سیستم های فازی و کاربردها
شاپای چاپی: ۲۷۱۷ـ۴۴۰۹
سال هشتم، شماره دوم (پائیز و زمستان ۱۴۰۴)
شاپای الکترونیکی: ۲۷۱۷ـ۳۹۸۴
صص. ۹۱ تا ۱۰۸
DOI:10.22034/jfsa.2025.523784.1273

خوشه‌بندی مبتنی بر جامعهٔ فازی؛ چه هنگام و چگونه؟

مرضیه رضائی، عباس پرچمی* و ایوب شیخی

بخش آمار، دانشکده ریاضی و کامپیوتر، دانشگاه شهید باهنر کرمان، کرمان، ایران

تاریخ دریافت: ۱۴۰۴/۰۲/۲۵ تاریخ پذیرش: ۱۴۰۴/۰۷/۰۹

نوع مقاله: علمی‌ـ پژوهشی

چکیده. تاکنون روش‌های متنوعی برای خوشه‌بندی داده‌های مستخرج از یک جامعهٔ دقیق مطرح شده است. اما ممکن است در عمل با مواردی روبرو شویم که جامعه به‌صورت نادقیق یا فازی باشد مانند جامعهٔ فازی خانوارهای پرمصرف، جامعهٔ فازی کالاهای باکیفیت یک کارخانه، جامعهٔ فازی افراد مسن، جامعهٔ فازی خانوارهای با درآمد بالا. درچنین وضعیت‌ـ هایی هر یک از داده‌های مستخرج از جامعه، علاوه بر مقدار مشاهده شده، شامل درجهٔ عضویت آن مشاهده به جامعهٔ فازی نیز است. در این مقاله روشی نوین برای خوشه‌بندی داده‌های مستخرج از یک جامعهٔ فازی پیشنهاد شده که تعمیمی از الگوریتم k-میانگین است. همچنین یک مثال عملی به‌منظور درک بهتر مفهوم جامعهٔ فازی و نحوهٔ عملکرد الگوریتم پیشنهادی مطرح شده است.

۱. مقدمه

هدف از خوشه‌بندی تقسیم‌بندی یک مجموعه از اشیاء به گروه یا خوشه‌هایی است که شباهت بین اشیاء درون هر خوشه نسبت به اشیاء بین خوشه‌ها خیلی بیشتر باشد. خوشه‌بندی ابزار مؤثری برای درک ساختار کلی مجموعه داده‌ها و تصمیم‌گیری درباره طبقه‌بندی نمونه‌های

2010 *Mathematics Subject Classification.* 03E72 ; 62A86
* Corresponding author
E-mails: Rezaei.Marzieh2021@math.uk.ac.ir, parchami@uk.ac.ir and sheikhy.a@uk.ac.ir.
عبارات و کلمات کلیدی. الگوریتم k‌ـ میانگین، تابع عضویت، داده‌های مستخرج از جامعهٔ فازی، وزن‌دهی.

جدید محسوب می‌شود. روش k-میانگین[1] یکی از روش‌های خوشه‌بندی است و علی‌رغم سادگی یک روش پایه برای بسیاری از روش‌های خوشه‌بندی دیگر محسوب می‌شود که برای داده‌های کیفی نیز قابل استفاده است. در این روش، ابتدا با استفاده از روش سلسله مراتبی، مشاهدات در تعداد مشخصی خوشه قرار داده می‌شوند. پس از آن، تمامی مشاهدات موجود تا جایی در خوشه‌ها جابجا می‌شوند که نتیجه امر منجر به کمترین فاصله بین مشاهدات درون خوشه‌ها و بیشترین فاصلهٔ بین خوشه‌ها شود. به‌منظور مطالعه بیشتر الگوریتم‌های خوشه‌ـ بندی و کاربردهای آن‌ها به [۷] مراجعه کنید. علاوه بر روش‌های خوشه‌بندی کلاسیک که هر مشاهده را به صورت قطعی به یک خوشه اختصاص می‌دهند، خوشه‌بندی فازی به عنوان یکی از شاخه‌های مهم الگوریتم‌های خوشه‌بندی، این امکان را فراهم می‌آورد که هر مشاهده با درجات عضویت مختلف، به چندین خوشه تعلق داشته باشد. در واقع، تمایز اصلی خوشه‌بندی فازی نسبت به روش‌های کلاسیک در این است که مشاهدات می‌توانند به صورت همزمان در بیش از یک خوشه مشارکت داشته باشند و این ویژگی، انعطاف‌پذیری بالاتری را در مدل‌سازی ساختار داده‌ها به ارمغان می‌آورد. از جمله الگوریتم‌های مطرح در این حوزه می‌توان به خوشه‌بندی فازی C ـ میانگین، خوشه‌بندی احتمالی فازی C ـ میانگین و الگوریتم‌های کاهشی اشاره نمود. این مجموعه الگوریتم‌ها به عنوان ابزارهای کلیدی در مسائل تشخیص الگوی بدون نظارت شناخته می‌شوند و قابلیت آنها در مدل‌سازی عدم قطعیت و اشتراک بین خوشه‌ها، موجب شده تا در کاربردهای متنوعی از جمله پردازش تصویر، داده‌کاوی، حوزه‌های پزشکی و شناسایی الگو مورد استفاده قرار گیرند؛ به ویژه در مسائلی که مرزهای تفکیک بین خوشه‌ها مشخص و واضح نیستند. برای مطالعه دقیق‌تر و بررسی جزئیات بیشتر در زمینه خوشه‌بندی فازی، خواننده می‌تواند به منابع [۱۶] و [۱۲] مراجعه نماید.

در داده‌های چندبُعدی، اهمیت متغیرها (ویژگی‌ها) یکسان نبوده و برخی از آن‌ها نقش مؤثرتری در تفکیک خوشه‌ها ایفا می‌کنند. فرآیند وزن‌دهی به متغیرها به معنای تخصیص درجهٔ اهمیت متفاوت به هر متغیر نسبت به سایرین است. به عبارت دیگر، انتخاب یک متغیر خاص برای ورود به تحلیل، نوعی وزن‌دهی ویژه به شمار می‌آید، به گونه‌ای که متغیرهای حذف‌شده وزن صفر دریافت می‌کنند. سنث و سوکلا [۱۸] انتقاداتی را نسبت به وزن‌دهی متغیرها مطرح کرده‌اند که مبتنی بر وجود پیش‌داوری در تخصیص وزن‌ها است و این پیش‌داوری خود بازتاب‌دهنده نوعی پیش‌رده‌بندی داده‌ها محسوب می‌شود. گوردن [۱۱] بیان می‌کند که در شرایطی که هیچ اطلاعاتی درباره اهمیت نسبی متغیرها در زمینه تحقیق موجود نباشد،

[1] k-means

تخصیص وزن‌های یکسان به تمام متغیرها معمولاً رویکرد مناسبی است؛ اما در غیر این صورت، وزن‌دهی برابر نتایج قابل اتکایی به‌دست نخواهد داد. برطبق نظر دان و اورایت [۶] چالش اصلی در زمینه وزن‌دهی پیشین صرفاً مسئله منطقی بودن این اقدام نیست، بلکه مشکل اساسی به نحوه تخصیص عملی وزن‌ها به متغیرها مربوط می‌شود. از این رو، پیشنهاد می‌شود که پیش از وزن‌دهی، مطالعات جزئی و دقیق روی داده‌ها انجام شود تا با استفاده از نتایج حاصل، اهمیت نسبی هر متغیر در جداسازی خوشه‌ها تعیین گردد. تعیین وزن متغیرها یکی از موضوعات مهم تحقیقاتی در حوزه تحلیل خوشه‌ای بوده و تحقیقات متعددی در این زمینه صورت گرفته است که می‌توان به [۸] و [۱۴] اشاره کرد.

علاوه بر وزن‌دهی به متغیرها، در برخی پژوهش‌ها اهمیت مشاهدات (نمونه‌ها) نیز می‌تواند متفاوت باشد. وزن‌دهی به مشاهدات به معنای تخصیص عددی مثبت به عنوان وزن به هر مشاهده است که نشان‌دهندهٔ اهمیت یا تأثیر آن مشاهده در فرآیند یادگیری یا خوشه‌بندی می‌باشد. این رویکرد به الگوریتم امکان می‌دهد تا برخی مشاهدات را نسبت به سایرین با اهمیت بیشتر یا کمتر در نظر بگیرد. در برخی مسائل، مشاهداتی که اطلاعات دقیق‌تر یا قابل‌اطمینان‌تری ارائه می‌دهند، بایستی وزن بیشتری دریافت کنند. وزن‌دهی به مشاهدات می‌تواند به شیوه‌های مختلفی از جمله استفاده از وزن‌های ثابت و از پیش تعیین شده براساس اهمیت هر مشاهده، و یا تخصیص وزن‌ها بر اساس معیارهای خارجی انجام شود. همچنین، در روش‌های وزن‌دهی خودکار، وزن‌ها به عنوان پارامترهایی در مدل لحاظ شده و در طول فرآیند یادگیری به‌روزرسانی می‌شوند. هاثوی و همکاران [۱۹] نیز یکی از نخستین پژوهش‌ها را در زمینه خوشه‌بندی فازی با وزن‌دهی به مشاهدات و متغیرها انجام دادند؛ در این مطالعه وزن‌ها به عنوان پارامترهایی در نظر گرفته می‌شوند تا نقش مشاهدات در فرایند خوشه‌بندی متناسب گردد. برای مطالعه بیشتر در این زمینه می‌توان به [۹] مراجعه نمود.

از طرفی، در بسیاری از مسائل کاربردی تحلیل آماری مبتنی بر سرشماری یا نمونه‌گیری از یک جامعهٔ فازی است [۲]. در روش‌های رایج ([۱۰]، [۱۵] و [۲۰])، ابتدا جامعهٔ فازی به کمک یک بازهٔ غیرفازی مدل‌بندی شده و سپس جمع‌آوری و بررسی داده‌ها انجام می‌شود. به عنوان چندین نمونه از این نوع مسائل کاربردی در علوم مختلف می‌توان به موارد زیر اشاره کرد: بررسی پوکی استخوان در افراد مسن، فشارخون افراد چاق، ارزش غذایی مواد پرکالری، میزان محصول درختان کم محصول، متوسط گردوغبار در روزهای طوفانی، بررسی الگوی مصرف در خانوارهای با درآمد پایین. در آمار کلاسیک خوشه‌بندی داده‌های حاصل از مسائل بالا بدون درنظر گرفتن درجهٔ عضویت هر مشاهده در جامعهٔ فازی انجام می‌شود. در چنین مواردی از

آنجاییکه از ابهام موجود در جامعهٔ فازی چشم‌پوشی شده است، بخشی مفید از اطلاعات نادیده گرفته می‌شود. در این حالت مناسب‌تر آن است که ابتدا جامعهٔ فازی موردنظر به‌وسیلهٔ یک تابع عضویت به طور دقیق مدل‌سازی شود تا با این کار درجهٔ عضویت هر داده در این جامعهٔ فازی مشخص شود. بنابراین در چنین حالتی بایستی روش‌های خوشه‌بندی در آمار کلاسیک به گونه‌ای مورد بازنگری و تعمیم قرار گیرند که نقش و میزان تاثیرگذاری تک‌تک داده‌ها در محاسبات آماری براساس میزان عضویت آن‌ها به جامعهٔ فازی مشخص و تعیین گردد. به عبارت دیگر، در چنین حالت‌هایی تحلیل آماری (از جمله خوشه‌بندی) باید به گونه‌ای صورت گیرند که داده‌های با درجات عضویت بزرگ‌تر، نقشی پررنگ‌تر و تاثیرگذارتر در محاسبات، نسبت به داده‌های با درجات عضویت کوچک‌تر داشته باشند [۲]. در این مقاله، الگوریتمی نوین برای خوشه‌بندی داده‌های مستخرج از یک جامعهٔ فازی پیشنهاد می‌شود که تعمیمی از الگوریتم k-میانگین در حالت کلاسیک است. برای محاسبه وزن مشاهدات در این الگوریتم از توابع عضویت مفاهیم فازی/نادقیق که تعریف کنندهٔ جامعهٔ فازی موردنظر هستند، استفاده می‌شود. بدین ترتیب وزن مشاهدات فقط یک مرتبه در ابتدای فرآیند خوشه‌بندی بدون نیاز به بروزرسانی در هر مرحله از تکرار الگوریتم، محاسبه می‌شوند.

ساختار مقاله به شرح زیر است. در بخش ۲ تعریف جامعهٔ فازی و میانگین داده‌های مستخرج از یک جامعهٔ فازی مرور شده است. الگوریتم پیشنهادی به منظور خوشه‌بندی داده‌های مستخرج از یک جامعهٔ فازی در بخش ۳ معرفی شده است. در پایان با ارائه یک مثال کاربردی در بخش ۴ نحوهٔ عملکرد روش پیشنهادی بررسی شده است.

۲. جامعهٔ فازی

در بررسی‌های آماری، با دو مجموعهٔ اساسی مجموعه (جامعهٔ) هدف و مجموعهٔ نمونه روبرو هستیم. در آمار کلاسیک این فرض وجود دارد که هر دوی این مجموعه‌ها (یعنی جامعه و نمونه) مشخص و دقیق هستند، یعنی عضویت یا عدم عضویت هر فرد در جامعه یا نمونه، واضح و روشن است. ولی در عمل با دو وضعیت زیر نیز روبرو هستیم [۲]:

(۱) مواردی که اعضای نمونه، مشخص و دقیق نیستند، یا اینکه مشاهدات نمونه، دقیق گزارش نمی‌شوند. به چنین نمونه‌هایی نمونهٔ فازی یا نمونهٔ نادقیق می‌گوییم. برای مثال هنگامی که می‌خواهیم شوری خاک را در نقاط مختلف یک منطقه اندازه‌گیری کنیم، ممکن است نتایج حاصل همراه با ابهام باشد [۱۳]. یا هنگامی که می‌خواهیم میزان درد را در نمونه‌ای

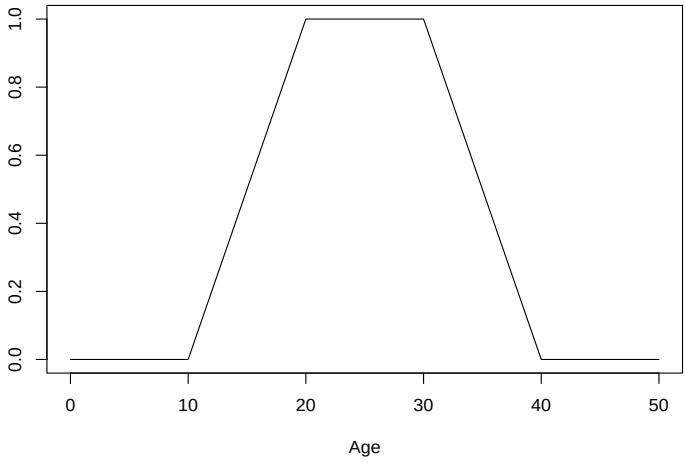
از افراد بیمار بررسی کنیم، پاسخ‌های آن‌ها معمولا به صورت «درد زیاد»، «درد متوسط»، «درد قابل تحمل»، «درد کم» و ... است [۱۷].

(۲) مواردی که جامعهٔ هدف به طور دقیق و واضح تعریف نشده است و با یک جامعهٔ هدف نادقیق روبرو هستیم. برای مثال اگر بخواهیم دربارهٔ میزان رطوبت هوا در روزهای بارانی بررسی انجام دهیم، جامعهٔ هدف «روزهای بارانی» است. در اینجا بدیهی است که روزهای مختلف با مقادیر مختلف بارندگی به یک اندازه مدنظر نیستند بلکه روزهای با بارندگی بیشتر، عضویت بالاتری در مجموعهٔ «روزهای بارانی» دارند و برعکس. به عنوان مثالی دیگر، اگر هدف بررسی وضع درآمد «جوانان کرمانی» باشد، آنگاه جامعهٔ هدف یعنی «جوانان کرمانی» یک مجموعهٔ دقیق و خوش تعریف نیست [۲].

تعریف ۱.۲. مجموعه‌ای فازی از افراد یا اشیاء را که یک یا چند ویژگی آن‌ها مورد بررسی است، یک جامعهٔ فازی می‌نامیم. هر جامعهٔ فازی به‌وسیلهٔ تابع عضویت مربوطه، به طور یکتا، مشخص می‌شود [۲].

مثال ۱.۲. فرض کنید می‌خواهیم متوسط درآمد کارمندان جوان شیرازی را محاسبه کنیم. از آنجا که «جوان» یک مفهوم فازی و نادقیق است لذا اعضای جامعهٔ جوانان شیرازی به طور دقیق مشخص نیست. در واقع هر شیرازی با درجه‌ای بین صفر و یک متعلق به مجموعهٔ فازی جوانان شیرازی است. در این حالت، ابتدا جامعهٔ فازی جوانان شیرازی را به کمک یک تابع عضویت، مثلا تابع عضویت زیر تعیین و تعریف می‌کنیم (شکل ۱ را ببینید).

$$\mu_{\text{جوانی}}(x) = \begin{cases} \dfrac{x - ۱۰}{۱۰} & ۱۰ \le x < ۲۰ \\ ۱ & ۲۰ \le x < ۳۰ \\ \dfrac{۴۰ - x}{۱۰} & ۳۰ \le x < ۴۰ \\ ۰ & \text{سایر نقاط} \end{cases}$$



شکل ۱: نمودار تابع عضویت «جوانی» برحسب سن

برای مثال، یک فرد که سن ۳۶ سال دارد، به اندازهٔ $\mu_{\text{جوانی}}(۳۶) = ۰/۴$ در جامعهٔ فازی جوانان شیرازی عضو است.

پس از اخذ نمونهٔ تصادفی از افراد این شهر (به حجم n)، اطلاعات مربوط به نمونه را به صورت $(x_۱, \mu_۱), ..., (x_n, \mu_n)$ ثبت می‌کنیم که در آن x_i درآمد فرد iام و μ_i درجهٔ عضویت او در مجموعهٔ فازی جوانان شیرازی است. در این مقاله چنین داده‌هایی را «داده‌های مستخرج از یک جامعهٔ فازی» می‌نامیم.

تعریف ۲.۲. فرض کنید داده‌های $(x_۱, \mu_۱), \ldots, (x_n, \mu_n)$ براساس نمونه‌ای تصادفی از جامعه‌ای فازی ثبت شده‌اند که در آن‌ها x_i دادهٔ مربوط به فرد iام و μ_i میزان عضویت فرد iام به جامعهٔ فازی موردنظر است. براساس داده‌های $(x_۱, \mu_۱), \ldots, (x_n, \mu_n)$ مقدار میانگین داده‌های مستخرج از جامعهٔ فازی با رابطهٔ

$$\bar{x} = \frac{\sum_{i=۱}^{n} x_i \mu_i}{\sum_{i=۱}^{n} \mu_i} \qquad (۱.۲)$$

تعریف می‌شود.

ملاحظه ۱.۲. اگر به‌جای تابع عضویت مجموعه فازی در رابطه (۱.۲) از تابع نشانگر جامعه (یعنی، I_Ω) استفاده شود، آنگاه $\sum_{i=۱}^{n} \mu_i = n$ و در نتیجه میانگین بیان شده در رابطه (۱.۲) به میانگین حسابی داده‌های $x_۱, ..., x_n$ تبدیل می‌شود. برای جزئیات بیشتر و آشنایی با سایر شاخص‌های آماری مربوط به جامعهٔ فازی به مرجع [۲] مراجعه کنید.

در بخش بعد، به ارائه روشی به منظور خوشه‌بندی داده‌های مستخرج از یک جامعهٔ فازی می‌پردازیم.

۳. خوشه‌بندی داده‌های مستخرج از جامعهٔ فازی

در این بخش به معرفی الگوریتم k-میانگین وزن‌دار فازی[1] (FWk-میانگین) که برای نخستین بار در این مقاله با هدف خوشه‌بندی داده‌های مستخرج از یک جامعهٔ فازی پیشنهاد شده است، می‌پردازیم. این الگوریتم درواقع نوع وزنی الگوریتم k-میانگین است، که در آن، مشاهدات به میزان عضویت در جامعهٔ فازی، وزن دارند. ماتریس مشاهدات $\mathbf{X}$ متشکل از n مشاهده و m متغیر را به صورت زیر درنظر بگیرید:

$$\mathbf{X} = \begin{bmatrix} x_{۱۱} & x_{۱۲} & \dots & x_{۱m} \\ x_{۲۱} & x_{۲۲} & \dots & x_{۲m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n۱} & x_{n۲} & \dots & x_{nm} \end{bmatrix} = [x_{ij}]_{n\times m}$$

ماتریس $\mathbf{X}$ را می‌توان برحسب ستون‌ها و همچنین سطرهایش به دو شکل

$$\mathbf{X} = \left[\mathbf{x}_{(۱)}, \mathbf{x}_{(۲)}, \dots, \mathbf{x}_{(m)}\right] = \left[\mathbf{x}_۱^T, \mathbf{x}_۲^T, \dots, \mathbf{x}_n^T\right]^T$$

نیز نشان داد که در آن $\mathbf{x}_{(j)} = \left[x_{۱j}, \dots, x_{nj}\right]^T$ بردار متغیر jام و $\mathbf{x}_i = \left[x_{i۱}, \dots, x_{im}\right]$ بردار مشاهده iام از ماتریس $\mathbf{X}$ است.

هدف خوشه‌بندی سطرهای ماتریس $\mathbf{X}$ (مشاهدات) در k خوشه است. از طرفی فرض کنید برخی از متغیرهای مورد بررسی در جامعه بصورت نادقیق و فازی باشند (مانند تورم کم، قد بلند، وزن مناسب و ...). در چنین حالتی ابتدا برای jامین متغیر تابع عضویت μ_j توسط متخصصین خبره درنظر گرفته می‌شود که مبین jامین ویژگی نادقیق جامعهٔ فازی است. از حاصلضرب درجات عضویت $\mu_j(\mathbf{x}_{(j)})$ به‌ازای $j = ۱, ۲, \dots, m$، که هریک برداری به طول n هستند، می‌توان بردار درجات تعلق هریک از n مشاهده را به جامعهٔ فازی بصورت زیر محاسبه کرد:

$$\boldsymbol{\mu} = (\mu_۱, \dots, \mu_n)^T = \prod_{j=۱}^{m} \mu_j(\mathbf{x}_{(j)}) = \mu_۱(\mathbf{x}_{(۱)}) \times \dots \times \mu_m(\mathbf{x}_{(m)}) \quad (۱.۳)$$

[1] Fuzzy Weighted k-means algorithm

ماتریس فاصله اقلیدسی[1] در اینجا بصورت

$$\mathbf{D} = \|\boldsymbol{\mu} \times (\mathbf{X} - \mathbf{Z})\|_۲ = [d_{il}]_{n\times k} \qquad (۲.۳)$$

تعریف می‌شود که در آن $\|.\|$ نماد $L_۲$−نُرم یک ماتریس است. همچنین $\mathbf{Z} = [z_{lj}]_{k\times m}$ به ماتریس مراکز خوشه‌ها و برای $i = ۱, ۲, \ldots, n$ و $l = ۱, ۲, \ldots, k$

$$d_{il} = \sqrt{\mu_i \times (\mathbf{x}_i - \mathbf{z}_l)^۲} = \sqrt{\mu_i \times \left((x_{i۱} - z_{l۱})^۲ + \cdots + (x_{im} - z_{lm})^۲\right)}$$

به فاصله اقلیدسی iامین مشاهده از مرکز lامین خوشه اشاره دارند.

تابع هدف در این روش به صورت

$$T(\mathbf{U}, \mathbf{Z}, \boldsymbol{\mu}) = \sum_{l=۱}^{k}\sum_{i=۱}^{n} u_{il}\, d_{il} \qquad (۳.۳)$$

تعریف می‌شود که در آن ، $\mathbf{U} = [u_{il}]_{n\times k}$ ماتریس تخصیص مشاهدات به k خوشه با درایه‌های

$$u_{il} = \begin{cases} ۱ & \mathbf{x}_i \in c_l \\ ۰ & \mathbf{x}_i \notin c_l \end{cases}$$

است که در آن c_l خوشه lام است. به عبارت دیگر u_{il} تابع نشانگر وجود iامین مشاهده در lامین خوشه به‌ازای $i = ۱, ۲, \ldots, n$ و $l = ۱, ۲, \ldots, k$ است بطوریکه برای هر مشاهده (مشاهدهٔ iام) رابطهٔ $\sum_{l=۱}^{k} u_{il} = ۱$ برقرار است. همچنین $\mathbf{z}_l$ مرکز خوشهٔ lام، که درواقع همان ترانهادهٔ سطر lام ماتریس $\mathbf{Z}$ است، از رابطهٔ زیر محاسبه می‌شود:

$$\mathbf{z}_l = \frac{\sum_{i=۱}^{n_l}\sum_{j=۱}^{m} \mu_i\, x_{ij}}{\sum_{i=۱}^{n_l} \mu_i}; l = ۱, ۲, \ldots k \qquad (۴.۳)$$

بطوریکه n_l تعداد مشاهدات در خوشهٔ lام است.

الگوریتم FWk-میانگین به منظور تخصیص n مشاهدهٔ $\left[\mathbf{x}_۱^T, \mathbf{x}_۲^T, \ldots, \mathbf{x}_n^T\right]^T$ به k خوشه و با هدف کمینه‌سازی تابع هدف (۸.۳) بصورت زیر معرفی می‌شود:

(۱) نقاط به طور تصادفی به k خوشه تفکیک می‌شوند.

(۲) مرکز هر یک از k خوشه براساس رابطهٔ (۴.۳) محاسبه می‌شود.

(۳) ماتریس فاصله با استفاده از رابطه (۲.۳) محاسبه شده و هر مشاهده به نزدیک‌ترین خوشه تخصیص داده می‌شود.

[1] Euclidean distance matrix

(۴) گام‌های ۲ و ۳ تا رسیدن به همگرایی تکرار می‌شوند بطوریکه تخصیص مشاهدات به خوشه‌ها دیگر تغییر نکند.

قضیه ۱.۳. *الگوریتم FWkـ میانگین تعمیمی از الگوریتم kـ میانگین است.*

اثبات. در الگوریتم FWkـ میانگین هدف مینیمم‌سازی تابع هدف زیر نسبت به $\mathbf{z}_l$ است،

$$T(\mathbf{U},\mathbf{Z},\boldsymbol{\mu})=\sum_{i=1}^{n}u_{il}\,d_{il}=\sum_{i=1}^{n}u_{il}\,\sqrt{\mu_i\times(\mathbf{x}_i-\mathbf{z}_l)^2} \tag{۵.۳}$$

برای مینیمم کردن تابع هدف (۵.۳) نسبت به $\mathbf{z}_l$، مشتق آن را نسبت به $\mathbf{z}_l$ محاسبه کرده و مساوی صفر قرار می‌دهیم:

$$\frac{\partial}{\partial\,\mathbf{z}_l}T(\mathbf{U},\mathbf{Z},\boldsymbol{\mu})=\sum_{i=1}^{n}u_{il}\,\mu_i\times(\mathbf{x}_i-\mathbf{z}_l)=0 \tag{۶.۳}$$

رابطهٔ (۶.۳) معادل است با:

$$\sum_{i=1}^{n}u_{il}\,\mu_i\,\mathbf{x}_i=\sum_{i=1}^{n}u_{il}\,\mu_i\,\mathbf{z}_l \tag{۷.۳}$$

است. با توجه به رابطهٔ (۷.۳) مقدار $\mathbf{z}_l$ که مینیمم کنندهٔ تابع هدف $T(\mathbf{U},\mathbf{Z},\boldsymbol{\mu})$ است به‌صورت،

$$\mathbf{z}_l=\frac{\sum_{i=1}^{n}u_{il}\,\mu_i\,x_i}{\sum_{i=1}^{n}u_{il}\,\mu_i};l=1,2,\dots k \tag{۸.۳}$$

نتیجه می‌شود. با فرض اینکه تعداد مشاهدات در خوشهٔ lام، n_l باشد، در این حالت برای مقادیر $i=1,\dots,n_l$ داریم $u_{il}=1$. بنابراین رابطهٔ (۸.۳) به‌صورت زیر ساده می‌شود:

$$\mathbf{z}_l=\frac{\sum_{i=1}^{n_l}\mu_i\,x_i}{\sum_{i=1}^{n_l}\mu_i};l=1,2,\dots k \tag{۹.۳}$$

با توجه به اینکه هر $\mathbf{x}_i$ یک بردار m بعدی است، می‌توان رابطهٔ (۹.۳) به شکل زیر بازنویسی نمود:

$$\mathbf{z}_l=\frac{\sum_{i=1}^{n_l}\sum_{j=1}^{m}\mu_i\,x_{ij}}{\sum_{i=1}^{n_l}\mu_i};l=1,2,\dots k \tag{۱۰.۳}$$

که همان رابطهٔ (۴.۳) برای بدست آوردن مرکز خوشهٔ lام است. در حالتی که جامعهٔ هدف یک جامعهٔ کلاسیک باشد درجات تعلق هر یک از n مشاهده به جامعه و مقدار μ_iها در معادلهٔ (۱۰.۳) برابر با عدد یک است لذا طبق ملاحظه ۱.۲ میانگین هر خوشه تبدیل به میانگین حسابی مشاهدات همان خوشه شده و در نتیجه الگوریتم پیشنهادی در این مقاله به الگوریتم k-میانگین

تبدیل می‌شود. بنابراین اثبات شد که الگوریتم FWk-میانگین؛ که به منظور خوشه‌بندی داده‌های مستخرج از یک جامعهٔ فازی معرفی شده است؛ تعمیمی از الگوریتم k-میانگین در حالت کلاسیک است. به عبارت دیگر الگوریتم k ـ میانگین حالت خاصی از الگوریتم FWk ـ میانگین است.

توجه به نکات زیر در اینجا ضروری است،

۱) معادلهٔ بیان شده در رابطه (۴.۳) در واقع تعمیمی از معادلهٔ (۱.۲) برای حالت چندمتغیره است.

۲) الگوریتم FWk-میانگین از سه دیدگاه با الگوریتم k-میانگین وزنی[1] هانگ و همکاران [۱۴] متفاوت است: (الف) محاسبه وزن‌ها در الگوریتم k-میانگین وزنی مبتنی بر فاصله است اما وزن‌ها در الگوریتم FWk-میانگین براساس تابع عضویت جامعهٔ فازی محاسبه می‌شوند، (ب) وزن‌دهی در الگوریتم k-میانگین وزنی برای متغیرها صورت می‌گیرد و این در حالی است که در الگوریتم FWk-میانگین وزن‌دهی برای مشاهدات انجام می‌شود، (ج) وزن‌های مشاهدات در الگوریتم FWk-میانگین براساس تابع عضویت متغیرها، که مفاهیمی ذهنی و نادقیق هستند، تعیین می‌شوند و لذا در هر دور از الگوریتم ثابت هستند. اما در روش k-میانگین وزنی، وزن متغیرها تابعی از فاصله مشاهده تا مرکز خوشه تخصیص یافته هستند و لذا در هر دور از الگوریتم تغییر می‌کنند.

۴. مثال کاربردی

ممکن است جامعهٔ هدف در عمل نادقیق باشد و هدف از خوشه‌بندی در چنین مواقعی، خوشه‌بندی داده‌های مستخرج از یک جامعهٔ فازی خواهد بود. به عنوان نمونه، در این بخش قصد داریم تا داده‌های مستخرج از جامعهٔ جوانان باهوش را با دو دیدگاه کلاسیک و فازی خوشه‌بندی کنیم.

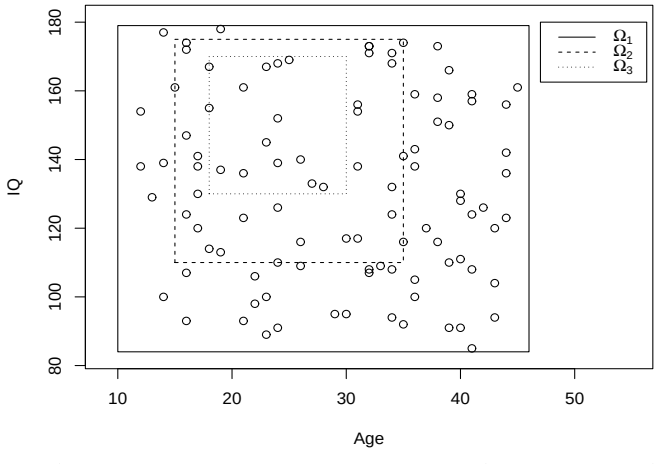
۱.۴. خوشه‌بندی به روش کلاسیک با الگوریتم k-میانگین. اگر خود را در صورت‌بندی مسئله محدود به تعریف جامعهٔ معمولی/دقیق کنیم، آنگاه با توجه به اینکه مفاهیم جوانی و هوش هر دو مبهم و نادقیق هستند تعاریف مختلفی از جامعهٔ جوانان باهوش را می‌توان ارائه داد که در ادامه به برخی از آنها اشاره می‌کنیم:

جامعهٔ Ω_1: متشکل از افراد با سن بین ۱۰ تا ۴۵ سال و ضریب هوشی بین ۸۵ تا ۱۸۰،

[1] weighted k-means algorithm

جامعهٔ $\Omega_۲$: متشکل از افراد با سن بین ۱۵ تا ۳۵ سال و ضریب هوشی بین ۱۱۰ تا ۱۷۵،
جامعهٔ $\Omega_۳$: متشکل از افراد با سن بین ۱۸ تا ۳۰ سال و ضریب هوشی بین ۱۳۰ تا ۱۷۰،

در این مثال از جامعهٔ $\Omega_۱$، ۱۰۰ داده به طور تصادفی با استفاده از نرم افزار R شبیه‌سازی شده و مجموعه این داده‌ها در جدول ۱ به طور خلاصه آورده شده است. ستون ۱ در جدول ۱ نشان‌دهندهٔ شماره مشاهدات، ستون‌های ۲ و ۳ به ترتیب بیانگر سن و ضریب هوشی مربوط به افراد و ستون‌های ۴ تا ۶ به ترتیب به وجود و عدم وجود هر یک از مشاهدات در جوامع کلاسیک تعریف شده در بالا اشاره دارند. در این مثال تعداد مشاهدات $n = ۱۰۰$، تعداد متغیرها $m = ۲$ و تعداد خوشه‌ها $k = ۲$ است. با توجه به تعاریف جوامع دقیق $\Omega_۱$، $\Omega_۲$ و $\Omega_۳$، بدیهی است که برخی از داده‌های مستخرج از جامعهٔ $\Omega_۱$ متعلق به $\Omega_۲$ و $\Omega_۳$ نیستند و لذا $n_{\Omega_۱} = ۱۰۰$، $n_{\Omega_۲} = ۴۴$ و $n_{\Omega_۳} = ۱۴$. نمودار پراکنش ۱۰۰ دادهٔ شبیه‌سازی شده از سه جامعهٔ $\Omega_۱$، $\Omega_۲$ و $\Omega_۳$ در شکل ۲ نشان داده شده است.

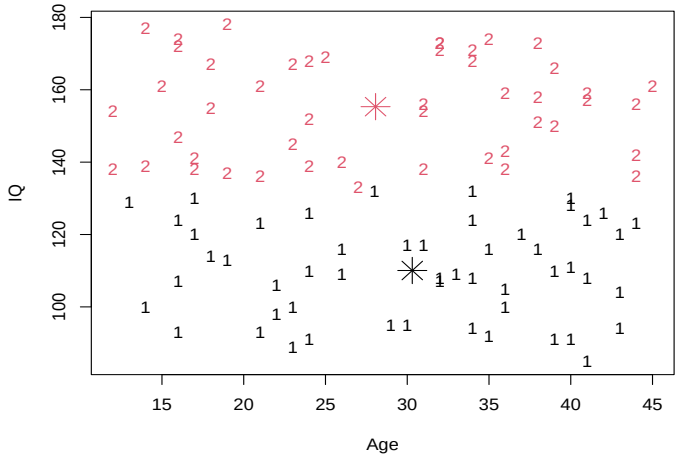


شکل ۲: نمودار پراکنش داده‌های شبیه‌سازی شده از سه جامعهٔ کلاسیک

تابع نشانگر $I_{\Omega r}$ در جدول ۱ نشان‌دهندهٔ وجود یا عدم وجود مشاهده مورد نظر در جامعهٔ کلاسیک Ω_r برای $r = ۱, ۲, ۳$ است. نتیجه خوشه‌بندی الگوریتم k-میانگین بر روی داده‌های مستخرج از جامعهٔ $\Omega_۱$ در دو خوشه در شکل ۳ نشان داده شده است. در دور آخر الگوریتم k-میانگین مراکز خوشه‌های ۱ و ۲ به‌ترتیب (۳۰٫۳۰, ۱۱۰٫۰۵) و (۲۸٫۰۶, ۱۵۵٫۳۳) و همچنین تعداد مشاهدات خوشه ۱ و خوشه ۲ در این دور از الگوریتم به‌ترتیب ۵۲ و ۴۸ است (جدول ۳ را ببینید).

جدول ۱: داده‌های مستخرج از سه جامعهٔ Ω_1، Ω_2 و Ω_3 و درجات تعلق آن‌ها

شماره مشاهدات (i)	سن (x_{i1})	ضریب هوشی (x_{i2})	$I_{\Omega 1}$	$I_{\Omega 2}$	$I_{\Omega 3}$
۱	۴۰	۱۱۱	۱	۰	۰
۲	۲۴	۱۲۶	۱	۱	۰
⋮	⋮	⋮	⋮	⋮	⋮
۲۹	۳۲	۱۷۳	۱	۱	۰
⋮	⋮	⋮	⋮	⋮	⋮
۴۶	۲۴	۱۳۹	۱	۱	۱
⋮	⋮	⋮	⋮	⋮	⋮
۹۹	۴۱	۱۵۷	۱	۰	۰
۱۰۰	۱۶	۱۰۷	۱	۰	۰



شکل ۳: خوشه‌بندی ۱۰۰ داده مستخرج از جامعهٔ Ω_1 با الگوریتم k-میانگین

۲.۴. خوشه‌بندی داده‌های مستخرج از جامعهٔ فازی جوانان باهوش با الگوریتم FWk-میانگین.

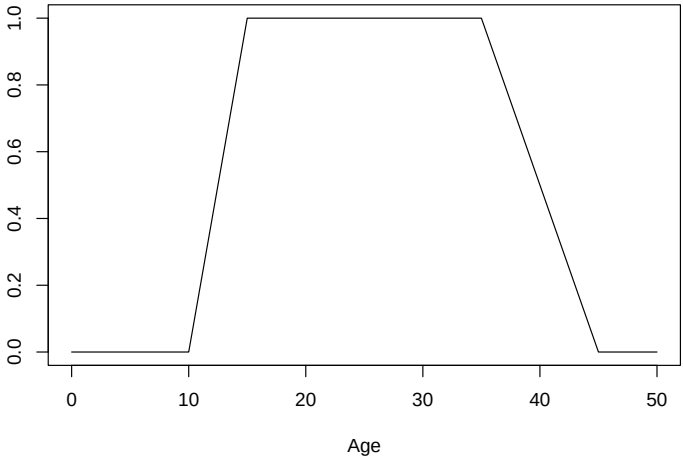
از آنجایی که مفاهیم جوانی و هوش هر دو مبهم و نادقیق هستند، بنابراین بهتر است در ابتدا برای هر یک از متغیرهای فازی جوانی و هوش یک تابع عضویت توسط متخصصین خبره تعریف گردد. با این کار دیگر نیازی به تعاریف مختلف از جامعهٔ موردنظر و تحلیل هر کدام به‌طور جداگانه نیست و همچنین جامعهٔ جوانان باهوش که متشکل از مفاهیم مبهم است بصورت نادقیق و فازی تعریف می‌شود. توابع عضویت دو مفهوم جوانی و هوش به ترتیب

بصورت زیر تعریف می‌شوند:

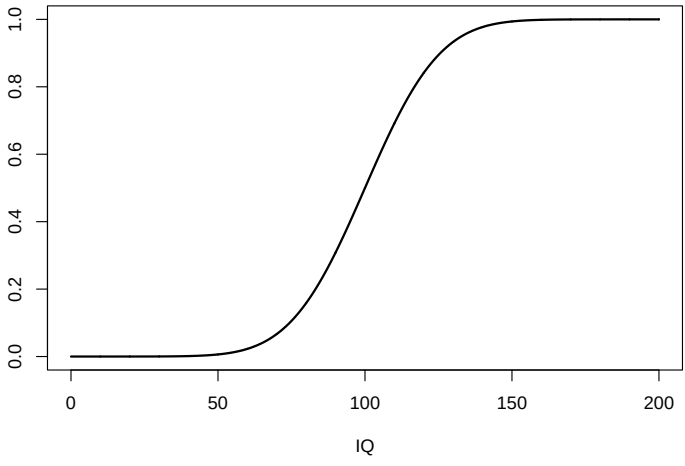
$$\mu_{۱}(x) = \begin{cases} \frac{x - ۱۰}{۱۰} & ۱۰ \leq x < ۱۵ \\ ۱ & ۱۵ \leq x < ۳۵ \\ \frac{۴۵ - x}{۱۰} & ۳۵ \leq x < ۴۵ \\ ۰ & \text{سایر نقاط} \end{cases}$$

$$\mu_{۲}(x) = \Phi\left(\frac{x - ۱۰۰}{۲۰}\right)$$

که در آن $\Phi(.)$ تابع توزیع تجمعی نرمال استاندارد است [۲].
همچنین نمودارهای مربوط به دو تابع عضویت جوانی و هوش به ترتیب در شکل ۴ و ۵ نشان داده شده است.



شکل ۴: نمودار تابع عضویت «جوانی» برحسب سن

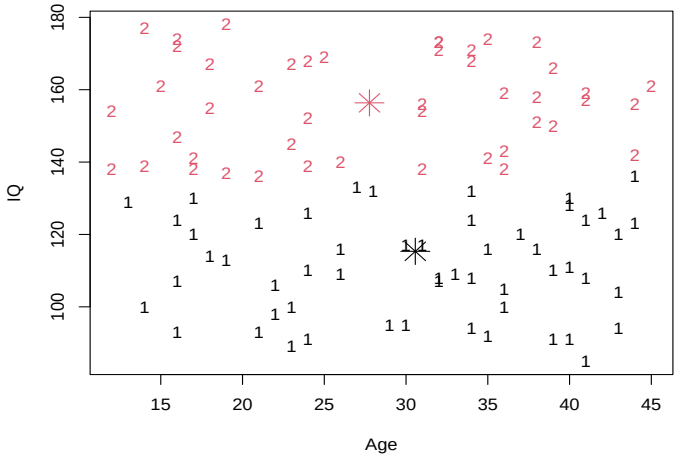


شکل ۵: نمودار تابع عضویت «هوش» برحسب IQ

ستون ۴، ۵ و ۶ در جدول ۲ به‌ترتیب بیانگر درجهٔ عضویت هر مشاهده در جامعهٔ فازی جوانان، درجهٔ عضویت هر مشاهده در جامعهٔ فازی افراد باهوش و درجهٔ عضویت هر مشاهده در جامعهٔ فازی جوانان باهوش می‌باشند. نتیجه خوشه‌بندی الگوریتم پیشنهادی بر روی داده‌های تولید شده از جامعهٔ کلاسیک اول در شکل ۶ نشان داده شده است. در دور آخر الگوریتم FWk-میانگین مراکز خوشه‌های ۱ و ۲ به‌ترتیب (۳۰٫۵۶, ۱۱۵٫۳۳) و (۲۷٫۷۶, ۱۵۶٫۳۷) و همچنین تعداد مشاهدات خوشه ۱ و خوشه ۲ در این دور از الگوریتم به‌ترتیب ۵۴ و ۴۶ است (جدول ۳ را ببینید). در جدول ۳ مراکز خوشه‌ها و تعداد مشاهدات هر یک از خوشه‌ها در هر دور از تکرار الگوریتم‌های k-میانگین و FWk-میانگین آورده شده است.

جدول ۲: درجات عضویت مشاهدات به جامعهٔ فازی جوانان باهوش

شماره مشاهدات (i)	سن (x_{i1})	ضریب هوشی (x_{i2})	$\mu_1(x_{i1})$	$\mu_2(x_{i2})$	μ_i
۱	۴۰	۱۱۱	۰٫۵	۰٫۷۰۸	۰٫۳۵۴
۲	۲۴	۱۲۶	۱	۰٫۹۰۳	۰٫۹۰۳
⋮	⋮	⋮	⋮	⋮	⋮
۲۹	۳۲	۱۷۳	۱	۰٫۹۹۹	۰٫۹۹۹
⋮	⋮	⋮	⋮	⋮	⋮
۴۶	۲۴	۱۳۹	۱	۰٫۹۷۴	۰٫۹۷۴
⋮	⋮	⋮	⋮	⋮	⋮
۹۹	۴۱	۱۵۷	۰٫۴	۰٫۹۹۷	۰٫۳۹۸
۱۰۰	۱۶	۱۰۷	۰٫۶	۰٫۶۳۶	۰٫۳۸۱



شکل ۶: خوشه‌بندی داده‌های مستخرج از جامعهٔ فازی جوانان باهوش با الگوریتم FWk-میانگین

جدول ۳: نتایج خوشه‌بندی ۱۰۰ مشاهده با الگوریتم‌های k-میانگین و FWk-میانگین

شماره دور الگوریتم	مرکز خوشه ۱		مرکز خوشه ۲	
	k-میانگین	FWk-میانگین	k-میانگین	FWk-میانگین
۱	(۲۹٫۶۴, ۱۲۷٫۸۲)	(۲۹٫۲۹, ۱۳۶٫۴۵)	(۲۸٫۸۲, ۱۳۵٫۷۶)	(۲۸٫۷۴, ۱۳۹٫۷۱)
۲	(۳۰٫۳۵, ۱۰۹٫۶۲)	(۳۰٫۴۷, ۱۱۶٫۹۶)	(۲۸٫۰۶, ۱۵۴٫۸۵)	(۲۷٫۶۵, ۱۵۷٫۶۶)
۳	(۳۰٫۳۰, ۱۱۰٫۰۵)	(۲۷٫۹۰, ۱۵۶٫۸۱)	(۲۸٫۰۶, ۱۵۵٫۳۳)	(۳۰٫۳۱, ۱۱۵٫۸۷)
۴	–	(۳۰٫۵۶, ۱۱۵٫۳۳)	–	(۲۷٫۷۶, ۱۵۶٫۳۷)
	تعداد مشاهدات خوشه ۱		تعداد مشاهدات خوشه ۲	
	k-میانگین	FWk-میانگین	k-میانگین	FWk-میانگین
۱	۵۱	۵۷	۴۹	۴۳
۲	۵۲	۵۵	۴۸	۴۵
۳	۵۲	۵۴	۴۸	۴۶
۴	–	۵۴	–	۴۶

در مثال کاربردی، ۱۰۰ داده به‌صورت تصادفی با استفاده از توابع مربوطه در نرم افزار R از جامعهٔ Ω_1 شبیه‌سازی شد. سپس، الگوریتم‌های خوشه‌بندی k ـ میانگین و FWk ـ میانگین به‌منظور تقسیم‌بندی داده‌های مستخرج از جامعهٔ Ω_1 در دو خوشه به کار گرفته شدند. همانطور که در جدول ۳ ملاحظه می‌شود، الگوریتم FWk ـ میانگین در مقایسه با الگوریتم k ـ میانگین در تعداد دور بیشتری همگرا شد.

هدف از ارائه این مثال، نشان دادن مزایای فرمول‌بندی مسئله مبتنی بر مفهوم جامعهٔ فازی به‌منظور خوشه‌بندی مناسب‌تر و عادلانه‌تر است. همانطور که ملاحظه شد، در خوشه‌بندی داده‌های مربوط به جامعه «جوانان باهوش» از دیدگاه کلاسیک، چالش اصلی در تعیین معیار عضویت افراد است؛ به گونه‌ای که مشخص شود کدام افراد با چه سن و ضریب هوشی باید به عنوان اعضای این جامعه در تحلیل خوشه‌بندی لحاظ شوند. از آنجا که تعاریف مختلف و متنوعی برای این جامعه وجود دارد، حل مسئله از دیدگاه کلاسیک، همانند آنچه در شکل ۲ نشان داده شده، در برخی موارد منجر به حذف تعدادی مشاهدات و از دست رفتن بخشی از اطلاعات می‌شود. در مقابل، رویکرد فازی امکان مدل‌سازی جامعهٔ «جوانان باهوش» را از طریق تعریف توابع عضویت مناسب برای متغیرهای سن و ضریب هوشی فراهم می‌آورد؛ به‌طوری که تمامی اعضای جامعه در تحلیل خوشه‌بندی مشارکت داده می‌شوند. اگرچه انتخاب توابع عضویت مناسب ممکن است تا حدی سلیقه‌ای باشد، اما این رویکرد منجر به یکپارچگی حل مسائل متعدد در قالب یک مسئله واحد و تسهیل فرآیند حل می‌گردد. علاوه بر این، میزان درجه عضویت و اهمیت هر مشاهده در جامعهٔ فازی، به صورت مؤثری در محاسبات تحلیل خوشه‌ای تأثیرگذار خواهد بود.

۵. نتیجه‌گیری

در این مقاله پس از مروری بر مفهوم «جامعهٔ فازی»، به معرفی تعمیمی خاص از الگوریتم k-میانگین برای خوشه‌بندی داده‌های مستخرج از یک جامعهٔ فازی پرداخته شد. در بسیاری از مسائل کاربردی داده‌های حاصل نتیجهٔ سرشماری/نمونه‌گیری از یک جامعهٔ فازی است مانند میزان فشارخون در جامعهٔ فازی افراد چاق، داده‌های مربوط به قندخون در جامعهٔ فازی افراد مسن، مقدار مصرف انرژی در جامعهٔ فازی خانوارهای پرمصرف و غیره. در مجموعه داده‌های مستخرج از یک جامعهٔ فازی بعضی از متغیرها، که در واقع همان متغیرهای تعریف کنندهٔ جامعهٔ فازی موردنظر می‌باشند، از مرز مشخصی در مقداردهی برخوردار نیستند. بنابراین برای خوشه‌بندی چنین داده‌هایی ممکن است شخص تحلیل‌گر با توجه به تعریف جامعهٔ فازی موردنظر و براساس سلیقهٔ شخصی فقط مشاهدات خاصی را برای خوشه‌بندی مدنظر قرار دهد که این امر منجر به نادیده گرفتن تعدادی از مشاهدات و در نتیجه از دست دادن بخشی از اطلاعات جامعه می‌شود. از طرفی برای خوشه‌بندی داده‌های مستخرج از یک جامعهٔ فازی، نادیده گرفتن برخی از مشاهدات و یا یکسان درنظر گرفتن درجهٔ اهمیت تمامی مشاهدات در خوشه‌بندی غیرمنصفانه و مسئله‌ساز است. در این مقاله، روشی جدید برای خوشه‌بندی داده‌های مستخرج

از یک جامعهٔ فازی بانام FWk-میانگین پیشنهاد شده است که در حقیقت تعمیمی از الگوریتم k-میانگین به‌حساب می‌آید.

برای آینده تحقیق می‌توان موارد زیر را انجام داد: ۱) الگوریتم پیشنهادی را برای خوشه‌بندی فازی داده‌های مستخرج از یک جامعهٔ فازی تعمیم داد. براساس روش پیشنهادی در این مقاله، تخصیص مشاهدات به خوشه‌ها به صورت قطعی و با مقادیر صفر و یک انجام می‌شود؛ به‌گونه‌ای که هر مشاهده تنها به یک خوشه تعلق دارد. با تعمیم الگوریتم FWk‌ـمیانگین به الگوریتم خوشه‌بندی فازی، امکان تخصیص مشاهدات به صورت نرم و با درجات عضویت بین صفر و یک به هر یک از خوشه‌ها فراهم می‌گردد، به طوری که هر مشاهده می‌تواند به طور هم‌زمان به چندین خوشه با درجات عضویت متفاوت تعلق داشته باشد. ۲) الگوریتم پیشنهادی در این مقاله را می‌توان به گونه‌ای توسعه داد که وزن مشاهدات تابعی از درجات عضویت مشاهدات در جامعهٔ فازی و فاصله مشاهدات از مرکز خوشه تخصیص یافته باشد و وزن‌ها در طول فرآیند خوشه‌بندی به‌روزرسانی و بهینه شوند.

مراجع

[۱] اسماعیلی، م. (۱۴۰۰) مفاهیم و تکنیک‌های داده‌کاوی، انتشارات نیاز دانش، ویراست سوم، چاپ پنجم.

[۲] پرچمی، ع، و طاهری، م. (۱۳۹۸) آمار توصیفی براساس داده‌های دقیق از یک جامعهٔ فازی. ریاضی و جامعه، شماره ۳، صص. ۴۹ تا ۵۹.

[۳] تیمورپور، ب، و نجفی، ح. (۱۳۹۴) داده‌کاوی با R به همراه متن‌کاوی و تحلیل شبکه‌های اجتماعی، انتشارات مرکز تحقیقات و توسعه سازمان اتکا، چاپ اول.

[۴] مشکانی، ع، و ناظمی، ع. (۱۳۸۸) مقدمه‌ای بر داده‌کاوی، انتشارات دانشگاه آزاد اسلامی واحد نیشابور، چاپ اول.

[۵] وزان، م. (۱۴۰۰) یادگیری ماشین و علم داده: مبانی، مفاهیم، الگوریتم‌ها و ابزارها، انتشارات میعاد اندیشه، چاپ اول.

[6] B.S. Everitt, and G. Dunn. (1991). Applied multivariate data analysis. London: Edward Arnold.

[7] C.C. Aggarwal, C.K. Reddy. (2014) Data clustering algorithms and applications, Data Mining and Knowledge Discovery Series.

[8] D.S. Modha and W.S. Spangler. (2003) Feature Weighting in k-Means Clustering, Machine Learning, 52, 217-237.

[9] D. Zhang, L. Wang, H. Wang, and C. Shi. (2020) Weighted fuzzy clustering with sample and feature weighting for high-dimensional data. Knowledge-Based Systems, 196, 105789.

[10] H.-J. Zimmermann. (2001) Fuzzy set theory and its applications, 418, Springer.

[11] I. J. Gordon. (1980) Statistical methods in social science, 2, New York: Wiley.

[12] J. C. Bezdek. (1981) Pattern recognition with fuzzy objective function algorithms, New York: Plenum Press.

[13] J. Mohammadi and S. M. Taheri. (2004) Pedomodels fitting with fuzzy least squares regression, Iranian Journal of Fuzzy Systems, 1, 45-61.

[14] J.Z. Huang, M.K.NG, H.Rong, Z.Li. (2008) Automated Variable Weighting in k-Means Type Clustering, Department of Decision Sciences And Information Management.

[15] K.A. Linderman, J.B. Sweeney, M.C. Ulrich, W. Li, Q. Song, and X. Yang. (2020) K-means clustering of overweight and obese population using quantile-transformed metabolic data. Journal of Biomedical Informatics, 109, 103511.

[16] M. Eftekhari, A. Mehrpooya, F. Saberi-Movahed, and V. Torra. (2022) How fuzzy concepts contribute to machine learning. Studies in Fuzziness and Soft Computing, 416, Springer.

[17] M. Namdari, J. H. Yoon, A. R. Abadi, S. M. Taheri and S. H. Choi. (2014) Fuzzy logistic regression with least absolute deviations estimators, Soft Computing, 19, 909-917.

[18] P. H. A. Sneath, and R. R. Sokal. (1973) Numerical taxonomy: The principles and practice of numerical classification, William H. Freeman and Company.

[19] R.J. Hathaway, and J.C. Bezdek. (1993) Weighted Fuzzy Clustering, Institute of Electrical and Electronics Engineers Transactions on Pattern Analysis and Machine Intelligence, 15(5), 522-531.

[20] V.P. Singh, and A.K. Mishra. (2021) Clustering diabetic patients based on their healthcare service utilization patterns, ResearchGate, Online avaliable.