مجله تحقیق در عملیات در کاربردهای آن، سال پانزدهم، شماره چهارم (پیاپی ۵۹)، زمستان ۹۷، ص ص ۱۴-۱
شاپا ۷۲۸۶-۲۲۵۱

ارایه شاخصی جدید جهت سنجش اعتبار خوشه‌بندی در الگوریتم‌های خوشه‌بندی فازی نوع-۲

ایمان مسگری[۱*]، وحیدرضا سلامت[۲]، بهروز مینائی بیدگلی[۳]

۱- دکتری مهندسی صنایع، دانشگاه علم و صنعت ایران، دانشکده مهندسی صنایع، تهران، ایران

۲- دانشجوی دکتری، دانشگاه علم و صنعت ایران، دانشکده مهندسی صنایع، تهران، ایران

۲- دانشیار، دانشگاه علم و صنعت ایران، دانشکده مهندسی کامپیوتر، تهران، ایران

رسید مقاله: ۴ مهر ۱۳۹۴

پذیرش مقاله: ۱۰ مهر ۱۳۹۶

چکیده

یکی از مسایل اصلی در خوشه‌بندی فازی تعیین تعداد خوشه‌هاست که باید پیش از خوشه‌بندی در اختیار باشد و انتخاب مقادیر متفاوت برای تعداد خوشه‌ها، به خوشه‌بندی‌های متفاوتی منجر خواهد شد؛ بنابراین لازم است تا خوشه‌های مختلفی را که از مقادیر متفاوت تعداد خوشه‌ها به دست می‌آید با یک شاخص، اعتبارسنجی نمود؛ اما تا کنون شاخصی مخصوص الگوریتم‌های خوشه‌بندی فازی نوع-۲ (IT2 FCM) معرفی نشده است و به هنگام استفاده از این الگوریتم، از شاخص‌های معمول جهت تعیین تعداد خوشه‌ها استفاده می‌شود و این مقادیر نیز به طور ثابت و عمومی در نظر گرفته می‌شود. در این مقاله بنا داریم تاشاخصی جهت سنجش اعتبار خوشه‌بندی در این الگوریتم‌هامعرفی نماییم. بدین منظور، ابتدا مروری بر شاخص‌های اعتبار خوشه‌بندی و تحقیقات مرتبط با آن نموده و سپس ناپایداری استفاده از شاخص‌های موجود در الگوریتم‌های خوشه‌بندی فازی نوع-۲، نشان داده می‌شود. نتایج پیاده‌سازی شاخص پیشنهادی بر روی چهار مجموعه داده نشان می‌دهد که ناپایداری و اشکالات موجود در استفاده از شاخص‌های معمول در الگوریتم IT2 FCM، در شاخص پیشنهادی به علت به دست آوردن بازه بهینه، وجود ندارد. استفاده از شاخص معرفی شده می‌تواند اثر چشمگیری در کنترلرهای نوع-۲ (سیستم‌های منطق فازی نوع-۲) داشته باشد و منجر به بهبود نتایج پیش‌بینی و کنترل در این سیستم‌ها گردد.

کلمات کلیدی: شاخص اعتبار خوشه‌بندی، خوشه‌بندی فازی، خوشه‌بندی فازی نوع دوم.

۱ مقدمه

خوشه‌بندی یک روش طبقه‌بندی غیر نظارت شده است وقتی که داده‌ها برچسبی ندارند و اطلاعات ساختاری در مورد داده‌ها وجود ندارد. از یک منظر، خوشه‌بندی را می‌توان به دو دسته قطعی و فازی تقسیم کرد. در خوشه-

* عهده‌دار مکاتبات

آدرس الکترونیکی: imesgari@iust.ac.ir

[DOR: 20.1001.1.22517286.2019.15.4.2.4]

بندی قطعی، هر عضو به شکل قطعی، تنها و تنها به یک خوشه تخصیص می‌یابد؛ در حالی که در خوشه‌بندی فازی، به هر عضو عددی مانند u_{ij} عددی در بازه صفر و یک را طوری نسبت می‌دهیم که بیانگر درجه عضویت عنصر jام به خوشه iام است. وقتی که خوشه‌های موجود در داده‌ها در طبیعت خود دارای اشتراکاتی هستند، خوشه‌بندی فازی می‌تواند اطلاعات بیش‌تری در اختیار قرار دهد.

خوشه‌بندی فازی به طور گسترده مورد مطالعه و استفاده قرار گرفته است؛ اما با این حال یک مساله اصلی در خوشه‌بندی فازی تعیین تعداد خوشه‌هاست که باید پیش از خوشه‌بندی در اختیار باشد. انتخاب مقادیر متفاوت برای تعداد خوشه‌ها به خوشه‌بندی‌های متفاوتی منجر خواهد شد؛ بنابراین لازم است تا خوشه‌های مختلفی که از مقادیر متفاوت تعداد خوشه‌ها به دست می‌آید با یک شاخص اعتبارسنجی شود. در این زمینه معیارهای متفاوتی برای اعتبارسنجی پیشنهاد شده است. بعضی از این شاخص‌ها فقط از مقادیر عضویت استفاده می‌کنند و دسته دیگری از شاخص‌ها علاوه بر مقادیر عضویت از ماتریس داده‌ها نیز بهره می‌برند.

محققان بر پایه الگوریتم FCM و به جهت برخورد با عدم قطعیت موجود در پارامترهای این الگوریتم، اقدام به توسعه و معرفی الگوریتم‌های FCM نوع ۲ نمودند که الگوریتم‌های T2 FCM نامیده می‌شود. الگوریتم‌های FCM نوع ۲ بازه‌ای نیز نوع خاصی از الگوریتم‌های T2 FCM هستند که به صورت IT2 FCM نشان داده می‌شود. الگوریتم‌های مزبور، قدرتمند و از نظر محاسباتی کاملا کارا هستند که این موضوع در مقالات مختلفی مورد بحث قرار گرفته است که در ادامه به آن‌ها اشاره خواهد شد.

آنچه ما در این تحقیق به آن علاقه‌مندیم آن است که در پیاده‌سازی الگوریتم‌های IT2 FCM نیز نیاز است تا از شاخصی برای تعیین تعداد مناسب خوشه‌ها استفاده گردد؛ اما تا به حال شاخصی تحت عنوان شاخص اعتبار خوشه‌بندی نوع-۲ معرفی نشده است و محققان از شاخص‌های مربوط به الگوریتم‌های FCM معمول، برای تعیین تعداد بهینه خوشه‌ها استفاده کرده‌اند. در ادامه ضمن بیان نحوه استفاده از شاخص‌های معمول اعتبار برای الگوریتم‌های خوشه‌بندی نوع ۲ و نارسایی‌های آن‌ها، به ارایه‌ای شاخصی خواهیم پرداخت که به طور خاص در الگوریتم‌های خوشه‌بندی IT2 FCM کاربرد خواهد داشت.

ادامه مقاله بدین ترتیب سازمان یافته است: در بخش دوم، مروری بر مطالعات صورت گرفته در رابطه با شاخص‌های اعتبار خوشه‌بندی خواهیم داشت. در بخش سوم، الگوریتم FCM نوع دوم و ناپایداری شاخص‌هایی که تا کنون جهت اعتبارسنجی در این الگوریتم به‌کار رفته است، نشان داده می‌شود و در انتها شاخص اعتبار خوشه‌بندی پیشنهادی را برای آن معرفی می‌نماییم. بخش چهارم به ارایه نتایج تجربی حاصل از به‌کارگیری شاخص معرفی شده بر روی چهار مجموعه داده اختصاص دارد و در نهایت، بخش پنجم، این مقاله را جمع‌بندی می‌نماید.

۲ پیشینه تحقیق

۲-۱ شاخص‌هایی که فقط از مقادیر عضویت استفاده می‌کنند

[DOR: 20.1001.1.22517286.2019.15.4.2.4]

بزدک [۱] تلاش کرد تا یک شاخص عملکرد تعریف کند که بر پایه حداقل کردن مقدار کلی اشتراکات دوتایی فازی در ماتریس U عمل می‌کرد. میزان بهینه تعداد خوشه‌ها با حداکثر کردن ضریب تقسیم (PC) حاصل می‌شد. بزدک [۲] همچنین شاخص آنتروپی تقسیم‌بندی (PE) را معرفی کرد. شاخص PE یک مقیاس عددی است که میزان فازی بودن رادر یک مجموعه داده شده U محاسبه می‌کند. محدودیت این شاخص در یکنواختی آشکار آن و طبیعت ابتکاری که اساس منطقی این فرمول را تشکیل می‌دهد، است. میزان بهینه تعداد خوشه‌ها c^* با حداکثر کردن آنتروپی تقسیم‌بندی حاصل می‌گردد.

ویندهام [۳] شاخص توان تقسیم (WPE) را ارایه نمود. این شاخص از n ماکزیمم در ستون‌های ماتریس U استفاده می‌کند. وی ادعا کرد که طبیعی است که بر n ماکزیمم مربوط به هر ستون تاکید کنیم؛ زیرا میزان زیاد ماکزیمم‌ها تمایز را به نحو بهتری در ساختار X نشان خواهند داد. میزان بهینه تعداد خوشه‌ها با حداکثر کردن شاخص توان تقسیم حاصل می‌گردد تا بهترین عملکرد به دست آید.

کیم [۴] شاخص اعتبار خود (KYI) را به صورت میانگین درجات نسبی به اشتراک‌گذاری جفت خوشه‌ها معرفی کرد، وقتی که درجه نسبی به اشتراک‌گذاری هر زوج از خوشه‌ها به عنوان مجموع وزنی درجات نسبی به اشتراکات در دو خوشه مورد بررسی باشد. در واقع تمام انتخاب‌های ۲ تایی خوشه‌ها مورد بررسی قرار گرفت، اشتراکات بین این دو خوشه فازی مبنای محاسبات قرار می‌گیرد و بنابراین اشتراک کم‌تر موجود در یک تقسیم‌بندی فازی و همچنین ابهام کم تر در مرز خوشه‌ها منجر به میزان کم‌تری از KYI می‌شود. به طور کلی میزان بهینه تعداد خوشه‌ها با حداکثر کردن شاخص KYI به دست می‌آید.

چِن و لینکنز [۵] نیز شاخصی برای اعتبار خوشه‌بندی ارایه کردند که شاخص P نام گرفت. این شاخص به صورت تفاضل دو عبارت بیان گردید. عبارت اول، تراکم و به هم فشردگی را در داخل یک خوشه نشان می‌دهد و بیان‌کننده میزان نزدیکی شی‌های تخصیص داده شده به یک خوشه به مرکز خوشه است و بنابراین مقدار بیش‌تر آن، نشان‌دهنده خوب بودن طبقه‌بندی عناصر خواهد بود. عبارت دوم میزان تمایز بین خوشه‌ها را نشان می‌دهد. در اینجا اشتراک دو مجموعه فازی برای ارزیابی میزان تمایز بین دو خوشه مورد استفاده قرار می‌گیرد و این عبارت سعی می‌کند تا اطلاعات مربوط به تراکم و تمایز در خوشه‌ها را با هم ترکیب نماید و اشتراکات بین دو خوشه به حداقل برسد؛ بنابراین میزان بهینه تعداد خوشه‌ها با توجه به ماکزیمم کردن مقدار شاخص P تعیین خواهد شد.

شاخص‌های ارایه شده در بالا فقط از میزان عضویت‌ها استفاده می‌کنند و این مساله می‌تواند به عنوان ضعف آن‌ها تلقی گردد. از جمله این ضعف‌ها عبارتند از:وابستگی یکنواخت آن‌ها به تعداد خوشه‌ها، حساسیت بالای آن‌ها به فازی‌ساز و کمبود دسترسی به اطلاعات مستقیم نسبت به موقعیت جغرافیایی داده‌ها به‌علت عدم استفاده از داده‌ها.

۲-۲ شاخص‌هایی که از مقادیر عضویت و اطلاعات داده‌ها استفاده می‌کنند

فوکویاما و سوگِنو [۶] شاخص اعتبار خود (FS) را به صورت تفاضل دو عبارت بیان کردند. عبارت اول، میزان فازی بودن در U را با میزان به هم فشردگی جغرافیایی داده‌ها ترکیب می‌کند و عبارت دوم، میزان فازی بودن در

[DOR: 20.1001.1.22517286.2019.15.4.2.4]

هر سطر i از ماتریس U را با فاصله مرکز خوشه i ام از میانگین مرکز خوشه‌ها ترکیب می‌کند. به طور کلی میزان بهینه تعداد خوشه‌ها با ماکزیمم کردن شاخص FS به دست می‌آید تا بهترین عملکرد در خوشه‌بندی حاصل شود.

ژی و بِنی [۷] یک تابع اعتبارسنجی دیگر ارایه کردند و بعدا توسط پال و بزدک [۸] اصلاح گردید. این شاخص بر دو مشخصه اصلی تاکید می‌کند: جدایی خوشه‌ها و به هم فشردگی داخل خوشه‌ها. در معادله ارایه شده برای این شاخص که به صورت کسری است، صورت کسر نشان‌دهنده به هم فشردگی خوشه‌های فازی است و مخرج آن نشان‌دهنده قدرت جدایی بین خوشه‌هاست. آن‌ها بیان کردند که در یک خوشه خوب، مقدار شاخص فشردگی، زیاد خواهد بود و مقدار شاخص جدایی خوشه‌ها، کم خواهد بود؛ بنابراین میزان بهینه تعداد خوشه‌ها با ماکزیمم کردن مقدار این کسر به دست خواهد آمد.

شاخص بعدی را کِوُن [۹] براساس شاخصی که ژی و بِنی [۷] ارایه کرده بودند، بیان کرد. او تمایل داشت تا میزان این شاخص را در هنگامی که تعداد خوشه‌ها به تعداد داده‌ها نزدیک می‌شود کاهش دهد و اعتقاد داشت این مساله ضعفی برای آن شاخص خواهد بود. به همین منظور او یک عبارت تنبیه‌کننده به صورت شاخص قبلی افزود و به شاخص جدیدی دست پیدا کرد.

زَهید [۱۰]، مفاهیم به هم فشردگی و تمایز فازی بین خوشه‌ها را به‌وسیله شاخص‌های اعتبارسنجی سنتی معرفی کرد. شاخص وی حاصل تفاضل دو عبارت است که عبارت اول، مشخصات جغرافیایی ساختار داده‌ها و تابع عضویت را در نظر می‌گیرد و عبارت دوم از اجتماع و اشتراک فازی برای دست‌یابی به درجه تمایز یا به هم فشردگی فازی استفاده می‌کند.

وو و یانگ [۱۱] شاخص اعتبار دیگری را معرفی کردند، بدین صورت که میزان بزرگ‌تر آن به این معناست که خوشه‌ها، فشرده‌تر و متمایزترند و مقدار کوچک تر به این معناست که بعضی از این خوشه‌ها کاملا فشرده و یا متمایز از یکدیگر نیستند. حداکثر کردن این شاخص با توجه به تغییر تعداد خوشه‌ها، می‌تواند برای شناسایی ساختار داده‌ها به نحوی که خوشه ها به هم فشرده و از یکدیگر متمایز باشند استفاده شود. در شرایط پیشنهادی برای این شاخص، یک نقطه نویز نمی تواند مشکلی در الگوریتم ایجاد کند و بنابراین این شاخص می‌تواند نتایج شگفت‌انگیزی در مورد چنین داده‌هایی ایجاد کند.

شاخص بعدی توسط کیم و همکاران [۱۲] به صورت نسبت درجه اشتراک بین خوشه‌های فازی به درجه جدایی خوشه‌ها از یکدیگر تعریف گردید. یک مقدار کم برای این شاخص، بخشی را نشان می‌دهد که در آن خوشه‌ها دارای اشتراک با درجه کم‌تر و جدایی بیش‌تر از یکدیگر هستند؛ بنابراین میزان بهینه تعداد خوشه‌ها با مینیمم کردن این شاخص با تغییر تعداد خوشه‌ها به‌دست می‌آید تا بهترین عملکرد در خوشه‌بندی حاصل شود.

پَخیرا و همکاران[۱۳] نیز شاخصی را تحت عنوان PBM معرفی کردند که برای خوشه‌بندی فازی و قطعی کابرد دارد و از حاصل‌ضرب سه عبارت شکل گرفته است. عبارت اول قابلیت تقسیم سیستم به c خوشه را نشان می‌دهد و مقدار آن با افزایش c کاهش می‌یابد. عبارت دوم شامل مجموع وزنی فواصل درون خوشه‌ای برای کل مجموعه داده به عنوان یک خوشه و در حالت c خوشه است. در این عبارت مخرج با افزایش تعداد خوشه‌ها کاهش می‌یابد و صورت تغییر نمی‌کند. ثابت بودن صورت باعث می‌شود تا شانس اینکه عبارت دوم مقدار خیلی

[DOR: 20.1001.1.22517286.2019.15.4.2.4]

کوچکی شود از بین برود. این عبارت معیاری برای سنجش فشردگی در سیستم c خوشه‌ای است. عبارت سوم، حداکثر جدایی بین خوشه‌ها را نشان می‌دهد و بیان‌کننده جدایی بین خوشه‌هاست؛ بنابراین وقتی که اولین عبارت کاهش می‌یابد دو عبارت دیگر با افزایش c افزایش خواهند یافت. بر پایه تحلیل‌های فوق میزان بهینه تعداد خوشه‌ها با حل مساله ماکزیمم سازی PBM با تغییر تعداد خوشه‌ها به دست می‌آید.

۲-۳ سایر رویکردها در مورد شاخص اعتبار خوشه‌بندی

بعضی از شاخص های پیشین که مورد بررسی قرار گرفت تنها بر فشردگی و پراکندگی درونی خوشه تمرکز می‌کنند. این موضوع باعث می‌شود که توانایی آن‌ها در ایجاد یک نمایش صحیح از ساختار داده با محدودیت مواجه شود؛ زیرا در مورد جدایی و تمایز خوشه‌ها، فقط فاصله بین خوشه‌ها را در نظر می‌گیرند. در واقع اگر تعداد خوشه‌ها به تعداد داده‌ها نزدیک شود، فاصله بین مرکز خوشه و نمونه‌ها صفر می‌شود و شاخص‌های سنتی، توانایی خود را برای اعتبار بخشی به تقسیم‌بندی فازی برای تعداد بزرگ خوشه‌ها از دست می‌دهند.

روش اعتبارسنجی بیزی فازی که از مفهوم بیزین در تئوری احتمال الهام گرفته شده است، یک تقسیم‌بندی فازی را انتخاب می‌کند که بیش‌ترین عضویت را در مجموعه داده داشته باشد (باراش و فریدمن [۱۴]). شاخص امتیاز بیزی (BS) که توسط چو و یو [۱۵] پیشنهاد گردید با انتقال اصول تئوری کلاسیک بیز به عضویت‌ها و به کار گیری ضرب و قاعده استقلال شکل گرفت.

چانگ و همکاران [۱۶] رویکردی ترکیبی برای حل مساله اعتبار خوشه‌بندی پیشنهاد کردند. تکنیک پیشنهادی از شاخص اعتبار خوشه‌بندی فازی (FS) در ترکیب با شاخص ادغام استفاده می‌کند تا تعداد بهینه خوشه را در مجموعه داده پیدا کند. تصمیم در مورد اینکه آیا مراکز ادغام شوند یا نه می‌تواند بر پایه نقطه میانی اتخاذ شود. تکنیک رویکرد ترکیبی دو گام دارد. در گام اول یک شاخص اعتبار خوشه‌بندی استفاده می‌شود تا یک تخمین غیر دقیق از تعداد بهینه خوشه‌ها به دست آید و این تعداد در گام بعدی به‌وسیله شاخص ادغام اصلاح خواهد شد.

وو و همکاران [۱۷] در پژوهشی که اخیرا انجام داده‌اند، مشکل ناپایداری شاخص اعتبارسنجی خوشه‌بندی را هنگامی که مراکز خوشه‌های به یکدیگر نزدیک هستند مد نظر قرار داده‌اند و شاخص جدیدی تحت عنوان WLI برای اعتبارسنجی خوشه‌بندی ارایه کرده‌اند. این شاخص، حداقل فاصله ومتوسط فاصله بین دو جفت مرکز خوشه را در نظر می‌گیرد و بدین ترتیب، تا حدودی وجود مراکز خوشه نزدیک به هم را مجاز می‌شمارد و از این طریق، ناپایداری موجود را تا حدی کاهش می‌دهد.

۳ شاخص اعتبار خوشه‌بندی برای الگوریتم FCM نوع-۲

۳-۱ الگوریتم FCM نوع ۲

[DOR: 20.1001.1.22517286.2019.15.4.2.4]

محققان بر پایه الگوریتم *FCM* اقدام به توسعه الگوریتم‌های FCM نوع ۲ نمودند که T2 FCM نامیده می‌شود. الگوریتم‌های FCM نوع ۲ بازه‌ای، نوع خاصی از الگوریتم‌های T2 FCM خواهند بود که به صورت IT2 FCM نشان داده می‌شود. در واقع در مواردی که استفاده از اعداد فازی نوع ۱ نتایج خوبی ندارد و یا نمی‌تواند عدم قطعیت موجود را در مساله به خوبی پوشش دهد، استفاده از مجموعه‌های فازی نوع دوم مورد توجه قرار می‌گیرد. مِلین و کاستلو [۱۸] مرور جامعی بر کاربردهای این الگوریتم در خوشه‌بندی، دسته‌بندی و تشخیص الگو انجام داده‌اند.

در الگوریتم FCM، پارامترهایی به عنوان پارامتر ورودی الگوریتم وجود دارند که عدم قطعیت‌های موجود در آن‌ها استفاده از مجموعه‌های نوع دوم را برای پوشش این عدم قطعیت توجیه‌پذیر می‌نماید زرندی و همکاران [۱۹] نشان می‌دهند مدل‌سازی عدم قطعیت موجود در پارامترهای ورودی با استفاده از مجموعه‌های فازی نوع ۲ می‌تواند باعث کارایی بیش‌تر الگوریتم شود.

پارامترهایی که منابع عدم قطعیت در آن‌ها مورد بررسی قرار گرفته و منجر به تولید الگوریتم‌های T2 FCM شده‌است عبارت است از: پارامتر فازی‌ساز (m) و «سنجه فاصله»[1]. در الگوریتم FCM از یک تابع برای سنجش فاصله بین نقاط و مراکز خوشه‌ها استفاده می‌شود. نوع این تابع با توجه به ساختار داده‌ها تعیین می‌شود و اثر زیادی بر نتایج الگوریتم خواهد داشت. از انواع این توابع که «سنجه فاصله» نامیده می‌شود می‌توان به نرم اقلیدسی، ماهالابونیس، نرم قطری و ... اشاره کرد. در مقالاتی که مورد اشاره قرار گرفت سعی شده است تا عدم قطعیت موجود در انتخاب از بین نرم‌های مختلف از طریق به‌کار گیری الگوریتم‌های حل شود. در این نوع از الگوریتم‌ها میزان عضویت هر نقطه در داده‌ها به هر خوشه با استفاده از مجموعه‌ای از «سنجه‌های فاصله» محاسبه خواهد شد.

بعضی از CVI ها از عبارت فازی‌ساز (m) استفاده می‌کنند در حالی که دسته دیگر از این عبارت استفاده نمی‌کنند. دسته اخیر، علاوه بر یافتن مقدار بهینه تعداد خوشه‌ها (c)، می‌توانند مقدار بهینه m را نیز معرفی کنند. به عبارت دیگر مقداری از c و m پاسخ مساله است که میزان شاخص مورد نظر را حداقل (یا حداکثر) نماید. برای یافتن مقدار m از این طریق، بازه‌ای به عنوان بازه محتمل برای m فرض می‌شود و از طریق شمارش، تمام ترکیبات m و c در شاخص CVI قرار داده شده و میزان حداقل (یا حداکثر) شاخص، مشخص کننده m و c بهینه خواهد بود. مقادیر m به صورت گسسته تقسیم‌بندی می‌شود، فرض می‌شود رفتار m در فواصل بین مقادیر گسسته متناسب با دو سر بازه خواهد بود که فرض درستی است.

تا کنون، محققان از شاخص‌های مربوط به الگوریتم‌های FCM معمول برای تعیین تعداد بهینه خوشه‌ها در الگوریتم‌های T2 FCM استفاده کرده‌اند ([۲۰]، [۲۱] و [۲۲]). نحوه استفاده از شاخص‌های معمول اعتبار برای الگوریتم‌های خوشه‌بندی نوع ۲ و محدودیت‌های آن در بخش بعد توضیح داده خواهد شد.

[1]Distance measure

[DOR: 20.1001.1.22517286.2019.15.4.2.4]

۳-۲ ناپایداری شاخص‌های ارایه شده برای الگوریتم T2 FCM

الگوریتم‌های T2 FCM سعی می‌کنند تا عدم قطعیت موجود در یکی از پارامترها را مورد توجه قرار دهند. الگوریتم‌هایی که عدم قطعیت موجود در میزان فازی‌ساز (m) را به عنوان منبع عدم اطمینان قرار می‌دهند به شدت تحت تاثیر مقادیر m هستند؛ لذا تعیین دو مقدار m نقش بسیار مهمی در کارایی الگوریتم خواهد داشت. محققان تا به حال از این موضوع به سادگی گذشته و دو مقدار عمومی به عنوان بازه اصلی و عمومی برای مقادیر m در نظر گرفته‌اند. مقادیر عمومی و فارغ از ساختار داده برای مقدار m در مقالات پژوهشگرانی مانند هوآنگ و همکاران [۲۳]، اُزکان و تُرکسن [۲۴] و فاضل زرندی و همکاران [۲۵] مورد بررسی قرار گرفته است. در این مقالات با توجه به روابط ریاضی موجود در الگوریتم FCM و بسط تیلور رابطه تولید کننده مقادیر عضویت، بازه‌ای برای m به دست می‌آید که فارغ از ساختار داده‌ها به طور کلی نشان می‌دهد که در چه بازه‌ای m بیش‌ترین تاثیرگذاری را خواهد داشت. این مقالات عموما بازه‌ای ۱/۵تا ۲/۵ را به عنوان بازه مورد نظر برای m معرفی می‌کنند و ادعا می‌کنند که تغییرات m در این بازه چشمگیر خواهد بود و در خارج از این بازه تغییرات m اثر چشمگیری بر مقادیر عضویت نخواهد داشت.

در نظر گرفتن مقادیر ثابت و عمومی برای مقدار m می‌تواند منجر به مشکلات متعددی در فرایند خوشه‌بندی گردد. عمده هدف استفاده از الگوریتم T2 FCM به جهت پوشش عدم قطعیت موجود در m است در حالی‌که با در نظر گرفتن مقادیر ثابت و عمومی و فارغ از ساختار داده برای m، این هدف نقض می‌شود؛ زیرا از یک طرف به دنبال پوشش عدم قطعیت موجود در مقادیر m هستیم و از طرف دیگر با در نظر گفتن مقادیر عمومی و ثابت برای m به راحتی از تاثیری که این مقادیر بر ادامه الگوریتم خواهد داشت چشم‌پوشی می‌کنیم.

بازه‌هایی مانند ۱/۵تا ۲/۵ که توسط محققان [۲۳] [۲۴] برای m در نظر گرفته شده تا در الگوریتم خوشه‌بندی T2 FCM مورد استفاده قرار گیرد، نمی‌تواند به خوبی عدم قطعیت موجود در m را پوشش دهد به خصوص هنگامی که این بازه فارغ از ساختار داده و برای هر نوع داده‌ای تجویز شده است. در واقع رفتار شاخص اعتبار خوشه‌بندی در این بازه نمی‌تواند منعکس کننده عدم قطعیت مورد نظر باشد. یکی از مسایل مهم در مورد بازه در نظر گرفته شده برای مقدار m پایداری شاخص CVI در این بازه و تولید نتایج سازگار است.

برای بررسی پایداری یک شاخص اعتبار خوشه‌بندی در بازه‌های مختلف m، ۴ مجموعه داده مختلف در نظر گرفته شدند که نتایج آن در جداول ۲ تا ۵ خلاصه شده است. خصوصیات این مجموعه داده‌ها در جدول ۱ بیان شده است. برای بررسی پایداری شاخص‌های مختلف اعتبار خوشه‌بندی و بررسی مناسب بودن بازه ۱/۵تا ۲/۵، ۶ شاخص خوشه‌بندی مختلف انتخاب گردید. این شاخص‌ها در جداول در ستون سمت راست نشان داده شده‌اند. هر کدام از این شاخص‌ها در بخش مرور ادبیات مورد اشاره قرار گرفته‌است و در همه آن‌ها عبارت فازی‌ساز یا m به‌کار رفته است. همچنین برای هر مجموعه داده، مقدار فازی‌ساز از ۱/۲ تا ۹ مورد نظر قرار داده شد و مقادیر این بازه با فواصل ۰/۲ در الگوریتم مورد نظر قرار گرفتند. مقادیر m در سرستون‌های جداول فوق نشان داده‌شده‌است. در هر یک از ستون‌های جداول تعداد بهینه خوشه‌ها با توجه به m و شاخص مورد نظر درج شده است.

[DOR: 20.1001.1.22517286.2019.15.4.2.4]

جدول ۱.مشخصات مجموعه داده‌های استفاده شده

عنوان مجموعه داده	تعداد نقاط	تعداد مشخصه	تعداد خوشه‌های پیش‌فرض	محل دسترسی
Seeds data set	۲۱۰	۷	۳	UCI machine
Example3	۶۰	۲	۳	ساخته شده
Ruspini	۷۵	۲	۴	UCI machine learning
Iris	۱۵۰	۴	۳	UCI machine learning

جدول ۲. نتایج شاخص‌های اعتبار خوشه‌بندی بر روی مجموعه داده Seeds

m	۱/۲	۱/۴	۱/۶	۱/۸	۲	۲/۲	۲/۴	۲/۶	۲/۸	۳	۳/۲	۳/۴	۳/۶	۳/۸	۴	۴/۲	۴/۴	۴/۶	۴/۸	۵
CVI																				
FS	۱۱	۱۳	۱۲	۱۲	۳	۳	۳	۳	۳	۳	۳	۳	۳	۲	۲	۲	۲	۲	۲	۲
XB	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۱۳	۲	۱۱	۳	۱۳	۱۴	۱۴	۱۳	۱۰	۱ ۳
FHV	۴	۴	۷	۴	۶	۵	۱۴	۷	۳	۲	۲	۴	۱۴	۳	۱۴	۱۱	۱۰	۱۴	۱۴	۱ ۲
APD	۷	۱۳	۵	۶	۱۳	۱۳	۱۴	۱۳	۱۳	۶	۷	۱۴	۱۴	۱۴	۱۴	۱۴	۱۴	۱۴	۱۴	۱ ۴
PD	۶	۳	۲	۶	۶	۳	۳	۶	۳	۳	۳	۲	۱۰	۱۴	۱۳	۱۴	۱۲	۲	۱۴	۲
SCG	۵	۶	۶	۶	۳	۳	۳	۳	۳	۳	۲	۲	۲	۲	۲	۲	۱۴	۱۴	۲	۲
m	۵/۲	۵/۴	۵/۶	۵/۸	۶	۶/۲	۶/۴	۶/۶	۶/۸	۷	۷/۲	۷/۴	۷/۶	۷/۸	۸	۸/۲	۸/۴	۸/۶	۸/۸	۹
CVI																				
FS	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲
XB	۱۲	۱۰	۹	۴	۴	۱۴	۱۳	۷	۷	۱۱	۱۴	۱۱	۱۳	۱۳	۱۴	۱۴	۱۳	۱۴	۱۲	۱۲
FHV	۸	۲	۲	۸	۹	۶	۲	۵	۵	۶	۲	۸	۸	۲	۶	۳	۲	۳	۴	۳
APD	۱۴	۱۳	۱۳	۱۳	۱۰	۱۰	۱۲	۸	۴	۷	۶	۲	۳	۲	۵	۲	۵	۲	۳	۵
PD	۵	۳	۵	۳	۴	۹	۲	۵	۲	۲	۳	۲	۲	۲	۲	۳	۲	۲	۲	۲
SCG	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲

جدول ۳. نتایج شاخص‌های اعتبار خوشه‌بندی بر روی مجموعه داده example

m	۱/۲	۱/۴	۱/۶	۱/۸	۲	۲/۲	۲/۴	۲/۶	۲/۸	۳	۳/۲	۳/۴	۳/۶	۳/۸	۴	۴/۲	۴/۴	۴/۶	۴/۸	۵
CVI																				
FS	۶	۵	۷	۷	۴	۳	۳	۳	۳	۳	۳	۳	۳	۳	۳	۳	۳	۳	۳	۳
XB	۳	۳	۳	۳	۳	۳	۳	۳	۳	۳	۳	۳	۳	۳	۳	۳	۳	۳	۳	۳
FHV	۴	۲	۴	۸	۸	۶	۳	۶	۷	۶	۴	۳	۴	۳	۸	۶	۵	۲	۳	۸
APD	۷	۸	۷	۷	۸	۷	۶	۷	۸	۸	۴	۳	۳	۶	۸	۵	۸	۵	۸	۷
PD	۸	۴	۸	۵	۲	۳	۷	۳	۵	۲	۸	۸	۸	۸	۸	۷	۴	۸	۸	۳
SCG	۸	۵	۶	۶	۸	۶	۷	۸	۳	۷	۳	۳	۳	۸	۳	۳	۳	۸	۸	۸

[1]www.ics.**uci**.edu/~mlearn/

[DOR: 20.1001.1.22517286.2019.15.4.2.4]

ادامه جدول ۳

m / *CVI*	۵/۲	۵/۴	۵/۶	۵/۸	۶	۶/۲	۶/۴	۶/۶	۶/۸	۷	۷/۲	۷/۴	۷/۶	۷/۸	۸	۸/۲	۸/۴	۸/۶	۸/۸	۹
FS	۳	۳	۳	۳	۳	۲	۸	۸	۸	۸	۸	۷	۸	۵	۵	۵	۷	۵	۲	۲
XB	۳	۳	۳	۳	۳	۸	۵	۵	۸	۵	۷	۷	۸	۸	۶	۸	۸	۷	۷	۸
FHV	۷	۸	۸	۸	۸	۶	۸	۷	۴	۷	۸	۵	۲	۶	۴	۷	۴	۴	۲	۳
APD	۳	۸	۸	۷	۸	۸	۸	۸	۸	۷	۸	۷	۷	۸	۷	۵	۷	۵	۸	۵
PD	۵	۸	۳	۶	۷	۷	۸	۲	۷	۸	۸	۶	۷	۵	۷	۶	۵	۳	۳	۵
SCG	۸	۸	۸	۵	۸	۸	۵	۷	۶	۶	۵	۷	۵	۵	۶	۵	۶	۶	۶	۵

جدول ٤. نتایج شاخص‌های اعتبار خوشه‌بندی بر روی مجموعه داده iris

m / *CVI*	۱/۲	۱/۴	۱/۶	۱/۸	۲	۲/۲	۲/۴	۲/۶	۲/۸	۳	۳/۲	۳/۴	۳/۶	۳/۸	۴	۴/۲	۴/۴	۴/۶	۴/۸	۵
FS	۱۱	۷	۵	۵	۵	۳	۳	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲
XB	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۷	۲	۲	۳
FHV	۲	۱۱	۸	۱۲	۱۰	۴	۲	۲	۳	۲	۶	۳	۱۱	۱۲	۴	۲	۲	۲	۲	۳
APD	۱۰	۹	۹	۷	۷	۷	۱۱	۱۱	۱۰	۴	۴	۶	۵	۱۲	۳	۱۲	۱۲	۷	۹	۷
PD	۲	۸	۲	۶	۲	۷	۲	۲	۲	۳	۳	۲	۲	۲	۲	۲	۲	۲	۲	۴
SCG	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۷	۲	۴
m / *CVI*	۵/۲	۵/۴	۵/۶	۵/۸	۶	۶/۲	۶/۴	۶/۶	۶/۸	۷	۷/۲	۷/۴	۷/۶	۷/۸	۸	۸/۲	۸/۴	۸/۶	۸/۸	۹
FS	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲	۲
XB	۴	۴	۴	۴	۴	۴	۴	۵	۴	۴	۱۲	۱۰	۱۲	۹	۱۱	۸	۱۲	۱۲	۱۰	۱۲
FHV	۳	۲	۶	۲	۲	۴	۱۱	۳	۳	۴	۲	۳	۳	۳	۳	۲	۲	۷	۲	۴
APD	۷	۸	۳	۳	۵	۹	۲	۲	۲	۸	۸	۴	۴	۲	۳	۲	۵	۳	۴	۵
PD	۲	۲	۳	۲	۲	۲	۲	۷	۲	۲	۲	۲	۲	۲	۲	۴	۵	۲	۲	۲
SCG	۴	۴	۴	۴	۴	۴	۵	۴	۴	۵	۱۲	۱۰	۱۲	۹	۱۱	۸	۱۲	۱۲	۱۰	۱۲

جدول ۵. نتایج شاخص‌های اعتبار خوشه‌بندی بر روی مجموعه داده ruspini

m / *CVI*	۱/۲	۱/۴	۱/۶	۱/۸	۲	۲/۲	۲/۴	۲/۶	۲/۸	۳	۳/۲	۳/۴	۳/۶	۳/۸	۴	۴/۲	۴/۴	۴/۶	۴/۸	۵
FS	۹	۷	۵	۷	۴	۴	۴	۴	۴	۴	۴	۴	۴	۴	۴	۳	۴	۴	۴	۴
XB	۴	۴	۴	۴	۴	۴	۴	۴	۴	۴	۴	۴	۴	۴	۴	۴	۴	۴	۴	۴
FHV	۹	۸	۲	۲	۶	۸	۲	۲	۳	۲	۳	۴	۹	۲	۲	۴	۸	۹	۲	۹
APD	۸	۵	۴	۹	۸	۸	۴	۴	۴	۴	۲	۲	۴	۹	۹	۵	۴	۲	۷	۸
PD	۳	۸	۹	۳	۴	۷	۳	۲	۳	۷	۵	۴	۵	۳	۴	۲	۹	۵	۲	۹
SCG	۵	۷	۳	۴	۴	۳	۸	۴	۳	۸	۳	۳	۵	۳	۳	۴	۳	۶	۸	۸

[DOR: 20.1001.1.22517286.2019.15.4.2.4]

ادامه جدول ۵

m / CVI	۵/۲	۵/۴	۵/۶	۵/۸	۶	۶/۲	۶/۴	۶/۶	۶/۸	۷	۷/۲	۷/۴	۷/۶	۷/۸	۸	۸/۲	۸/۴	۸/۶	۸/۸	۹
FS	۳	۳	۹	۹	۹	۹	۹	۹	۹	۹	۹	۹	۹	۹	۹	۹	۹	۸	۸	۸
XB	۴	۴	۴	۴	۴	۴	۴	۸	۶	۷	۸	۵	۷	۸	۷	۸	۸	۷	۶	۶
FHV	۵	۷	۲	۹	۶	۳	۷	۹	۳	۹	۹	۲	۹	۹	۸	۹	۸	۶	۶	۷
APD	۹	۹	۸	۹	۸	۹	۹	۹	۹	۸	۹	۸	۹	۸	۹	۹	۷	۹	۷	۹
PD	۳	۹	۹	۹	۹	۲	۲	۹	۹	۹	۹	۹	۳	۹	۹	۹	۹	۹	۸	۵
SCG	۶	۹	۶	۹	۹	۹	۹	۹	۹	۹	۹	۶	۹	۸	۹	۹	۹	۶	۸	۸

نتایج نشان می‌دهد که تغییرات در تعداد بهینه خوشه‌ها به شدت از میزان m تاثیر پذیر است. به علاوه نتایج به خوبی نشان می‌دهد که تعداد بهینه خوشه‌ها در بازه ۱/۵ تا ۲/۵ به شدت متغیر است و شاخص‌های مورد بررسی در این بازه از سازگاری و پایداری مناسبی برخوردار نیستند. لازم به ذکر است تمامی شاخص‌های معرفی شده قادر به یافتن تعداد بهینه (تعداد پیش‌فرض در مورد هر مجموعه داده) خوشه‌ها در میزانی از m هستند. این تعداد c در کنار m مرتبط، میزان m و c بهینه را با توجه به آن شاخص نشان می‌دهند؛ ولی در بازه ۱/۵ تا ۲/۵ شاخص‌های عملکرد، پایداری ندارند و مقادیر مختلفی را به عنوان میزان c بهینه معرفی می‌کنند که این ناپایداری می‌تواند منجر به آسیب‌های جدی به الگوریتم IT2 FCM شود که از مقدار واحد c و مقادیر بازه‌ای m (۱/۵ تا ۲/۵) استفاده می‌کند.

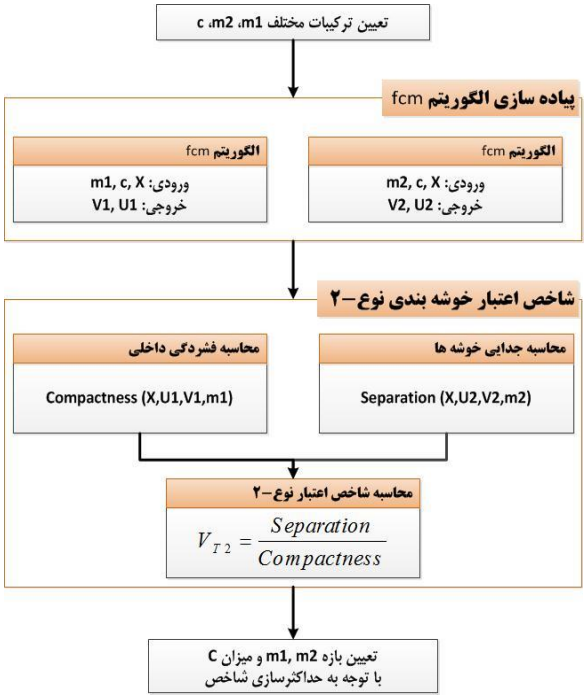
با توجه به مسایلی که مطرح شد باید به معرفی شاخص اعتباری پرداخت که مشکلات فوق در آن مرتفع شده باشد. در بخش بعد به معرفی این شاخص که ما پیشنهاد داده ایم، می‌پردازیم.

۳-۳ شاخص پیشنهادی برای الگوریتم IT2 FCM

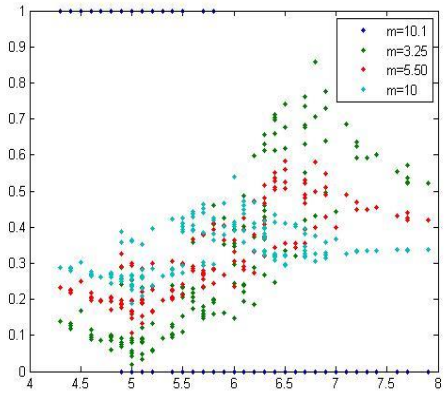
شکل ۱، مراحل استفاده و به‌کار گیری این شاخص را نشان می‌دهد. با این شاخص می‌توان مقادیر m۱ و m۲ و همچنین c را به طور یک‌جا معرفی کرد. به این صورت که با ماکزیمم‌سازی رابطه مربوط به شاخص با استفاده از شمارش حالات ممکن، مقدار بهینه هر سه مقدار m۱ و m۲ و c تعیین می‌گردد.

قبل از توضیح در مورد نحوه تولید شاخص نوع-۲ بهتر است به اثر مقدار m بر روی مقادیر عضویت و همچنین شاخص اعتبار خوشه‌بندی توجه شود. در شکل اثر مقدار m در خوشه‌بندی بر روی داده‌های $iris$ نشان داده شده است. به این منظور مجموعه داده $iris$ در ۳ خوشه، خوشه‌بندی شده و مقادیر عضویت مربوط به خوشه اول بر روی متغیر اول تصویر شده است. همان‌طور که در شکل ۲ مشخص است هر چه میزان m زیاد می‌شود مقادیر تصویر شده عضویت از صفر و یک فاصله گرفته و به مقادیر میانی نزدیک می‌شود. در حالت حدی همه مقادیر در مقدار $1/c$ همگرا خواهند شد. در مقادیر نزدیک به ۱ خوشه‌بندی تقریبا به حالت قطعی انجام می‌شود و با بزرگ شدن m خوشه‌بندی کاملا فازی خواهد بود.

[DOR: 20.1001.1.22517286.2019.15.4.2.4]



شکل ۱. مراحل پیاده سازی شاخص پیشنهادی اعتبار خوشه‌بندی نوع-۲



شکل ۲. اثر m بر روی مقادیر عضویت در مجموعه داده *iris*

برای طراحی شاخص نوع-۲ جهت سنجش اعتبار خوشه‌بندی مطابق با اکثر شاخص‌های معمول، از دو عبارت «به هم فشردگی داخلی»[1] و «جدایی خوشه‌ها از یکدیگر»[2] استفاده گردیده است. برای تعیین فرمول‌هایی برای محاسبه میزان این عبارات، سعی شده است تا به نوعی عمل شود که رفتار شاخص‌ها با تغییرات m، رفتار یکنواختی داشته باشد به نحوی که صعودی یا نزولی بودن رفتار فرمول «به هم فشردگی داخلی» و «جدایی خوشه‌ها از یکدیگر» با توجه به کاهش یا افزایش میزان m برقرار باشد و با تغییر ساختار داده و سایر متغیرها از بین نرود.

با توجه به نکات بالا، برای طراحی شاخص نوع-۲ باید به این صورت عمل شود که در ابتدا الگوریتم FCM برای مقدار m۱ اجرا می‌شود. همچنین این الگوریتم در مرحله بعد برای مقدار m۲ اجرا خواهد شد و در هر دو مرحله از یک مقدار یکسان برای c استفاده می‌شود. بدین ترتیب، مقادیر U۱ و V۱ از اجرای اول الگوریتم و مقادیر U۲ و V۲ از اجرای دوم الگوریتم به دست می‌آیند. مقدار U نشان دهنده ماتریس مقادیر عضویت و مقدار V نشان‌دهنده ماتریس مراکز خوشه‌هاست. حال از مقادیر U۲، V۲، m۲ برای محاسبه عبارت «جدایی خوشه‌ها از یکدیگر» و از مقادیر U۱، V۱، m۱ و ماتریس اطلاعات داده‌ها X برای محاسبه عبارت «به هم فشردگی داخلی» استفاده خواهد شد. هر چه عبارت «جدایی خوشه‌ها از یکدیگر» بیش‌تر باشد به معنای تفکیک مطلوب خوشه‌ها از یکدیگر بوده و مطلوب‌تر است. همچنین هر چه عبارت «به هم فشردگی داخلی» کم‌تر باشد نشان‌دهنده فشردگی بیش‌تر بوده و مطلوب‌تر است؛ لذا از تقسیم این دو شاخصی به دست

[1] compactness
[2] separation

[DOR: 20.1001.1.22517286.2019.15.4.2.4]

می‌آید که حداکثر بودن آن به بیش‌ترین مقدار مطلوبیت منجر خواهد شد؛ لذا مقادیر $m۱$، $m۲$ و c از این رابطه به دست می‌آیند: $\max_{\substack{۲\le c\le n-۱\\ m۱\\ m۲}} V_{T۲}$. روابط مربوط به این شاخص در زیر آمده است:

$$V_{T۲}=\frac{Separation\left(U۲,V۲,m۲\right)}{Compactness\left(X,U۱,V۱,m۱\right)} \quad (۱)$$

$$Separation\left(U۲,V۲,m۲\right)=\sum_{i=۱}^{n}\sum_{j=۱}^{c}\sum_{k=۱}^{p}U۲_{ij}^{m۲}\left(V۲_{ik}-\overline{V۲}\right)^{۲} \quad (۲)$$

$$Compactness\left(X,U۱,V۱,m۱\right)=\sum_{j=۱}^{c}\sum_{k=۱}^{p}\frac{\sum_{i=۱}^{n}U۱_{ij}^{m۱}\left(x_{ik}-V۱_{jk}\right)^{۲}}{\sum_{i=۱}^{n}U۱_{ij}^{m۱}} \quad (۳)$$

$$V_{T۲}\left(X,U۱,U۲,V۱,V۲,m۱,m۲\right)=\frac{\sum_{i=۱}^{n}\sum_{j=۱}^{c}\sum_{k=۱}^{p}U۲_{ij}^{m۲}\left(V۲_{ik}-\overline{V۲}\right)^{۲}}{\sum_{j=۱}^{c}\sum_{k=۱}^{p}\frac{\sum_{i=۱}^{n}U۱_{ij}^{m۱}\left(x_{ik}-V۱_{jk}\right)^{۲}}{\sum_{i=۱}^{n}U۱_{ij}^{m۱}}} \quad (۴)$$

در مورد روابط فوق باید گفت استفاده از مقادیر $m۱$و $U۱$ در مخرج و $m۲$ و $U۲$ در صورت کسر به این علت است که رفتار شاخص نهایی در فاصله زیاد بین مقادیر $m۱$و $m۲$ دچار رفتار غیر واقعی نشود. در روابط فوق i شمارنده تعداد نقاط داده است. j شمارنده تعداد خوشه‌هاست و k شمارنده خصوصیات[1] مربوط به داده‌ها و مرکز خوشه‌هاست. همچنین در این روابط n تعداد نقاط داده‌ها، c نشان‌دهنده تعداد خوشه‌ها و p نشان دهنده تعداد خصوصیات (ابعاد داده‌ها) مربوط به داده‌هاست. عبارت $\overline{V}$ در روابط فوق نشان‌دهنده میانگین مرکز خوشه‌هاست. همچنین $m۱$ نشان‌دهنده حد پایین فازی‌ساز و $m۲$ نشان‌دهنده حد بالای فازی‌ساز است.

در جدول ۶، شبه کدی از نحوه پیاده‌سازی این شاخص آورده شده است. حداکثر تابع $\boldsymbol{V_{T۲}}$ نشان‌دهنده $m۱$، $m۲$ و c بهینه است. برای پیاده‌سازی الگوریتم نیاز به دو مقدار α و β خواهد بود که در شبه کد از آن‌ها استفاده شده است. مقدار α برابر با حداقل فاصله بین $m۱$ و $m۲$ است و مقدار β برابر با حداکثر مقدار تفاوت بین $m۱$ و $m۲$ خواهد بود. در اینجا α و β به ترتیب برابر ۰/۴ و ۲ قرار داده شده‌است.

جدول ٦. شبه کد شاخص پیشنهادی

```
for i = ۱ : √n                          % i for number of clusters
    for j = ۱ : step : ۱۰               % j for m۱
        for k = (j + α) : step : (j + β)    %k for m۲
            run fcm(X, i, j)
            run fcm(X, i, k)
            evaluate V_T۲(X, U۱, U۲, V۱, V۲, m۱, m۲)
```

[1] feature

[DOR: 20.1001.1.22517286.2019.15.4.2.4]

```
        end
    end
end
find m1, m2, c when V_T2 is max
```

٤ نتایج تجربی

در جدول ۷، نتایج حاصل از پیاده‌سازی شاخص معرفی شده بر روی ۴ مجموعه داده توضیح داده شده در جدول ۱، نمایش داده شده است. ستون دوم این جدول مقدار بهینه‌ای است که برای تعداد خوشه‌ها به دست آمده است. همچنین بازه مورد نظر برای m و میزان شاخص در این نقطه نشان داده شده است. تا به حال شاخصی برای IT2 FCM معرفی نشده است و شاخص دیگری نیز وجود ندارد که بتوان این شاخص را با آن مقایسه نمود؛ اما می‌توان بیان کرد که اشکالات اشاره شده در استفاده از شاخص‌های معمول در الگوریتم IT2 FCM، در اینجا به‌علت به دست آوردن بازه بهینه، وجود ندارد.

جدول ۷. نتایج حاصل از شاخص نوع-۲ پیشنهادی

مجموعه داده	تعداد خوشه	m۱	m۲	تابع هدف
Seed	۳	۱/۶	۲/۴	۱۰۲۵۱۸
example	۳	۱/۶	۲/۷	۲۹۷۳۲
Iris	۴	۱/۲	۲/۲	۹۰۵۴۲
ruspini	۳	۱/۶	۳/۲	۴۶۳۹۴

۵ نتیجه‌گیری

در این تحقیق تلاش شد تا شاخصی برای سنجش اعتبار خوشه‌بندی نوع-۲ معرفی شود. به دلیل عدم قطعیتی که در پارامتر فازی سازی در الگوریتم مورد نظر وجود دارد، از دو مقدار فازی‌ساز (m) برای خوشه‌بندی در آن استفاده می‌شود. این امر منجر به تولید توابع فازی نوع-۲ به عنوان عضویت هر عنصر داده در هر خوشه می‌شود و چنین شرایطی منجر به الگوریتمی می‌شود که الگوریتم IT2 FCM نام دارد. تا کنون شاخصی برای سنجش اعتبار خوشه‌بندی در چنین الگوریتمی ارایه نشده است و به هنگام استفاده از این الگوریتم از شاخص‌های معمول جهت تعیین تعداد خوشه‌ها استفاده می‌شود و مقادیر m نیز به طور ثابت و عمومی در نظر گرفته می‌شود. در اینجا شاخصی را معرفی نمودیم که مقادیر m۱ و m۲ و c را به صورت یکجا تعیین نماید. این کار از طریق حل یک مساله ماکزیمم‌سازی بر روی شاخص پیشنهاد شده و شمارش مقادیر مختلف متغیرهای تصمیم مساله انجام می‌شود. شاخص پیشنهادی و شاخص‌های معمولی که در این الگوریتم‌ها استفاده می‌شود بر روی ۴ مجموعه داده پیاده‌سازی شده و نتایج نشان داده شد. استفاده از شاخص معرفی شده می‌تواند اثر چشمگیری در کنترلرهای نوع-۲ (سیستم‌های منطق فازی نوع-۲) داشته باشد و منجر به بهبود نتایج پیش‌بینی و کنترل در این سیستم‌ها گردد.

[DOR: 20.1001.1.22517286.2019.15.4.2.4]

منابع

[1] Bezdek, J.C.,(1981).Objective function clustering. In Pattern recognition with fuzzy objective function algorithms, 43-93, Springer, Boston, MA.

[2] Bezdek, J.C.,(1973). Cluster validity withfuzzy sets. 58-73.

[3] Windham, M.P., (1981). Cluster validity for fuzzy clustering algorithms. Fuzzy Sets and Systems, 5(2), 177-185.

[4] Kim, Y.I., et al., (2004). A cluster validation index for GK cluster analysis based on relative degree of sharing. Information Sciences, 168(1), 225-242.

[5] Chen, M.Y. and. Linkens, D.A., (2004). Rule-base self-generation and simplification for data-driven fuzzy models. Fuzzy Sets and System,142(2), 243-265.

[6] Fukuyama, Y. and Sugeno, M., (1989). A new method of choosing the number of clusters for the fuzzy c-means method. in Proc. 5th Fuzzy Syst. Symp.

[7] Xie, X.L. and Beni, G., (1991). A validity measure for fuzzy clustering. Pattern Analysis and Machine Intelligence. IEEE Transactions,13(8), 841-847.

[8] Pal, N.R. and Bezdek, J.C., (1995). On cluster validity for the fuzzy c-means model. Fuzzy Systems. IEEE Transactions,3(3), 370-379.

[9] Kwon, S.H., (1998). Cluster validity index for fuzzy clustering. Electronics Letters, 34(22), 2176-2177.

[10] Zahid, N., Limouri, M., and Essaid, A., (1999). A new cluster-validity for fuzzy clustering. Pattern Recognition, 32(7), 1089-1097.

[11] Wu, K.L. and Yang, M.S., (2005). A cluster validity index for fuzzy clustering. Pattern Recognition Letters, 26(9), 1275-1291.

[12] Kim, D.W., Lee, K.H., and Lee, D., (2004). On cluster validity index for estimation of the optimalnumber of fuzzy clusters. Pattern Recognition, 37(10), 2009-2025.

[13] Pakhira, M.K., Bandyopadhyay, S., and Maulik, U., (2004). Validity index for crisp and fuzzy clusters. Pattern Recognition, 37(3), 487-501.

[14] Barash, Y. and Friedman, N., (2002). Context-specific Bayesian clustering forgene expression data. Journal of Computational Biology, 9(2), 169-191.

[15] Cho, S.B. and Yoo, S.H., (2006). Fuzzy Bayesian validation for cluster analysis of yeast cell-cycle data. Pattern Recognition, 39(12), 2405-2414.

[16] Chong, A., Gedeon, T., and Koczy, L., (2002). A hybrid approach for solving the cluster validity problem. in Digital Signal Processing, DSP 2002. 2002 14th International.

[17] Wu, C. H., Ouyang, C. S., Chen, L. W., & Lu, L. W., (2015). A new fuzzy clustering validity index with a median factor for centroid-based clustering. IEEE Transactions on Fuzzy Systems, 23(3), 701-718.

[18] Melin, P., and Castillo, O., (2014). A review on type-2 fuzzy logic applications in clustering, classification and pattern recognition. Applied soft computing, 21, 568-577.

[19] Zarandi, M., Faraji, M., and Karbasian, M., (2012). Interval type-2 fuzzy expert system for prediction of carbon monoxideconcentration in mega-cities. Applied Soft Computing, 12(1), 291-301.

[20] Sanchez, M. A., Castillo, O., and Castro, J. R., (2015). Information granule formation via the concept of uncertainty-based information with Interval Type-2 Fuzzy Sets representation and Takagi–Sugeno–Kang consequents optimized with Cuckoo search. Applied Soft Computing, 27, 602-609.

[21] Faraji, M. R., and Qi, X., (2016). Face recognition under varying illuminations using logarithmic fractal dimension-based complete eight local directional patterns. Neurocomputing, 199, 16-30.

[22] Wang, J., Chen, Q. H., Zhang, H. Y., Chen, X. H., and Wang, J. Q., (2017). Multi-criteria decision-making method based on type-2 fuzzy sets. Filomat, 31(2), 431-450.

[23] Huang, M., et al., (2012). The range of the value for the fuzzifier of the fuzzy c-means algorithm. Pattern Recognition Letters, 33(16), 2280-2284.

[24] Ozkan, I. and Turksen, I., (2007). Upper and lower values for the level of fuzziness in FCM. Information Sciences, 23(177), 5143-5152.

[25] Fazel Zarandi, M., Faraji, M., and Karbasian, M., (2010). An Exponential Cluster Validity Index for Fuzzy Clustering with Crisp and Fuzzy Data. Scientia Iranica, Transaction E, Industrial Engineering, 17(2), 95.

[DOR: 20.1001.1.22517286.2019.15.4.2.4]