

# خوشه بندی داده های فازی با به کارگیری الگوریتم FCM بر مبنای یک معیار اندازه فاصله پارامتری

مدینه فرنام و مجید دره میرکی\*

گروه مهندسی برق، دانشگاه شهید چمران اهواز- پردیس صنعتی شهدای هویزه، دشت آزادگان،  
ایران

گروه ریاضی و آمار، دانشگاه صنعتی خاتم الانبیاء بهبهان، بهبهان، ایران

تاریخ پذیرش: ۱۴۰۱/۷/۲۸

تاریخ دریافت: ۱۴۰۱/۴/۲۲

نوع مقاله: علمی-پژوهشی

## چکیده

الگوریتم خوشه بندی  $C$  میانگین فازی (FCM) متداول‌ترین روش خوشه‌بندی فازی است، که تاکنون شکل‌های مختلفی از آن ارائه شده است. یکی از موثرین عوامل در بهبود عملکرد الگوریتم‌های خوشه‌بندی تعیین معیار فاصله و تشابه کارآمد برای بهره‌گیری در آن است. به طور کلی اندازه فاصله بین دو عدد فازی می‌تواند به صورتی قطعی یا پارامتری بیان شود. در این بین، معیارهای پارامتری انعطاف پذیری بیشتری برای حل مسئله فراهم می‌سازند. از این رو، در این مقاله ابتدا یک معیار فاصله پارامتری جدید معرفی می‌شود. در ادامه ضمن بررسی اصول موضوعی اندازه برای معیار پیشنهادی، الگوریتم FCM را بر مبنای آن و به عنوان یک روش خوشه بندی کارا و قوی برای داده‌های فازی ارائه می‌دهیم. با توجه به این که معیار بیان شده در این مقاله بر اساس  $\alpha$ -برش‌ها (پارامتر مورد نظر) است، توانایی تصمیم‌گیری در سطوح مختلف را برای تصمیم‌گیرنده فراهم می‌سازد. در پایان دو مثال عددی و یک مثال کاربردی برای نشان دادن کارایی روش پیشنهادی ارائه شده است.

عبارات و کلمات کلیدی: خوشه بندی، عدد فازی، آلفا برش، الگوریتم FCM.

Email(s): m.farnam@scu.ac.ir, darehmiraki@bkatu.ac.ir.

۱۴۰۱ انجمن سیستم‌های فازی ایران

Mathematics Subject Classification:

## ۱ مقدمه

یکی از ابزارهای مهم در تجزیه و تحلیل داده‌ها، خوشه‌بندی است. در خوشه‌بندی، داده‌ها به رده‌هایی با بیشترین تشابه در همان خوشه و بیشترین تمایز نسبت به خوشه‌های دیگر تقسیم می‌شوند. از خوشه‌بندی سلسله مراتبی به عنوان اولین روش خوشه‌بندی یاد می‌شود که توسط بسیاری از محققان به خصوص زیست‌شناسان و دانشمندان علوم اجتماعی مورد استفاده قرار می‌گیرد. علاوه بر این، خوشه‌بندی به عنوان یکی از شاخه‌های تجزیه و تحلیل چند متغیره آماری مطرح شده است و از این رو، بسیاری از نظریه‌ها و روش‌های مختلف برای آن بررسی می‌شود. از سوی دیگر، شناخت الگوهای پنهان در داده‌ها به طور کلی از خوشه‌بندی آغاز می‌شود که به عنوان یکی از مهمترین موضوعات در یادگیری ماشین شناخته می‌شود. در واقع خوشه‌بندی به عنوان نوعی از یادگیری بدون ناظر، به شناخت الگو منتهی می‌شود [۱، ۲، ۳، ۴، ۵].

خوشه‌بندی به دو دسته کلی خوشه‌بندی سخت و نرم تقسیم می‌شود. اگر در مواردی که میزان عضویت یک عنصر به یک خوشه به وضوح تعریف نشده یا نامعلوم باشد، خوشه‌بندی اشیاء را با استفاده از روش‌های خوشه‌بندی سخت، همچون الگوریتم  $k$  میانگین انجام دهیم، نتایجی با دقت کمتر را تولید خواهیم کرد. بسیاری از الگوریتم‌های خوشه‌بندی سخت، با معیارهای ریاضی که برای شناسایی خوشه بهینه استفاده می‌شوند توسط دودا و هارت در سال ۱۹۷۳ [۶] مورد بحث قرار گرفته است. این الگوریتم‌های خوشه بندی سخت دارای محدودیت‌های ذاتی هستند که هر عضو مجبور است عضو یک خوشه باشد یا نباشد، با وجود اینکه ممکن است کاملاً مشابه اعضای آن خوشه نباشد. در تکنیک خوشه‌بندی نرم، مجموعه‌های فازی برای خوشه‌بندی داده‌ها استفاده می‌شوند، به طوری که هر نقطه می‌تواند به دو یا چند خوشه با درجه‌های عضویت مختلف تعلق داشته باشد. در این مورد، داده‌ها به یک مقدار عضویت مناسب مرتبط خواهند شد. در بسیاری از موارد، نتایج خوشه‌بندی فازی طبیعی‌تر از خوشه‌بندی سخت است. اشیاء در مرزهای بین چندین کلاس مجبور نیستند به طور کامل متعلق به یکی از کلاسها باشند، بلکه درجه اعطایی بین ۰ و ۱ را نشان می‌دهند که معرف میزان عضویت جزئی آنها است. در حالی که، در تکنیک‌های خوشه‌بندی سخت، داده‌ها به روش منحصر به فرد گروه‌بندی می‌شوند، به طوری که اگر یک داده خاص متعلق به یک خوشه قطعی باشد، آن را نمی‌توان در خوشه

دیگری قرار داد.

منطق فازی توسط لطفی زاده معرفی شد، تا از عدم اطمینان و ابهام ناشی از مسائل دنیای واقعی در مدلسازی استفاده شود. در حقیقت، منطق فازی می‌تواند به عنوان تلاش بشر برای مکانیزه کردن دو قابلیت انسانی مورد توجه قرار گیرد: توانایی تفکر و تصمیم‌گیری منطقی در محیط غیر دقیق؛ و توانایی انجام انواع مختلفی از وظایف فیزیکی و روانی بدون هیچ‌گونه اندازه‌گیری و محاسبات. از زمانی که لطفی زاده نظریه مجموعه‌های فازی را مطرح کرد، مبحث خوشه‌بندی فازی که شامل ایده درجه عضویت نسبت به خوشه‌ها است نیز مطرح و بسیار با موفقیت در تجزیه و تحلیل خوشه‌ها استفاده شد. تا به امروز، خوشه‌بندی فازی به طور گسترده مورد مطالعه و کاربرد قرار گرفته است. در بیشتر نسخه‌های فازی از الگوریتم‌های خوشه‌بندی، عضویت اختصاص داده شده به یک خوشه فازی است، اما خود داده‌ها و مراکز فازی نیستند. تاکنون در این زمینه، روش‌های خوشه‌بندی گوناگونی برای داده‌هایی متنوع، ارائه شده است. به عنوان مثال برای منعکس کردن واقعیت کلی روش‌های خوشه‌بندی فازی، روش‌های مبتنی بر افراز، خوشه‌بندی سلسله‌مراتبی، خوشه‌بندی مبتنی بر تراکم، خوشه‌بندی بردار پشتیبان و خوشه‌بندی طیفی [۷، ۸، ۹، ۱۰، ۱۱]، پیشنهاد شده‌اند.

با استفاده از منطق فازی در خوشه‌بندی، مشاهدات می‌توانند به بیش از یک خوشه متعلق باشند. این ترکیب، انعطاف‌پذیری بیشتری را برای رسیدگی به عدم قطعیت در مشاهدات دنیای واقعی ارائه می‌دهد و منجر به ابداع بسیاری از روش‌های خوشه‌بندی فازی شده است. به عنوان نمونه و در راس آنها FCM؛ معرف توسعه‌ی نرم از الگوریتم  $K$  میانگین [۱۲] است. بعدها، مدل رگرسیون فازی (FCRM)؛ برای غلبه بر مشکل ناهمگنی مشاهدات فازی و بر مبنای روش‌های خوشه‌بندی فازی ارائه [۱۳] گردید. در سال ۲۰۱۱، روش امکان‌سنجی  $c$  (PCM)؛ به عنوان توسعه‌ی برای FCM معرفی شد، اما به علت نداشتن خاصیت خوددوگانگی دارای مشکل خوشه‌های منطبق بود. [۱۴] لازم به ذکر است که روش‌های مبتنی بر هسته؛ برای داده‌های غیرخطی [۱۵] و الگوریتم پوسته‌های کروی فازی (FCSS)؛ برای داده‌هایی که الزاماً نظم خاصی ندارند [۱۶] خوشه‌بندی‌های مناسبی ایجاد می‌کنند. روش حداکثر آنتروپی خوشه‌بندی [۱۷]؛ که خطای ناشی از اطلاعات ناکامل و ناکافی را کاهش می‌دهد در سال ۲۰۱۴ توسط لی و

لی معرفی شد. خوشه بندی مبتنی بر دانش [۱۸]؛ به عنوان الگوریتمی مستقل از نظم می تواند به طور طبیعی به اشیاء خوشه ای به صورت افزایشی گسترش یابد توسط سریدر و مورتی بیان شد. این دسته از روش های خوشه بندی فازی و بسیاری از روش های دیگر در زمینه های مختلف از جمله پزشکی [۱۹، ۲۰]، پردازش تصویر [۲۱]، تشخیص گفتار [۲۲] و غیره کاربرد دارند.

اخیرا بورلیا و همکاران [۲۳] الگوریتم خوشه بندی فرم واحد را به عنوان یک روش جدید خوشه بندی فازی پیشنهاد کردند که در آن الگوریتم های فازی FCM و  $K$ -Means به عنوان یک الگوریتم منفرد قابل تنظیم در نظر گرفته شده است. سورونو و همکاران [۲۴] از یک روش جدید استفاده به منظور خوشه بندی داده های فازی استفاده کردند که در آن ترکیب فاصله مینکوفسکی با فاصله چیشف را به عنوان معیار تشابه جدید در فرآیند خوشه بندی در FCM استفاده می کند. کوشال و همکاران در [۲۵] یک الگوریتم مبتنی بر فازی شهودی، یعنی الگوریتم خوشه بندی فازی شهودی تعمیم یافته را پیشنهاد کردند که از مجموعه های فازی شهودی آتاناسوف تطبیقی اندازه گیری فاصله اقلیدسی در تابع معیار خود برای خوشه بندی مجموعه داده ها در محیط فازی شهودی استفاده می کند. روش فازی سازی شهودی پیشنهادی شامل تکنیکی برای تبدیل مجموعه داده ها به مجموعه های فازی شهودی آتاناسوف است و ساختار اصلی خود را حفظ می کند که تمایل به تغییر در طول هر فرآیند فازی سازی دارد. الگوریتم FCM دارای اشکالاتی مانند انتخاب تعداد خوشه ها و مقداردهی اولیه خوشه است. عبدالهوم و همکاران [۲۶] رویکردهای جدیدی را برای مقابله با این دو اشکال پیشنهاد کردند. برای مسئله اول دو رویکرد پیشنهاد نمودند. رویکرد پیشنهادی اول از شبکه های عصبی و شاخص Xie و Beni بهره برداری می کند، در حالی که روش دوم از هیستوگرام بهره برداری می کند. در مورد مشکل دوم نیز یک رویکرد ترکیبی فراابتکاری جدید با استفاده از الگوریتم ژنتیک، الگوریتم مبتنی بر جغرافیای زیستی، و الگوریتم کرم شب تاب را پیشنهاد نمودند.

با توجه به کاربردهای گسترده الگوریتم های خوشه بندی فازی، در این پژوهش قصد داریم با معرفی یک اندازه فاصله پارامتری بهبودی در اجرای الگوریتم FCM داشته باشیم. معیار جدید با استفاده از  $\lambda$  برش ها فاصله بین دو عدد فازی محاسبه می شود که پس از آن برای محاسبه فاصله بین داده ها و مراکز خوشه ها در روش FCM به کار گرفته خواهد شد. از

ویژگی‌های برجسته فاصله پیشنهادی می‌توان به این مطلب اشاره نمود که اولاً نظر تصمیم‌گیرنده، با انتخاب مقدار  $\lambda$  نزدیک به ۱ (یک تصمیم‌گیری سطح بالا) و با انتخاب مقدار  $\lambda$  نزدیک به صفر (یک تصمیم‌گیری سطح پایین)، را هنگام محاسبه فاصله بین دو عدد فازی منظور می‌کند. ثانیاً فرمول فاصله پیشنهادی قوی است به این منظور که با ورود یک داده پرت به مجموعه داده‌ها جواب بهینه مساله زیاد تغییر نمی‌کند. برای این منظور سایر بخش‌های این مقاله به صورت زیر سازماندهی می‌شود:

در بخش ۲: مروری بر مفاهیم مقدماتی و روش خوشه بندی FCM آورده شده است.

در بخش ۳: با معرفی یک معیار متریک جدید، الگوریتم خوشه بندی پیشنهاد خود را ارائه می‌دهیم.

در بخش ۴: ضمن بررسی ۳ مثال، نتایج حاصل از اجرای الگوریتم پیشنهادی در بخش ۳ را بیان می‌کنیم.

و در بخش ۵: یافته‌های حاصل از پژوهش به همراه پیشنهادات تحقیقات آینده آورده شده است.

## ۲ مروری بر مفاهیم مقدماتی و روش خوشه بندی FCM

در این قسمت به منظور معرفی الگوریتم پیشنهادی برخی از مفاهیم مقدماتی از اعداد فازی به همراه مقدمه‌ای از روش FCM به اختصار آورده می‌شود.

### ۱.۲ مروری بر مفاهیم مقدماتی

تعاریف فازی مورد نیاز برای این قسمت از منابع [۲۷، ۲۸] آورده شده است.

تعریف ۱.۲. مجموعه‌ی فازی  $\tilde{A}$  روی مجموعه‌ی مرجع (جهانی)  $X$ ، اغلب با یک مجموعه از زوج مرتب‌ها به صورت زیر نمایش داده می‌شود:

$$\tilde{A} = \{(x, \mu_{\tilde{A}}(x)) | x \in X\},$$

$$(\mu_{\tilde{A}}(x) : X \rightarrow [0, 1]),$$

که در آن  $\mu_{\bar{A}}(x)$  به عنوان تابع عضویت، درجه عضویت (میزان تعلق) عنصر  $x$  را با عددی در بازه  $[0, 1]$  نشان می‌دهد.

**تعریف ۲.۲.** یک عدد فازی مجموعه‌ای فازی مانند  $A : R \rightarrow I = [0, 1]$  است که روابط زیر را برقرار می‌سازد:

(۱)  $A$  نیم پیوسته بالایی است،

(۲)  $A(x) = 0$  برای نقاط خارج بازه  $[a, d]$ ،

(۳) اعداد حقیقی  $b, c$  وجود دارند که  $a \leq b \leq c \leq d$  و

-  $A(x)$  روی بازه  $[a, b]$  یکنوای صعودی است.

-  $A(x)$  روی بازه  $[c, d]$  یکنوای نزولی است.

-  $A(x) = 1$  برای  $b \leq x \leq c$ .

تابع عضویت  $A$  را می‌توان به صورت زیر بیان کرد:

$$A(x) = \begin{cases} A_L(x), & a \leq x \leq b \\ 1, & b \leq x \leq c \\ A_R(x), & c \leq x \leq d \\ 0, & e.w \end{cases}$$

که در آن  $A_L : [a, b] \rightarrow [0, 1]$  و  $A_R : [c, d] \rightarrow [0, 1]$  توابع عضویت چپ و راست از عدد فازی  $A$  هستند.

**تعریف ۳.۲.** در فرم پارامتری عددی فازی به صورت زوج مرتب  $(\underline{A}(r), \bar{A}(r))$  از توابع  $\underline{A}(r)$  و  $\bar{A}(r)$  که در آن  $0 \leq r \leq \omega$ ، تشکیل می‌شود که گزاره‌های زیر را برقرار می‌سازند:

(۱)  $\underline{A}(r)$  تابع صعودی یکنوای کراندار و از سمت چپ پیوسته روی بازه  $[0, \omega]$  است.

(۲)  $\bar{A}(r)$  تابع نزولی یکنوای کراندار و از سمت راست پیوسته روی بازه  $[0, \omega]$  است.

(۳)  $\underline{A}(r) \leq \bar{A}(r)$  برای  $0 \leq r \leq \omega$ . مقدار ثابت دلخواهی و  $0 < \omega \leq 1$  دارد.

یک عدد قطعی (غیرفازی) مانند "k" به سادگی با  $\underline{A}(r) = \bar{A}(r) = k$  که در آن  $0 \leq r \leq \omega$  می‌باشد نشان داده می‌شود.

اگر در تعریف بالا  $\omega = 1$  فرض شود. عدد فازی تعریف شده، عدد فازی نرمال می‌باشد.

در اینجا  $\tilde{A}_\omega$  نشان‌دهنده عددی فازی است که در آن "w" ماکسیم درجه عضویت روی آن می‌باشد. هرگاه منظور عدد فازی نرمال باشد برای سادگی نمایش آن را با  $\tilde{A}$  نشان می‌دهیم.

تعریف ۴.۲. زیر مجموعه‌ای از عناصر مجموعه‌ی مرجع (ثابت)  $X$ ، که درجه عضویت آنها روی مجموعه فازی  $\tilde{A}$ ، حداقل به اندازه  $\alpha$  باشد را  $-\alpha$  برش نامیم که نمایش آن به صورت مجموعه‌ی قطعی زیر می‌باشد:

$$\tilde{A}_\alpha = \{x \in X | \mu_{\tilde{A}}(x) \geq \alpha\},$$

در حالتی که نامساوی به صورت اکید برقرار باشد،  $-\alpha$  برش قوی داریم که با  $\tilde{A}_{\alpha+}$  نمایش داده می‌شود.

نمایش  $-\alpha$  برش برای عدد فازی  $\tilde{A}$ ، بازه  $[\tilde{A}_\alpha^L, \tilde{A}_\alpha^R]$  است.

تعریف ۵.۲. یک نوع پرکاربرد و معروف از اعداد فازی به نام عدد فازی دوزنقه‌ای تعمیم یافته است که به صورت زیر نمایش داده می‌شود:

$$\tilde{A}_\omega^{Tra} = (a, b, c, d; \omega),$$

و تابع عضویت آن عبارت است از:

$$\mu_{\tilde{A}}(x) = \begin{cases} \frac{(x-a)}{(b-a)}\omega, & a \leq x \leq b \\ \omega, & b \leq x \leq c \\ \frac{(d-x)}{(d-c)}\omega, & c \leq x \leq d \\ 0, & o.w, \end{cases}$$

که در آن  $\omega \in [0, 1]$  و  $a, b, c, d \in \mathbb{R}$ .

اگر در رابطه مربوط به تابع عضویت، دو عدد میانی برابر باشند، عدد فازی دوزنقه‌ای تعمیم یافته به عدد فازی مثلثی تعمیم یافته  $\tilde{A}_\omega^{Tri} = (a, b, c; \omega)$  تبدیل می‌شود.

تعریف ۶.۲. اگر  $\mathcal{F}$  مجموعه‌ای از اعداد فازی باشد، آنگاه  $d_{\mathcal{F}}$  یک متر روی  $\mathcal{F}$  تعریف می‌کند هرگاه خواص زیر را برقرار سازد:

$$(1) \quad d_{\mathcal{F}}(X, X) = 0, \quad X \in \mathcal{F}$$

$$(2) \quad d_{\mathcal{F}}(X, Y) \geq 0, \quad X, Y \in \mathcal{F}$$

(3) برای هر دو عدد فازی دلخواه  $\tilde{u}_\omega$  و  $\tilde{v}_\omega \in \mathcal{F}$  داریم:

$$d_{\mathcal{F}}^\alpha(\tilde{u}_\omega, \tilde{v}_\omega) = d_{\mathcal{F}}^\alpha(\tilde{v}_\omega, \tilde{u}_\omega), \quad 0 \leq \alpha \leq 1.$$

(4) برای هر سه عدد فازی دلخواه  $\tilde{u}_\omega, \tilde{v}_\omega, \tilde{w}_\omega \in \mathcal{F}$  داریم:

$$d_{\mathcal{F}}^\alpha(\tilde{u}_\omega, \tilde{w}_\omega) \leq d_{\mathcal{F}}^\alpha(\tilde{u}_\omega, \tilde{v}_\omega) + d_{\mathcal{F}}^\alpha(\tilde{v}_\omega, \tilde{w}_\omega).$$

## ۲.۲ روش خوشه بندی FCM

در خوشه بندی فازی، وابستگی هر یک از اشیاء به خوشه‌ها به صورت یک عدد بین صفر و یک بیان می‌شود. این نوع خوشه‌بندی برای توصیف ماهیت اعضای خوشه‌ها مناسب‌تر



به نظر می‌رسند. تکنیک‌های بسیاری برای بهبود الگوریتم‌های خوشه‌بندی فازی پیشنهاد شده‌اند. در میان همه روش‌های خوشه‌بندی فازی، FCM یکی از روش‌هایی است که به طور گسترده در بسیاری از برنامه‌های کاربردی مورد استفاده قرار گرفته است. الگوریتم خوشه‌بندی FCM با به حداقل رساندن تابع هدف FCM با فرمولی که توسط دان [۲۹] و بزدک [۳۰] معرفی شده است به کار می‌رود.

فرض کنید  $\mathbb{R}$  فضای اقلیدسی یک بعدی با نرم  $\|\cdot\|$  باشد. یک عدد فازی یک تابع نیمه پیوسته بالایی و محدب  $[\cdot, 1]$   $\mu: \mathbb{R} \rightarrow [0, 1]$  می‌باشد. علاوه بر این، فرض کنید  $\mathcal{F}$  مجموعه همه اعداد فازی باشد به طوری که  $\alpha$  برش‌ها برای  $0 \leq \alpha \leq 1$  غیرتهی و فشرده هستند. فضای  $\mathcal{F}$  با متر  $d_{\mathcal{F}}$  (مترهای گوناگونی تعریف شده است که فضای  $\mathcal{F}$  با آن مترها یک فضای متریک می‌شود، برای اطلاعات بیشتر می‌توان به [۳۱] رجوع کرد.) را در نظر می‌گیریم. اکنون فرض کنید  $\mathcal{G}$  یک زیر مجموعه با متر مورد نظر و  $c$  یک عدد صحیح بزرگتر از یک باشد. یک افراز از  $\mathcal{G}$  به  $c$  بخش می‌تواند به وسیله مجموعه‌های دوبه‌دو مجزای  $g_i, i = 1, 2, \dots, c$  نمایش یابد، به طوری که  $g = \bigcup_{i=1}^c g_i$  باشد. یا به طور معادل، برای همه  $x \in g$  و برای همه  $i = 1, 2, \dots, c$  به وسیله توابع مشخصه  $\mu_1, \dots, \mu_c$  نمایش یابد، به طوری که  $\mu_i(x) = 1$  اگر  $x \in g_i$  و  $\mu_i(x) = 0$  اگر  $x \notin g_i$ . این نوع خوشه‌بندی در واقع همان خوشه‌بندی سخت است. حال توسیعی از آن را که همان خوشه‌بندی نرم است در نظر می‌گیریم به این صورت که اجازه می‌دهیم توابع عضویت  $\mu_i$ ها، برای همه  $x$ ها در  $\mathcal{G}$  علاوه بر صفر و یک مقادیر بین آنها را نیز انتخاب کنند، به شرطی که  $\sum_{i=1}^c \mu_i(x) = 1$  باشد.

فرض کنید  $\mathcal{G} = \{X_1, \dots, X_n\}$  یک مجموعه از اعداد فازی باشد. در این قسمت ما علاقه‌مندیم که  $\mathcal{G}$  را به  $c$  خوشه تقسیم کنیم. بر اساس ایده روش‌های خوشه‌بندی FCM، هدف کمینه کردن تابع زیر می‌باشد:

$$F_m(\mu, V) = \sum_{j=1}^n \sum_{i=1}^c \mu_{ij}^m d_{\mathcal{F}}^m(X_j, V_i), \quad (1)$$

که  $m > 1$  یک شاخص وزن است که فازی پذیری خوشه‌بندی بدست آمده را کنترل می‌کند و  $\mu_{ij}$  درجه عضویت یا میزان تعلق مربوط به  $i$  امین نمونه در  $j$  امین خوشه

را تعیین و مشخص می‌کند. برای این تابع هدف، محدودیت‌هایی به این صورت  $\mu_{ij} \in [0, 1]$  و  $0 < \sum_{i=1}^n \mu_{ij} < n$  و  $\sum_{j=1}^c \mu_{ij} = 1$  وجود دارند. با استفاده از روش ضرایب لاگرانژ به جواب های زیر برای مساله بهینه‌سازی ذکر شده می‌رسیم که در هر مرحله از روش FCM باید بروزرسانی شوند:

$$V_i = \frac{\sum_{j=1}^n \mu_{ij}^m X_j}{\sum_{j=1}^n \mu_{ij}^m}, \quad (2)$$

$$\mu_{ij} = \left( \sum_{k=1}^c \left( \frac{d_{\mathcal{F}}^2(X_j, V_i)}{d_{\mathcal{F}}^2(X_j, V_k)} \right)^{\frac{1}{m-1}} \right)^{-1}, \quad (3)$$

در هر تکرار امکان تغییر میزان عضویت و مراکز خوشه‌ها وجود دارد و الگوریتم می‌تواند زمانی که این مقادیر به یک آستانه از پیش تعریف شده رسیدند، متوقف شود. در پایان اجرای الگوریتم یک خوشه‌بندی نرم از فضای ورودی به دست می‌آید.

### ۳ الگوریتم خوشه‌بندی پیشنهادی

در بخش قبل روش خوشه‌بندی فازی FCM را به صورتی مختصر معرفی کردیم. در این بخش با معرفی یک متر برای فضای اعداد فازی یک حل کارا برای مساله خوشه‌بندی اعداد فازی مبتنی بر روش FCM معرفی می‌کنیم. به این منظور ابتدا تعریف اعداد فازی پارامتری را یاد آوری می‌کنیم. یک عدد فازی  $\tilde{u}_\omega$  در شکل پارامتری عبارت است از یک جفت توابع  $(\underline{u}, \bar{u})$  که در شرایط زیر صدق می‌کنند:

$$(1) \quad \underline{u} \text{ یک تابع کراندار از چپ پیوسته و به طور یکنواخت صعودی است.}$$

$$(2) \quad \bar{u} \text{ یک تابع کراندار از چپ پیوسته و به طور یکنواخت نزولی است.}$$

$$(3) \quad \underline{u} \leq \bar{u} \text{ و } 0 \leq r \leq \omega$$

که  $\omega$  یک ثابت دلخواه بین صفر و یک است. عدد قطعی  $k$  به صورت  $0 \leq r \leq \omega$  و  $\underline{u}(r) = \bar{u}(r) = k$  نمایش داده می‌شود. اگر  $u$  یک عدد فازی دلخواه باشد، آنگاه  $-\alpha$

برش‌های آن به صورت  $0 \leq \alpha \leq \omega$  و  $[u]_\alpha = [\underline{u}(\alpha), \bar{u}(\alpha)]$  خواهند بود. اگر  $\omega = 1$ ، عدد فازی تعریف شده در بالا یک عدد فازی معمولی نامیده می‌شود. در این جا به طور کلی  $\omega$  نشانگر بیشترین درجه عضویتی است که یک عدد فازی می‌تواند اختیار کند. به عنوان نمونه برای یک عدد فازی ذوزنقه‌ای  $\tilde{u}_\omega = (a, b, c, d; \omega_{\tilde{u}})$  می‌توان نمایش متناظر زیر را بر مبنای تعریف فوق ارائه نمود:

$$\tilde{u}_\omega = (\underline{u}(r), \bar{u}(r)) = \left( a + \frac{r}{\omega}(b - a), d + \frac{r}{\omega}(c - d) \right),$$

که در آن  $0 \leq \alpha \leq r \leq \omega$

به منظور استفاده از روش FCM در فضای اعداد فازی  $\mathcal{F}$ ، یک متر جدید را برای این فضا معرفی می‌کنیم. فرض کنیم که  $(0 \leq r \leq \omega)$  و  $\tilde{u}_\omega = (\underline{u}(r), \bar{u}(r))$  و  $(0 \leq r \leq \omega)$  و  $\tilde{v}_\omega = (\underline{v}(r), \bar{v}(r))$  دو عدد فازی دلخواه در شکل پارامتری باشند. در این صورت نرم پیشنهادی به صورت زیر است:

$$d_{\mathcal{F}}^{\alpha}(\tilde{u}_\omega, \tilde{v}_\omega) = \int_{\alpha}^{\omega} (|\underline{u}(r) - \underline{v}(r)|^2 + |\bar{u}(r) - \bar{v}(r)|^2) dr, \quad 0 \leq \alpha \leq 1 \quad (4)$$

اگر بیشترین درجه عضویت ممکن ( $\omega$ ) برای دو عدد فازی یکسان نباشد از بین آنها، مقدار کوچکتر را در کران بالای انتگرال قرار می‌دهیم. به عنوان نمونه، دو عدد فازی ذوزنقه‌ای نمایش داده شده در شکل ۱ را در نظر بگیرید. نمایش این دو عدد فازی در شکل پارامتری به صورت زیر است:

$$\tilde{u}_{\omega_1} = (\underline{u}(r), \bar{u}(r)) = \left( a_1 + \frac{r}{\omega_1}(b_1 - a_1), d_1 + \frac{r}{\omega_1}(c_1 - d_1) \right),$$

و

$$\tilde{v}_{\omega_2} = (\underline{v}(r), \bar{v}(r)) = \left( a_2 + \frac{r}{\omega_2}(b_2 - a_2), d_2 + \frac{r}{\omega_2}(c_2 - d_2) \right),$$

فاصله این دو با استفاده از فرمول داده شده در (۴) به صورت زیر محاسبه می‌شود:

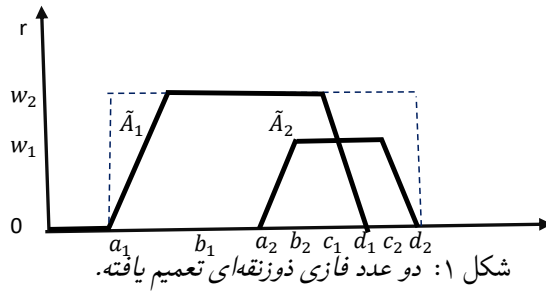
$$\begin{aligned}
 d_{\mathcal{F}}^2(A, B) &= \int_{\alpha}^{\omega_2} (|\underline{u}(r) - \underline{v}(r)|^2 + |\bar{u}(r) - \bar{v}(r)|^2) dr \\
 &= \int_{\alpha}^{\omega_2} \left| \left( a_1 + \frac{r}{\omega_1} (b_1 - a_1) \right) - \left( a_2 + \frac{r}{\omega_2} (b_2 - a_2) \right) \right|^2 \\
 &\quad + \left| \left( d_1 + \frac{r}{\omega_1} (c_1 - d_1) \right) - \left( d_2 + \frac{r}{\omega_2} (c_2 - d_2) \right) \right|^2 dr \\
 &= \int_{\alpha}^{\omega_2} \left[ ((a_1 - a_2)^2 + \frac{r^2}{\omega_1^2} (b_1 - a_1 - b_2 + a_2)^2 + 2(a_1 - a_2) \frac{r}{\omega_2} \right. \\
 &\quad (b_1 - a_1 - b_2 + a_2)) + ((d_1 - d_2)^2 + \frac{r^2}{\omega_1^2} (c_1 - d_1 - c_2 + d_2)^2 \\
 &\quad \left. + 2(d_1 - d_2) \frac{r}{\omega_2} (c_1 - d_1 - c_2 + d_2)) \right] dr \\
 &= ((a_1 - a_2)^2 (\omega_2 - \alpha) + \frac{\omega_2^3 - \alpha^3}{3\omega_1^2} (b_1 - a_1 - b_2 + a_2)^2 \\
 &\quad + 2(a_1 - a_2) \frac{\omega_2^2 - \alpha^2}{\omega_2} (b_1 - a_1 - b_2 + a_2)) + ((d_1 - d_2)^2 \\
 &\quad (\omega_2 - \alpha) + \frac{\omega_2^3 - \alpha^3}{3\omega_1^2} (c_1 - d_1 - c_2 + d_2)^2 \\
 &\quad + 2(d_1 - d_2) \frac{\omega_2^2 - \alpha^2}{\omega_2} (c_1 - d_1 - c_2 + d_2))
 \end{aligned}$$

البته لازم به ذکر است که در بسیاری از مسائل واقعی اعداد فازی نرمال هستند یعنی مقادیر  $\omega_1 = \omega_2 = 1$  و همچنین مقدار  $\alpha$  نیز می‌تواند صفر اختیار شود. با تعیین مقدار  $\alpha$  می‌توان سطح تصمیم‌گیری را مشخص کرد. مثلا مقادیر  $\alpha$  نزدیک یک نشان‌دهنده یک تصمیم‌گیری سطح بالا و مقادیر  $\alpha$  نزدیک صفر نشان‌دهنده یک تصمیم‌گیری سطح پایین است، که البته این از ویژگی‌های کارآمد روش ارائه شده در این مقاله می‌باشد. در اینجا  $\omega_2 = \min\{\omega_1, \omega_2\}$  در نظر گرفته شده است.

قضیه ۱.۳.  $(\mathcal{F}, d_{\mathcal{F}})$  یک فضای متریک است.

اثبات. شرایط لازم و کافی برای متریک بودن یک فضا را بررسی می‌کنیم.

(۱) واضح است که  $d_{\mathcal{F}}(X, X) = 0$  برای هر  $X \in \mathcal{F}$ .



(۲) چون مقدار داخل انتگرال همواره بزرگتر از صفر است لذا  $d_{\mathcal{F}}(X, Y) \geq 0$  برای هر  $X, Y \in \mathcal{F}$ .

(۳) برای هر دو عدد فازی دلخواه  $\tilde{u}_\omega, \tilde{v}_\omega \in \mathcal{F}$  داریم:

$$\begin{aligned} d_{\mathcal{F}}^{\alpha}(\tilde{u}_\omega, \tilde{v}_\omega) &= \int_{\infty}^{\omega} (|\underline{u}(r) - \underline{v}(r)|^{\alpha} + |\bar{u}(r) - \bar{v}(r)|^{\alpha}) dr \\ &= \int_{\infty}^{\omega} (|\underline{v}(r) - \underline{u}(r)|^{\alpha} + |\bar{v}(r) - \bar{u}(r)|^{\alpha}) dr \\ &= d_{\mathcal{F}}^{\alpha}(\tilde{v}_\omega, \tilde{u}_\omega) \quad 0 \leq \alpha \leq 1. \end{aligned}$$

(۴) برای هر سه عدد فازی دلخواه  $\tilde{u}_\omega, \tilde{v}_\omega, \tilde{w}_\omega \in \mathcal{F}$  داریم:

$$\begin{aligned} d_{\mathcal{F}}^{\alpha}(\tilde{u}_\omega, \tilde{w}_\omega) &= \int_{\infty}^{\omega} (|\underline{u}(r) - \underline{w}(r)|^{\alpha} + |\bar{u}(r) - \bar{w}(r)|^{\alpha}) dr \\ &= \int_{\infty}^{\omega} (|\underline{u}(r) - \underline{v}(r) + \underline{v}(r) - \underline{w}(r)|^{\alpha} \\ &\quad + |\bar{u}(r) - \bar{v}(r) + \bar{v}(r) - \bar{w}(r)|^{\alpha}) dr \\ &\leq \int_{\infty}^{\omega} (|\underline{v}(r) - \underline{u}(r)|^{\alpha} + |\bar{v}(r) - \bar{u}(r)|^{\alpha}) dr \\ &\quad + \int_{\infty}^{\omega} (|\underline{v}(r) - \underline{w}(r)|^{\alpha} + |\bar{v}(r) - \bar{w}(r)|^{\alpha}) dr \\ &= d_{\mathcal{F}}^{\alpha}(\tilde{u}_\omega, \tilde{v}_\omega) + d_{\mathcal{F}}^{\alpha}(\tilde{v}_\omega, \tilde{w}_\omega), \end{aligned}$$

که نامساوی بالا با توجه به خاصیت مثلثی نرم  $\|\cdot\|_{\alpha}$  نوشته شده است. بنابراین متر داده

شده در خاصیت ۴ نیز صدق می‌کند. با توجه به (۱) تا (۴) فضای  $\mathcal{F}$  همراه با متر  $d_{\mathcal{F}}$  یک فضای متریک است. □

در ادامه قوی بودن نرم پیشنهادی را بررسی می‌نماییم. برای این منظور، مجموعه داده‌های زیر را در نظر می‌گیریم:

$$\begin{aligned} & \{(13, 0, 0/27, 1), (14, 0, 1/95, 1/93), (14/4, 0, 0/56, 1/17), \\ & (14/7, 0, 0/89, 0/88), (14/9, 0, 0/12, 1/21), (15, 0, 1/19, 0/41), \\ & (15/1, 0, 1/82, 0/9), (15/3, 0, 1/9, 1/85), (15/6, 0, 1/79, 1/95), \\ & (16, 0, 1/47, 0/42), (17, 0, 0/38, 1/38)\} \end{aligned}$$

اگر تابع  $\sum_{j=1}^n d_{\mathcal{F}}(X_j, W)$  را مینیمم کنیم مقدار  $W$  برابر می‌شود با  $(55, 0, 0/38, 1/38)$ . حال اگر داده پرت  $(55, 0, 0/38, 1/38)$  را به این مجموعه داده اضافه کنیم و دوباره تابع داده شده را مینیمم کنیم مشاهده می‌کنیم که مقدار  $W$  برابر با مقدار قبلی آن می‌شود که نشان‌دهنده قوی بودن نرم پیشنهادی می‌باشد. اکنون به بررسی روش FCM که در آن داده‌ها و مراکز فازی هستند می‌پردازیم. فرض کنیم این داده‌ها و به دست آوردن مراکز خوشه‌ها که با  $\tilde{X} = \{\tilde{x}_1, \dots, \tilde{x}_n\}$  این داده‌ها و به دست آوردن مراکز خوشه‌ها که با  $\tilde{V} = \{\tilde{v}_1, \dots, \tilde{v}_n\}$  نمایش داده می‌شوند باید مساله بهینه‌سازی زیر را حل نماییم:

$$\begin{aligned} \min J_m(X, V) &= \sum_{j=1}^n \sum_{i=1}^c \mu_{ij}^m d_{\mathcal{F}}^m(\tilde{X}_j, \tilde{V}_i) \\ \text{to. Subject} & \\ \sum_{i=1}^c \mu_{ij} &= 1 \quad j = 1, 2, \dots, n, \\ 0 &< \sum_{j=1}^n \mu_{ij} < n \quad i = 1, 2, \dots, c. \end{aligned}$$

در ادامه این مقاله فرض می‌کنیم که تمامی داده‌های مورد بحث در این مقاله اعداد فازی دوزنقه‌ای هستند. بنابراین داریم  $n, 2, \dots, 1$  و  $\tilde{x}_j = (x_{j1}, x_{j2}, x_{j3}, x_{j4}, \omega_{\tilde{x}_j})$

شرایط لازم برای بهینگی را به کمک روش ضرایب لاگرانژ استخراج می‌نماییم. قبل از آن لازم است که قیدهای دسته دوم در محدودیت‌های مساله را به تساوی تبدیل نماییم بدین منظور از متغیرهای کمکی  $q_i$  و  $p_i$  استفاده می‌کنیم. از این رو داریم:

$$\sum_{j=1}^n \mu_{ij} - n + p_i^{\vee} = 0, \quad i = 1, 2, \dots, c \quad (5)$$

$$- \sum_{j=1}^j \mu_{ij} + q_i^{\vee} = 0, \quad i = 1, 2, \dots, c. \quad (6)$$

حال با توجه به روش ضرایب لاگرانژ داریم

$$L(\tilde{X}, \tilde{V}, \delta, \gamma, \rho) = \sum_{j=1}^n \sum_{i=1}^c \mu_{ij}^m d_{\mathcal{F}}^{\vee}(\tilde{x}_j, \tilde{v}_i) + \sum_{j=1}^n \delta_j \left(1 - \sum_{i=1}^c \mu_{ij}\right) + \sum_{i=1}^c \gamma_i \left(\sum_{j=1}^n \mu_{ij} - n + p_i^{\vee}\right) + \sum_{i=1}^c \rho_i \left(- \sum_{j=1}^n \mu_{ij} + q_i^{\vee}\right).$$

با مشتق گرفتن از  $L(\tilde{X}, \tilde{V}, \delta, \gamma, \rho)$  نسبت به همه متغیرها و مساوی صفر قرار دادن آنها شرایط لازم برای بهینگی به صورت زیر به دست می‌آیند ولی قبل از آن لازم است که با توجه به رابطه (۴) و این مطلب که همه داده‌های فازی موجود در مساله اعداد فازی دوزنقه‌ای هستند مقدار  $d_{\mathcal{F}}^{\vee}(\tilde{x}_j, \tilde{v}_i)$  را محاسبه کنیم:

$$\begin{aligned} d_{\mathcal{F}}^{\vee}(\tilde{x}_j, \tilde{v}_i) &= ((x_{j1} - v_{i1})^{\vee}(1 - \alpha) + \frac{1 - \alpha^{\vee}}{3}(x_{j2} - x_{j1} - v_{i2} + v_{i1})^{\vee})^{\vee} \\ &+ 2(x_{j1} - v_{i1})(1 - \alpha^{\vee})(x_{j2} - x_{j1} - v_{i2} + v_{i1}) \\ &+ ((x_{j4} - v_{i4})^{\vee}(1 - \alpha) + \frac{1 - \alpha^{\vee}}{3}(x_{j3} - x_{j4} - v_{i3} + v_{i4})^{\vee})^{\vee} \\ &+ 2(x_{j4} - v_{i4})(1 - \alpha^{\vee})(x_{j3} - x_{j4} - v_{i3} + v_{i4}) \end{aligned}$$

حال با مشتق‌گیری از تابع لاگرانژ داریم:

$$\frac{\partial L}{\partial \mu_{ij}} = m\mu_{ij}^{m-1} d_{\mathcal{F}}^{\gamma}(\tilde{x}_j \tilde{v}_i) - \delta_j + \gamma_i - \rho_j = 0, \quad (7)$$

$$\frac{\partial L}{\partial \delta_j} = 1 - \sum_{i=1}^c \mu_{ij} = 0, \quad (8)$$

$$\frac{\partial L}{\partial \gamma_i} = \sum_{j=1}^n \mu_{ij} - n + p_i^{\gamma} = 0, \quad (9)$$

$$\frac{\partial L}{\partial \rho_i} = - \sum_{j=1}^n \mu_{ij} + q_i^{\rho} = 0, \quad (10)$$

$$\frac{\partial L}{\partial p_i} = \gamma p_i \gamma_i = 0, \quad (11)$$

$$\frac{\partial L}{\partial q_i} = \rho q_i \rho_i = 0, \quad (12)$$

$$\begin{aligned} \frac{\partial L}{\partial v_{i1}} = \sum_{j=1}^n \mu_{ij}^m \left( -\gamma(x_{j1} - v_{i1})(1 - \alpha) \right. \\ \left. - \left( \frac{1}{\gamma} - \alpha + \frac{1}{\gamma} \alpha^{\gamma} \right) (v_{i3} - v_{i4} - x_{j3} + x_{j4}) \right) = 0, \end{aligned} \quad (13)$$

$$\begin{aligned} \frac{\partial L}{\partial v_{i2}} = \sum_{j=1}^n \mu_{ij}^m \left( (-x_{j2} + v_{i2})(1 - \alpha) \right. \\ \left. + \left( \frac{1}{\gamma} - \alpha + \frac{1}{\gamma} \alpha^{\gamma} \right) (v_{i3} + v_{i4} - x_{j3} - x_{j4}) \right) = 0, \end{aligned} \quad (14)$$

$$\begin{aligned} \frac{\partial L}{\partial v_{i3}} = \sum_{j=1}^n \mu_{ij}^m \left( (\gamma x_{j1} - x_{j2} - \gamma v_{i1} + v_{i2}) \left( \frac{1}{\gamma} - \alpha + \frac{1}{\gamma} \alpha^{\gamma} \right) \right. \\ \left. + \gamma (v_{i3} - x_{i3}) \left( \frac{1}{\gamma} - \alpha + \alpha^{\gamma} - \frac{1}{\gamma} \alpha^{\gamma} \right) \right) = 0, \end{aligned} \quad (15)$$

$$\begin{aligned} \frac{\partial L}{\partial v_{i4}} = \sum_{j=1}^n \mu_{ij}^m \left( (\gamma x_{j1} + x_{j2} - \gamma v_{i1} - v_{i2}) \left( \frac{1}{\gamma} - \alpha + \frac{1}{\gamma} \alpha^{\gamma} \right) \right. \\ \left. + \gamma (x_{j4} - v_{i4}) \left( \frac{1}{\gamma} - \alpha + \alpha^{\gamma} - \frac{1}{\gamma} \alpha^{\gamma} \right) \right) = 0, \end{aligned} \quad (16)$$

با توجه به معادلات (۶) و (۷) و اینکه  $q_i > 0$  و  $p_i$  داریم  $\rho_i = \gamma_i = 0$ . از معادلات (۸)-(۱۱) پارامترهای مربوط به مراکز خوشه‌ها به صورت زیر به دست می‌آیند:



$$v_{i1} = \frac{\sum_{j=1}^n \mu_{ij}^m \left( \frac{1}{2(1-\alpha)} \left( \frac{1}{\alpha} - \alpha + \frac{1}{\alpha} \alpha^2 \right) (v_{i3} - v_{i4} - x_{j3} + x_{j4}) + x_{j1} \right)}{\sum_{j=1}^n \mu_{ij}^m}, \quad (17)$$

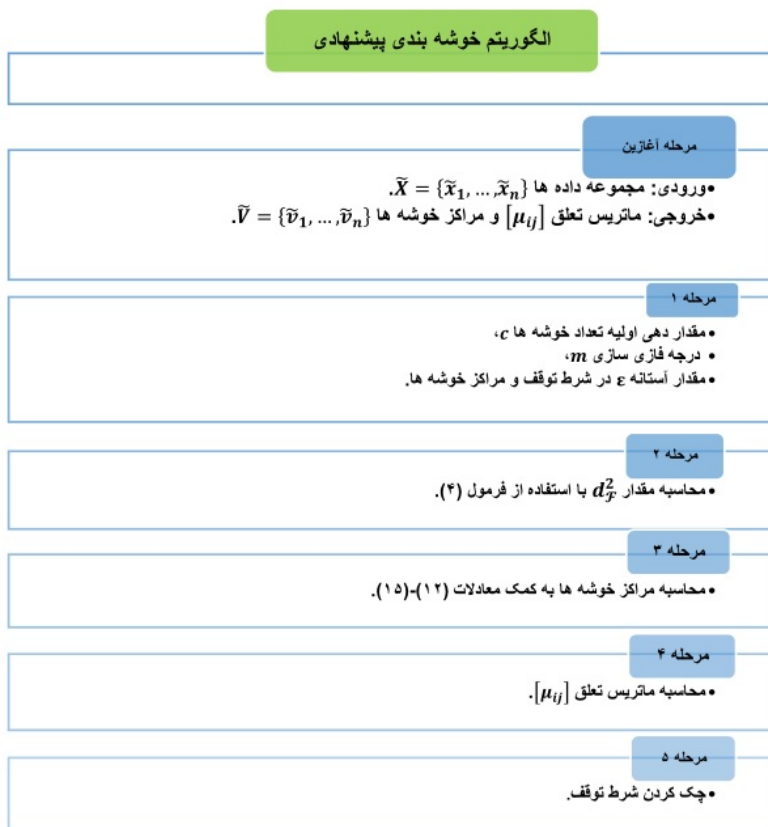
$$v_{i2} = \frac{\sum_{j=1}^n \mu_{ij}^m \left( \frac{-1}{(1-\alpha)} \left( \frac{1}{\alpha} - \alpha + \frac{1}{\alpha} \alpha^2 \right) (v_{i3} + v_{i4} - x_{j3} - x_{j4}) + x_{j2} \right)}{\sum_{j=1}^n \mu_{ij}^m}, \quad (18)$$

$$v_{i3} = \sum_{j=1}^n \mu_{ij}^m \left( \frac{-1}{2 \left( \frac{1}{\alpha} - \alpha + \alpha^2 - \frac{1}{\alpha} \alpha^3 \right)} \right. \\ \left. (2x_{j1} - x_{j2} - 2v_{i1} + v_{i2}) \left( \frac{1}{\alpha} - \alpha + \frac{1}{\alpha} \alpha^2 \right) + x_{j3} \right) / \sum_{j=1}^n \mu_{ij}^m, \quad (19)$$

$$v_{i4} = \sum_{j=1}^n \mu_{ij}^m \left( \frac{1}{2 \left( \frac{1}{\alpha} - \alpha + \alpha^2 - \frac{1}{\alpha} \alpha^3 \right)} \right. \\ \left. (2x_{j1} + x_{j2} - 2v_{i1} - v_{i2}) \left( \frac{1}{\alpha} - \alpha + \frac{1}{\alpha} \alpha^2 \right) + x_{j4} \right) / \sum_{j=1}^n \mu_{ij}^m, \quad (20)$$

مراحل الگوریتم پیشنهادی با توجه به شرایط لازم بهینگی به صورت زیر می‌باشند:

در اینجا منظور از شرط توقف بررسی رابطه  $\|J_m^t - J_m^{t-1}\| \leq \epsilon$  می‌باشد. الگوریتم FCM یکی از روش‌هایی است که به طور گسترده‌ای در خوشه‌بندی فازی به کار می‌رود. با توجه به تعداد مشخصی از خوشه‌ها، می‌توان ساختار پنهان یک مجموعه داده را از طریق بهینه‌سازی تابع هدف پیدا کرد. بنابراین تعداد خوشه‌ها و مقدار اولیه تاثیر زیادی بر نتیجه خوشه‌بندی دارد که منجر به دو مسئله در بهبود FCM می‌شود. در واقع یکی انتخاب تعداد خوشه‌ها و دیگری کاهش حساسیت به مقدار اولیه است. روش اصلی تعیین تعداد خوشه تعریف یک تابع ارزیاب است. با ارزیابی نتیجه خوشه‌بندی با تعداد خوشه‌های مختلف، روش به ما بهترین تعداد خوشه‌ها را نتیجه می‌دهد. انواع مختلفی از توابع ارزیاب وجود دارند [۳۲]. از جمله مهم‌ترین آنها می‌توان به شاخص افراز، آنتروپی افراز، شاخص نسبت



شکل ۲: الگوریتم پیشنهادی

وینگهام [۳۳، ۳۴، ۳۵]، شاخص فوکویاما-سگنو، شاخص بن سای و فشردگی و جدایی رضایی [۳۶، ۳۷، ۳۸] اشاره کرد.

## ۴ مثال‌های عددی

در این بخش، دو مثال عددی و یک مثال کاربردی را برای نشان دادن کارایی روش پیشنهادی ارائه می‌کنیم. در دو مثال عددی، دو مجموعه اعداد فازی به ترتیب ۵ و ۶ عضوی مورد بررسی قرار می‌گیرند. به منظور بررسی اثربخشی الگوریتم خوشه‌بندی پیشنهادی، الگوریتم پیشنهاد شده و روش‌های HFCM [۳۹] و MVFCM [۴۰] برای

خوشه‌بندی مجموعه داده‌های فازی ذکر شده مورد استفاده قرار می‌گیرند. مثال‌های این بخش از مرجع [۴۱] انتخاب شده‌اند.

مثال ۱.۰۴. یک مجموعه داده با پنج عدد فازی مثلثی که در شکل ۳ نشان داده شده را در نظر می‌گیریم. MVFCM، HFCM و الگوریتم‌های خوشه‌بندی FCM پیشنهاد شده در این مجموعه داده با  $c = 2$  اجرا شده و درجات عضویت به دست آمده در جدول ۱ و مراکز به دست آمده در جدول ۲ نشان داده شده‌اند.

جدول ۱: درجات عضویت برای داده‌های فازی مثال ۱.۰۴ و مقایسه روش پیشنهادی با الگوریتم HFCM و الگوریتم MVFCM.

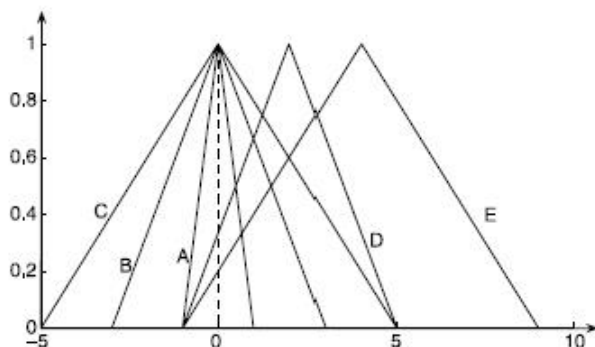
روش پیشنهادی				H FCM		MVFCM		داده‌های فازی
$\alpha = 0.8$		$\alpha = 0.3$		$u_{2k}$	$u_{1k}$	$u_{2k}$	$u_{1k}$	
$u_{2k}$	$u_{1k}$	$u_{2k}$	$u_{1k}$	$u_{2k}$	$u_{1k}$	$u_{2k}$	$u_{1k}$	
0.173	0.9827	0.1074	0.8926	0.334	0.9666	0.3879	0.6121	$A = (0, 0, 1, 1)$
0.0055	0.9945	0.0426	0.9574	0.338	0.662	0.2978	0.7022	$B = (0, 0, 3, 3)$
0.002	0.998	0.0166	0.9834	0.8256	0.1744	0.1663	0.8337	$C = (0, 0, 5, 5)$
0.615	0.385	0.3221	0.6779	0.6847	0.3153	0.6908	0.3092	$D = (2, 0, 3, 3)$
0.972	0.028	0.8669	0.1331	0.8766	0.1234	0.7868	0.2132	$E = (4, 0, 5, 5)$

جدول ۲: مراکز خوشه‌ها برای داده‌های فازی مثال ۱.۰۴ و مقایسه روش پیشنهادی با الگوریتم HFCM و الگوریتم MVFCM.

روش	خوشه ۱	خوشه ۲
MVFCM	(0.6793, 0.0095, 3/0.017, 3/0.017)	(3/7.078, 1/9977, 2/2269, 2/2269)
FCM H	(0.8791, 0.2135, 3/1890, 3/1890)	(3/8399, 1/9600, 1/7830, 1/7830)
روش پیشنهادی ( $\alpha = 0.8$ )	(0.0966, 0.4748, 3/0.262, 3/0.212)	(3/4281, 0.7629, 4/4577, 4/4132)
روش پیشنهادی ( $\alpha = 0.3$ )	(0.3135, 1/6095, 3/1112, 3/1168)	(3/6978, 3/0614, 4/6851, 4/6949)

همانطور که در جدول ۱ مشاهده می‌شود هنگام استفاده از الگوریتم MVFCM اعداد فازی A و B با درجه عضویت بالاتری عضو خوشه ۱ و اعداد C و D و E با درجه عضویت بالاتری عضو خوشه ۲ می‌باشند. در حالیکه هنگام استفاده از الگوریتم HFCM روش پیشنهادی هنگامی که یک تصمیم سطح بالا را اختیار می‌کنیم (در حالت  $\alpha = 0.8$ ) اعداد فازی A و B و C با درجه عضویت بالاتری عضو خوشه ۱ و اعداد فازی D و E با درجه عضویت بالاتری عضو خوشه ۲ می‌باشند. ولی هنگامی که یک تصمیم سطح پایین را به کمک روش پیشنهادی اختیار می‌کنیم اعداد فازی A، B، C و D عضو خوشه ۱

ولی عدد فازی  $E$  عضو خوشه ۲ می شود.



شکل ۳: اعداد فازی مثلثی مثال ۱.۴.

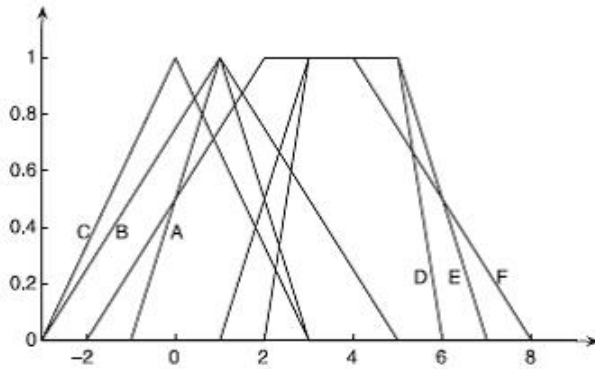
مثال ۲.۴. در این مثال ۶ عدد فازی داده شده در شکل ۴ را به کمک روش های ذکر شده خوشه بندی می کنیم.

جدول ۳: درجات عضویت برای داده های فازی مثال ۱.۴ و مقایسه روش پیشنهادی با الگوریتم HFCM و الگوریتم MVFCM.

روش پیشنهادی			H FCM		MVFCM		داده های فازی
$\alpha = 0.8$		$\alpha = 0.3$	$u_{1k}$	$u_{2k}$	$u_{1k}$	$u_{2k}$	
$u_{2k}$	$u_{1k}$	$u_{2k}$	$u_{1k}$	$u_{2k}$	$u_{1k}$	$u_{2k}$	$A = (1, 0, 2, 2)$
0/0.231	0/0.9769	0/1.210	0/0.8790	0/0.1940	0/0.8060	0/0.266	0/0.9734
0/0.143	0/0.9857	0/0.363	0/0.9637	0/0.595	0/0.9405	0/0.279	0/0.9721
0/0.233	0/0.9670	0/0.652	0/0.9348	0/0.397	0/0.9603	0/0.310	0/0.9690
0/0.9830	0/0.170	0/0.8848	0/0.1152	0/0.9473	0/0.527	0/0.9769	0/0.231
0/0.9880	0/0.120	0/0.9371	0/0.629	0/0.9924	0/0.076	0/0.9922	0/0.078
0/0.9270	0/0.0730	0/0.9151	0/0.849	0/0.4608	0/0.5392	0/0.8866	0/0.1134

همانطور که از جدول ۳ مشاهده می شود اعداد فازی  $A$ ،  $B$  و  $C$  در هر سه روش در خوشه ۱ و اعداد فازی  $D$ ،  $E$  و  $F$  در خوشه ۲ قرار می گیرند.

مثال ۳.۴. این مثال یک مثال کاربردی در مورد خوشه بندی خودروها بر اساس سه ویژگی قیمت، راحتی و ایمنی است [۴۲]. با توجه به اینکه قیمت یک خودرو یک پارامتر کمی است آن را با یک عدد قطعی نمایش می دهیم اما برای نمایش دو پارامتر دیگر از اعداد فازی ذوزنقه ای استفاده می کنیم. جدول ۵ مقدار هر کدام از این سه ویژگی را



شکل ۴: اعداد فازی مثلثی و ذوزنقه‌ای مثال ۲.۴.

جدول ۴: مراکز خوشه‌ها برای داده‌های فازی مثال ۲.۴ و مقایسه روش پیشنهادی با الگوریتم *HFCM* و الگوریتم *MVFCM*

خوشه ۲	خوشه ۱	روش
(۳/۷۰۷۸, ۱/۹۹۷۷, ۲/۲۲۶۹, ۲/۲۲۶۹)	(۰/۶۷۹۳, ۰/۰۰۹۵, ۳/۰۰۱۷, ۳/۰۰۱۷)	MVFCM
(۳/۸۳۹۹, ۱/۹۶۰۰, ۱/۷۸۳۰, ۱/۷۸۳۰)	(۰/۸۷۹۱, ۰/۲۱۳۵, ۳/۱۸۹۰, ۳/۱۸۹۰)	FCM H
(۳/۶۹۳۱, ۲/۳۶۵۲, ۲/۲۹۳۰, ۲/۲۳۲۱)	(۰/۶۷۸۵, ۰/۴۲۷۱, ۳/۰۱۵۰, ۳/۰۰۸۵)	روش پیشنهادی ( $\alpha = ۰.۸$ )
(۳/۶۳۹۱, ۳/۳۰۹۲, ۲/۳۴۴۹, ۲/۳۶۸۱)	(۰/۶۸۸۶, ۱/۵۱۰۷, ۳/۰۴۴۱, ۳/۰۴۷۰)	روش پیشنهادی ( $\alpha = ۰.۳$ )

برای خودروهای مختلف نمایش می‌دهد. در جدول ۶ مراکز خوشه‌های برای هر دو سطح تصمیم‌گیری نمایش داده شده است. با توجه به داده‌های جدول ۷ مشاهده می‌شود که خودروهای استون مارتین، میتسوبیسی لنسر، تیرا آتیوا، تویوتا ترسل و نیسا میکرو متعلق به خوشه ۱ و مابقی خودروها متعلق به خوشه ۲ هستند.

## ۵ نتیجه‌گیری

در این مقاله یک الگوریتم FCM بر پایه یک فاصله جدید پارامتری برای خوشه‌بندی داده‌های فازی با مراکز فازی ارائه شده است. از ویژگی‌های حائز اهمیت فاصله پیشنهاد شده این است که با استفاده از این فاصله، الگوریتم FCM نسبت به داده‌های پرت حساس نیست و ثانیاً نظر تصمیم‌گیرنده را در خوشه‌بندی داده‌ها دخالت می‌دهد. به طوری که اگر تصمیم‌گیرنده بخواهد یک تصمیم سطح بالا را بگیرد می‌بایست مقدار  $\alpha$  را نزدیک به یک

جدول ۵: داده‌های مربوط به ویژگی‌های ۱۰ خودرو

نوع خودرو	قیمت	راحتی	ایمنی
استون مارتین	(۰, ۰, ۹/۶۳, ۹/۶۳)	(۲, ۲, ۱۰, ۱۰)	(۳, ۳, ۹, ۹)
میتسوبیshi لنسر	(۰, ۰, ۹/۵۱, ۹/۵۱)	(۲, ۲, ۶, ۶)	(۳, ۳, ۶, ۶)
میتسوبیshi گالانت	(۰, ۰, ۸/۷۱, ۸/۷۱)	(۰, ۲, ۱۴, ۱۰)	(۰, ۳, ۵/۱۷, ۵/۱۲)
تیرا آتیوا	(۰, ۰, ۹/۴۶, ۹/۴۶)	(۲, ۲, ۶, ۶)	(۳, ۳, ۶, ۶)
ام ۲۰۰۰	(۰, ۰, ۶/۶۴, ۶/۶۴)	(۲, ۲, ۸, ۸)	(۳, ۳, ۹, ۹)
تویوتا ترسل	(۰, ۰, ۸/۴۵, ۸/۴۵)	(۲, ۰, ۶, ۲)	(۳, ۳, ۶, ۶)
تویوتا کرولا	(۰, ۰, ۳/۷۴, ۳/۷۴)	(۰, ۲, ۱۴, ۱۰)	(۳, ۳, ۱۲, ۱۲)
تویوتا پرمیو	(۰, ۰, ۹/۷۲, ۹/۷۲)	(۲, ۲, ۱۰, ۱۰)	(۰, ۳, ۵/۱۷, ۵/۱۲)
نیسان لورل آلتیما	(۰, ۰, ۹/۶۹, ۹/۶۹)	(۲, ۲, ۸, ۸)	(۳, ۳, ۱۲, ۱۲)
نیسان میکرو	(۰, ۰, ۹/۳۹, ۹/۳۹)	(۲, ۰, ۶, ۲)	(۳, ۰, ۵/۵, ۵/۰)

جدول ۶: مراکز خوشه‌ها برای مثال ۳.۴

نوع خودرو	قیمت	راحتی	امنیت
خوشه ۱ ( $\alpha = 0.3$ )	(۴۶/۰۴, ۴۶/۰۲, -۰/۰۸, ۰/۲۵)	(۹/۴۲, ۱۱/۷۱, ۱/۹۷, ۱/۲۹)	(۵/۵۹, ۷/۴۲, ۲/۴۸, ۳/۰۲)
خوشه ۲ ( $\alpha = 0.3$ )	(۶۹/۷۲, ۶۹/۷۲, -۰/۱۵, ۰/۴۶)	(۴/۱۷, ۶/۶۲, ۰/۹۷, ۲/۰۱)	(۱۱/۷۵, ۱۵/۲۲, ۲/۹۵, ۱/۶۸)
خوشه ۱ ( $\alpha = 0.8$ )	(۴۶/۱۰, ۴۶/۰۷, ۰/۲۲, -۰/۶۶)	(۹/۳۹, ۱۱/۰۱, ۲/۰۴, ۱/۱۳)	(۵/۴۲, ۶/۵۸, ۲/۴۸, ۲/۹۲)
خوشه ۲ ( $\alpha = 0.8$ )	(۶۹/۸۲, ۶۹/۸۲, ۰/۴۰, -۱/۲۰)	(۳/۷۶, ۶/۲۰, ۰/۸۳, ۱/۹۴)	(۱۱/۷۸, ۱۴/۳۰, ۳/۰۴, ۱/۴۹)

جدول ۷: مراکز خوشه‌ها برای مثال ۳.۴

نوع خودرو	خوشه ۱ ( $\alpha = 0.3$ )	خوشه ۲ ( $\alpha = 0.3$ )	خوشه ۱ ( $\alpha = 0.8$ )	خوشه ۲ ( $\alpha = 0.8$ )
استون مارتین	۰/۵۴۲۵	۰/۴۵۷۵	۰/۵۲۵۹	۰/۴۷۴۱
میتسوبیshi لنسر	۰/۶۷۴۹	۰/۳۲۵۱	۰/۷۰۳۴	۰/۲۹۶۶
میتسوبیshi گالانت	۰/۳۳۲۵	۰/۶۶۷۵	۰/۳۳۲۵	۰/۶۶۷۵
تیرا آتیوا	۰/۷۰۵۹	۰/۲۹۴۱	۰/۷۳۵۲	۰/۲۶۴۸
ام ۲۰۰۰	۰/۴۵۹۵	۰/۵۴۰۵	۰/۴۶۳۲	۰/۵۳۶۸
تویوتا ترسل	۰/۶۸۹۳	۰/۳۱۳۷	۰/۶۸۰۲	۰/۳۱۹۸
تویوتا کرولا	۰/۳۴۸۵	۰/۶۵۱۵	۰/۳۳۵۳	۰/۶۶۷۴
تویوتا پرمیو	۰/۳۳۲۵	۰/۶۶۶۵	۰/۳۴۴۴	۰/۶۵۵۶
نیسان لورل آلتیما	۰/۲۶۴۶	۰/۷۳۵۴	۰/۲۸۲۲	۰/۷۱۷۸
نیسان میکرو	۰/۶۲۶۶	۰/۳۷۳۴	۰/۶۲۲۱	۰/۳۷۷۹

اختیار کند و اگر بخواهد یک تصمیم سطح پایین را بگیرد باید مقدار  $\alpha$  را نزدیک به صفر اختیار کند. با مقایسه نتایج الگوریتم جدید برای دو مجموعه داده فازی و نتایج به دست آمده با استفاده از دو روش HFCM و MVFCM مشاهده می‌شود که تقریب جدید دارای کارایی قابل قبولی است. این ایده جدید می‌تواند در بسیاری از برنامه‌های کاربردی مورد استفاده قرار گیرد و ما امیدواریم نتایج خوشه‌بندی را با استفاده از فاصله فازی جدید در

## مراجع

- [1] Jafari, H., and M. J. Ebadi. "Malliavin calculus in statistical inference: Cramer-Rao lower bound for fuzzy random variables.", *Journal of Decisions and Operations Research*, 5(2) (2020): 124-132,
- [2] Farahani, H., and M. J. Ebadi, "Finding Fuzzy Inverse Matrix Using Wu's Method", *Journal of Mahani Mathematical Research Center*, 10(1) (2021), 37-52.
- [3] Farahani, H., M. J. Ebadi, and H. Jafari. "Finding Inverse of a Fuzzy Matrix using Eigenvalue Method.", *International Journal of Innovative Technology and Exploring Engineering*, 9(2) (2019): 3030-3037.
- [4] Ebadi, M. J., and M. S. Shiri Shahraki. "Determination of scale elasticity in the existence of non-discretionary factors in performance analysis." *Knowledge-Based Systems* 23.5 (2010): 434-439.
- [5] Ebadi, M. J., Hosseini, M.M., Karbassi, S.M.: An efficient one-layer recurrent neural network for solving a class of nonsmooth pseudoconvex optimization problems. *Journal of Theoretical and Applied Information Technology* 96.7 (2018).
- [6] R.O. Duda, P.E. Hart, (1973). *Pattern Classification and Scene Analysis*, Wiley, New York.
- [7] Na, S., Xumin, L., & Yong, G. (2010, April). Research on k-means clustering algorithm: An improved k-means clustering algorithm. In *Intelligent Information Technology and Security Informatics (IITSI), 2010 Third International Symposium on* (pp. 63-67). IEEE.

- [8] Srinivas, M., & Mohan, C. K. (2010, July). Efficient clustering approach using incremental and hierarchical clustering methods. In *Neural Networks (IJCNN), The 2010 International Joint Conference on* (pp. 1-7). IEEE.
- [9] Kim, Y., Shim, K., Kim, M. S., & Lee, J. S. (2014). DBCURE-MR: an efficient density-based clustering algorithm for large data using MapReduce. *Information Systems*, 42, 15-35.
- [10] Lee, J., & Lee, D. (2005). An improved cluster labeling method for support vector clustering. *IEEE Transactions on pattern analysis and machine intelligence*, 27(3), 461-464.
- [11] Liu, B., Wan, C., & Wang, L. (2006). An efficient semi-supervised gene selection method via spectral biclustering. *IEEE Transactions on nanobioscience*, 5(2), 110-114.
- [12] Pedrycz, W., & Rai, P. (2008). Collaborative clustering with the use of Fuzzy C-Means and its quantification. *Fuzzy Sets and Systems*, 159(18), 2399-2427.
- [13] Staiano, A., Tagliaferri, R., & Pedrycz, W. (2006). Improving RBF networks performance in regression tasks by means of a supervised fuzzy clustering. *Neurocomputing*, 69(13), 1570-1581.
- [14] Ji, Z. X., Sun, Q. S., & Xia, D. S. (2011). A modified possibilistic fuzzy c-means clustering algorithm for bias field estimation and segmentation of brain MR image. *Computerized Medical Imaging and Graphics*, 35(5), 383-397.
- [15] Ding, Y., & Fu, X. (2016). Kernel-based fuzzy c-means clustering algorithm based on genetic algorithm. *Neurocomputing*, 188, 233-238.
- [16] Song, Q., Yang, X., Soh, Y. C., & Wang, Z. M. (2010). An information-theoretic fuzzy C-spherical shells clustering algorithm. *Fuzzy Sets and Systems*, 161(13), 1755-1773.



- [17] Li, K., & Li, P. (2014). Fuzzy clustering with generalized entropy based on neural network. In *Unifying Electrical Engineering and Electronics Engineering* (pp. 2085-2091). Springer, New York, NY.
- [18] Sridhar, V., & Murty, M. N. (1991). A knowledge-based clustering algorithm. *Pattern recognition letters*, 12(9), 511-517.
- [19] Dovžan, D., & Škrjanc, I. (2011). Recursive fuzzy c-means clustering for recursive fuzzy identification of time-varying processes. *ISA transactions*, 50(2), 159-169.
- [20] Ji, Z. X., Sun, Q. S., & Xia, D. S. (2011). A modified possibilistic fuzzy c-means clustering algorithm for bias field estimation and segmentation of brain MR image. *Computerized Medical Imaging and Graphics*, 35(5), 383-397.
- [21] Alpaydin, E. (2014). *Introduction to machine learning*. MIT press.
- [22] Webb, A. (2002). *Statistical Pattern Recognition*. Wiley, New Jersey.
- [23] Borlea, I. D., Precup, R. E., Borlea, A. B., & Iercan, D. (2021). A unified form of fuzzy C-means and K-means algorithms and its partitional implementation. *Knowledge-Based Systems*, 214, 106731.
- [24] Surono, S., & Putri, R. D. A. (2021). Optimization of fuzzy c-means clustering algorithm with combination of minkowski and chebyshev distance using principal component analysis. *International Journal of Fuzzy Systems*, 23(1), 139-144.
- [25] Kaushal, M., & Lohani, Q. M. (2022). Generalized intuitionistic fuzzy c-means clustering algorithm using an adaptive intuitionistic fuzzification technique. *Granular Computing*, 7(1), 183-195.

- [26] Abdellahoum, H., Mokhtari, N., Brahim, A., & Boukra, A. (2021). CSFCM: An improved fuzzy C-Means image segmentation algorithm using a cooperative approach. *Expert Systems with Applications*, 166, 114063.
- [27] Zimmermann, (1995) H. J. *Fuzzy set theory and its Application*, kluwer Academic publishers, Boston / Derdecht / London.
- [28] Ma, M., Friedman, M., and Kandel, A. (1999). A new fuzzy arithmetic, fuzzy sets and systems, vol. 108, no. 1, pp. 83-90.
- [29] Dunn, J. C. (1973). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters.
- [30] Bezdek JC (1981) Models for pattern recognition. In: *Pattern recognition with fuzzy objective function algorithms*, 1-13. US: Springer.
- [31] Castro-Company, F., & Tirado, P. (2014). On Yager and Hamacher t-Norms and Fuzzy Metric Spaces. *International Journal of Intelligent Systems*, 29(12), 1173-1180.
- [32] Zhang, D., Ji, M., Yang, J., Zhang, Y., & Xie, F. (2014). A novel cluster validity index for fuzzy clustering based on bipartite modularity. *Fuzzy Sets and Systems*, 253, 122-137.
- [33] Bezdek, J. C. (1973). Cluster validity with fuzzy sets.
- [34] Bezdek, J. C. (1974). Numerical taxonomy with fuzzy sets. *Journal of Mathematical Biology*, 1(1), 57-71.
- [35] Windham, M. P. (1982). Cluster validity for the fuzzy C-means clustering algorithm. *IEEE transactions on pattern analysis and machine intelligence*, (4), 357-363.

- [36] Fukuyama, Y. (1989). A new method of choosing the number of clusters for the fuzzy c-mean method. In Proc. 5th Fuzzy Syst. Symp., 1989 (pp. 247-250).
- [37] Xie, X. L., & Beni, G. (1991). A validity measure for fuzzy clustering. IEEE Transactions on pattern analysis and machine intelligence, 13(8), 841-847.
- [38] Rezaee, B. (2010). A cluster validity index for fuzzy clustering. Fuzzy Sets and Systems, 161(23), 3014-3025.
- [39] Yang, M. S., Hung, W. L., & Cheng, F. C. (2006). Mixed-variable fuzzy clustering approach to part family and machine cell formation for GT applications. International Journal of Production Economics, 103(1), 185-198.
- [40] Hathaway, R. J., Bezdek, J. C., & Pedrycz, W. (1996). A parametric model for fusing heterogeneous fuzzy data. IEEE transactions on Fuzzy Systems, 4(3), 270-281.
- [41] Effati, S., Yazdi, H. S., & Sharahi, A. J. (2013). Fuzzy clustering algorithm for fuzzy data based on  $\alpha$ -cuts. Journal of Intelligent & Fuzzy Systems, 24(3), 511-519.
- [42] Coppi, R., D'Urso, P., & Giordani, P. (2012). Fuzzy and possibilistic clustering for fuzzy data. Computational Statistics & Data Analysis, 56(4), 915-927.