

خوشه‌بندی به روش نزدیک‌ترین همسایه با اولویت انتخاب خوشه‌های متراکم‌تر در آنالیز ماتریس تشابه جهت تفکیک پالس‌های راداری

محمدجواد قلندری^۱، مجید آقابابایی^۲، احسان مرادی^۳

Jghalandari@ihu.ac.ir

۱- دانشجوی دکتری مهندسی برق دانشگاه جامع امام حسین (ع)

۲- استادیار دانشگاه علوم دریایی امام خمینی (ره) نوشهر

۳- دانشجوی دکتری الکترونیک دانشگاه تهران

چکیده

دسته‌بندی و میانبرداری پالس‌های راداری، از کاربردهای خوشه‌بندی است. از مشکلات اساسی در بیشتر روش‌ها، انتخاب مناسب مراکز اولیه جهت خوشه‌بندی است. در الگوریتم پیشنهادی، ابتدا ماتریس تشابه برای داده‌های ورودی بر اساس شعاع همسایگی تعریف شده، تولید می‌گردد. سپس با آنالیز کدهای ماتریس تشابه و انتخاب سطرهای با بیشترین کد تشابه، دسته‌های متراکم‌تر به ترتیب جدا می‌شوند. در این روش با توجه به جریان زیاد داده‌های راداری، بدون نیاز به تکرار پردازش، رشته پالس‌ها به صورت بهینه جدا می‌شوند. الگوریتم حاصل نسبت به توابع موجود، داده را بر اساس شعاع همسایگی تعریف شده به تعداد خوشه‌های از پیش تعیین نشده با اولویت انتخاب دسته‌های متراکم و با دقت در اختصاص داده‌های واقع در مرز مشترک تفکیک می‌نماید. به‌طور خاص، مزیت استفاده از این الگوریتم در تفکیک پالس راداری، انتخاب نقاط متراکم به‌عنوان مراکز خوشه‌ها است. روش پیشنهادی می‌تواند برای تفکیک داده‌ها در زمینه‌های مختلف نیز به کار رود. از محاسن این روش نسبت به الگوریتم‌های مبتنی بر روش k -mean، دقت در انتخاب اولیه مراکز خوشه‌ها است. نتایج خوشه‌بندی روش پیشنهادی برای نمونه داده‌های راداری شامل ۲۰۰ پالس با نتایج روش خوشه‌بندی حول رهبر که یکی از الگوریتم‌های مطرح در زمینه خوشه‌بندی پالس‌های راداری است، مقایسه شده است. بررسی شاخص‌های ارزیابی و اعتبارسنجی خوشه‌بندی مانند Dunn، Silhouette و RMSSD، بهینه بودن الگوریتم پیشنهادی را نشان می‌دهد.

واژگان کلیدی: خوشه‌بندی پالس راداری، تفکیک پالس راداری، جنگ الکترونیک، اعتبارسنجی خوشه‌ها.

تاریخ دریافت مقاله : ۹۵/۰۴/۱۳

تاریخ پذیرش مقاله : ۹۶/۰۲/۲۳

۱- مقدمه

خوشه‌بندی، طبقه‌بندی نمونه داده‌های مشابه در گروه‌هایی است که خوشه نامیده می‌شوند. برای خوشه‌بندی موفق اطلاع از تکنیک‌های جمع‌آوری داده نیز مهم است. روش خوشه‌بندی نزدیک‌ترین همسایه به دلیل آن که مجاورت، نق‌ش کلییدی در الگوریتم‌های خوشه‌بندی دارد، می‌تواند به‌عنوان یکی از فرایندهای خوشه‌بندی مطرح شود. در خوشه‌بندی هوشمند، تعداد خوشه‌های موردنظر مشخص نبوده و خود الگوریتم باید با توجه به تعریف شباهت داده‌ها، آن را تعیین نماید.

یکی از موارد خوشه‌بندی، دسته‌بندی جریان پالس راداری در سیستم‌های جنگ الکترونیک (EW) است. هدف تشکیل خوشه‌هایی است که پارامترهای آن‌ها تا حد امکان به هم شبیه بوده و نسبت به خوشه‌های دیگر متفاوت باشند.

خوشه‌بندی، طبقه‌بندی نمونه داده‌های مشابه در گروه‌هایی است که خوشه نامیده می‌شوند. برای خوشه‌بندی موفق اطلاع از تکنیک‌های جمع‌آوری داده نیز مهم است. روش خوشه‌بندی نزدیک‌ترین همسایه به دلیل آن که مجاورت، نق‌ش کلییدی در الگوریتم‌های خوشه‌بندی دارد، می‌تواند به‌عنوان یکی از فرایندهای خوشه‌بندی مطرح شود. در خوشه‌بندی هوشمند، تعداد خوشه‌های موردنظر مشخص نبوده و خود الگوریتم باید با توجه به تعریف شباهت داده‌ها، آن را تعیین نماید.

یکی از موارد خوشه‌بندی، دسته‌بندی جریان پالس راداری در سیستم‌های جنگ الکترونیک (EW) است. هدف تشکیل خوشه‌هایی است که پارامترهای آن‌ها تا حد امکان به هم شبیه بوده و نسبت به خوشه‌های دیگر متفاوت باشند.

روش‌های مختلف تفکیک پالس‌های راداری در مراجع [۱]-۴، ۶ و ۷ توضیح داده شده است. الگوریتم‌های متعدد طراحی شده سعی در افزایش دقت، کاهش زمان محاسبه و پاسخ سیستم دارند. روش‌های پایه تفکیک پالس‌های راداری را می‌توان به دو دسته کلی زیر تقسیم نمود:

الف- روش استخراج ویژگی شکل موج

در این روش تفکیک پالس راداری بر اساس استخراج ویژگی شکل موج با توجه به ویژگی‌های فرستنده‌های راداری (امضای راداری) و تشخیص نوع مدولاسیون پالس با روش‌های مبتنی بر شبکه‌های عصبی بوده و در حال تحقیق و بررسی است [۸-۱۱].

ب- روش پارامتری

این روش بر اساس اندازه‌گیری‌های به‌دست‌آمده از پالس‌های ورودی است. روش تفکیک پارامتری خود به سه روش زیر تقسیم می‌شود:

روش اول - تفکیک زمانی رشته پالس‌های راداری تفکیک زمانی رشته پالس‌های راداری شامل الگوریتم‌های مبتنی بر آنالیز زمان ورود پالس (TOA^۱) و مدولاسیون به‌کاررفته در زنجیره تولید پالس است [۱۰، ۱۲ و ۱۳].

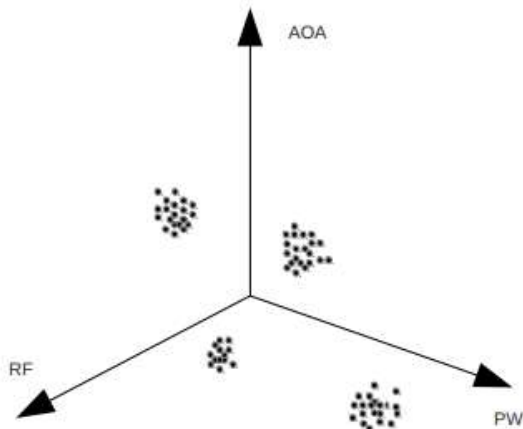
روش دوم - خوشه‌بندی^۲ پالس‌های راداری خوشه‌بندی پالس‌های راداری شامل الگوریتم‌های مبتنی بر پردازش پارامترهای هر تک پالس راداری است. با یافتن تشابه بین پارامترهای پالس‌های مختلف در محدوده تغییرات تعریف‌شده به دسته‌بندی و جداسازی رشته پالس‌ها پرداخته می‌شود. در این روش‌ها با محاسبه فاصله اقلیدسی و شعاع همسایگی، پالس‌ها در خوشه‌های مختلف مانند شکل (۱) دسته‌بندی می‌گردند. پارامترهای اصلی عبارت‌اند از زاویه، فرکانس و عرض پالس و فاصله اقلیدسی می‌تواند به‌صورت زیر تعریف شود:

$$\bar{M}_{ij} = \frac{(ang_i - ang_j)^2}{ang_{no}^2} + \frac{(f_i - f_j)^2}{f_{no}^2} + \frac{(pw_i - pw_j)^2}{pw_{no}^2} \quad (1)$$

حاصل M_{ij} ماتریسی متقارن است که فواصل بین تمام نقاط را با توجه به پارامترها محاسبه می‌کند. مقادیر Ano ، Fno و $PWno$ جهت نرمالیزه کردن هر یک از پارامترها، با توجه به درجه اهمیت آن‌ها تعیین می‌شود. روش‌های خوشه‌بندی عبارت‌اند از:

- خوشه‌بندی منفرد (روش سلسله مراتبی)
- خوشه‌بندی به روش دسته‌بندی زاویه‌ای
- روش آنالیز گروهی
- روش سلولی یا خوشه‌بندی حول رهبر (روش ترتیبی)

² Clustering¹ Time Of Arrival



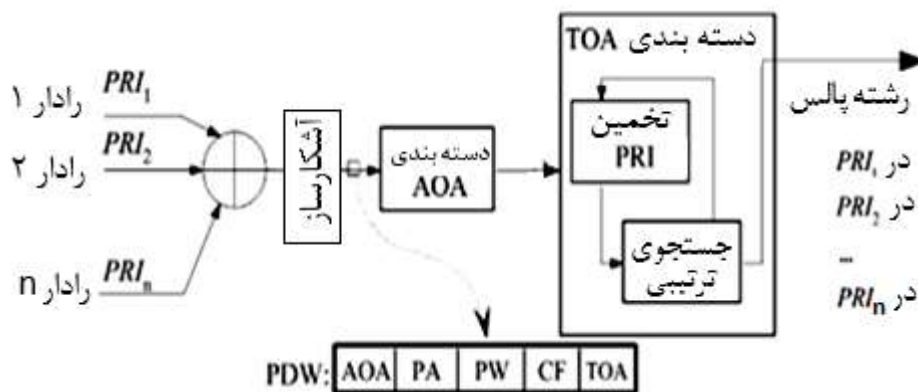
شکل (۱) خوشه‌بندی در فضای سه‌بعدی زاویه، فرکانس و عرض پالس مرجع [۳].

روش سوم - روش‌های مرکب خوشه‌بندی و تفکیک زمانی در این روش از تکنیک‌های پردازش زمان ورود پالس جهت تصحیح نتایج حاصل از خوشه‌بندی استفاده می‌گردد. به عبارتی با ترکیب دو روش فوق نقایص هر یک با مزایای دیگری پوشانده می‌شود. مطابق شکل (۲) در بلوک اول پالس‌های هم زاویه، هم فرکانس و با پهنای پالس تقریباً مساوی در یک گروه یا خوشه قرار می‌گیرد. در بلوک بعدی، اطلاعات پالس‌های مربوط به هر لانه یا رشته پالس از نظر پررود زمانی مورد تجزیه و تحلیل قرار می‌گیرد. آنالیز زمان ورود پالس مطابق شکل (۳) به دنبال یافتن ارتباط زمانی بین پالس‌های یک مجموعه است. دلیل این عمل تأیید صحت تفکیک مرحله قبل، تعیین خوشه‌های دقیق‌تر و همچنین تشخیص نوع PRI و مقدار آن است که خود از مشخصات راداری است [۱۵-۱۸].

در روش خوشه‌بندی منفرد، برای N پالس دریافتی، فاصله اقلیدسی داده‌ها محاسبه و یک ماتریس فاصله متقارن $N \times N$ به دست می‌آید. سپس داده‌های با کمترین فاصله در یک دسته قرار می‌گیرند. برای N پالس ورودی تعداد محاسبات برابر N^2 بوده و بار پردازشی زیاد است. در علوم خوشه‌بندی به این روش سلسله مراتبی می‌گویند [۵].

در خوشه‌بندی به روش دسته‌بندی زاویه‌ای به علت پایداری نسبی پارامتر زاویه نسبت به پارامترهای دیگر، ابتدا پالس‌ها بر اساس زاویه دسته‌بندی می‌شوند. سپس آنالیز خوشه‌بندی منفرد در هر دسته زاویه‌ای انجام می‌شود؛ زیرا در زوایای نزدیک به هم احتمال هم‌قطاری رشته پالس‌ها بیشتر است.

در روش آنالیز گروهی با توجه به ثبات زاویه، ابتدا دسته‌بندی اولیه بر اساس زاویه ورود صورت می‌گیرد. در داخل هر دسته، دسته‌های دیگر بر اساس تقسیم‌بندی فرکانسی به دست آمده و پارامترهای عرض پالس و دامنه نیز می‌توانند در دسته‌بندی شرکت نمایند. در روش سلولی ابتدا اولین پالس ورودی به‌عنوان رهبر انتخاب‌شده و فاصله متریک پالس‌های بعدی با آن محاسبه می‌شود. در صورتی که این فاصله کمتر از حد آستانه باشد، پالس‌ها در این لانه قرار گرفته و در غیر این صورت خود رهبر یک لانه جدید می‌گردد. در این روش زنجیره پالس‌های مختلف در لانه‌های متفاوت ذخیره می‌گردد. مشکلات عمده این روش تعیین صحیح پالس رهبر و حد آستانه پذیرش است [۴]. این روش در علوم خوشه‌بندی روش ترتیبی نام دارد [۵].



شکل (۲) تفکیک اهداف با خوشه‌بندی و آنالیز زمانی مربوط به مرجع [۱۶].

۲-۲- شاخص‌های اعتبارسنجی ریشه میانگین مربع انحراف از معیار (RMSSDT¹)

در شاخص اعتبارسنجی RMSSDT از واریانس خوشه‌ها مطابق رابطه (۳) استفاده می‌شود [۵].

$$RMSSDT = \sqrt{\frac{\sum_{j=1..d} \sum_{i=1..n_d} \sum_{k=1}^{n_d} (x_k - \bar{x}_j)^2}{\sum_{j=1..d} (n_{ij}-1)}} \quad (3)$$

این رابطه معیار میزان همگنی خوشه‌ها را اندازه می‌گیرد و هرچه مقدار آن کمتر باشد، نشان‌دهنده خوشه‌بندی بهتر داده‌ها است.

۳-۲- شاخص دیویس بولدین (Davies Bouldin Index)

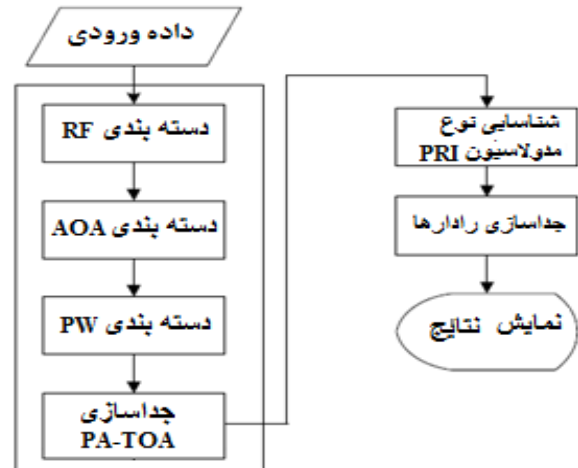
این معیار از شباهت بین دو خوشه استفاده می‌کند که بر اساس پراکندگی یک خوشه (si) و عدم شباهت بین دو خوشه (dij) تعریف می‌شود.

۴-۲- شاخص سیلوت (Silhouette Index)

شاخص سیلوت، یکی از معیارهای متداول اعتبارسنجی خوشه‌بندی است و دو معیار فواصل درون خوشه‌ای و برون خوشه‌ای را هم‌زمان در نظر می‌گیرد.

جدول (۱) علائم مورداستفاده در شاخص‌های ارزیابی و اعتبارسنجی خوشه‌ها.

مفهوم	علامت	مفهوم	علامت
تعداد داده‌های درون j امین بعد از خوشه i ام	n_{ij}	تعداد خوشه‌ها	n_c
تعداد عناصر j امین بعد از تمام داده‌ها	n_j	تعداد ابعاد	d
نقطه مرکز خوشه i ام	v_i	فاصله بین دو داده	$d(x, y)$
i امین خوشه	c_i	امید ریاضی j امین بعد داده‌ها	\bar{x}_j
تعداد داده‌های درون i امین خوشه	$\ c_i\ $	یک بردار ستونی $\sqrt{X^T X}$ است.	$\ x\ $



شکل (۳) خوشه‌بندی پالس‌های راداری و تفکیک با آنالیز زمانی مربوط به مرجع [۱۷].

۲- اعتبارسنجی الگوریتم‌های خوشه‌بندی

هدف از اعتبارسنجی خوشه‌ها، یافتن خوشه‌هایی است که بهترین تناسب را با داده‌های موردنظر داشته باشند. دو معیار پایه اندازه‌گیری پیشنهاد شده برای ارزیابی و انتخاب خوشه‌های بهینه عبارت‌اند از:

تراکم (Compactness): داده‌های متعلق به یک خوشه بایستی تا حد ممکن به یکدیگر نزدیک باشند. معیار رایج برای تعیین میزان تراکم داده‌ها واریانس داده‌ها است. جدایی (Separation): خوشه‌ها خود بایستی به اندازه کافی از یکدیگر جدا باشند. در جدول (۱) مجموعه‌ای از علائم استفاده شده در ادامه این بخش ارائه شده است:

۱-۲- شاخص دون (Dunn Index)

این معیار طبق مرجع [۵] توسط رابطه (۲) تعریف می‌شود:

$$D = \min_{i=1..n_c} \left\{ \min_{j=i+1..n_c} \left(\frac{d(c_i, c_j)}{\max_{k=1..n_c} (diam(c_k))} \right) \right\} \quad (2)$$

که $d(c_i, c_j)$ و $diam(c_k)$ با روابط زیر محاسبه می‌شوند:

$$d(c_i, c_j) = \min_{x \in c_i, y \in c_j} \{d(x, y)\}$$

$$diam(c_k) = \max_{x, y \in c_k} \{d(x, y)\}$$

اگر مجموعه داده، دارای خوشه‌های جدا پذیر باشد، انتظار می‌رود فاصله بین خوشه‌ها زیاد و قطر خوشه‌های آن کوچک باشد. در نتیجه مقداری بزرگ‌تر برای این معیار مقداری مطلوب‌تر است.

¹ Root Mean Square Standard Deviation

۳- روش پیشنهادی برای تفکیک پالس

داده‌های ورودی گیرنده ESM داده‌های جریانی هستند. با توجه به زمان مورد نیاز برای جمع‌آوری اطلاعات و پردازش اولیه برای محاسبه پارامترهای راداری، یک بازه زمانی برای دسته‌های ورودی تعیین می‌شود. داده‌های این فریم ورودی با هم مقایسه و تفکیک شده و نتیجه تفکیک این فریم با خوشه‌های مجزا شده قبل، بررسی می‌شود.

در روش پیشنهادی برای تفکیک پالس‌های ورودی از خوشه‌بندی به روش نزدیک‌ترین همسایه با اولویت انتخاب مراکز خوشه‌های متراکم‌تر در آنالیز ماتریس تشابه استفاده شده است. در این روش بعد از بررسی اولیه و تعیین نقاط متراکم، پالس رهبر به همراه گروه خود از میانگین خوشه متراکم انتخاب شده و تفکیک از دسته‌های متراکم‌تر آغاز می‌شود. به دلیل تمرکز بر انتخاب دقیق مراکز اولیه خوشه‌ها در این روش، تکرار پردازش با میانگین‌های جدید خوشه‌ها (برحسب دقت مورد نیاز) حذف و به ازای آن، بار پردازشی در تعیین ماتریس تشابه خواهد بود. این مشکل با پیش‌پردازش به کمک تابع مرتب‌سازی اولیه داده‌ها برحسب شعاع اقلیدسی برطرف می‌گردد.

این روش در ظاهر مشابه روش الگوریتم خوشه‌بندی DBSCAN^۱ بوده ولی روش اجرای آن بر اساس آنالیز ماتریس کد تشابه است. در الگوریتم‌های خوشه‌بندی بر اساس چگالی، تعداد خوشه‌ها هم‌زمان با عمل خوشه‌بندی تعیین شده و در تشخیص نویز بسیار کارا هستند.

در اجرای این روش و به‌طور خاص برای تفکیک پالس‌های راداری با توجه به مسائل مربوط به نویز، تمرکز بر خوشه‌های متراکم به‌عنوان رادارهای فعال در منطقه است. برای رسیدن به این هدف از حداکثر کدهای ماتریس تشابه و انتخاب بیشترین همسایه در شعاع آستانه تعریف شده، استفاده شده است. در ادامه، مراحل بهینه‌سازی الگوریتم توضیح داده می‌شود.

۳-۱- شبیه‌سازی ورودی

داده‌های ورودی مطابق اطلاعات ماتریس PDW^۲ شبیه‌سازی و برای آزمایش الگوریتم به کار می‌روند. هر

سطر ماتریس، پارامترهای PDW مربوط به یک نمونه داده یا یک پالس راداری است. ستون‌های اصلی ماتریس ورودی به‌عنوان خصیصه‌هایی که تفکیک بر اساس آن‌ها انجام می‌شود، به ترتیب از چپ: زاویه برحسب درجه، فرکانس برحسب مگاهرتز، عرض پالس برحسب میکروثانیه و باقی ستون‌ها شامل زمان ورود، دامنه، نوع مدولاسیون و ... می‌باشند. یک نمونه از داده ورودی در پیوست الف بر اساس مشخصات رادارهای تجاری و نظامی متعارف در محدوده ۲GHZ تا ۱۰GHZ نشان داده شده است.

۳-۲- محاسبه کد تشابه

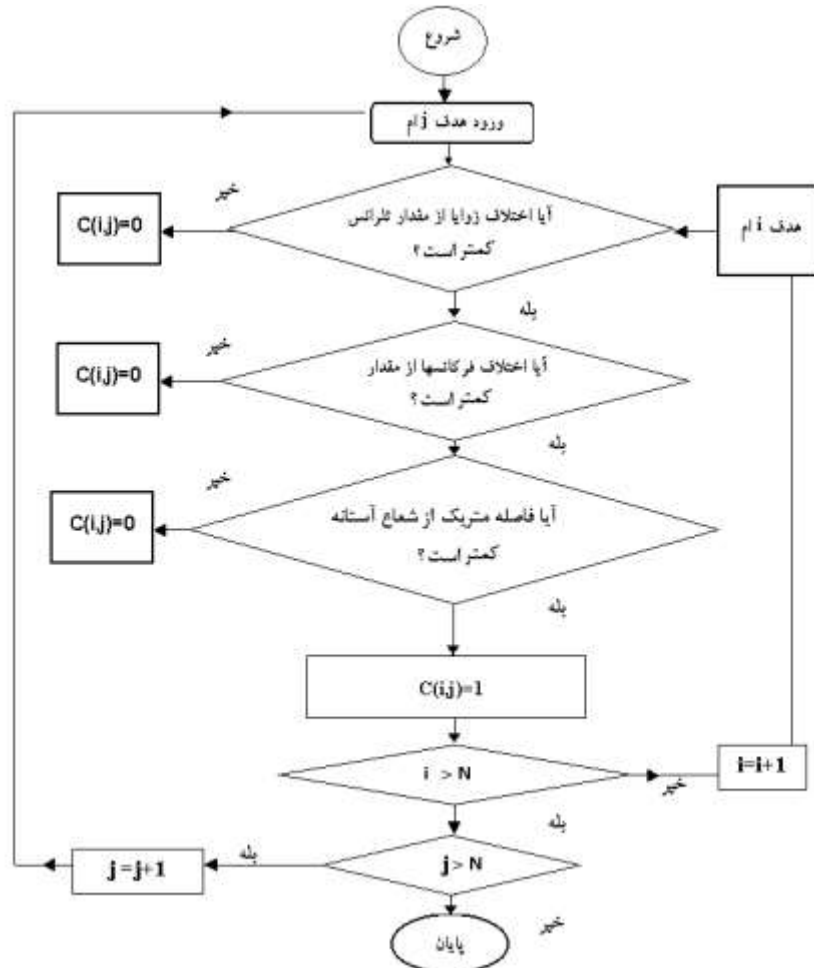
برای محاسبه کدهای تشابه از فلوچارت شکل (۴) استفاده می‌شود. با گروه‌بندی اولیه چنانچه اختلاف زوایا در سطرهای ماتریس کمتر از محدوده قابل قبول، اختلاف فرکانس‌ها و در نهایت اختلاف عرض پالس نیز قابل قبول باشد، کد تشابه مربوطه یک وگرنه صفر خواهد شد. در نسخه نهایی به جای مقایسه عرض پالس، فاصله متریک مناسب جایگزین و با شعاع همسایگی مقایسه می‌گردد. اگر این فاصله کمتر باشد، به‌عنوان همسایه پذیرفته می‌شود.

به‌عنوان مثال، به ازای ورودی p پیوست - که یک ماتریس داده با $i=6$ نمونه ورودی است - کدهای ماتریس تشابه به نام c با ابعاد 6×6 به صورت زیر نتیجه می‌شود. بدون توجه به قطر اصلی عدد هر آرایه این ماتریس متقارن، وضعیت تشابه نمونه‌های ردیف یک تا ۶ را با بقیه ورودی‌ها با توجه به شعاع همسایگی نشان می‌دهد. در این مثال سطرهای اول و دوم بیشترین کد ۱ را دارند.

$$c_{(i \times i)} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 2 & 1 & 1 & 0 & 0 & 0 \\ 3 & 0 & 1 & 1 & 0 & 0 \\ 4 & 0 & 0 & 1 & 1 & 0 \\ 5 & 0 & 0 & 0 & 1 & 0 \\ 6 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

² PDW: Pulse Description Word

¹ Density-Based Clustering



شکل (۴) الگوریتم محاسبه ماتریس شباهت.

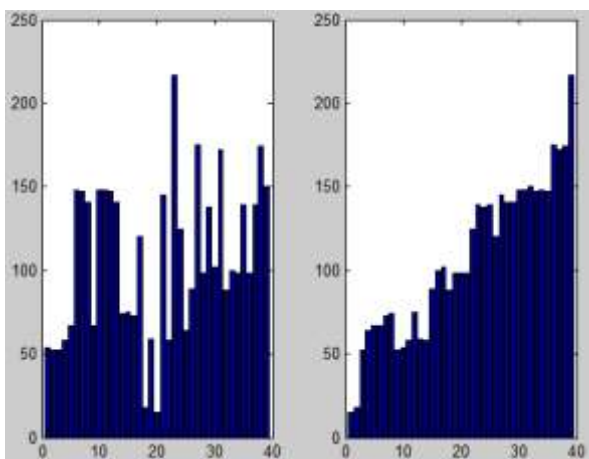
ورودی تقریباً کنار هم قرار گرفته و در مراحل بعدی تفکیک، سبب کاهش محاسبات می گردد.

آرایه‌های ۱، تشابه نمونه‌های (۲ و ۱) و (۴ و ۱) در سطر اول و (۱ و ۲) و (۳ و ۲) را در سطر دوم نشان می‌دهند. محدوده همسایگی تعریف شده در برنامه برای زاویه ۵ درجه، فرکانس ۰/۵٪ و برای عرض پالس ۲۰٪ است [۶].

۳-۳- مرتب‌سازی اولیه داده‌ها

در روش پیشنهادی مانند خوشه‌بندی منفرد، نیاز به N^2 پردازش برای محاسبه ماتریس تشابه است که حجم پردازشی بالایی دارد. با مرتب‌سازی اولیه داده‌ها بر اساس وزن تعریف شده که همان شعاع همسایگی از مرکز دستگاه مختصات چندبعدی است، مقدار پردازش به نحو مطلوب کاهش می‌یابد [۶].

از تابع presorting برای مرتب‌سازی اولیه داده‌ها استفاده شده است. بعد از اجرای این تابع، سطرهای مشابه



شکل (۵) نمایش هیستوگرام زاویه ماتریس ورودی (ستون اول ماتریس p) و خروجی حاصل از تابع presorting.

۴- از بین خوشه‌های متراکم، انتخاب خوشه‌ای که در مجموع کمترین انحراف معیار را در مقدار شعاع همسایگی دارد.

۵- محاسبه میانگین پارامترها به‌عنوان نماینده خوشه و جدا کردن سطرهای مربوطه در لانه مجزا

۶- صفر نمودن کد مربوطه در ماتریس تشابه بعد از اختصاص به یک خوشه

۷- تعیین لانه متراکم بعدی و شروع از مرحله سوم

۸- ادامه تا پایان داده یا رسیدن به حداقل تراکم تعریف‌شده

برای سادگی توضیح الگوریتم، مراحل ارتقای آن که در دو فاز در نرم‌افزار MATLAB نوشته شده، بیان می‌شود.

۴-۱- تابع CLUSTER_PHAS1

در فاز اول سطری که بیشترین کد تشابه را با همسایه‌های خود دارد، تقریباً مرکز خوشه فرض شده و اعضای مرتبط به این رهبر که در شعاع همسایگی آن قرار دارند، انتخاب و در این گروه قرار می‌گیرند. برای دو سطر با تعداد کد تشابه یکسان، فاصله متریک تعریف‌شده که در ماتریس اختلاف ذخیره شده، مقایسه شده و هر کدام که مجموع شعاع همسایگی یا انحراف معیار آن کمتر باشد، به‌عنوان خوشه متراکم‌تر انتخاب می‌شود.

تابع CLUSTER_PHAS1 بیشترین کدهای مشابه یکسان و متراکم‌تر را انتخاب می‌کند. برای مراحل بعد، سطرهای مرتبط با سطر متراکم دارای کد تشابه صفر می‌شوند.

۴-۲- تابع CLUSTER_PHAS2

در فاز دوم طراحی الگوریتم و تکمیل برنامه تابع CLUSTER_PHAS2 هم تعداد سطرهای مشابه و هم تراکم آن‌ها مدنظر است. در این تابع ابتدا بیشترین کد شباهت، انتخاب شده و اگر خوشه چگال‌تری بیش از نصف تعداد اعضای این خوشه (با تعریف مناسب از طرف کاربر) عضو داشته باشد، جایگزین می‌شود. تا پایان سطر ماتریس تشابه داده‌ها، این روند بررسی می‌گردد.

با این روش، اگر سطری با دو یا چند سطر دیگر در تعریف تشابه قرار گرفته باشد، سطر یا سطرهای مشترک به خوشه متراکم‌تر سپرده می‌شود. به عبارتی اگر شرط زیر برقرار

تابع presorting می‌تواند در الگوریتم‌های مختلف تفکیک نیز به کار رود. ایده پیش‌پردازش در مقالات دیگر نیز به‌کاررفته است [۲۲].

۴- توضیح الگوریتم برنامه تفکیک

بعد از محاسبه کد تشابه، ردیف‌هایی که بیشترین تشابه را با بقیه سطرها دارند، با جمع تعداد یک سطرها به دست می‌آید. به این ترتیب سطری که بیشترین کد تشابه را با همسایه‌های خود دارد، تعیین می‌شود. این سطر تقریباً مرکز خوشه خواهد شد. اگر چند خوشه متراکم پیدا شود، ابتدا خوشه‌ای که کمترین انحراف معیار را در شعاع همسایگی دارد، جدا می‌گردد. برای این کار از فاصله متریک در ماتریس اختلاف‌ها و محاسبه مجموع آن‌ها استفاده شده است.

در برنامه مجزا جهت نمایش این الگوریتم همه پارامترهای سطرهایی که طبق شعاع آستانه، با این سطر (سطری تقریباً میانی و در وسط خوشه متراکم) مشابه هستند، جمع نموده و میانگین پارامترهای آن‌ها به‌عنوان نماینده خوشه معرفی می‌گردد.

بعد از تعیین اولین خوشه متراکم و حذف کدهای خوشه متراکم مربوطه در ماتریس تشابه، جمع کدهای هر سطح ماتریس تشابه محاسبه و همین روال دنبال می‌شود. برای این کار از همان کدهای تولیدشده در مقایسه اولیه استفاده شده تا مقدار پردازش کم شود. درواقع هدف این است که خوشه‌های متراکم‌تر در ابتدا به‌عنوان خوشه‌های مورد تأیید و بااهمیت جدا شود.

این خود یکی از محاسن این روش تفکیک نسبت به الگوریتم‌های مبتنی بر روش k-mean است و انتخاب اولیه مراکز خوشه‌ها در آن با دقت انجام می‌گیرد. روش پیشنهادی می‌تواند برای تفکیک انواع دیگر داده نیز به کار رود. مراحل اجرای الگوریتم به‌صورت زیر است:

۱- محاسبه کد تشابه

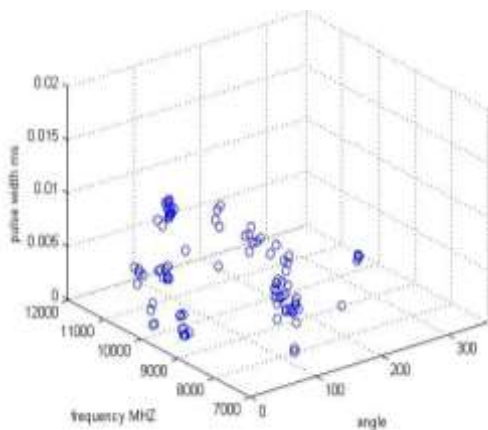
۲- محاسبه ماتریس اختلاف‌ها برای سطرهای مشابه هم‌زمان با محاسبه کد تشابه

۳- تعیین ردیف‌هایی که بیشترین تشابه را با بقیه سطرها دارند (با محاسبه جمع کدهای هر سطر).

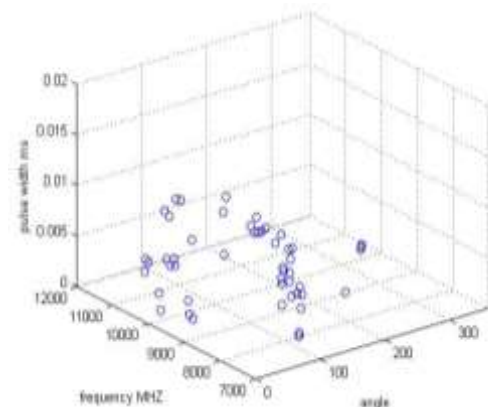
۴- با تعیین نسبت تراکم در اولویت دادن جداسازی خوشه‌ها مطابق شکل زیر دو خوشه‌ای که در محدوده $X < R$ قرار دارند، از هم تفکیک می‌شوند. این مورد در الگوریتم تفکیک با خوشه‌بندی حول رهبر باز به ترتیب ورود داده بستگی دارد و ممکن است به یک یا دو خوشه تقسیم گردد.

۵- با توجه به این که نوع داده راداری داده جریانی است، در اولین مرتبه خوشه‌بندی نتایج مطلوبی حاصل شده و نیاز به تکرار خوشه‌بندی تا حصول نتیجه مناسب مانند روش‌های معمول برطرف می‌گردد.

نتایج خوشه‌بندی داده نمونه توسط الگوریتم پیشنهادی با یک الگوریتم کلاسیک مطرح مانند خوشه‌بندی به روش سلولی یا حول رهبر مقایسه و نتایج ارزیابی کیفیت خوشه‌بندی در جدول (۲) نشان داده شده است. در دو الگوریتم فوق از یک شعاع آستانه مناسب طبق مرجع [۶] استفاده شده است.



شکل (۶) پراکنگی ۲۰۰ پالس راداری در یک محیط پرتراکم، برای یک بسته از اطلاعات قبل از خوشه‌بندی.



شکل (۷) پراکنگی ۴۰ هدف مجزا (با توجه به حد آستانه تعریف شده) حاصل از خوشه‌بندی.

باشد، سطر مربوط به مرکز خوشه جدید، جایگزین سطر قبل می‌شود. این شرط به صورت زیر تعریف شده است:

$$\frac{\text{جمع فاصله سطر } i \text{ با سطرهای مشابه خود}}{\text{تعداد سطرهای مشابه با سطر } i} < \frac{\text{جمع فاصله سطر } j \text{ با سطرهای مشابه خود}}{\text{تعداد سطرهای مشابه با سطر } j}$$

در شکل (۶) پراکنده‌گی تعداد ۲۰۰ پالس راداری در یک محیط متراکم، برای یک بسته از اطلاعات پالس‌ها قبل از خوشه‌بندی، نشان داده شده است. داده ورودی که طبق توضیح بخش (۳-۱) بر اساس مشخصات پالس در یک سیستم ESM شبیه سازی شده بر اساس سه پارامتر زاویه، فرکانس و عرض پالس تفکیک گردیده است. بعد از خوشه‌بندی مطابق شکل (۷) پراکنده‌گی ۴۰ هدف یا خوشه مجزا نمایش داده شده است. در ادامه روند الگوریتم، خوشه‌های متراکم با چگالی و انحراف معیار کمتر، به ترتیب انتخاب می‌شوند.

۵- ارزیابی و مقایسه الگوریتم پیشنهادی با روش‌های کلاسیک خوشه‌بندی

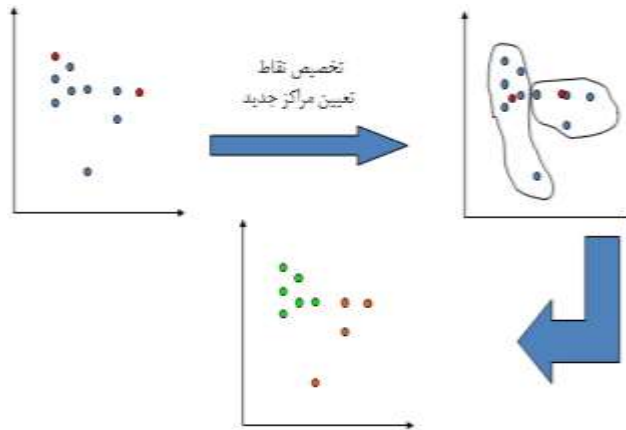
انتخاب اولیه مراکز خوشه‌ها با آنالیز ماتریس تشابه از محاسن این روش تفکیک نسبت به الگوریتم‌های مبتنی بر روش k-mean است.

در روش‌های معمول مانند شکل (۸) خوشه‌بندی با تخصیص مجدد مراکز خوشه‌ها و تکرار خوشه‌بندی تا حصول نتیجه مطلوب نشان داده شده است. در شکل (۹) خوشه‌بندی به روش پیشنهادی نمایش داده شده که با تکیه بر چگالی و انتخاب مناسب مراکز خوشه‌ها، نیاز به تکرار خوشه‌بندی تا حصول نتیجه مناسب تر تا حد زیادی رفع می‌گردد. در مقایسه با روش تفکیک خوشه‌بندی حول رهبر و الگوریتم‌های مشابه خوشه‌بندی، روش پیشنهادی مزایای زیر را دارد:

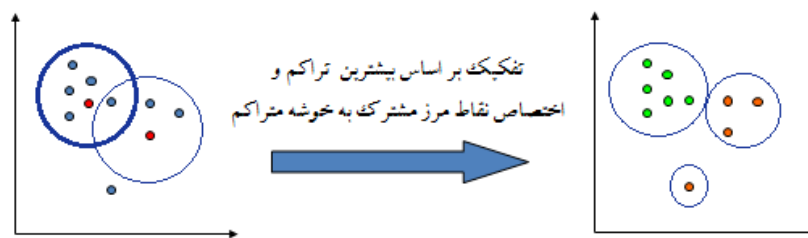
۱- انتخاب اولیه خوشه‌های متراکم با توجه به نوع داده راداری.

۲- در روش خوشه‌بندی حول رهبر، مرکز خوشه (میانگین داده‌های خوشه) با توجه به شکل (۱۰) به ترتیب ورود داده‌ها بستگی دارد و جایی بین سه محدوده معین شده است. ولی با این روش با جدا کردن زیر خوشه متراکم اول مرکز بعدی زیرپوشه دوم است.

۳- داده‌های مرز مشترک به خوشه متراکم و با انحراف معیار کمتر اختصاص می‌یابد.

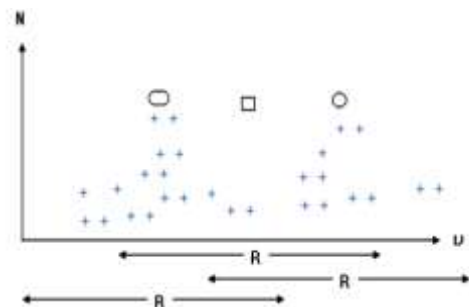


شکل (۸) خوشه‌بندی با تخصیص مجدد مراکز خوشه‌ها و تکرار آن در روش‌های معمول.



شکل (۹) خوشه‌بندی با محاسبه ماتریس شباهت و بیشترین کد شباهت از سطر i ام با توجه به شعاع همسایگی.

جدول مقایسه، حصول نتایج بهینه‌تر در شاخص Dunn، شاخص Silhouette و RMSD را تأیید می‌کند. علت نتیجه بهتر شاخص دیویس بولدین در روش رهبر این است که تعداد خوشه‌های آن در تفکیک داده بیشتر است (درحالی‌که بعضی از این خوشه‌ها طبق شاخص Dunn و RMSD به خوبی تفکیک نشده و شعاع آستانه تعریف‌شده در آن لحاظ نشده است). افزایش شاخص Dunn به‌خصوص به‌عنوان یکی از شاخص‌های مهم اعتبار سنجی خوشه‌بندی در روش پیشنهادی نسبت به روش خوشه‌بندی حول رهبر قابل توجه است. شاخص Dunn و Silhouette بالاتر و انحراف میانگین کمتر به معنای خوشه‌های متراکم‌تر و دقت بالاتر الگوریتم خوشه‌بندی است. با اجرای تابع presorting، تغییر جزئی به سمت بهینه‌تر شدن نتایج دیده می‌شود؛ در نتیجه الگوریتم پیشنهادی به پیش‌پردازش وابسته نیست. داده‌های راداری دارای ترافیک بالا و پیوسته - داده جریان‌ی - هستند و زمان در تفکیک پالس بسیار اهمیت دارد. روش خوشه‌بندی منفرد که در علوم خاص خوشه‌بندی به روش سلسله مراتبی افزایشی معروف است، زمان بر و دارای حجم پردازشی بالا است.



شکل (۱۰) مقایسه خوشه‌بندی روش پیشنهادی با روش رهبر، که در آن خوشه‌های نهایی به ترتیب ورود داده بستگی دارند.

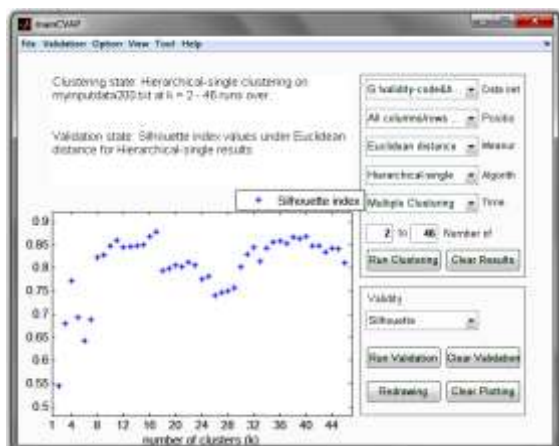
علت مقایسه روش پیشنهادی با روش رهبر آن است که روش رهبر یک روش بهینه در مبحث خوشه‌بندی پالس‌های راداری نسبت به روش‌های دیگر خوشه‌بندی مناسب داده‌های راداری است [۴، ۶ و ۱۸]. همچنین در روش سلولی حتی با فرض عدم تکمیل پردازش و الحاق همه داده‌ها به رهبران مربوطه، نتایج خوشه‌بندی لحظه‌ای می‌تواند مورد استفاده قرار گرفته و با تکمیل خوشه‌بندی، صحت میانگین خوشه‌ها افزایش می‌یابد؛ ولی در روش‌های دیگر تا پایان خوشه‌بندی یک فریم، نتایج خوشه‌ها قابل اعتماد نیستند.

جدول (۲) مقایسه الگوریتم پیشنهادی با یک الگوریتم کلاسیک مانند روش سلولی یا حول رهبر.

پارامتر	روش رهبر	روش پیشنهادی با presorting	روش پیشنهادی بدون presorting	حالت بهینه
ni	200	200	200	-----
clusters	43	38	38	-----
DB	0.47307	0.37541	0.33453	max
Dunn	0.026582	0.089618	0.10082	max
Silhouette	0.28738	0.31019	0.3358	max
RMSD	4.8186	4.1864	4.1864	min

حداقل رسیده است. سپس با اختصاص بیشترین تعداد همسایه‌ها به یک خوشه بر اساس ماتریس تشابه، نیاز به تکرار خوشه‌بندی در مقایسه با روش‌های معمول که ماتریس فاصله در آنها محاسبه می‌گردد، تا حد زیادی کاهش می‌یابد. با توجه به پیوستگی داده راداری این تکنیک می‌تواند در پردازش سریع‌تر داده‌ها مؤثر گردد. همچنین در این روش، نقاط متراکم‌تر به‌عنوان مراکز خوشه‌های با اولویت بالاتر انتخاب و خوشه‌های متراکم به‌عنوان رشته پالس‌های راداری جدا می‌شوند.

از محاسن این روش تفکیک نسبت به الگوریتم‌های مبتنی بر روش k-mean آن است که انتخاب اولیه مراکز خوشه‌ها در آن با دقت انجام می‌گیرد.



شکل (۱۱) متلب در جعبه‌ابزار Cvap، چند روش خوشه‌بندی متداول و شاخص‌های معتبر برای ارزیابی خوشه‌های به‌دست‌آمده را فراهم نموده است.

در این الگوریتم تعداد خوشه‌ها به شعاع همسایگی و دفعات تکرار به تعداد خوشه‌ها بستگی دارد. در کاربردهایی که فقط دقت بالا مدنظر باشد، این روش مفید است. لذا از دقت این الگوریتم بدون توجه به مسئله زمان، برای بررسی و مقایسه تعداد بهینه خوشه‌های مجزا نتایج جدول (۲) به کمک برنامه^۱ Cvap متلب استفاده شده است. متلب در این جعبه‌ابزار، چند روش خوشه‌بندی متداول و شاخص‌های معتبر برای ارزیابی خوشه‌های به‌دست‌آمده را مطابق شکل (۱۱) فراهم نموده است.

همان داده شبیه‌سازی‌شده مربوط به ۲۰۰ پالس راداری با سه پارامتر زاویه، فرکانس و عرض پالس به Cvap داده شده و مرحله خوشه‌بندی به روش سلسله‌مراتبی انجام شد. از شاخص‌های Dunn و silhouette برای تعیین تعداد بهینه خوشه‌ها و بررسی تعداد خوشه‌های جدول مقایسه مطابق شکل (۱۲) استفاده شده است. نتایج مقایسه بالاتر بودن شاخص Dunn و مقدار بهینه تعداد خوشه‌های روش پیشنهادی را نشان می‌دهد. همچنین شاخص silhouette برای ۳۸ خوشه روش پیشنهادی از روش رهبر (ترتیبی) که به ۴۳ خوشه تفکیک شده، بالاتر است.

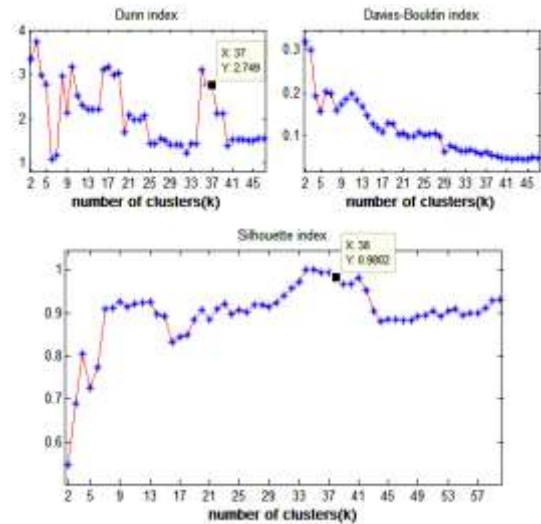
۶- نتایج و مزایای روش پیشنهادی

در روش پیشنهادی برای تفکیک پالس‌های راداری از خوشه‌بندی به روش نزدیک‌ترین همسایه با اولویت انتخاب خوشه‌های متراکم‌تر در آنالیز ماتریس تشابه به‌عنوان یک ایده جدید استفاده شده است. پس از محاسبه و آنالیز کدهای ماتریس تشابه، پردازش‌های بعدی بر اساس آن به

¹ Cluster Validity Analysis Platform

در این روش دقت خوشه‌بندی مستقل از ترتیب پردازش داده‌های یک فریم است که از معایب روش رهبر در تعیین مراکز خوشه‌ها است. حسن دیگر آن است که تفکیک داده‌های مشترک بین دو لانه بر مبنای انتساب به خوشه متراکم‌تر با فاصله کمتر صورت می‌گیرد.

مقایسه نتایج خوشه‌بندی یک نمونه داده راداری الگوریتم پیشنهادی با روش رهبر که یکی از الگوریتم‌های مطرح در خوشه‌بندی پالس‌های راداری است، نشان می‌دهد که این الگوریتم دارای شاخص Dunn و Silhouette بالاتر و انحراف میانگین کمتر است؛ یعنی خوشه‌ها متراکم‌تر و دقت خوشه‌بندی بالاتر است. همچنین با اعمال خوشه‌بندی به روش سلسله مراتبی، تعداد خوشه‌های روش پیشنهادی بررسی و دقت بالاتر آن با معیارهای اعتبارسنجی نشان داده شد. استفاده از این الگوریتم در سیستم‌های پردازنده ESM می‌تواند به‌طور مؤثری سبب افزایش دقت و سرعت تفکیک پالس‌های راداری گردد.



شکل (۱۲) نتایج مقایسه شاخص Dunn و silhouette برای مقایسه دقت تعداد خوشه‌های جدول (۲) بهینه بودن تعداد خوشه‌ها در روش پیشنهادی را نشان می‌دهد.

۷- پیوست

پیوست الف- مقایسه نتایج فایل خوشه‌بندی شده p1 خروجی دو تابع CLUSTER_PHAS1 و CLUSTER_PHAS2 با داده‌های فایل نمونه ورودی نامرتب p:

p =							
20	9320	0.001	5500	400	1	0.002	45
23	9310	0.0011	5010	20	1	0.002	25
25	9300	0.0012	5010	20	1	2.001	25
28	9310	0.0011	2010	20	1	0.001	25
29	9310	0.0011	2010	20	1	0.001	25
33	9300	0.0012	4500	100	1	0.001	20

۱- خروجی با بالاترین کدهای مشابه CLUSTER_PHAS1:

p1 =							
23	9310	0.0011	5010	20	1	0.002	25
25	9300	0.0012	5010	20	1	2.001	25
29	9310	0.0011	2010	20	1	0.001	25
33	9300	0.0012	4500	100	1	0.001	20
28	9310	0.0011	2010	20	1	0.001	25
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
20	9320	0.001	5500	400	1	0.002	45

۲- خروجی با کدهای مشابه متراکم CLUSTER_PHAS2:

خوشه متراکم ابتدایی با مرکزیت زاویه ۲۵ انتخاب شده و سطر با زاویه ۳۳ مجزا می‌شود.

p1 =					
20	9320	0.001	5500	400	0.002
23	9310	0.0011	5010	20	0.002
28	9310	0.0011	2010	20	0.001
29	9310	0.0011	2010	20	0.001
25	9300	0.0012	5010	20	2.001
0	0	0	0	0	0
0	0	0	0	0	0
33	9300	0.0012	4500	100	0.001

- [13] Naseri, A., Shahhoseini, H. S. and Naderi, M. "Data Clustering by Minimum Difference Tree and PRI Transform", IEEE proceedings, MELECON, pp.183-187, 2002.
- [14] Lee, D. W., Hun, J. W. and lee W. D. "Adaptive Radar Pulses Clustering Based on Density Cluster Window", 23rd ITS-CSCC, pp.1377-1380, 2008.
- [۱۵] احمدی، م. و محامدپور، ک.، "تشخیص مدولاسیون فاصله تکرار پالس‌های راداری"، مجله انجمن مهندسی برق و الکترونیک ایران، سال ششم، شماره دوم، پائیز و زمستان ۱۳۸۸.
- [16] Yujun, K., Gingbo, S. and Gianbin C. "A Simply Way to de-Interleave Repetitive Pulse Sequences", 7th Wseas int Conf. on Mathematical Methods and Computational Techniques in Electrical Engineering, Sofia, 2005.
- [17] He, A. L., Zeng, D. G., Wang, J. and Tang, B. "Multi-Parameter Signal Sorting Algorithm Based on Dynamic Distance Clustering". Journal of Electronic Science and Technology of China, Vol.7, No.3, 2009.
- [18] Bailie, S. "An FPGA Implementation of Incremental Clustering for Radar Pulse De-interleaving", Master of Science at Northeastern University, Boston, 2010.
- [19] Maier, M. W. "Processing Throughput Estimation for Radar Intercept Receivers", IEEE Transactions on Aerospace and Electronic System, Vol.34, No.1, 1998.
- [۲۰] آدمی، د.، "جنگ الکترونیک"، نایبی م.م، موسسه انتشارات علمی دانشگاه صنعتی شریف، ۱۳۸۵.
- [۲۱] یقینی، م و ورد، م.، "خوشه‌بندی خودکار داده‌های مختلط با استفاده از الگوریتم ژنتیک"، نشریه بین‌المللی مهندسی صنایع و مدیریت تولید، دانشگاه علم و صنعت ایران، شماره ۲، جلد ۲۳، شهریور ۱۳۹۱.
- [۲۲] نوری، و.، اکبرزاده توتونچی، م و روحانی منش، ع.، "پیش‌پردازش داده‌های ورودی الگوریتم‌های دسته‌بندی با ترکیب خوشه‌بندی فازی نوع اول و نوع دوم تعمیم‌یافته و K میانگین"، همایش ملی مهندسی کامپیوتر و توسعه پایدار، ۱۳۹۲.
- [23] Bolshakova, N. and Azuaje, F. "Cluster Validation Techniques for Genome Expression Data". Signal Processing Vol.83. No.4, pp.825-833, 2003.
- [24] Dimitriadou, E., Dolnicar, S. and Weingessel, A. "An Examination of Indexes for Determining the Number of Cluster in Binary Data Sets". Psychometrika, Vol.67, No.1, pp.137-160, 2002.
- ۸- مراجع
- [1] Kilincarsla, A. M. "A Study on Identification of Radar Emitters", M.S.c Thesis, Atılım University, April 2011.
- [2] Hansson, P. "Analysis of Some Methods for Deinterleaving of Pulse Trains", M.S.c Thesis, stockholm, sweden, 2007.
- [3] Kilincarslan, A. M., Uekae, B. and Kara, A. "Identification of Highly Jittered Radar Emitters: Issues on Low Cost Embedded Design". ICCIT, Ankara, Turkey, pp 739-744, 2012.
- [۴] نادری م. و موسوی ا.، "جداسازی و نمایش پالس‌های راداری در سیستم‌های پشتیبانی جنگ الکترونیک"، پایان‌نامه کارشناسی ارشد، دانشگاه علم و صنعت، ۱۳۷۹.
- [۵] مختاری، و. و کنگاوری، م.، "خوشه‌بندی داده‌های جریانی با استفاده از موازی‌سازی الگوریتم‌های ترکیبی"، پایان‌نامه کارشناسی ارشد، دانشگاه آزاد اسلامی قزوین، ۱۳۸۶.
- [۶] قلندری، م.، "طراحی و شبیه‌سازی الگوریتم تفکیک پالس و دسته‌بندی رادارها در پردازشگر ESM با امکان شناسایی رادارهای ECCM" پایان‌نامه کارشناسی ارشد، دانشگاه آزاد واحد نور، ۱۳۹۳.
- [7] Wiley, R. "ELINT: The Interception and Analysis of Radar Signals", Artech House, 2006.
- [۸] ناصری، ع. و سعادت‌ی مقدم، گ.، "ارائه یک الگوریتم هوشمند برای جداسازی سیگنال‌های متداخل راداری"، هجدهمین کنفرانس مهندسی برق ایران، دانشگاه صنعتی اصفهان، ص ۱۱ تا ۱۲۱، اردیبهشت ۱۳۸۹.
- [۹] نصرالهی، پ.، خداپرست، ف.، حسام پور، ک. و دادگر نیا، ا.، "خوشه‌بندی پالس‌های متداخل راداری با استفاده از مدل مخلوط گوسی در محیط‌های متراکم سیگنال"، ششمین کنفرانس ملی جنگ الکترونیک ایران، مهر ۱۳۹۲.
- [۱۰] ناصری، ع.، "ارائه الگوریتم مناسب برای بخش پردازش سیستم‌های شناسایی رادار ELINT و ESM"، نشریه علمی پژوهشی انجمن کامپیوتر ایران، مجلد ۱۰، ۱۳۹۱.
- [11] Ryoo Y. and Song K. "Recognition of PRI Modulation Types of Radar Signals Using the Autocorrelation", IEICE Transactions on Communications, Vol.90, No.6, pp.1290-1294, 2007.
- [12] Shahhoseini, H. S., Naseri, A. and Naderi, M. "A New Matrix Method for Pulse Train Identification", IEEE proceedings, MELECON, pp.183-187, 2002.