال این هستیم که با معرفی چهار حالت منطقی از میان حالت‌های مختلف ممکن، حالتی که به بالاترین دقت در خوشه‌بندی جزءبندی منتهی می‌شود را مشخص کنیم. در این مقاله سعی شده به یکی از سوالات پرتکرار که چه تعداد پارامتر و به چه دلیل برای بهینه‌سازی خوشه‌بندی داده‌های فراطیفی مناسب هستند پاسخ داده شود. در واقع، هدف اساسی و اصلی این تحقیق بررسی تاثیر وارد کردن هر یک از پارامترهای تعداد و موقعیت مراکز خوشه‌ها و تعداد و موقعیت باندها در الگوریتم بهینه‌سازی توده ذرات (در قالب حالت‌های مختلف) جهت بهبود دقت خوشه‌بندی و نهایتاً ارائه یک روند نوین، قابل اطمینان، پایدار و رو به جلو برای خوشه‌بندی تصاویر ماهواره‌ای فراطیفی است.

در ادامه پس از تشریح الگوریتم بهینه‌سازی توده ذرات، به بهینه‌سازی خوشه‌بندی داده‌های فراطیفی بر اساس این الگوریتم در قالب حالات مختلف پرداخته شده است. سپس داده‌های مورد استفاده جهت پیاده‌سازی الگوریتم‌ها، پیش‌پردازش‌های موردنیاز و معیار ارزیابی الگوریتم‌ها معرفی شده‌اند. پس از آن، بخش نتایج و بحث به ارائه‌ی نتایج حاصل از پیاده‌سازی الگوریتم‌ها بر روی داده‌ها و بحث و بررسی در این خصوص اختصاص می‌یابد. در انتها، نتیجه‌گیری تحقیق ارائه شده است.

۲- الگوریتم‌ها و داده‌های مورد استفاده

۲-۱- الگوریتم بهینه‌سازی توده ذرات

الگوریتم بهینه‌سازی توده ذرات^۱ یک روش بهینه‌سازی جمعیت مینا، الهام گرفته شده از رفتار جمعی پرندگان یا

ماهی‌ها است که بوسیله Eberhart و Kennedy توسعه داده شده است [۱۶]. در الگوریتم بهینه‌سازی توده ذرات، هر ذره پتانسیل کامل حل مسئله را دارد و یک جمعیت از تعدادی ذره تشکیل شده است که بیانگر کاندیداهای مختلف حل مسئله است. فضای راه حل مسئله به عنوان فضای جستجو به گونه‌ای فرمول‌دهی می‌شود که هر موقعیت در فضای جستجو یک راه‌حل برای مسئله می‌باشد. ذرات برای پیدا کردن بهترین موقعیت در فضای جستجو با هم همکاری می‌کنند تا نهایتاً بر اساس تابع سازگاری^۲ در حل مسئله به بهترین جواب برسند. در این الگوریتم هر ذره مطابق با معادلات سرعت و موقعیت، اشاره شده در روابط ۱ و ۲، در فضای جستجو حرکت می‌کند.

$$v_{id}(t+1) = wv_{id}(t) + c_1r_1(p_{id}(t) - x_{id}(t)) + c_2r_2(p_{gd}(t) - x_{id}(t)) \quad (1)$$

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1) \quad (2)$$

در رابطه ۱ و ۲، $x_{id}(t)$ موقعیت ذره i در زمان t ، $v_{id}(t)$ سرعت ذره i در زمان t ، $p_{id}(t)$ بهترین موقعیت پیدا شده برای ذره i تا زمان t و $p_{gd}(t)$ بهترین موقعیت پیدا شده برای کل جمعیت^۴ تا زمان t است، در تمامی متغیرهای تعریف شده در بالا d نشان دهنده ابعاد داده است. همچنین w ، یک وزن اینرسی مقیاس دهنده به سرعت مرحله قبل، c_1 و c_2 دو ضریب شتاب برای مقیاس‌دهی به مولفه شناختی فردی^۵ و اجتماعی^۶ در محاسبه سرعت می‌باشند و r_1 و r_2 مقادیر تصادفی بین صفر و یک هستند.

لازم به ذکر است که اگر ذرات داری ماهیت باینری باشند از روابط ۳ و ۴ برای به‌روزرسانی موقعیت استفاده می‌شود.

$$s(v_{id}) = \frac{1}{1 + \exp(-v_{id})} \quad (3)$$

$$x_{id}(t) = \begin{cases} 1, & \text{if } \rho_{id} \leq s(v_{id}) \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

^۲ Fitness Function

^۳ Pbest (Personal Best)

^۴ Gbest (Global Best)

^۵ Personal Component

^۶ Social Component

^۱ PSO (Particle Swarm Optimization)

جدول ۱- حالت‌های مختلف خوشه‌بندی مورد بررسی در این مقاله

حالت‌ها	موقعیت خوشه‌ها	تعداد باندها	موقعیت باندها
حالت اول	نامعلوم	معلوم-کل باندها	معلوم-کل باندها
حالت دوم	نامعلوم	معلوم-کاهش باند	معلوم-کاهش باند
حالت سوم	نامعلوم	معلوم-بعد ذاتی	نامعلوم
حالت چهارم	نامعلوم	نامعلوم	نامعلوم

۲-۲-۱- حالت اول

همانطور که در جدول ۱ مشخص است، در این حالت همه پارامترها بجز موقعیت مراکز خوشه‌ها ثابت در نظر گرفته شد و هدف در بهینه‌سازی خوشه‌بندی پیدا کردن موقعیت بهینه‌ی مراکز خوشه‌هاست. در این حالت هر یک از معیارهای بین کلاسی یا معیارهای داخل کلاسی می‌توانند به‌عنوان تابع سازگاری استفاده شوند. در این راستا، در همه‌ی حالات، رابطه‌ی ۵ به‌عنوان تابع سازگاری در نظر گرفته شد.

$$intra = \frac{1}{N_p} \sum_{k=1}^K \sum_{u \in C_k} dist(u, m_k) \quad (5)$$

در رابطه‌ی بالا، N_p تعداد کل داده‌ها، K تعداد کل خوشه‌ها، C_k خوشه‌ی k ام، $u \in C_k$ داده‌ی متعلق به خوشه‌ی k ام است. همچنین m_k موقعیت مرکز خوشه‌ی k ام و $dist$ معیار اندازه‌گیری است. در ارتباط با ماهیت ذرات همانطور که مشخص است هر یک از ذرات، موقعیت مراکز خوشه‌ها را نشان می‌دهند که در ابتدای الگوریتم به صورت تصادفی از میان داده‌ها مقداردهی اولیه شدند. شکل ۱ نحوه‌ی معرفی ذرات را نشان می‌دهد.

$$x_{id} = [CC_{i1} \quad \dots \quad CC_{ip} \quad \dots \quad CC_{ik} \quad \dots \quad \dots \quad CC_{i(k+p)}]$$

شکل ۱- معرفی ذره در بهینه‌سازی خوشه‌بندی داده‌های فراطیفی (حالت اول)

در شکل فوق k و p به ترتیب بیانگر تعداد خوشه‌ها و باندها هستند. نحوه‌ی عملکرد الگوریتم بهینه‌سازی توده ذرات در این حالت در شکل ۲ نشان داده شده است.

در مسایل بهینه‌سازی مبتنی بر الگوریتم بهینه‌سازی توده ذرات، دو جزء اساسی با نام‌های جمعیت و تابع سازگاری وجود دارد. جمعیت شامل تعدادی ذره می‌باشد که هر یک از این ذرات پتانسیل کامل حل مسئله را دارد. تابع سازگاری نیز میزان مطلوبیت هر ذره را مشخص می‌کند. لازم به ذکر است که برای هر مسئله، نوع تابع سازگاری و ماهیت ذرات متفاوت است [۱۷].

۲-۲-۲- بهینه‌سازی خوشه‌بندی داده‌های فراطیفی

مبتنی بر الگوریتم بهینه‌سازی توده ذرات

از آنجایی که هدف این تحقیق بهینه‌سازی و بهبود خوشه‌بندی داده‌های فراطیفی است، در ابتدا باید نوع خوشه‌بندی مورد استفاده مشخص شود. همانگونه که توضیح داده شد از میان روش‌های مختلف خوشه‌بندی، روش جزءبندی پرکاربردترین روش برای خوشه‌بندی داده‌های فراطیفی است. برای بهبود این خوشه‌بندی باید پارامترهای آن را شناخته و سعی در یافتن مقدار بهینه آنها داشت. پارامترهای این خوشه‌بندی را می‌توان به پنج دسته پارامتر (۱) معیار شباهت، (۲) موقعیت بهینه مراکز خوشه‌ها، (۳) تعداد بهینه خوشه‌ها (۴) تعداد بهینه باندها و (۵) موقعیت بهینه باندها تقسیم کرد. از میان این پنج پارامتر، تابع بهینه شباهت برای این نوع داده‌ها برابر با معیار SAM در نظر گرفته شد. بعلاوه، پارامتر تعداد خوشه‌ها نیز ثابت در نظر گرفته شد. دلیل این کار، وابستگی سایر پارامترها به تعداد خوشه‌ها است. بنابراین در این مطالعه به دنبال یافتن مقادیر بهینه برای سایر پارامترها و ارائه بهترین حالت برای خوشه‌بندی داده‌های فراطیفی خواهیم بود.

با توجه به توضیحات، همانگونه که مشخص است برای بهبود خوشه‌بندی داده‌های فراطیفی بایستی سه پارامتر موقعیت خوشه‌ها، تعداد باندها و موقعیت باندها بهینه شوند که این بهینه‌سازی می‌تواند به صورت همزمان یا جداگانه صورت پذیرد. برای این منظور در این تحقیق حالات مختلفی در نظر گرفته شد تا بررسی شود کدامیک منجر به بهبود خوشه‌بندی با دقت بالاتری خواهند شد. حالت‌های مختلف مورد بررسی در این تحقیق در جدول ۱ نشان داده شده است.

۲-۲-۲-۱ دیدگاه اول: کاهش باند مبتنی بر خوشه‌بندی داده‌ها

در این حالت بعد از تعیین تعداد باندها توسط روش RACPA، موقعیت باندها با استفاده از الگوریتم خوشه‌بندی k-means مشخص گردید. به بیان دیگر کاهش باند در فضای اصلی داده‌ها رخ داده است. در این حالت نیز عملیات بهینه‌سازی با استفاده از الگوریتم بهینه‌سازی توده ذرات و مشابه با حالات اول صورت گرفت. ذرات نیز همانند حالت قبل طبق شکل ۱ به دست می‌آیند.

۲-۲-۲-۲ دیدگاه دوم: کاهش باند بر پایه‌ی PCA

در این حالت بعد از تعیین تعداد باندها توسط روش RACPA، ابعاد داده‌ها با تکیه بر روش PCA کاهش داده شد. تابع سازگاری و ذرات الگوریتم بهینه‌سازی در این حالت همانند حالت ۱، ۲، ۲، ۲ معرفی گردید.

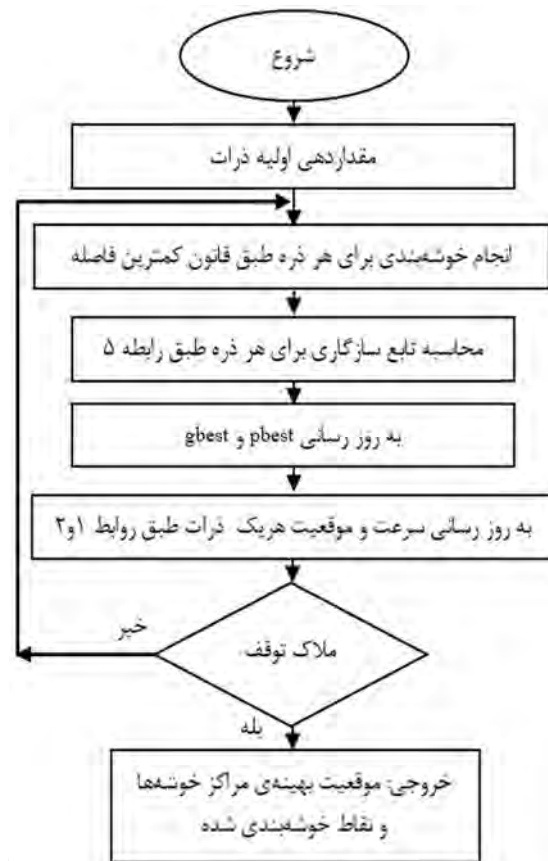
۲-۲-۳ حالت سوم

طبق جدول ۱، در این حالت علاوه بر تعداد خوشه‌ها، تعداد باندها نیز معلوم در نظر گرفته شد. تفاوتی که این حالت با حالت‌های اول و دوم دارد این است که در این حالت موقعیت بهینه‌ی باندها همزمان با موقعیت بهینه‌ی مراکز خوشه‌ها به دست می‌آید. در این حالت نیز از تابع سازگاری ارائه شده در رابطه‌ی ۵ استفاده گردید. لازم به ذکر است که هر یک از ذرات به صورت شکل ۳ معرفی شدند.

$$x_{id} = [V_{i1} \dots V_{iD} \quad CC_{i1} \dots CC_{ip} \dots CC_{ik} \dots CC_{i(k+p)}]$$

شکل ۳- نحوه‌ی معرفی ذرات در بهینه‌سازی خوشه‌بندی در داده‌های فراطیفی (حالت سوم)

در شکل ۳، V_{iD} نشان‌دهنده‌ی بعد ذاتی داده‌ها- تعداد باندها- است. شکل ۴ نحوه‌ی بهینه‌سازی خوشه‌بندی را در این حالت نشان می‌دهد.



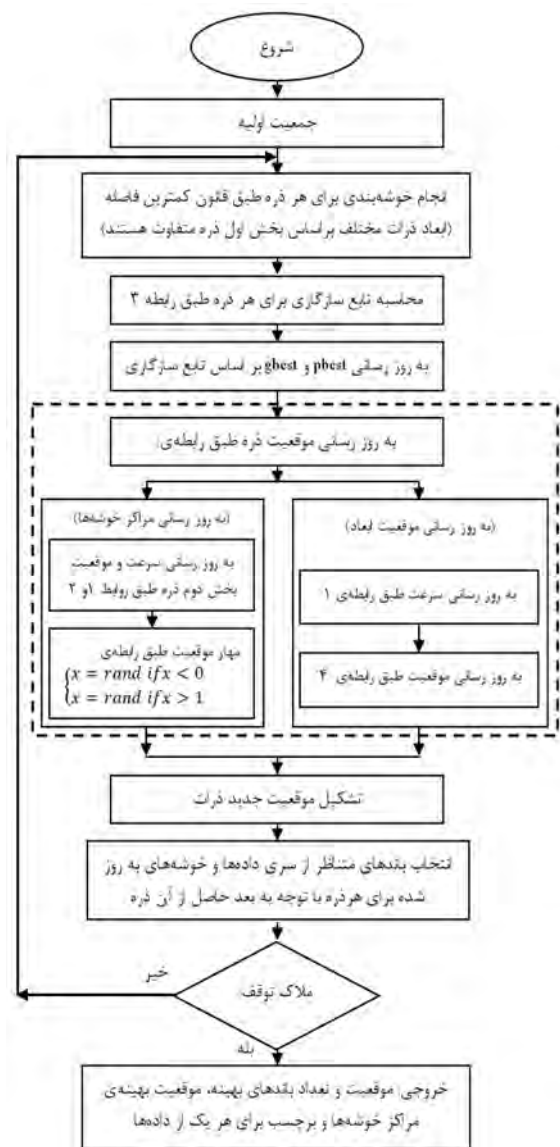
شکل ۲- بهینه‌سازی خوشه‌بندی داده‌های فراطیفی در حالت اول

۲-۲-۲-۲ حالت دوم

همانطور که از جدول ۱ مشخص است برای این حالت نیز مشابه حالت اول، تعداد و موقعیت باندها ثابت در نظر گرفته شد، با این تفاوت که در اینجا از روش کاهش باند استفاده شده است. در ارتباط با تعداد باندها از هر روش مبتنی بر بعد ذاتی می‌توان استفاده کرد که در این مطالعه روش نوین RACPA^۱ مورد استفاده قرار گرفت [۵]. در واقع، RACPA یک روش جدید مبتنی بر آنالیز تغییرات برای تعیین تعداد عناصر خالص موجود در تصاویر فراطیفی است. بعد از اینکه تعداد باندها با تکیه بر روش‌های مبتنی بر بعد ذاتی پیدا شد موقعیت باندها در یک مرحله‌ی پیش‌پردازشی با توجه به تعداد آنها به دست آمد. البته این موقعیت می‌تواند در فضای داده‌ها یا فضای دیگری مثل فضای حاصل از PCA به دست آید. بر این اساس و با توجه به دو فضای مختلف مطرح شده، این حالت خود به دو زیر بخش تقسیم گردید که هر یک از آنها در ادامه توضیح داده خواهد شد.

۱ Residual Analysis- Change Point Analysis

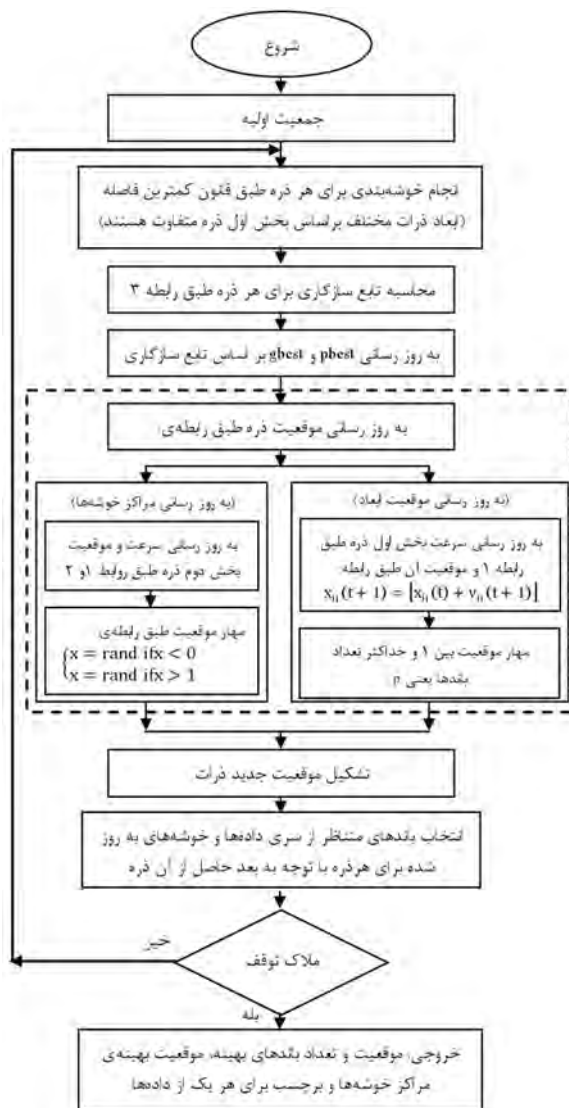
ذره شامل k مرکز خوشه است که هر یک از این مراکز خوشه‌ها خود دارای d بعد هستند. لازم به ذکر است که d متغیر نهایی هر ذره، دارای مقادیر دودویی ۰ یا ۱، بیانگر تعداد باندهای انتخاب شده یا نشده می‌باشد، به این ترتیب که اگر مقدار متغیر برابر ۰ بود آن بعد در نظر گرفته نمی‌شود و اگر مقدار متغیر ۱ بود آن بعد برای آن ذره در نظر گرفته می‌شود. نحوه عملکرد این الگوریتم در شکل ۶ نشان داده شده است.



شکل ۶- حل همزمان خوشه‌بندی و انتخاب باندهای بهینه تصویر فراطیفی توسط الگوریتم بهینه‌سازی توده ذرات

۳-۲-۳- داده‌های فراطیفی

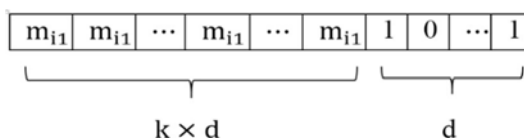
جهت ارزیابی صحت روش‌های اجرایی مختلف در این تحقیق، ابتدا از یک داده شبیه‌سازی شده استفاده شد. همچنین دو داده واقعی شناخته‌شده یعنی Pavia



شکل ۴- حل همزمان خوشه‌بندی و انتخاب باندهای بهینه تصویر فراطیفی توسط الگوریتم بهینه‌سازی توده ذرات در حالت سوم

۲-۲-۴- حالت چهارم

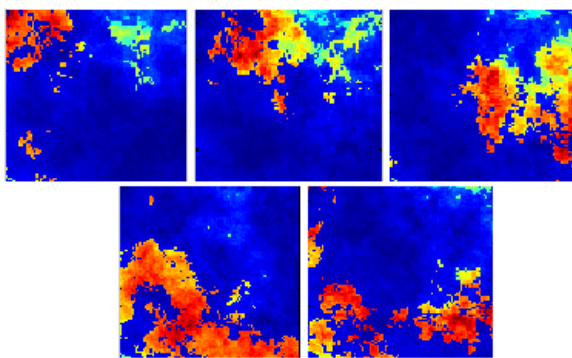
در این حالت هر سه پارامتر خوشه‌بندی جزءبندی طی فرآیند بهینه‌سازی به‌دست آمدند. نحوه معرفی ذرات در این حالت طبق شکل ۵ صورت گرفت.



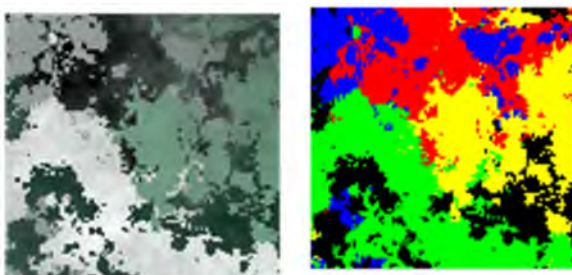
شکل ۵- نمایش یک ذره در بهینه‌سازی خوشه‌بندی داده‌های فراطیفی (حالت چهارم)

اگر به تعداد k مرکز خوشه داشته باشیم، تعداد متغیرهای هر ذره برابر $(k + 1) \times d$ می‌باشد؛ چون هر

http://www.ehu.es/ccwintco/index.php/Hyperspectral_Imagery_Synthesis_tools_for_MATLAB ایجاد شدند [۱۸]، که تمامی ویژگی‌های فراوانی عناصر خالص، اعم از نامنفی بودن فراوانی‌ها و نیز یک بودن مجموع فراوانی‌ها در هر پیکسل را دارا می‌باشند. شکل ۸ نقشه‌های فراوانی مورد استفاده در شبیه‌سازی داده‌های فراطیفی را نشان می‌دهد. این نقشه‌ها شامل پنج نقشه فراوانی متناظر با پنج عنصر خالص و دارای ابعاد ۱۰۰ سطر در ۱۰۰ ستون هستند. در شکل ۹ یک ترکیب رنگی کاذب از تصویر مورد مطالعه و همچنین نقشه واقعیت زمینی منطقه ارائه شده است.



شکل ۸- نقشه فراوانی پنج عنصر خالص شبیه‌سازی شده



الف	ب
Muscovite_GDS116 (MG)	Desert_Vamish_GD141 (DVG)
Montmorillonite_CM20 (MCM)	Kaolinite_CM9 (KCM)
Andradite_WS488 (AWS)	

شکل ۹- الف) نقشه واقعیت زمینی ب) ترکیب رنگی منطقه برای داده شبیه‌سازی شده

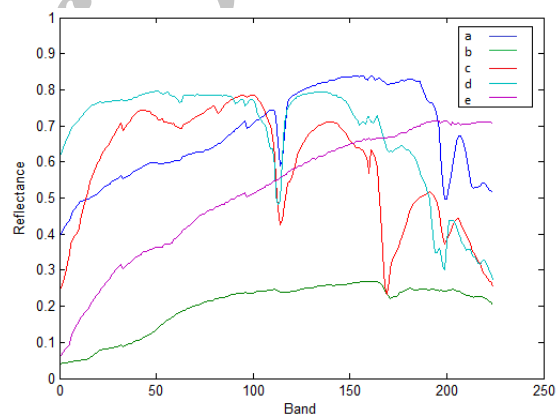
۲-۳-۲- داده‌های واقعی

الف: داده Pavia University: این داده در سال ۲۰۰۳ توسط سنجنده ROSIS با پوشش طیفی متغیر از ۰/۴۳ تا ۰/۸۶ میکرومتر و توان تفکیک مکانی در حدود ۱/۳ متر اخذ شد. مجموعه‌ای که در این مطالعه مورد استفاده قرار گرفته است محدوده‌ای با ۶۱۰ در ۳۴۰ پیکسل از دانشگاه

University و Salinas-A از دو سنجنده‌ی مختلف فراطیفی مورد استفاده قرار گرفت تا کارایی الگوریتم‌های مورد استفاده در این تحقیق را نشان دهند.

۲-۳-۱- داده‌ی شبیه‌سازی

از آنجاییکه داده شبیه‌سازی شده از عناصر خالص با مشخصه طیفی و نقشه‌های فراوانی معلوم ساخته شده است، بنابراین می‌توان از آن برای ارزیابی نتایج الگوریتم‌ها استفاده نمود. این داده‌ی شبیه‌سازی شده شامل یک مکعب داده با ابعاد ۱۰۰ سطر در ۱۰۰ ستون در ۲۲۴ باند طیفی است. ماتریس مشخصه طیفی عناصر خالص دارای پنج عنصر خالص متمایز در ۲۲۴ باند طیفی است. مشخصه‌های طیفی و نام این مواد که از کتابخانه طیفی سازمان زمین‌شناسی ایالات متحده انتخاب شده‌اند به ترتیب در شکل ۷ و جدول ۲ نشان داده شده‌اند.



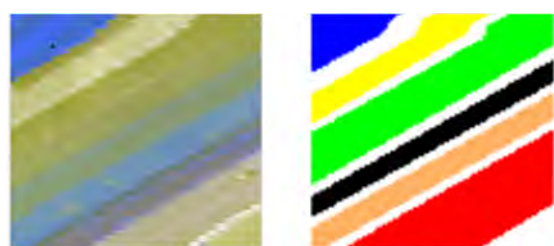
شکل ۷- مشخصه طیفی پنج عنصر خالص

جدول ۲- نام عناصر خالص مورد استفاده در داده شبیه‌سازی شده

نام عنصر خالص	برچسب‌ها در شکل ۷
Muscovite_GDS116	(a)
Desert_Vamish_GD141	(b)
Montmorillonite_CM20	(c)
Kaolinite_CM9	(d)
Andradite_WS488	(e)

از عناصر خالص در جدول ۲ با وزن‌های مختلف، برای ایجاد پیکسل‌های مختلط استفاده شد. وزن‌های اعمال شده برای هر عنصر خالص در هر پیکسل، همان نقشه فراوانی عناصر خالص است. در اینجا نقشه‌های فراوانی توسط جعبه ابزار تولید نقشه فراوانی قابل دسترس در

محاسبات به دلیل تنظیم پارامترهای الگوریتم توده ذرات است. برای مثال می‌توان به پارامترهای مربوط به محدودیت مکان اشاره کرد که در صورت همسان‌سازی داده‌ها به راحتی می‌توان از خروج موقعیت مراکز خوشه‌ها از محدوده‌ی مورد بررسی جلوگیری کرد. شایان ذکر است که همسان‌سازی داده‌ها به صورت خطی انجام شده است به طوری که مقادیر در بازه‌ی [0,1] قرار بگیرند. همچنین در داده‌های واقعی، پیکسل‌های مربوط به پس‌زمینه در نظر گرفته نشده‌اند، زیرا اطلاعاتی از آنها در دسترس نبود و طبیعتاً در صورت استفاده از آنها، امکان ارزیابی الگوریتم‌های پیشنهادی فراهم نبود.



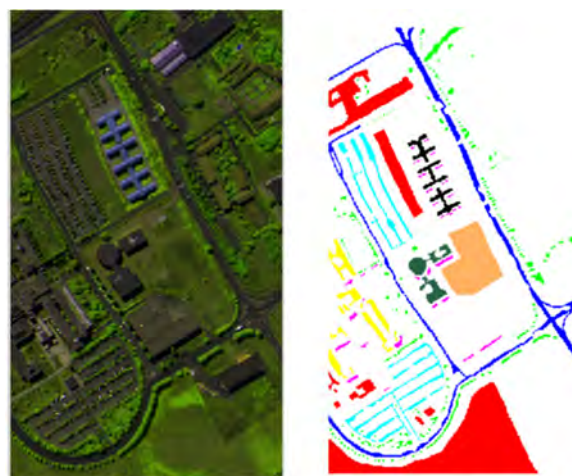
شکل (۱- الف) نقشه واقعیت زمینی (ب) ترکیب رنگی منطقه برای داده‌ی Salinas-A

۲-۵- معیار ارزیابی

برای ارزیابی نتایج روش پیشنهادی در این تحقیق، به منظور برآورد صحت خوشه‌بندی از ماتریس خطا به کمک داده‌های واقعیت زمینی استفاده شد. اگرچه با استفاده از ماتریس خطا می‌توان معیارهای مختلفی برای ارزیابی طبقه‌بندی ارائه کرد، در این تحقیق مولفه‌ی ضریب کاپا که در رابطه ۶ ارائه شده به عنوان یکی از پرکاربردترین معیارهای ارزیابی مورد استفاده قرار گرفت [۱۰].

$$\hat{k} = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} \times x_{+i})}{N^2 - \sum_{i=1}^r (x_{i+} \times x_{+i})} \quad (6)$$

پایا در ایتالیا است. لازم به ذکر است که این داده‌ها از لحاظ اتمسفری تصحیح شده و بعد از حذف باندهای نویزی ۱۰۳ باند طیفی دارد. شکل ۱۰ یک ترکیب رنگی کاذب از تصویر مورد مطالعه و همچنین نقشه واقعیت زمینی مربوط به آن را نشان می‌دهد [۱۹].



شکل (۱۰- الف) نقشه واقعیت زمینی (ب) ترکیب رنگی منطقه برای داده Pavia university

ب: داده Salinas-A: دومین داده واقعی مورد استفاده در این مطالعه داده‌ی فراطیفی Salinas-A است که توسط سنجنده AVIRIS از منطقه‌ی Salinas در کالیفرنیا آمریکا اخذ شده است. توان تفکیک مکانی این داده ۳/۷ متر و تعداد باندهای آن برابر ۲۰۴ است. تصویر مورد نظر شامل ۸۳ در ۸۶ پیکسل و متشکل از ۶ کلاس است. در شکل ۱۱ یک ترکیب رنگی کاذب از تصویر مورد مطالعه و همچنین نقشه واقعیت زمینی منطقه داده شده است. دو داده فراطیفی واقعی اشاره شده در بالا در آدرس <http://www.ehu.es/ccwintco/index.php/Hyperspectr> al_Remote_Sensing_Scenes قابل دسترسی هستند.

۲-۴- پیش‌پردازش داده‌ها

پیش از پیاده‌سازی مدل‌های خوشه‌بندی، همسان‌سازی داده‌ها بسیار مهم و ضروری است. مزیت اصلی این کار در این است که از تاثیر داده در باندها با محدوده عددی بالاتر بر داده‌ها با محدوده عددی پایین‌تر جلوگیری می‌شود. مزیت دیگر آن جلوگیری از پیچیده شدن

۳- نتایج و بحث

۳-۱- تنظیم پارامترها

در ارتباط با الگوریتم‌های بهینه‌سازی توده ذرات اعم از تک‌هدفه و چند هدفه چندین پارامتر باید از قبل تنظیم شوند. اولین پارامتر این الگوریتم‌ها تعداد تکرار یا معیار توقف الگوریتم است. معیارهای ارزیابی مختلفی برای توقف الگوریتم‌های فراابتکاری مثل توده ذرات معرفی شده‌اند [۲۱]. در این مطالعه اگر تفاضل تابع سازگاری مربوط به بهترین ذره در ۲۰ تکرار متوالی تغییر نکند، الگوریتم متوقف خواهد شد. برای سایر پارامترهای این الگوریتم تنظیمات مندرج در جدول ۳ استفاده شده‌اند. الگوریتم مورد استفاده‌ی دیگر در این مطالعه، الگوریتم RACPA است که پارامترها و مقادیر پیش‌فرض مربوط به آن نیز در جدول ۳ آورده شده است.

جدول ۳- مقادیر پارامترهای مورد استفاده در این تحقیق

الگوریتم	پارامتر	مقدار
SOPSO	P_size	20
	W	0.72
	$C1$	1
	$C2$	1
k-means	حداکثر تعداد تکرار	100
RACPA	سطح اطمینان برای تعیین نقاط تغییر	0.95
	بازه اطمینان اطراف هر نقطه تغییر	0.95
	روش جایگزینی بدون جایگزینی	روش نمونه‌برداری

۳-۲- ارزیابی نتایج در حالات مختلف

در ادامه نتایج هر یک از حالتها به صورت دقیق مورد بحث و بررسی قرار گرفته است.

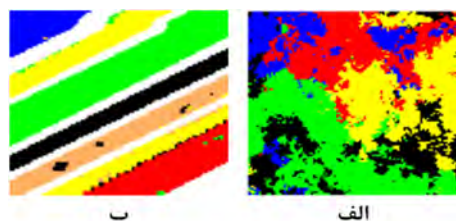
۳-۲-۱- ارزیابی نتایج در حالت اول

با توجه به جدول ۱، هدف در این حالت بهبود خوشه‌بندی داده‌های فراطیفی مبتنی بر یافتن موقعیت بهینه‌ی مراکز خوشه‌ها می‌باشد. نتایج حاصل از این خوشه‌بندی بر روی سه داده مورد مطالعه در این تحقیق

در جدول ۴ نشان داده شده‌اند. نقشه‌های پوشش زمین حاصل از انجام خوشه‌بندی در این حالت در شکل ۱۲ آورده شده‌اند.

جدول ۴- بهبود خوشه‌بندی جزءبندی داده‌های فراطیفی در حالت اول

نام داده	جواب با بالاترین دقت	جواب با بیشترین تکرار	بیشترین تکرار	جواب با پایین‌ترین دقت
Salinas-A	78.68	78.68	17	68.00
Pavia University	55.27	47.00	5	32.15
داده‌ی شبیه‌سازی	98.63	98.63	30	98.63



شکل ۱۲- نقشه‌ی خوشه‌بندی در حالت اول برای الف) شبیه‌سازی، ب) Salinas-A، ج) Pavia University

در این حالت نتایج الگوریتم k-means در ۳۰ بار تکرار در جدول ۵ آورده شده تا با نتایج این حالت مقایسه شود. لازم به ذکر است که در این حالت مقادیر اولیه الگوریتم k-means از بین داده‌ها به صورت تصادفی انتخاب شدند. همانطور که از جداول ۴ و ۵ مشخص است الگوریتم ارائه شده در این حالت در داده‌ی Pavia به دقت بهتری نسبت به k-means منجر شد. در ارتباط با داده‌های Salinas-A و شبیه‌سازی اگرچه دقت هر دو روش در ۳۰ بار تکرار یکسان است ولی روش ارائه شده در این حالت دارای پایداری بسیار بیشتری نسبت به الگوریتم k-means

جدول ۷- بهبود خوشه‌بندی جزءبندی داده‌های فراطیفی در حالت دوم (با تکیه بر انتقال داده‌ها به فضای حاصل از PCA)

نام داده	جواب با	جواب با	بیشترین تکرار
	بالاترین دقت	بیشترین تکرار	
Salinas-A	83.50	80.08	7
Pavia University	43.79	37.83	5
داده‌ی شبیه‌سازی	99.97	99.97	30

همانگونه که در جداول ۶ و ۷ دیده می‌شود در حالی که ابعاد داده‌ها کاهش پیدا کند الگوریتم بهینه‌سازی توده ذرات برای همی داده‌ها به پایداری و صحت بالاتری منجر خواهد شد. البته در ارتباط با داده‌های Pavia این امر صادق نبود؛ به این معنا که روش بهینه‌سازی خوشه‌بندی در این حالت به دقت بالاتری نسبت به حالت اول منتهی نشد. دلیل آنکه خوشه‌بندی داده‌های کاهش باند یافته توسط روش خوشه‌بندی باندها در این حالت به جواب بهتری منجر نشده است را می‌توان ناشی از دو عامل مهم دانست که در ادامه ذکر خواهند شد. اولاً خود خوشه‌بندی باندها با استفاده از روش *k-means*، تابعی از مقدار-دهی اولیه مراکز خوشه‌ها می‌باشد که عملکرد الگوریتم را تحت تاثیر قرار خواهد داد. ثانیاً بعد ذاتی داده‌ها یک مسئله کاربردی می‌باشد که بسته به کاربرد می‌تواند تغییر کند. به بیان دیگر ممکن است این روش با تعداد باندهای دیگر به جواب بهتری منجر شود. از آنجایی که در این تحقیق الگوریتم پیشنهادی ما برای تعیین بعد ذاتی یک سری نماینده برای این مقدار معرفی می‌کند ما مورد دوم را بررسی خواهیم کرد. در این راستا وقتی برای داده‌ی Pavia بعد ذاتی برابر با ۱۸ در نظر گرفته شد ضریب کاپا برای خوشه‌بندی این داده برابر با ۵۵/۲۱ به دست آمد. در ارتباط با مشکل اول نیز می‌توان از الگوریتم‌های مختلف بهینه‌سازی یا الگوریتم‌هایی که برای تعیین مقادیر اولیه خوشه‌بندی طراحی شده‌اند [۲۲]، استفاده کرد که در بخش پیشنهادات نیز به این مهم اشاره شده است. همانطور که در جدول ۸ نشان داده شده است حالت دوم عموماً نه‌تنها به نتایج بهتری نسبت به حالت اول منجر خواهد شد بلکه باعث کاهش حجم داده‌ها و بالا رفتن سرعت الگوریتم‌های مورد استفاده نیز خواهد شد.

بود که این امر را می‌توان در تعداد جوابها با بیشترین تکرار و جواب با پایین‌ترین دقت به‌وضوح دید.

جدول ۵- نتایج حاصل از خوشه‌بندی داده‌های مورد مطالعه در این تحقیق توسط *k-means*

نام داده	جواب با	جواب با	بیشترین تکرار	پایین‌ترین دقت
	بالاترین دقت	بیشترین تکرار		
Salinas-A	78.68	62.49	8	33.44
Pavia University	40.38	27.14	12	25.92
داده‌ی شبیه‌سازی	98.63	98.63	19	37.78

با توجه به جداول ۴ و ۵، الگوریتم پیشنهادی در این بخش برای داده‌ی شبیه‌سازی با SNR^2 بالا و الگوی سیگنال مشخص به بالاترین دقت و پایداری منتهی شد. در نقطه مقابل، الگوریتم برای داده‌های واقعی بخصوص *pavia* به دلیل حجم بالای آنها از یک طرف و الگوهای اطلاعاتی مختلف و بعضاً نامشخص از طرف دیگر به دقت و پایداری پایین‌تری منجر شد.

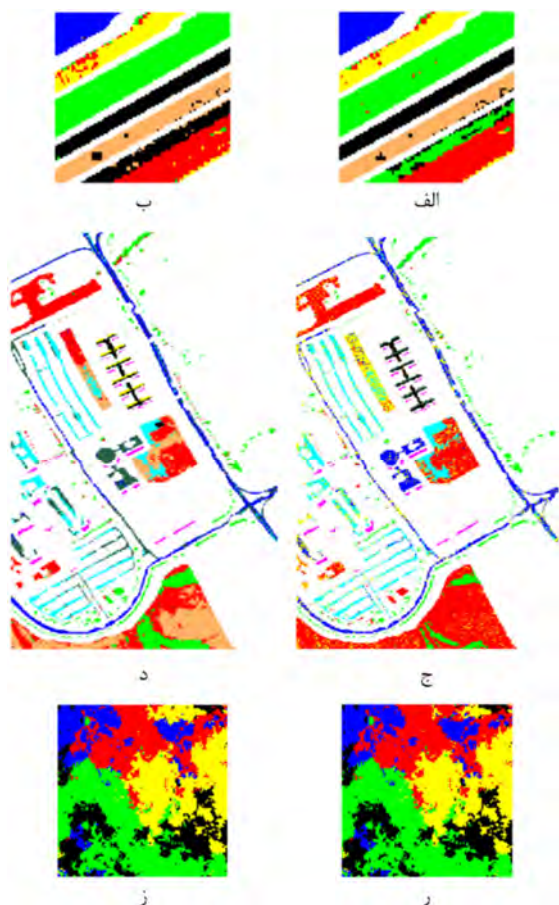
۳-۲-۲-۳- ارزیابی نتایج در حالت دوم

با توجه به جدول ۱، در این حالت بعد از اینکه در یک مرحله‌ی پیش‌پردازشی ابعاد داده کاهش پیدا کرد بهبود خوشه‌بندی با تکیه بر یافتن موقعیت بهینه‌ی خوشه‌ها انجام شد. در جداول ۶ و ۷ نحوه‌ی عملکرد خوشه‌بندی در حالت دوم برای سه داده‌ی *Salina-A*، *Pavia University* و شبیه‌سازی شده نشان داده شده است. جدول ۶ بر اساس کاهش بعد داده‌ها به روش خوشه‌بندی باندها و جدول ۷ بر اساس کاهش بعد داده‌ها به روش PCA می‌باشند.

جدول ۶- بهبود خوشه‌بندی جزءبندی داده‌های فراطیفی در حالت دوم (با تکیه بر مفهوم خوشه‌بندی باندها)

نام داده	جواب با	جواب با	بیشترین تکرار
	بالاترین دقت	بیشترین تکرار	
Salinas-A	82.08	79.42	4
Pavia University	43.38	43.38	9
داده‌ی شبیه‌سازی	97.97	97.97	30

^۲ Signal-to-Noise Ratio



شکل ۱۳- نقشه‌ی خوشه‌بندی در حالت دوم برای الف: Salinas-A با تکیه بر تبدیل PCA (ب) Salinas-A با تکیه بر مفهوم خوشه‌بندی باندها (ج) Pavia با تکیه بر تبدیل PCA (د) Pavia با تکیه بر مفهوم خوشه‌بندی باندها (ر) داده‌ی شبیه‌سازی با تکیه بر تبدیل PCA (ز) داده شبیه‌سازی با تکیه بر مفهوم خوشه‌بندی باندها

همانگونه که از نتایج نیز مشخص است نتایج در این حالات داری دقت و صحت بالایی نسبت به حالات قبل نیستند که البته انتظار آن به دلیل وابستگی بین پارامترهای موقعیت باند و تابع سازگاری می‌رفت. به این معنا که یک مجموعه از باندها به خاطر مقادیر کمتر بازتاب در آن محدوده داری مقدار کمتری بوده و در نتیجه میزان تابع سازگاری را به صورت کاذب کم خواهند کرد. لازم به ذکر است که همسان‌سازی داده‌ها این مشکل را نمی‌تواند حل کند به این دلیل که در همان بازه بین صفر تا ۱ باز همان باندهایی که بازتابش کمتری داشتند مقادیر کمتری را به خود خواهند گرفت. یکی از راه‌حلهایی که می‌توان برای بهبود صحت در این بخش متصور شد استفاده از یک معیار خارجی مثل ضریب کاپا به‌عنوان تابع سازگاری است [۲۴]. البته با اینکار روش خوشه‌بندی از حالت خودکار به نیمه‌خودکار تبدیل می‌شود چون برای

جدول ۸- مقایسه حالت دوم با حالت اول

نام داده	روش کاهش حجم داده‌ها	بهبود دقت	درصد کاهش
Salinas-A	خوشه‌بندی باندها	0.04	98%
	انتقال به فضای حاصل از PCA	0.06	98%
Pavia University	خوشه‌بندی باندها	تقریباً برابر	77%
	خوشه‌بندی باندها	تقریباً برابر	98%
داده‌ی شبیه‌سازی شده	انتقال به فضای حاصل از PCA	0.01	98%
	خوشه‌بندی باندها	0.01	98%

با توجه به نقشه‌های مربوط به خوشه‌بندی در شکل ۱۳ نه تنها در این حالت بلکه در سایر حالات خوشه‌بندی قادر نبوده تا کلاس سنگریزه^۳ را در داده‌های Pavia تشخیص دهد. یکی از مهمترین دلایلی که این کلاس به صورت درست تشخیص داده نشده را می‌توان معیار شباهت مورد استفاده دانست. در این راستا (۱) معیار شباهت دیگر، (۲) خوشه‌بندی در فضای فاصله‌ها [۲۳] و (۳) اطلاعات اضافی مانند اطلاعات بافت یا ساختار می‌توانند مورد استفاده قرار گیرند تا این مشکل رفع شود. همانطور که در ماتریس خطاهای مربوط به داده‌ی Pavia دیده می‌شود این داده با کلاس مربوط به آجر^۴ معمولاً اشتباه می‌شود که به نظر می‌رسد با تکیه بر مفاهیم بالا قابل حل باشد.

۳-۲-۳- ارزیابی نتایج در حالت سوم

با توجه به جدول ۱، هدف در این حالت بهبود خوشه‌بندی داده‌های فراطیفی مبتنی بر بهینه‌سازی همزمان موقعیت خوشه‌ها و باندها می‌باشد. نتایج حاصل از خوشه‌بندی در این مرحله بر روی سه داده مورد مطالعه، در جدول ۹ آورده شده است.

جدول ۹- بهبود خوشه‌بندی جزءبندی داده‌های فراطیفی در حالت سوم

نام داده	جواب با بالاترین دقت	جواب با بیشترین تکرار	بیشترین تکرار
Salinas-A	69.74	57.62	5
Pavia University	29.50	27.63	3
داده‌ی شبیه‌سازی	65.16	56.65	4

^۳ Gravel

^۴ Brick

از این حالت‌ها برای سه نوع داده‌ی مورد استفاده، در جدول ۱۰ به نمایش گذاشته شده است.

بر اساس جدول ۱۰، از میان حالت‌های مختلف، حالت دوم که در آن تعداد باندها در یک مرحله‌ی پیش‌پردازشی به روش خوشه‌بندی باندها کاهش داده شده (دیدگاه ۱) به طور کلی دارای بالاترین دقت است.

جدول ۱۰- مقایسه‌ی نتایج حالت‌های مختلف بر اساس بالاترین دقت‌ها

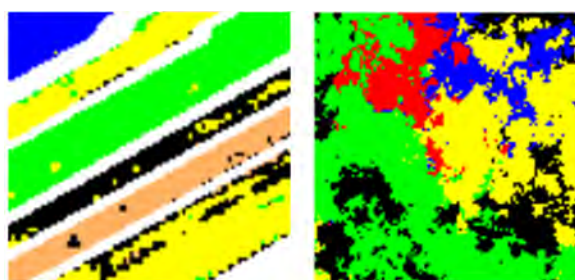
نام داده	حالت ۲			حالت ۳	k-means
	حالت ۱	دیدگاه			
		۱	۲		
Salinas-A	78.68	83.50	82.08	69.74	78.68
Pavia University	55.27	43.79	43.38	29.50	40.38
داده‌ی شبیه‌سازی	98.63	99.97	97.97	65.16	98.63

۴- نتیجه‌گیری

همان‌طور که بیان شد، خوشه‌بندی جزءبندی داده‌های فراطیفی تابع پنج پارامتر تعداد و موقعیت خوشه‌ها، تعداد و موقعیت باندها و معیار اندازه‌گیری است. در این پژوهش، تاثیر وارد کردن هر یک از سه پارامتر تعداد و موقعیت باندها و موقعیت مراکز خوشه‌ها در الگوریتم بهینه‌سازی توده ذرات جهت بهبود دقت خوشه‌بندی مورد بررسی قرار گرفت. این بررسی در قالب چهار حالت مطابق جدول ۱ انجام شد. همان‌طور که ملاحظه شد، اگرچه خوشه‌بندی جزءبندی داده‌های فراطیفی به‌طور کلی تابع پنج پارامتر است، ولی نتایج نشان دادند که حل همزمان همه پارامترها به دلیل وابستگی بعضی پارامترها به هم و پدیده‌ی فزونی پارامترها از یک طرف و وابستگی شدید بعضی از آنها با تابع سازگاری از طرف دیگر، نه تنها نمی‌تواند به بهبود خوشه‌بندی داده‌های فراطیفی منجر شود بلکه ممکن است باعث کاهش دقت آن نیز بشوند.

برای حل مشکل وابستگی بین بعضی از پارامترهای خوشه‌بندی جزءبندی و تابع سازگاری می‌توان از یک تابع سازگاری استفاده کرد که مستقل از داده‌ها باشد. در این راستا، ضریب کاپا یا هر معیار خارجی دیگر می‌تواند به عنوان تابع سازگاری استفاده شود. البته لازم به ذکر است که در این حالت دیگر روش طبقه‌بندی، نظارت نشده نبوده و به یک روش نظارت شده تبدیل می‌شود. از میان

هدایت داده‌ها به ضریب کاپا نیاز است. نقشه‌های پوشش زمین حاصل از انجام خوشه‌بندی در حالت سوم در شکل ۱۴ نمایش داده شده است.



الف ب



ج

شکل ۱۴- نقشه‌ی خوشه‌بندی در حالت سوم برای الف) داده‌ی شبیه‌سازی، ب) Salinas-A و ج) Pavia University

۳-۲-۴- ارزیابی نتایج در حالت چهارم

همان‌طور که از لحاظ نظری نیز انتظار می‌رفت در این حالت به دلیل وابستگی شدید تابع سازگاری به تعداد باندها مسئله اصلاً همگرا نشد. به این معنا که برای هر سه داده بهترین راه‌حل، راه‌حلی است که داده‌ها حداکثر یک باند داشته باشند که از لحاظ عملی منطقی نیست. اگرچه مشکل این راه‌حل را می‌توان با استفاده از ضریب کاپا به عنوان تابع سازگاری حل کرد، اما همان‌طور که گفته شد اینکار مسئله را از حالت نظارت نشده خارج و به حالت نیمه نظارت شده تبدیل می‌کند.

در نهایت به منظور مقایسه‌ی کلی سه حالت ۱ الی ۳ و همچنین روش k-means، بیشترین دقت‌ها برای هر یک

کرد از لحاظ محاسباتی توصیه نمی‌شود. به علاوه نتایج بر روی داده‌های واقعی Pavia university نشان داد که بعد ذاتی یک مسئله وابسته به کاربرد است، به این معنا که بسته به هدف و نوع داده‌ها، بعد ذاتی می‌تواند مقادیر مختلفی داشته باشد.

چهار حالت مذکور برای بهبود خوشه‌بندی داده‌های فراطیفی، حالت دوم که در آن تعداد باندها در یک مرحله-ی پیش‌پردازشی کاهش پیدا می‌کند به دقت بالاتری منجر شد. در رتبه‌ی بعدی حالتی که فقط موقعیت مراکز خوشه‌ها بهینه می‌شد دارای دقت و صحت بالاتری بود. اگرچه این حالت دقت و صحت قابل قبولی داشت ولی به دلیل اینکه از تمام حجم داده‌های فراطیفی استفاده می-

مراجع

- [1] J.A. Richards, X. Jia, Remote sensing digital image analysis-hardback, Springer, Berlin/Heidelberg, 2006.
- [2] M.G. Omran, Particle swarm optimization methods for pattern recognition and image processing, University of Pretoria, 2004.
- [۳] ا.ع. نایینی, خوشه بندی داده های فراطیفی با استفاده از تکنیک بهینه سازی خرد جمعی, دانشگاه تهران, ۱۳۸۹.
- [4] C.-I. Chang, Hyperspectral imaging: techniques for spectral detection and classification, Springer, 2003.
- [5] A.A. Naeini, S. Homayouni, M. Saadatseresht, Intrinsic Dimensionality Estimation in Hyperspectral Imagery Using Residual and Change-Point Analyses, IEEE Geoscience and Remote Sensing Letter DOI: 10.1109/LGRS.2014.2317352 (To be published).
- [6] A. Villa, J. Chanussot, J.A. Benediktsson, C. Jutten, R. Dambreville, Unsupervised methods for the classification of hyperspectral images with low spatial resolution, Pattern Recogn. 46 (2013) 1556-1568.
- [7] X. Sun, L. Yang, B. Zhang, L. Gao, L. Zhang, Hyperspectral image clustering method based on Artificial Bee Colony algorithm, Advanced Computational Intelligence (ICACI), 2013 Sixth International Conference on, IEEE, 2013, pp. 106-109.
- [۸] ج.ع.آ.س.ه.ر. شاه‌حسینی, به کارگیری معیارهای جدید شباهت طیفی برای افزایش کارایی خوشه بندی تصاویر فراطیفی به کمک الگوریتم FCM, نشریه مهندسی نقشه برداری و اطلاعات مکانی ۳ (۱۳۸۹) ۶۳-۷۵.
- [9] S.R. Namin, A.A. Naeini, F. Samadzadegan, Evaluating the Potential of Particle Swarm Optimization for Hyperspectral Image Clustering in Minimum Noise Fraction Feature Space, Computational Intelligence and Decision Making, Springer, 2013, pp. 69-79.
- [10] S. Niazmardi, A.A. Naeini, S. Homayouni, A. Safari, F. Samadzadegan, Particle swarm optimization of kernel-based fuzzy c-means for hyperspectral data clustering, Journal of Applied Remote Sensing 6 (2012) 063601-063601.
- [11] D.S. Shulman, I. Roth, S.R. Rotman, Automatic clustering of multidimensional data (ACMD) applied to hyperspectral images, European Symposium on Optics and Photonics for Defence and Security, International Society for Optics and Photonics, 2004, pp. 117-125.
- [12] A. Paoli, F. Melgani, E. Pasolli, Clustering of hyperspectral images based on multiobjective particle swarm optimization, Geoscience and Remote Sensing, IEEE Transactions on 47 (2009) 4175-4188.
- [13] F. Melgani, E. Pasolli, Multiobjective PSO for Hyperspectral Image Clustering, Computational Intelligence in Image Processing, Springer, 2013, pp. 265-280.
- [14] J. Senthilnath, S. Omkar, V. Mani, N. Karnwal, P. Shreyas, Crop stage classification of hyperspectral data using unsupervised techniques, Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of 6 (2013) 861-866.
- [15] G. Bilgin, S. Erturk, T. Yildirim, Segmentation of hyperspectral images via subtractive clustering and cluster validation using one-class support vector machines, Geoscience and Remote Sensing, IEEE Transactions on 49 (2011) 2936-2944.

- [16] J. Kennedy, R. Eberhart, Particle swarm optimization, Proceedings of IEEE international conference on neural networks, Perth, Australia ,1995 ,pp. 1942-1948.
- [17] Y. Del Valle, G.K. Venayagamoorthy, S. Mohagheghi, J.-C. Hernandez, R.G. Harley, Particle swarm optimization: basic concepts, variants and applications in power systems, Evolutionary Computation, IEEE Transactions on 12 (2008) 1195-71.
- [18] Computational Intelligence Group, Basque University (2010). Hyperspectral Imagery Synthesis toolbox for MATLAB. Available: http://www.ehu.es/ccwintco/index.php/Hyperspectral_Imagery_Synthesis_tools_for_MATLAB.
- [19] A. Plaza, J.A. Benediktsson, J.W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, Recent advances in techniques for hyperspectral image processing, Remote Sensing of Environment 113 (2009) S110-S122.
- [20] Hyperspectral Remote Sensing Scenes. Available: http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes.
- [21] A.P. Engelbrecht, Computational intelligence: an introduction, John Wiley & Sons, 2007.
- [22] Y. Zhang, Q. Jiang, An Improved initialization method for clustering high-dimensional data, Database Technology and Applications (DBTA), 2010 2nd International Workshop on, IEEE, 2010, pp. 1-4.
- [23] R.S. Hosseini, S. Homayouni, A SVMs-based hyperspectral data classification algorithm in a similarity space, Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, 2009. WHISPERS'09. First Workshop on, IEEE, 2009, pp. 1-4.

[۲۴] ا.ع. نایینی، ف. صمدزادگان، س. نیازمردی، حل همزمان خوشه‌بندی و تعیین باندهای بهینه داده‌های فراطیفی با استفاده از الگوریتم بهینه‌سازی توده ذرات، سنجش از دور و GIS ایران ۳ (۱۳۹۰).