سیستم های فازی و کاربردها
سال نهم، شماره اول (بهار و تابستان ۱۴۰۵)
صص. ۹۳ تا ۱۱۴
شاپای چاپی: ۲۷۱۷ـ۴۴۰۹
شاپای الکترونیکی: ۲۷۱۷ـ۳۹۸۴
DOI:10.22034/jfsa.2026.563812.1295

خوشه‌بندی داده‌های مستخرج از جامعهٔ فازی با استفاده از الگوریتم سلسله‌مراتبی و کاربرد آن در رباتیک مدرن

مرضیه رضائی، عباس پرچمی* و ایوب شیخی

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

تاریخ دریافت: ۱۴۰۴/۰۹/۱۱ تاریخ پذیرش: ۱۴۰۵/۰۲/۱۳

نوع مقاله: علمی‌ـ پژوهشی

چکیده. خوشه‌بندی سلسله‌مراتبی یکی از روش‌های یادگیری بدون نظارت در داده‌کاوی است که با ارزیابی شباهت میان داده‌ها، آن‌ها را در قالب یک ساختار سلسله‌مراتبی گروه‌بندی کرده و خوشه‌هایی با بالاترین میزان شباهت درون‌خوشه‌ای ایجاد می‌کند. روش‌های موجود در این حوزه، برای داده‌های مستخرج از جوامع دقیق طراحی شده‌اند. با این حال، در بسیاری از کاربردهای واقعی، جوامع مورد مطالعه ماهیتی نادقیق یا فازی دارند؛ مواردی از این قبیل جوامع را می‌توان به دانشجویان بااستعداد، محصولات کشاورزی سالم، دوره‌های آموزشی مفید و شغل‌های پردرآمد اشاره کرد. در چنین جوامعی، هر مشاهده علاوه بر مقدار عددی خود، دارای درجه‌ای از عضویت در جامعهٔ فازی نیز است. در این مقاله، روش خوشه‌بندی سلسله‌مراتبی به الگوریتم خوشه‌بندی داده‌های حاصل از جوامع فازی تعمیم داده شده است. همچنین با تعمیم دو شاخص سیلوئیت و دان از حالت کلاسیک به جامعهٔ فازی، تعداد خوشه‌های بهینه بر اساس معیارهای تشابه میان خوشه‌ها تعیین شده است و کیفیت خوشه‌بندی در استفاده کاربردی روش پیشنهادی در داده‌های رباتیک مدرن تشریح می‌شود.

2010 *Mathematics Subject Classification.* 62H30; 68T37
* Corresponding author
E-mails: Rezaei.Marzieh2021@math.uk.ac.ir, parchami@uk.ac.ir and sheikhy.a@uk.ac.ir.
عبارات و کلمات کلیدی. خوشه‌بندی، الگوریتم سلسله‌مراتبی، تابع عضویت، داده‌های مستخرج از جامعهٔ فازی، وزن‌دهی.

۱. مقدمه

خوشه‌بندی یکی از روش‌های مهم داده‌کاوی در تحلیل‌های چندمتغیره است که با هدف گروه‌بندی داده‌ها بر اساس بیشترین شباهت به کار می‌رود. انواع روش‌های خوشه‌بندی شامل خوشه‌بندی سلسله‌مراتبی[۱]، خوشه‌بندی مبتنی بر افراز[۲]، خوشه‌بندی مبتنی بر چگالی[۳] و روش‌های مدل‌محور[۴] هستند که هر یک با بهره‌گیری از معیارهای متفاوت، ساختار پنهان داده‌ها را آشکار می‌کنند [۵، ۶]. خوشه‌بندی به پژوهشگران این امکان را می‌دهد که الگوهای مشابه را در داده‌های پیچیده شناسایی کنند و تحلیل‌های تصمیم‌گیری و پیش‌بینی را بهبود بخشند [۱، ۳].

برخلاف تکنیک‌های خوشه‌بندی مبتنی بر افراز که در آن‌ها تعداد خوشه‌ها به‌عنوان پارامتر ورودی توسط کاربر تعیین می‌شود، در روش‌های خوشه‌بندی سلسله‌مراتبی، مجموعه داده‌ها و معیاری برای ارزیابی تشابه، به‌عنوان ورودی در نظر گرفته می‌شوند. بسته به این‌که تحلیل سلسله‌مراتبی از پایین به بالا[۵] یا از بالا به پایین[۶] انجام گیرد، عملیات اصلی در الگوریتم‌های خوشه‌بندی سلسله‌مراتبی را می‌توان در دو ردهٔ ادغام[۷] و تقسیم[۸] طبقه‌بندی کرد. فرآیند ادغام و تقسیم خوشه‌ها نقشی اساسی در عملکرد این الگوریتم‌ها ایفا می‌کند؛ زیرا در صورت اتخاذ تصمیمی نادرست در هر مرحله از ادغام یا تقسیم، الگوریتم امکان بازگشت و اصلاح تصمیم را ندارد و این امر می‌تواند موجب کاهش اعتبار خوشه‌های نهایی شود. در الگوریتم‌های خوشه‌بندی سلسله‌مراتبی پایین به بالا، هر نمونهٔ داده در آغاز در یک خوشهٔ مجزا قرار می‌گیرد و با پیشرفت الگوریتم، خوشه‌ها به‌تدریج با یکدیگر ادغام می‌شوند تا در نهایت یا تمامی نمونه‌ها در یک خوشه قرار گیرند، یا شرط ازپیش‌تعیین‌شده‌ای برای توقف فرآیند برآورده شود. در مقابل، دستهٔ دیگری از الگوریتم‌های سلسله‌مراتبی وجود دارند که از یک رویکرد بالا به پایین پیروی می‌کنند. در این روش‌ها، برخلاف روش پیشین، تمامی نمونه‌ها در آغاز در یک خوشهٔ کلی قرار می‌گیرند و سپس با بهره‌گیری از معیار تشابه، در چند مرحله به‌صورت سلسله‌مراتبی به خوشه‌های کوچک‌تر تقسیم می‌شوند. این روند نیز می‌تواند تا زمانی ادامه یابد که هر نمونه در

[1] Hierarchical Clustering
[2] Partition-based Clustering
[3] Density-based Clustering
[4] Model-based Clustering
[5] Bottom-up
[6] Top-down
[7] Agglomerative
[8] Divisive

یک خوشهٔ مجزا قرار گیرد یا شرط خاصی برای توقف اجرای الگوریتم تعیین شود. در هر دو نوع الگوریتم سلسله‌مراتبی، کاربر می‌تواند تعداد خوشه‌های نهایی را به‌عنوان شرط توقف مشخص کند. معمولاً فرآیند خوشه‌بندی سلسله‌مراتبی توسط نموداری به نام درخت‌واره‌نگار[1] (در برخی از متون دارنگاشت یا نمودار درختی) نمایش داده می‌شود. این نمودار فرآیند ادغام یا تقسیم خوشه‌ها را در هر مرحله با سطحی از شباهت مصورسازی می‌کند. خوشه‌بندی سلسله‌مراتبی در حوزه‌های مختلفی مانند بیوانفورماتیک، بازاریابی، علوم اجتماعی، متن‌کاوی و امنیت اطلاعات برای تحلیل داده‌ها، گروه‌بندی و شناسایی الگوها کاربرد دارد [۱۴، ۱۵].

در برخی مطالعات، اهمیت نمونه‌ها یا مشاهدات ممکن است یکسان نباشد. در چنین شرایطی، به هر مشاهده یک وزن عددی مثبت تخصیص داده می‌شود تا میزان تأثیر یا اهمیت آن در فرآیند یادگیری یا خوشه‌بندی مشخص گردد. این روش به الگوریتم اجازه می‌دهد تا نقش مشاهدات را به‌صورت تفکیک‌شده ارزیابی کرده و برخی از آن‌ها را با اهمیت بیشتری در خوشه‌بندی لحاظ کند. فرآیند وزن‌دهی می‌تواند به روش‌های گوناگونی انجام شود؛ از جمله استفاده از وزن‌های ثابت و ازپیش‌تعیین‌شده بر اساس میزان اهمیت هر مشاهده، یا تعیین وزن‌ها بر مبنای شاخص‌ها و معیارهای بیرونی. برای مطالعه بیشتر در این زمینه می‌توان به [۱۱، ۸] مراجعه کرد. اگر جامعهٔ مورد بررسی به‌صورت دقیق تعریف نشده باشد (اصطلاحاً جامعه فازی باشد)، آنگاه منجر به شیوه‌ای جدید برای وزن‌دهی به داده‌ها براساس تابع عضویت جامعهٔ فازی می‌شود. در بسیاری از مسائل کاربردی ، داده‌های حاصل از سرشماری یا نمونه‌گیری، مستخرج از یک جامعهٔ فازی هستند. در چنین جوامعی، هر مشاهده با درجات عضویت متفاوتی به جامعهٔ مورد مطالعه تعلق دارد. به‌عنوان مثال، در تحلیل داده‌های زیستی می‌توان به جامعهٔ فازی گونه‌های گیاهی با شباهت‌های ژنتیکی اشاره کرد در این مورد شباهت گونه‌های گیاهی نسبی است و نمی‌توان به‌طور قطعی تعیین کرد دو گونه «کاملاً مشابه» یا «کاملاً متفاوت» هستند بلکه هر دو گونه با درجات عضویت مشخصی و متعلق به بازهٔ صفر و یک به جامعهٔ فازی گونه‌های گیاهی با شباهت ژنتیکی تعلق دارند. در آمار کلاسیک، روش‌های متداول و کنونی، خوشه‌بندی سلسله‌مراتبی بدون در نظر گرفتن میزان عضویت هر مشاهده در جامعهٔ فازی انجام می‌شوند. در نتیجه، با نادیده گرفتن عدم‌قطعیت موجود در عضویت داده‌ها، بخشی از اطلاعات ارزشمند از بین می‌رود. برای رفع این مشکل، مناسب‌تر است که ابتدا جامعهٔ فازی مورد نظر به‌وسیلهٔ یک تابع عضویت مدل‌سازی گردد تا درجهٔ تعلق هر مشاهده به جامعه مشخص شود. سپس، روش‌های خوشه‌بندی کلاسیک بایستی به‌گونه‌ای تعمیم یابند

[1] Dendrogram

که وزن یا تأثیر هر داده در محاسبات سلسله‌مراتبی بر اساس میزان عضویت آن تنظیم شود. به عبارت دیگر، در تحلیل‌های آماری مبتنی بر جوامع فازی، داده‌هایی با درجات عضویت بالاتر باید نقش پررنگ‌تر و تأثیر بیشتری در ساختار نهایی خوشه‌ها ایفا کنند [۲، ۴]. در این مقاله، روش خوشه‌بندی سلسله‌مراتبی با رویکرد پایین به بالا به منظور خوشه‌بندی داده‌های مستخرج از یک جامعهٔ فازی گسترش یافته است. برای این منظور، معیارهای سنجش شباهت بین خوشه‌ها در چارچوب کلاسیک به صورت معیارهای وزنی تعمیم یافته‌اند، به گونه‌ای که وزن‌ها همان درجات عضویت هر مشاهده هستند که از تابع عضویت جامعهٔ فازی مورد نظر استخراج می‌شوند. با این حال، این وزن‌ها ساختار فاصلهٔ اولیه را تغییر نمی‌دهند و ماتریس فاصله میان مشاهدات همانند حالت کلاسیک محاسبه می‌شود. تفاوت صرفاً در مرحلهٔ محاسبهٔ فاصله میان خوشه‌ها ظاهر می‌شود؛ از میان چهار روش، دو روش پیوند تکی و پیوند کامل بدون تغییر باقی می‌مانند. در مقابل، نسخهٔ تعمیم‌یافتهٔ روش‌های پیوند میانگین و وارد به جامعهٔ فازی در فرآیند خوشه‌بندی به‌کار گرفته می‌شود.

ساختار این مقاله به شرح زیر است. در بخش ۲ تعریف جامعهٔ فازی و میانگین داده‌های مستخرج از یک جامعهٔ فازی مرور شده است. الگوریتم پیشنهادی در بخش ۳ معرفی شده است. در بخش ۴ دو شاخص سیلوئیت و دان به منظور تعیین تعداد خوشه‌های بهینه و ارزیابی کیفیت خوشه بندی الگوریتم پیشنهادی از حالت کلاسیک به جامعهٔ فازی تعمیم داده شده است. در پایان با ارائه یک مثال کاربردی در بخش ۵ نحوهٔ عملکرد روش پیشنهادی مورد بررسی قرار گرفته 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 [۹]. 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 [۱۲].

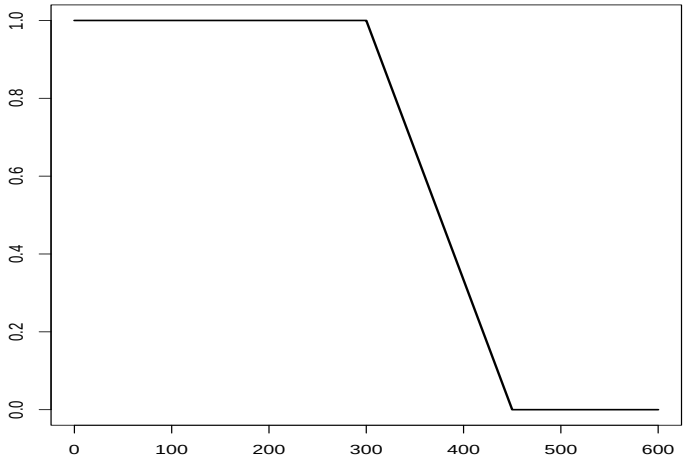
(۲) مواردی که جامعهٔ هدف به طور دقیق و واضح تعریف نشده است و با یک جامعهٔ هدف نادقیق روبرو هستیم. برای مثال اگر بخواهیم دربارهٔ میزان رطوبت هوا در روزهای بارانی بررسی انجام دهیم، جامعهٔ هدف «روزهای بارانی» است. در اینجا بدیهی است که روزهای مختلف با مقادیر مختلف بارندگی به یک اندازه مدنظر نیستند بلکه روزهای با بارندگی بیشتر، عضویت بالاتری در مجموعهٔ «روزهای بارانی» دارند و برعکس. به‌عنوان مثالی دیگر، اگر هدف بررسی وضع درآمد «جوانان زاهدانی» باشد، آنگاه جامعهٔ هدف یعنی «جوانان زاهدانی» 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 ۲.۲. فرض کنید می‌خواهیم متوسط تعداد اعضای خانواده‌های کرمانی در خانوارهای کم‌مصرف انرژی برق در تابستان را برحسب کیلووات ساعت محاسبه کنیم. از آنجا که «خانوارهای کم‌مصرف» یک مفهوم فازی و نادقیق است لذا اعضای جامعهٔ خانوارهای کم‌ـ مصرف به طور دقیق مشخص نیست. در واقع هر خانواده کرمانی با درجه‌ای بین صفر و یک متعلق به مجموعهٔ فازی خانوارهای کم‌مصرف است. در این حالت، ابتدا جامعهٔ فازی خانوارهای کم‌مصرف را به کمک یک تابع عضویت، به‌عنوان مثال، به‌صورت زیر تعیین و تعریف می‌کنیم (شکل ۱ را ببینید).

$$\mu_{\text{خانوار کم‌مصرف}}(x) = \begin{cases} ۱ & ۰ \le x \le ۳۰۰, \\ \dfrac{۴۵۰ - x}{۱۵۰} & ۳۰۰ < x < ۴۵۰, \\ ۰ & x \ge ۴۵۰. \end{cases} \qquad (۱.۲)$$

بنابراین طبق این تعریف از خانوارهای کم‌مصرف، مثلاً یک خانواده با میزان مصرف انرژی برق $۳۵۰kwh$، به اندازهٔ ۰/۶۶ در جامعهٔ فازی خانوارهای کم‌مصرف عضو است زیرا $\mu_{\text{خانوار کم‌مصرف}}(۳۵۰) = ۰/۶۶$. پس از اخذ نمونهٔ تصادفی از افراد این شهر (به حجم n)، اطلاعات مربوط به نمونه را به‌صورت $(x_۱, \mu_۱), ..., (x_n, \mu_n)$ ثبت می‌کنیم که در آن x_i



شکل ۱. تابع عضویت «خانوارهای کم‌مصرف» برحسب میزان مصرف انرژی برق برحسب کیلووات ساعت

تعداد اعضای خانوادهٔ iام و μ_i درجهٔ عضویت او در مجموعهٔ فازی «خانوارهای کم‌مصرف» است. در این مقاله چنین داده‌هایی را «داده‌های مستخرج از یک جامعهٔ فازی» می‌نامیم و در بخش ۳ به ارائه چندین رویکرد خوشه‌بندی سلسله‌مراتبی برای این نوع از داده‌ها می‌پردازیم.

تعریف ۳.۲. فرض کنید داده‌های $(x_1,\mu_1),\ldots,(x_n,\mu_n)$ 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 $(x_1,\mu_1),\ldots,(x_n,\mu_n)$ مقدار میانگین داده‌های مستخرج از جامعهٔ فازی با رابطهٔ

$$\bar{x}=\frac{\sum_{i=1}^{n} x_i\mu_i}{\sum_{i=1}^{n}\mu_i} \qquad (۲.۲)$$

تعریف می‌شود [۲] .

ملاحظه ۴.۲. اگر به‌جای تابع عضویت مجموعه فازی در رابطه (۲.۲) از تابع نشانگر جامعه (یعنی، I_Ω) استفاده شود، آنگاه $\sum_{i=1}^{n}\mu_i=n$ و در نتیجه میانگین بیان شده در رابطه (۲.۲) به میانگین حسابی داده‌های $x_1,...,x_n$ تبدیل می‌شود. برای جزئیات بیشتر و آشنایی با سایر شاخص‌های آماری مربوط به جامعهٔ فازی به مرجع [۲] مراجعه کنید.

جدول ۱. کاربردهای خوشه‌بندی بر اساس جامعهٔ فازی در حوزه‌های علمی مختلف

حوزه علمی	کاربرد خوشه‌بندی	جامعهٔ فازی بالقوه	معیار عضویت
روانشناسی شناختی	برنامه‌های شناسایی استعداد	افراد با استعداد	نمرات هوش، سوابق تحصیلی
پزشکی	برنامه‌های درمان	بیماران دیابتی حاد	سطح قند خون، شاخص توده بدنی
بازاریابی	بخش‌بندی مشتریان	مشتریان وفادار	تعداد خرید، بازخورد مشتری
علوم غذایی	گروه‌بندی محصولات	محصولات با کیفیت	رنگ، بافت، نقص‌ها
دامپزشکی	خوشه‌بندی اپیدمی‌ها	دام‌های بزرگ	علائم بیماری حیوانات
مدیریت مالی	خوشه‌بندی متقاضیان وام	متقاضیان پرریسک وام	سوابق اعتباری، درآمد
جامعه‌شناسی	برنامه‌های رفاه اجتماعی	محله‌های کم‌برخوردار	درآمد، دسترسی به خدمات رفاهی
هوش مصنوعی	شناسایی وضعیت احساسی	کاربران ناامید	الگوهای رفتاری کاربران
علم سنجی	دسته‌بندی مقالات پژوهشی	مقالات پر استناد	تعداد ارجاعات، شاخص‌های ارجاع
زمین‌شناسی	آمادگی در برابر زمین‌لرزه	مناطق لرزه‌ای	تعداد، بزرگی، عمق زمین‌لرزه‌ها
اقلیم‌شناسی	برنامه‌ریزی کشاورزی	مناطق اقلیمی مطلوب	دما، بارش، رطوبت، باد
توان‌بخشی گفتار	برنامه‌های توانبخشی	بیماران حاد گفتار	لکنت، اختلالات صدا
تحقیقات بازار	مدیریت موجودی	محصولات پرفروش	نرخ فروش هفتگی، بازگشت محصول
بیمه	تعیین دقیق‌تر حق بیمه	رانندگان با رفتار پرخطر	سرعت، ترمز ناگهانی، تعداد تصادفات
اپیدمیولوژی	استراتژی‌های واکسیناسیون	ناقلان پرمخاطره	نرخ ابتلا، تراکم جمعیت، سن

۳. روش سلسله‌مراتبی برای خوشه‌بندی داده‌های مستخرج از جامعهٔ فازی

در این بخش، با الهام گرفتن از [۵] تعمیم معیارهای سنجش شباهت بین خوشه‌ها در چارچوب کلاسیک به جامعهٔ فازی ارائه می‌شود. در این تعمیم، هر مشاهده دارای یک وزن است که متناظر با درجه عضویت آن در جامعهٔ فازی مورد نظر می‌باشد. لازم به ذکر است که این وزن‌ها تغییری در ساختار فاصله‌های اولیه ایجاد نمی‌کنند و ماتریس فاصله بین مشاهدات مشابه حالت کلاسیک باقی می‌ماند. اختلاف اصلی تنها در محاسبه فاصله میان خوشه‌ها مشاهده می‌شود، جایی که وزن‌ها در معیارهای شباهت لحاظ می‌شوند. بکارگیری این مفاهیم منجر به معرفی الگوریتم سلسله‌مراتبی تجمیعی جدیدی در این مقاله می‌شود که در خوشه‌بندی داده‌های مستخرج از یک جامعهٔ فازی کاربرد دارد و لذا این الگوریتم را به اختصار الگوریتم خوشه‌بندی سلسله‌مراتبی تجمیعی وزن‌دهی فازی[1] $(FWAHC)$ می‌نامیم.

[1] Fuzzy Weighted Agglomerative Hierarchical Clustering

۱.۳. ماتریس فاصلهٔ بین مشاهدات. ماتریس مشاهدات $\mathbf{X}$ متشکل از n مشاهده و m متغیر را به صورت زیر درنظر بگیرید:

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

ماتریس $\mathbf{X}$ را می‌توان برحسب سطرها به شکل

$$\mathbf{X} = \left[\mathbf{x}_۱^T, \mathbf{x}_۲^T, \dots, \mathbf{x}_n^T\right]^T$$

نیز نشان داد که در آن $\mathbf{x}_r = \left[x_{r۱}, \dots, x_{rm}\right]$، $r = ۱, \dots, n$، بردار مشاهده rام از ماتریس $\mathbf{X}$ است. با فرض اینکه $\mu(x)$ تابع عضویت جامعهٔ فازی موردنظر باشد $\boldsymbol{\mu} = (\mu_۱, \dots, \mu_n)^T$ بردار درجات عضویت هر یک از n مشاهده به جامعهٔ فازی است. ماتریس فاصله اقلیدسی[1] به‌صورت

$$\mathbf{D} = \begin{bmatrix} d_{۱۱} & d_{۱۲} & \dots & d_{۱n} \\ d_{۲۱} & d_{۲۲} & \dots & d_{۲n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n۱} & d_{n۲} & \dots & d_{nn} \end{bmatrix} = [d_{ij}]_{n\times n}$$

تعریف می‌شود. لازم به ذکر است که درایهٔ (i,j)ام ماتریس $\mathbf{D}$ به فاصله اقلیدسی بین مشاهدات iام و jام از یکدیگر اشاره دارد که به‌ازای $i = ۱, ۲, \dots, n$ و $j = ۱, ۲, \dots, n$ برحسب $L_۲-$نُرم به‌صورت زیر تعریف می‌شود

$$\begin{aligned} d_{ij} &= d(\mathbf{x}_i, \mathbf{x}_j) \\ &= \|\mathbf{x}_i - \mathbf{x}_j\|_۲ = \sqrt{(x_{i۱} - x_{j۱})^۲ + \dots + (x_{im} - x_{jm})^۲}. \end{aligned} \tag{۱.۳}$$

۲.۳. معیارهای سنجش تشابه خوشه‌ها. در روش‌های خوشه‌بندی مبتنی بر افراز، معمولاً معیارهای شباهت میان مشاهدات مورد محاسبه قرار می‌گیرند. با این حال، در روش‌های خوشه‌بندی سلسله‌مراتبی، لازم است معیارهایی تعریف شوند که علاوه بر محاسبه میزان شباهت بین مشاهدات، میزان نزدیکی و هم‌پوشانی میان خوشه‌ها را نیز ارزیابی کنند. اغلب

[1] Euclidean distance matrix

الگوریتم‌های خوشه‌بندی سلسله‌مراتبی از مجموعه‌ای متنوع از این معیارها بهره می‌برند که در ادامه به تعمیم برخی از آن‌ها برای بکارگیری در خوشه‌بندی سلسله‌مراتبی تجمیعی داده‌های مستخرج از یک جامعهٔ فازی پرداخته می‌شود.

الف‌ـ انتخاب دو مشاهده با بیشترین شباهت. یکی از ساده‌ترین و متداول‌ترین روش‌ها برای سنجش میزان شباهت میان دو خوشه، در نظر گرفتن دو مشاهده با بیشترین میزان شباهت (یا به‌عبارتی کم‌ترین فاصله) است؛ با این شرط که هر مشاهده متعلق به یک خوشه نباشند. این روش که در متون علمی با نام‌های دیگری همچون پیوند تکی (منفرد)[۱] یا نزدیک‌ترین همسایه[۲] نیز شناخته می‌شود، بر اساس حداقل فاصلهٔ موجود میان تمامی زوج مشاهدات در دو خوشه عمل می‌کند؛ به‌گونه‌ای که برای هر خوشه، یک مشاهده از آن انتخاب شده و فاصلهٔ میان آن‌ها محاسبه می‌گردد. بیشترین تشابه میان خوشه‌های C_i و C_j به‌صورت زیر تعریف می‌شود

$$d_{min}(C_i, C_j) = \min_{x_k \in C_i,\ x_l \in C_j} d_{kl}, \quad i, j = ۱, \ldots, n. \tag{۲.۳}$$

ب‌ـ انتخاب دو مشاهده با کم‌ترین شباهت. در این روش معیار تصمیم‌گیری، دو مشاهده‌ای هستند که کم‌ترین میزان شباهت (یا به‌عبارتی بیشترین فاصله) را با یکدیگر دارند. بدین ترتیب، ادغام دو خوشه تنها زمانی انجام می‌شود که حتی دورترین مشاهدات متعلق به آن‌ها نیز از لحاظ فاصله در محدوده‌ای قابل قبول قرار گیرند. در این روش نیز، هر دو مشاهدهٔ مورد بررسی باید متعلق به خوشه‌های متفاوت باشند. این تکنیک با نام‌های پیوند کامل[۳] یا دورترین همسایه[۴] نیز شناخته می‌شود. با فرض اینکه دو مشاهده متعلق به خوشه‌های C_i و C_j باشند معیار کم‌ترین تشابه میان دو خوشه با نماد $d_{max}(C_i, C_j)$ نشان داده شده و به‌صورت زیر تعریف می‌شود

$$d_{max}(C_i, C_j) = \max_{x_k \in C_i,\ x_l \in C_j} d_{kl}, \quad i, j = ۱, \ldots, n. \tag{۳.۳}$$

پ‌ـ میانگین تشابه. در روش پیوند میانگین[۵] معیار شباهت میان دو خوشه بر اساس میانگین تمامی فواصل میان زوج مشاهدات تعیین می‌شود؛ به‌گونه‌ای که هر جفت مشاهده شامل یک مشاهده از هر خوشه باشد. این رویکرد، بر خلاف دو روش پیشین که تنها بر فواصل حدی تمرکز دارند، تصویر جامع‌تری از نزدیکی کلی خوشه‌ها ارائه می‌دهد. معیار میانگین تشابه میان

[1] Single Linkage
[2] Nearest Neighbour
[3] Complete Linkage
[4] Furthest Neighbour
[5] Average Linkage

دو خوشهٔ C_i و C_j را با نماد $d_{avg}(C_i, C_j)$ نشان داده و به‌صورت زیر تعریف می‌شود

$$d_{\text{avg}}(C_i, C_j) = \frac{۱}{|C_i|\,|C_j|} \sum_{k:\ \boldsymbol{x}_k \in C_i} \sum_{l:\ \boldsymbol{x}_l \in C_j} \mu_k \mu_l d_{kl}, \qquad i, j = ۱, \ldots, n. \tag{۴.۳}$$

که در آن $|C_i| = \sum_{k:\ x_k \in C_i} \mu_k$.

ت‌ـ روش وارد. یکی از تکنیک‌های مشهور خوشه‌بندی سلسله‌مراتبی روش وارد[1] است که بر پایهٔ کاهش واریانس درون خوشه‌ای عمل می‌کند. در این روش، در هر مرحله از ادغام، دو خوشه با حداقل افزایش در مجموع مربعات فاصله‌ها برای ترکیب انتخاب می‌شوند. به بیان دیگر، هدف روش وارد این است که ادغام خوشه‌ها به گونه‌ای انجام شود که بیشترین همگنی درون هر خوشه حفظ شود و افزایش واریانس داخلی نیز به حداقل برسد.

فرض کنید مجموعهٔ داده‌ها به‌صورت $\mathbf{X} = \left[\mathbf{x}_۱^T, \mathbf{x}_۲^T, \ldots, \mathbf{x}_n^T\right]^T$ در فضای mـ بعدی باشد. اگر در مرحله‌ای از الگوریتم، دو خوشهٔ C_i و C_j درنظر گرفته شوند، افزایش مجموع مربعات خطا پس از ادغام آن‌ها به‌صورت $\Delta(C_i, C_j) = w_{ij}\ d^۲(\bar{\mathbf{x}}_i, \bar{\mathbf{x}}_j)$ تعریف می‌شود که در آن $w_{ij} = \frac{|C_i||C_j|}{|C_i|+|C_j|}$ است به‌طوری که $|C_i| = \sum_{k:\ x_k \in C_i} \mu_k$. همچنین $\bar{\mathbf{x}}_i$ و $\bar{\mathbf{x}}_j$ بردارهای میانگین iامین و jامین خوشه و $d(\bar{\mathbf{x}}_i, \bar{\mathbf{x}}_j)$ فاصله اقلیدسی بین دو بردار میانگین را نشان می‌دهند. هر یک از مؤلفه‌های بردار میانگین خوشهٔ iام مطابق رابطهٔ زیر محاسبه می‌شود:

$$\bar{x}_{ic} = \frac{\sum_{k:\ \mathbf{x}_k \in C_i} \mu_k x_{kc}}{\sum_{k:\ x_k \in C_i} \mu_k}, \quad c = ۱, \ldots, m. \tag{۵.۳}$$

معادلهٔ بیان شده در رابطه (۵.۳) در واقع تعمیمی از معادلهٔ (۲.۲) برای حالت چندمتغیره است. همچنین $d^۲(\bar{\mathbf{x}}_i, \bar{\mathbf{x}}_j)$ مربع فاصلهٔ اقلیدسی بین میانگین‌ها (/مراکز) دو خوشهٔ C_i و C_j است که به‌صورت $d^۲(\bar{\mathbf{x}}_i, \bar{\mathbf{x}}_j) = \|\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j\|_۲^۲$ محاسبه می‌شود. در هر تکرار، دو خوشه‌ای که کمترین مقدار $\Delta(C_i, C_j)$ را دارند با یکدیگر ادغام می‌شوند. این فرآیند تا زمانی ادامه می‌یابد که تمامی مشاهدات در یک خوشه ادغام شوند و یا اینکه شرط توقف دلخواهی (مانند تعداد خوشه‌های ازپیش‌تعیین‌شده) برقرار گردد.

مراحل اجرای الگوریتم $FWAHC$ به منظور خوشه‌بندی سلسله‌مراتبی با رویکرد از پایین به بالا (تجمیعی) ماتریس مشاهدات $\mathbf{X} = \left[\mathbf{x}_۱^T, \mathbf{x}_۲^T, \ldots, \mathbf{x}_n^T\right]^T$ را می‌توان در قالب یک الگوریتم به‌صورت زیر خلاصه کرد:

[1] Ward squared Euclidean distance

الگوریتم ۱ الگوریتم خوشه‌بندی سلسله‌مراتبی تجمیعی وزنی فازی $(FWAHC)$

ورودی: ماتریس مشاهدات و بردار وزن مربوط به درجات عضویت مشاهدات

خروجی: ساختار نهایی خوشه‌بندی

۱: محاسبهٔ ماتریس فاصلهٔ بین خوشه‌ها با استفاده از رابطهٔ (۱.۳)

۲: **تا زمان رسیدن به شرط توقف تکرار کن:**

۳: تشخیص دو خوشه با بیشترین شباهت یا کمترین فاصله

۴: ادغام دو خوشهٔ انتخاب‌شده و تشکیل خوشهٔ جدید

۵: بروزرسانی ماتریس فاصلهٔ بین خوشه‌ها براساس معیار فاصلهٔ انتخابی توسط کاربر برای سنجش میزان تشابه بین خوشه‌ها با استفاده از روش‌های پیوند تکی، پیوند کامل، پیوند میانگین و وارد

شرط توقف: ادغام تمامی مشاهدات در یک خوشهٔ نهایی

ملاحظه ۱.۳. در صورتی که جامعه کلاسیک باشد وزن هر مشاهده برابر با عدد یک است در نتیجه روش پیوند میانگین در رابطهٔ (۴.۳) که برای خوشه‌بندی داده‌های مستخرج از جامعهٔ فازی تبیین شده به پیوند میانگین مبتنی بر جامعهٔ کلاسیک تبدیل می‌شود. همچنین در این حالت طبق ملاحظهٔ ۴.۲ با تبدیل درایه‌های میانگین وزنی هر خوشه در رابطهٔ (۵.۳) به میانگین حسابی، روش وارد مبتنی بر جامعهٔ فازی نیز همان روش وارد در جامعهٔ کلاسیک است.

۴. شاخص‌های ارزیابی الگوریتم $FWAHC$

شاخص‌های ارزیابی کیفیت خوشه‌بندی ابزارهایی هستند که کیفیت ساختار خوشه‌ها را بدون نیاز به برچسب‌های واقعی داده‌ها ارزیابی می‌کنند. این شاخص‌ها با ارزیابی همزمان نزدیکی نقاط درون هر خوشه و میزان فاصله یا جدایی میان خوشه‌ها، امکان سنجش میزان انسجام داخلی خوشه‌ها و تمایز آن‌ها از یکدیگر را فراهم می‌کنند. محاسبهٔ این شاخص‌ها در زمینهٔ جامعهٔ فازی تا حدی با محاسبهٔ آن‌ها در جامعهٔ کلاسیک متفاوت است. در این بخش، دو شاخص برای تعیین تعداد خوشه‌های بهینه و ارزیابی کیفیت خوشه‌بندی الگوریتم پیشنهادی معرفی می‌شوند.

۱.۴. میانگین امتیاز سیلوئیت وزن‌دار فازی. امتیاز سیلوئیت وزن‌دار فازی[1] (FWSS) به‌طور همزمان میزان انسجام مشاهدات در خوشهٔ مربوط به خود و میزان جدایی آن‌ها از خوشه‌های دیگر را اندازه‌گیری می‌کند. برای هر مشاهدهٔ $\mathbf{x}_r$، مقدار $FWSS(r)$ با استفاده از دو کمیت $a_{FW}(r)$ و $b_{FW}(r)$ محاسبه می‌شود. به‌طوری که برای هر نمونهٔ $\mathbf{x}_r$، شاخص $a_{FW}(r)$ به‌صورت زیر تعریف می‌شود:

$$a_{FW}(r) = \frac{\sum_{k,l:\ \mathbf{x}_k, \mathbf{x}_l \in C_i} \mu_l d_{kl}}{\sum_{k:\ \mathbf{x}_k \in C_i} \mu_k}, \qquad r = ۱, ۲, \ldots, n.$$

این شاخص، میانگین وزنی فاصلهٔ نمونهٔ $\mathbf{x}_r$ با سایر نقاط موجود در خوشهٔ خود را اندازه‌گیری می‌کند و میزان فشردگی درون‌خوشه‌ای را نشان می‌دهد. همچنین شاخص $b_{FW}(r)$ به‌صورت زیر محاسبه می‌شود:

$$b_{FW}(r) = \min_{i \neq j} \left(\frac{\sum_{k,l:\ \mathbf{x}_k \in C_i, \mathbf{x}_l \in C_j} \mu_l d_{kl}}{\sum_{\mathbf{x}_l \in C_j} \mu_l} \right), \qquad r = ۱, ۲, \ldots, n.$$

این معیار، کمترین میانگین وزنی فاصلهٔ نمونهٔ $\mathbf{x}_r$ با نقاط موجود در سایر خوشه‌ها را نشان داده و میزان جدایی بین خوشه‌ها را ارزیابی می‌کند. سپس، مقدار امتیاز سیلوئیت وزن‌دار فازی برای مشاهدهٔ rام به‌صورت

$$FWSS(r) = \frac{b_{FW}(r) - a_{FW}(r)}{\max\{a_{FW}(r),\ b_{FW}(r)\}}, \qquad r = ۱, ۲, \ldots, n.$$

تعریف می‌شود. در نهایت، میانگین امتیاز سیلوئیت وزن‌دار فازی[2] به‌صورت زیر تعریف می‌شود

$$MFWSS = \frac{\sum_{r=۱}^{n} \mu_r FWSS(r)}{\sum_{r=۱}^{n} \mu_r}.$$

این شاخص به‌طور متداول برای تعیین تعداد بهینهٔ خوشه‌ها و همچنین ارزیابی عملکرد خوشه‌بندی کاربرد دارد. بیشترین مقدار برای این شاخص به‌ازای تعداد خوشه‌های مختلف نشان‌دهندهٔ تعداد بهینه خوشه‌ها است.

---

[1] Fuzzy Weighted Silhouette Score
[2] Mean Fuzzy Weighted Silhouette Score

۲.۴. شاخص دان وزن‌دار فازی. شاخص دان وزن‌دار فازی[1] کیفیت خوشه‌بندی را با تأکید بر خوشه‌هایی که هم متراکم و هم جدا از یکدیگر هستند، ارزیابی می‌کند. فرض کنید $d(C_i, C_j)$ نشان‌دهندهٔ فاصلهٔ بین خوشه‌های iام و jام باشد که به‌صورت فاصلهٔ بین مراکز خوشه‌های C_i و C_j محاسبه می‌شود. همچنین δ_i به قطر خوشهٔ iام اشاره دارد که به‌عنوان بیشترین فاصله بین هر دو نقطه در داخل خوشه تعریف می‌شود

$$\delta_i = \max_{\mathbf{x}_k, \mathbf{x}_l \in C_i} d_{kl}, \quad i = ۱, ۲, \ldots, n.$$

مختصات مراکز تمام خوشه‌ها که برای محاسبهٔ این شاخص استفاده شده است، از رابطهٔ (۵.۳) به‌دست می‌آیند. شاخص دان وزن‌دار فازی به‌صورت زیر تعریف می‌شود. مقادیر بالاتر FWDI نشان‌دهندهٔ عملکرد بهتر خوشه‌بندی است، به این معنا که خوشه‌ها هم متراکم‌تر و همچنین از یکدیگر بهتر جدا شده‌اند

$$FWDI = \frac{\min_{۱ \le i < j \le n} d(C_i, C_j)}{\max_{۱ \le i \le n} \delta_i}.$$

ملاحظه ۱.۴. شاخص‌های خوشه‌بندی معرفی‌شده در این بخش، نسخه‌های تعمیم‌یافتهٔ شاخص‌های پرکاربرد در چارچوب کلاسیک به‌شمار می‌آیند. در حالتی که جامعهٔ مورد بررسی یک جامعهٔ کلاسیک باشد، درجات عضویت تمامی مشاهدات برابر با $\mu_i = ۱$ به‌ازای $i = ۱, \ldots, n$ است. همچنین براساس ملاحظهٔ ۴.۲، درایه‌های مراکز خوشه محاسبه شده از رابطهٔ (۵.۳) به میانگین حسابی مشاهدات موجود در هر خوشه تبدیل می‌شوند. ازاین‌رو، شاخص‌های $MFWSS$ و $FWDI$ که برای ارزیابی کیفیت خوشه‌بندی در یک جامعهٔ فازی تعریف شده‌اند، به‌ترتیب با شاخص‌های سیلوئت و دان در چارچوب کلاسیک متناظر خواهند شد.

با توجه به مباحث مطرح‌شده دربارهٔ خوشه‌بندی سلسله‌مراتبی مبتنی بر یک جامعهٔ فازی، تعریف مسئله و شاخص‌های ارزیابی، می‌توان چارچوب کلی این روش را بررسی کرد. با این حال، شناخت بهتر عملکرد این الگوریتم نیازمند بررسی آن در مسائل واقعی است. به همین منظور، در ادامه یک نمونهٔ کاربردی از استفادهٔ خوشه‌بندی سلسله‌مراتبی در حوزهٔ رباتیک ارائه می‌شود تا ارتباط میان مباحث نظری و پیاده‌سازی عملی آن به‌صورت واضح‌تر نشان داده شود.

[1] Fuzzy Weighted Dunn Index

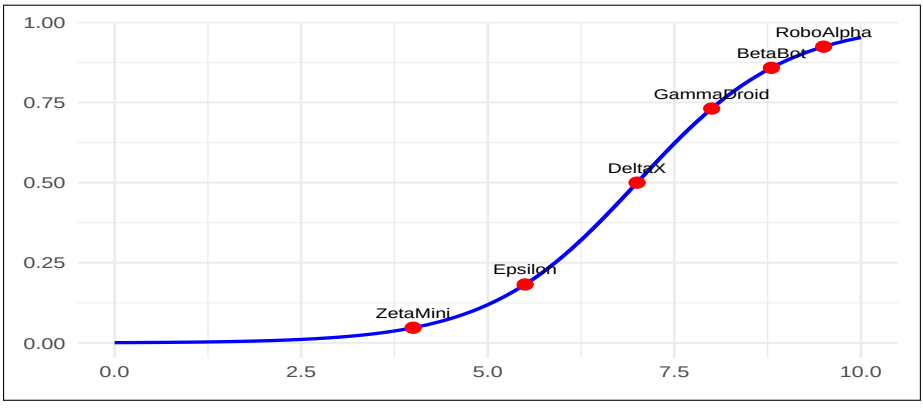
۵. مثال کاربردی در رباتیک مدرن

قابلیت برنامه‌پذیری ربات‌ها یکی از مهم‌ترین ویژگی‌های عملکردی در رباتیک مدرن است. این ویژگی به معنای توانایی ربات در یادگیری، انجام و سازگاری با مجموعه‌ای از برنامه‌ها و وظایف مختلف است. در محیط‌های صنعتی و تحقیقاتی، ربات‌ها با وظایف متنوع و شرایط محیطی متفاوت روبه‌رو هستند؛ بنابراین، ارزیابی قابلیت برنامه‌پذیری آن‌ها برای طراحی سیستم‌های تصمیم‌یار، تخصیص منابع و بهینه‌سازی عملکرد حیاتی است. ویژگی برنامه‌پذیری بطور طبیعی فازی است؛ زیرا ربات‌ها نمی‌توانند صرفاً در دو دستهٔ «برنامه‌پذیر» یا «غیر برنامه‌پذیر» قرار گیرند. برخی ربات‌ها توانایی اجرای تعداد زیادی وظیفه را دارند، در حالی که برخی دیگر محدودیت بیشتری دارند؛ همچنین برخی ربات‌ها می‌توانند برنامه‌های پیچیده را با سرعت بالا اجرا کنند، در حالی که برخی دیگر تنها وظایف ساده را به خوبی انجام می‌دهند.

جدول ۲ داده‌های مربوط به ۶ ربات با قابلیت برنامه‌پذیری را نشان می‌دهد. قابلیت برنامه‌پذیری یک ربات معمولاً بر اساس ترکیبی از مشخصات عملکردی ربات مانند تعداد عملکرد قابل برنامه‌ریزی، سرعت پردازش و غیره محاسبه می‌شود. تعداد عملکرد قابل برنامه‌ریزی نشان‌دهندهٔ میزان انعطاف‌پذیری ربات در اجرای وظایف مختلف و توانایی آن در انطباق با نیازهای متنوع است، به‌گونه‌ای که هرچه این تعداد بیشتر باشد، دامنه کاربردهای ربات نیز وسیع‌تر خواهد بود. سرعت پردازش نیز نقش حیاتی در تعیین کیفیت برنامه‌پذیری دارد، زیرا سرعت بالای پردازش امکان اجرای سریع و همزمان چندین عملکرد را فراهم می‌کند و پاسخگویی ربات به تغییرات محیطی را بهبود می‌بخشد. در مجموع، ترکیب مشخصات عملکردی ربات‌ها امکان رتبه‌بندی آن‌ها را در یک مقیاس عددی از ۱ تا ۱۰ به‌عنوان امتیاز برنامه‌پذیری فراهم می‌کند و معیار مناسبی برای مقایسه قابلیت برنامه‌پذیری آن‌ها ارائه می‌دهد.

جدول ۲. مشخصات عملکردی ربات‌ها و درجات تعلق هر یک از آن‌ها به جامعهٔ فازی «ربات‌های برنامه‌پذیر»

نام ربات	تعداد عملکرد قابل برنامه‌ریزی	سرعت پردازش	امتیاز برنامه‌پذیری	درجه عضویت
RoboAlpha	۱۵	۲۰۰	۹/۵	۰/۹۲
BetaBot	۱۲	۱۸۰	۸/۸	۰/۸۵
GammaDroid	۱۰	۱۵۰	۸/۰	۰/۷۳
DeltaX	۸	۱۲۰	۷/۰	۰/۵
Epsilon	۵	۱۰۰	۵/۵	۰/۱۸
ZetaMini	۳	۸۰	۴/۰	۰/۰۴



شکل ۲. تابع عضویت ربات‌های برنامه‌پذیر برحسب امتیاز برنامه‌پذیری

در این مثال کاربردی قصد داریم به خوشه‌بندی داده‌های مستخرج از جامعهٔ فازی «ربات‌های برنامه‌پذیر» در جدول ۲ براساس دو ویژگی تعداد عملکرد قابل برنامه‌ریزی و سرعت پردازش با استفاده از الگوریتم $FWAHC$ بپردازیم. برای این منظور ابتدا با استفاده از تابع عضویت سیگموید[1]، جامعهٔ فازی مورد نظر توسط یک شخص خبره در این حوزه به‌صورت زیر مدل‌سازی شده است. همچنین در شکل ۲ تابع عضویت قابلیت برنامه‌پذیری ربات براساس امتیاز برنامه‌پذیری نشان داده شده است.

$$\mu_{\text{ربات‌های برنامه‌پذیر}}(x) = \frac{۱}{۱ + e^{-\frac{x-۷}{۱}}}. \qquad (۱.۵)$$

[1] Sigmoid

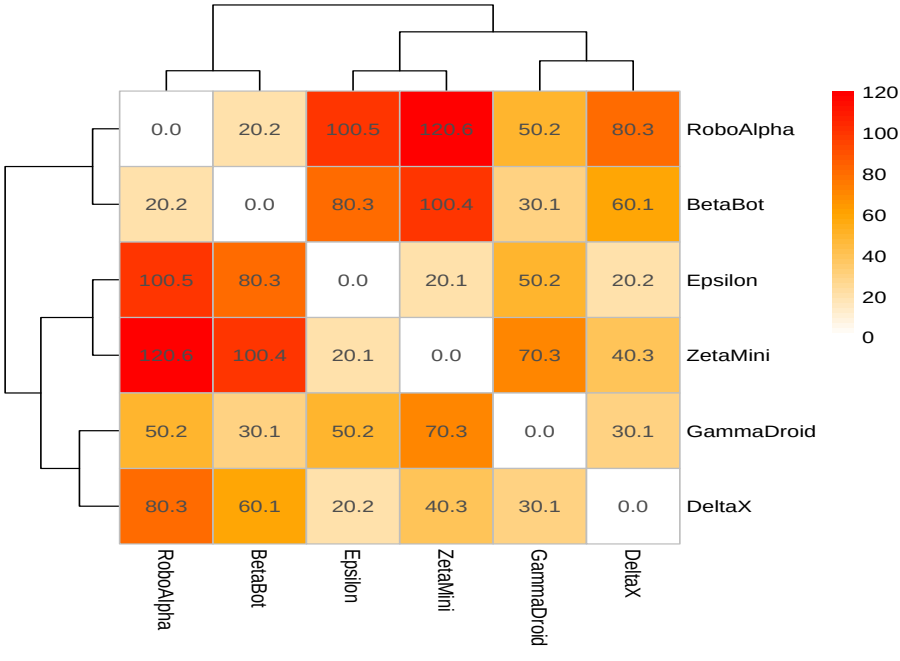
با درنظر گرفتن تابع عضویت (۱.۵)، به‌عنوان مثال درجه عضویت ربات $RoboAlpha$ در جامعهٔ فازی ربات‌های برنامه‌پذیر با امتیاز برنامه‌پذیری $x = ۹/۵$ به‌صورت زیر به‌دست می‌آید:

$$\mu_{\text{ربات‌های برنامه‌پذیر}}(۹/۵) = \frac{۱}{۱ + e^{-(۹/۵-۷)}} = \frac{۱}{۱ + e^{-۲/۵}} \approx ۰/۹۲$$

به‌طور مشابه، برای ربات $ZetaMini$ با امتیاز برنامه‌پذیری $x = ۴$ داریم:

$$\mu_{\text{ربات‌های برنامه‌پذیر}}(۴) = \frac{۱}{۱ + e^{-(۴-۷)}} = \frac{۱}{۱ + e^{۳}} \approx ۰/۰۴$$

با توجه به ستون آخر در جدول ۲ ربات‌های $RoboAlpha$ و $ZetaMini$ به‌ترتیب با داشتن درجات عضویت $\mu(۹/۵) = ۰/۹۲$ و $\mu(۴) = ۰/۰۴$ از بیشترین و کم‌ترین میزان عضویت در جامعهٔ فازی ربات‌های برنامه‌پذیر برخوردارند. همانطور که در شکل ۲ ملاحظه می‌شود به‌عنوان مثال ربات $DeltaX$ با درجهٔ عضویت $\mu(۷) = ۰/۵$ به اندازهٔ ۰/۵ به‌عنوان یک ربات با قابلیت برنامه‌پذیری شناخته می‌شود. شکل ۳ نمودار حرارتی ماتریس فاصلهٔ میان مشاهدات



شکل ۳. نمودار حرارتی ماتریس فاصلهٔ بین مشخصات عملکردی ربات‌ها

(مشخصات عملکردی ربات‌ها) را نشان می‌دهد که بر اساس رابطه‌ی (۱.۳) محاسبه شده است.

به عنوان مثال فاصلهٔ اقلیدسی میان جفت ربات‌های $Epsilon$ و $ZetaMini$، $RoboAlpha$ و $BetaBot$ و همچنین $GammaDroid$ و $DeltaX$ به‌صورت زیر محاسبه می‌شود

$$d(Epsilon, ZetaMini) = \sqrt{(۵-۳)^۲ + (۱۰۰-۸۰)^۲} \approx ۲۰/۱$$

$$d(RoboAlpha, BetaBot) = \sqrt{(۱۵-۱۲)^۲ + (۲۰۰-۱۸۰)^۲} \approx ۲۰/۲$$

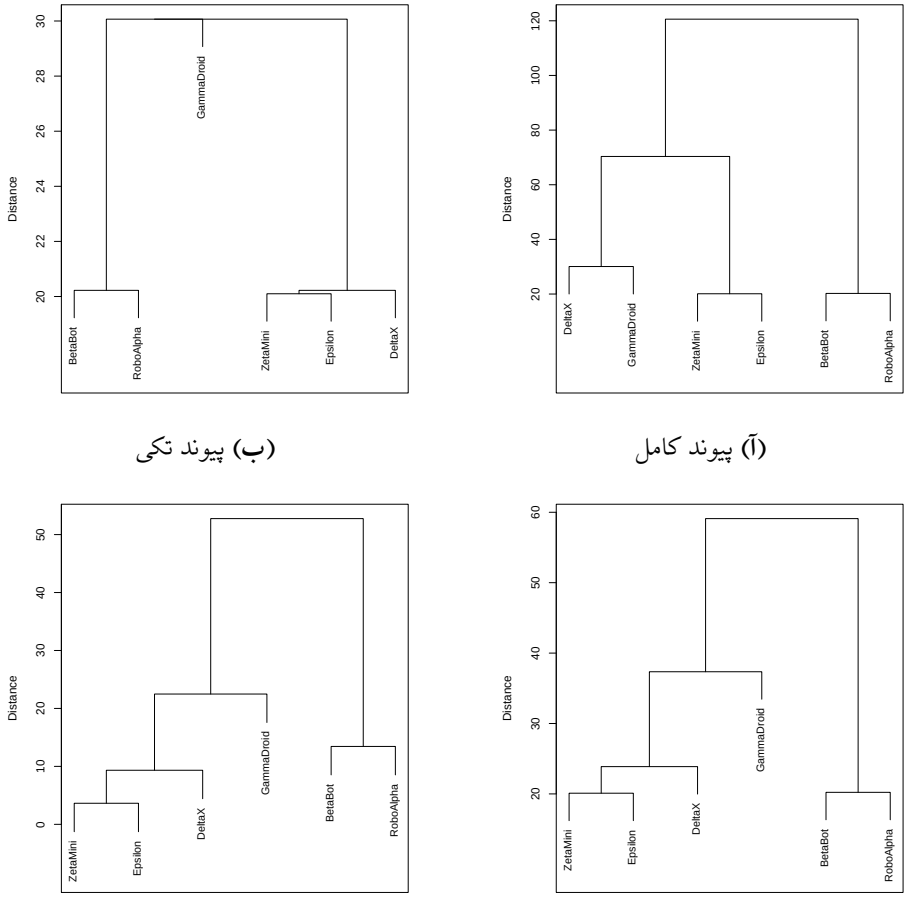
$$d(GammaDroid, DeltaX) = \sqrt{(۱۰-۸)^۲ + (۱۵۰-۱۲۰)^۲} \approx ۳۰/۱$$

همان‌گونه که ملاحظه می‌شود، ربات‌های $Epsilon$ و $ZetaMini$ بیشترین شباهت را از نظر قابلیت برنامه‌پذیری دارند. به‌طور مشابه، ربات‌های $RoboAlpha$ و $BetaBot$ نیز در یک خوشه قرار گرفته‌اند و شباهت رفتاری قابل‌توجهی را نشان می‌دهند. در نهایت، ربات‌های $GammaDroid$ و $DeltaX$ نیز از نظر ویژگی برنامه‌پذیری، نزدیکی بیشتری نسبت به سایر ربات‌ها دارند.

شکل ۴ نتایج خوشه‌بندی داده‌های مستخرج از جامعهٔ فازی «ربات‌های برنامه‌پذیر» با استفاده از الگوریتم $FWAHC$ براساس چهار روش محاسبهٔ تشابه بین خوشه‌ها را نشان می‌دهد. به‌منظور ارزیابی نحوهٔ گروه‌بندی نمونه‌ها، الگوریتم $FWAHC$ با استفاده از معیارهای سنجش تشابه خوشه‌ها شامل پیوند تکی، پیوند کامل، پیوند میانگین و روش وارد به‌کار گرفته شد. با توجه به نمودارهای حاصل از بکارگیری معیارهای تشابه خوشه‌ها، تفاوت میان نتایج به‌طور عمده ناشی از نحوهٔ تعریف فاصلهٔ بین خوشه‌ها در هر معیار است. برای مشاهدهٔ نحوهٔ ادغام خوشه‌ها براساس هر یک از معیارهای تشابه در هر مرحله شکل ۴ را ببینید.

در شکل ۵ مقادیر $MFWSS$ در برابر تعداد خوشه‌ها برای چهار روش پیوند کامل، پیوند تکی، پیوند میانگین و وارد ترسیم شده است. همان‌گونه که ملاحظه می‌شود، مقدار $MFWSS$ با تغییر تعداد خوشه‌ها رفتار متفاوتی را در هر یک از روش‌ها نشان می‌دهد و این موضوع بیانگر تأثیر نوع معیار ادغام بر عملکرد الگوریتم پیشنهادی است. در این نمودار، نقاط متناظر با بیشترین مقدار $MFWSS$ با رنگ قرمز مشخص شده‌اند.

در این مطالعه، عملکرد چهار روش خوشه‌بندی سلسله‌مراتبی شامل پیوند کامل، پیوند تکی، پیوند میانگین و وارد با استفاده از دو شاخص اعتبارسنجی $MFWSS$ و $FWDI$ در تعداد خوشه‌های مختلف از ۲ تا ۵ مورد بررسی قرار گرفت. از آنجایی که شاخص $MFWSS$ انسجام و کیفیت تخصیص هر نقطه داده به خوشه خود را با مقایسه فاصله آن از نقاط هم‌خوشه‌ای و فاصله از نقاط خوشه‌های مجاور می‌سنجد و شاخص $FWDI$ میزان جدایی بین خوشه‌ها را نسبت به تراکم درون‌خوشه‌ای اندازه‌گیری می‌کند مقادیر بالاتر



(آ) پیوند کامل

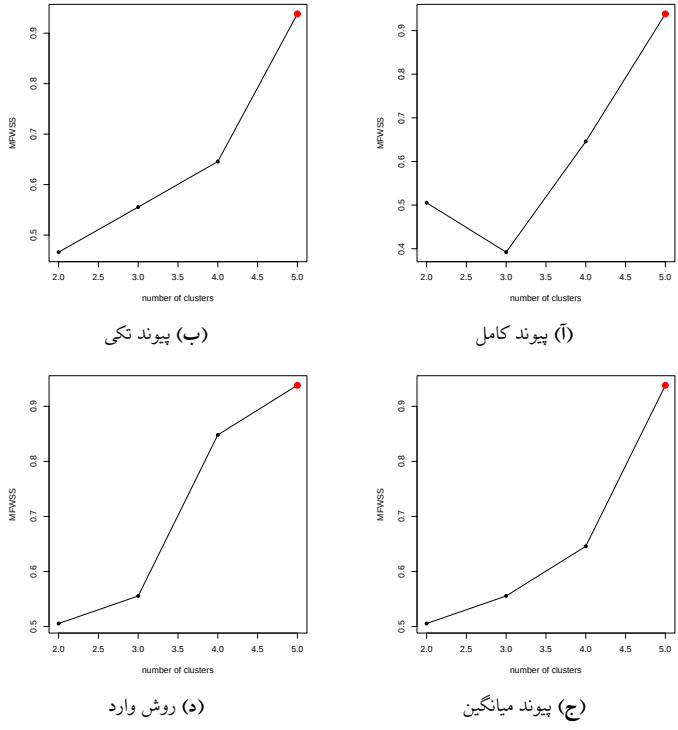
(ب) پیوند تکی

(ج) پیوند میانگین

(د) روش وارد

شکل ۴. نمودار دارنگاشت برای داده‌های مستخرج از جامعهٔ فازی «ربات‌های برنامه‌پذیر» با استفاده از الگوریتم $FWAHC$ براساس چهار روش (آ) پیوند کامل، (ب) پیوند تکی، (ج) پیوند میانگین و (د) روش وارد

هر دو شاخص نشان‌دهندهٔ تفکیک‌پذیری و انسجام بیشتر خوشه‌ها است. نتایج ارائه شده در جداول ۳ و ۴ نشان می‌دهد که در روش پیوند کامل، بیشترین مقدار $FWDI$ برابر با ۱/۳۸۷ در $k = ۳$ مشاهده شده است، در حالی که در روش پیوند تکی بیشینه این شاخص در $k = ۲$ و مقدار ۱/۳۱۵ به‌دست آمد. در روش پیوند میانگین، بیشترین مقدار $FWDI$ برابر با ۱/۱۸۱ در $k = ۴$ و در روش وارد، بیشترین مقدار این شاخص در $k = ۵$ و برابر با ۱/۰۰۶ گزارش



(آ) پیوند کامل (ب) پیوند تکی

(ج) پیوند میانگین (د) روش وارد

شکل ۵. نمودار $MFWSS$ در برابر تعداد خوشه‌ها براساس چهار روش (آ) پیوند کامل، (ب) پیوند تکی، (ج) پیوند میانگین و (د) روش وارد

جدول ۳. مقادیر شاخص $MFWSS$ در الگوریتم پیشنهادی برای روش‌های مختلف ادغام

تعداد خوشه‌ها	۲	۳	۴	۵
پیوند کامل	۰/۵۰۵	۰/۳۹۲	۰/۶۴۶	۰/۹۳۸
پیوند تکی	۰/۴۶۶	۰/۵۵۵	۰/۶۴۶	۰/۹۳۸
پیوند میانگین	۰/۵۰۵	۰/۵۵۵	۰/۶۴۶	۰/۹۳۸
روش وارد	۰/۵۰۵	۰/۵۵۵	۰/۸۴۸	۰/۹۳۸

شده است. از سوی دیگر، شاخص $MFWSS$ برای تمامی روش‌ها در $k = ۵$ به بیشترین مقدار یعنی ۰/۹۳۸ دست یافته است که نشان‌دهندهٔ انسجام بالای خوشه‌ها، جدایی مناسب میان آن‌ها و پایداری مطلوب ساختار خوشه‌بندی در این تعداد خوشه است. تحلیل همزمان

جدول ۴. مقادیر شاخص $FWDI$ در الگوریتم پیشنهادی برای روش‌های مختلف ادغام

تعداد خوشه‌ها	۲	۳	۴	۵
پیوند کامل	۰/۸۴۰	۱/۳۸۷	۱/۱۸۱	۱/۰۰۶
پیوند تکی	۱/۳۱۵	۰/۹۲۶	۱/۱۸۱	۱/۰۰۶
پیوند میانگین	۰/۸۴۰	۰/۹۲۶	۱/۱۸۱	۱/۰۰۶
روش وارد	۰/۸۴۰	۰/۹۲۶	۰/۵۰۲	۱/۰۰۶

دو شاخص نشان می‌دهد که اگرچه در برخی مقادیر کوچک‌تر k، $FWDI$ مقادیر بالاتری دارد، اما کاهش همزمان $MFWSS$ نمایانگر کاهش انسجام خوشه‌ها است؛ در مقابل، در $k = ۵$ هر دو شاخص به مقادیر مطلوب خود رسیده‌اند، به طوری که انسجام و تفکیک‌پذیری خوشه‌ها بهینه شده است. بر اساس این تحلیل چندشاخصه و با در نظر گرفتن توازن میان فشردگی درون‌خوشه‌ای، جدایی بین خوشه‌ها و پایداری ساختار خوشه‌بندی، تعداد $k = ۵$ به‌عنوان تعداد بهینه خوشه‌ها برای تمامی روش‌های مورد بررسی انتخاب شده است، که عملکرد مطلوب الگوریتم پیشنهادی در استخراج ساختار داده‌ها را تأیید می‌کند. بررسی عملکرد روش‌ها نیز نشان می‌دهد که روش وارد با تمرکز بر کمینه‌سازی واریانس درون‌خوشه‌ای، بالاترین مقادیر $MFWSS$ و مقادیر قابل قبول $FWDI$ را ارائه می‌دهد و از این نظر بهترین عملکرد را در میان چهار روش دارد، در حالی که پیوند میانگین با روند یکنواخت و پایدار در هر دو شاخص، نیز به‌عنوان گزینه‌ای قابل اعتماد و پایدار مطرح می‌شود. در مقابل، روش پیوند تکی به دلیل نوسانات شاخص‌ها و حساسیت بالا به تغییر تعداد خوشه‌ها، کمترین اعتبار را از خود نشان می‌دهد.

۶. نتیجه‌گیری

در این مقاله، پس از مرور مفهوم جامعهٔ فازی، الگوریتم سلسله‌مراتبی تجمیعی به‌منظور خوشه‌بندی داده‌های استخراج‌شده از این جوامع گسترش داده شد. با توجه به کاربرد گستردهٔ جوامع فازی در مسائل دنیای واقعی، از جمله بیماران در معرض خطر ابتلا به دیابت، مناطق شهری آلوده، دانش‌آموزان موفق، زمین‌های کشاورزی حاصلخیز و مشتریان وفادار فروشگاه‌های اینترنتی، که در آن‌ها هر مشاهده دارای درجه‌ای از عضویت بین صفر و یک در جامعهٔ فازی مورد

نظر است، ضرورت توسعه روش‌های خوشه‌بندی متناسب با این ساختار بیش از پیش احساس می‌شود. در این راستا، نوآوری اصلی این پژوهش در گسترش ساختار کلاسیک خوشه‌بندی سلسله‌مراتبی از طریق وارد کردن درجات عضویت مشاهدات در جامعهٔ فازی به‌عنوان وزن هر مشاهده در فرآیند تجمیع نهفته است. این رویکرد امکان تحلیل مؤثرتر داده‌هایی را فراهم می‌سازد که از جوامعی با ماهیت فازی و نادقیق استخراج شده‌اند. بر این اساس، الگوریتم سلسله‌مراتبی تجمیعی $FWAHC$ برای خوشه‌بندی داده‌های حاصل از جوامع فازی معرفی شد. از دیگر دستاوردهای مهم این پژوهش، معرفی دو شاخص ارزیابی ویژه برای الگوریتم پیشنهادی $FWAHC$ است که امکان تعیین تعداد بهینهٔ خوشه‌ها بر اساس روش‌های پیوند کامل، پیوند تکی، پیوند میانگین و وارد را فراهم می‌سازد. این شاخص‌ها همچنین زمینه مقایسهٔ نظام‌مند عملکرد روش‌های مختلف پیوند و ارزیابی دقیق کیفیت خوشه‌بندی را متناسب با ماهیت فازی جوامع مورد مطالعه فراهم کرده‌اند.

برای پژوهش‌های آینده، می‌توان رفتار الگوریتم پیشنهادی را در مواجهه با داده‌های پرت بررسی کرد و استواری هر یک از معیارهای تشابه میان خوشه‌ها را در حضور این داده‌ها ارزیابی نمود. علاوه بر این، با توجه به پیچیدگی و عدم قطعیت موجود در داده‌های واقعی، کاربرد الگوریتم در تحلیل داده‌های فازی و مدل‌سازی جوامع فازی بر اساس توابع عضویت نوع‌ـ۲ نیز می‌تواند مورد مطالعه قرار گیرد.

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] رضائی، م، پرچمی، ع، و شیخی، ا. (۱۴۰۴) خوشه‌بندی مبتنی بر جامعهٔ فازی؛ چه هنگام و چگونه؟. سیستم‌ـ های فازی و کاربردها، شماره ۲، صص. ۹۱ تا ۱۰۸.

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

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

[7] G. J. Klir, & B. Yuan. (1995) Fuzzy sets and fuzzy logic: theory and applications, Upper Saddle River, NJ, USA: Prentice Hall.

[8] J. Byrd, & Z. Lipton. (2019) What is the effect of importance weighting in deep learning? Proceedings of the 36th International Conference on Machine Learning (PMLR 97), 872–881.

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

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

[11] M. Kimura, & H. Hino. (2024) A short survey on importance weighting for machine learning. arXiv preprint arXiv:2403.10175.

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

[13] M. S. Yang, & W. L. Hung. (2021) Advances in Fuzzy Clustering and Its Applications, John Wiley Sons.

[14] P. N. Tan, M. Steinbach, & V. Kumar. (2018) Introduction to Data Mining, (2nd ed.), Pearson.

[15] S. M. S. Hussain, & T. S. B. Sudhakar. (2019) A Comparative Analysis of Hierarchical Clustering Algorithms, International Journal of Computer Sciences and Engineering (IJCSE), 7(5), 266-271.