پانزدهمین سمینار احتمال و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان

15th Seminar On Probability and Stochastic Processes
August 30-31, 2025
University of Kurdistan

انجمن آمار ایران

خوشه‌بندی دوطرفه فازی با استفاده از تابع فاصله اطلاعاتی

زهرا احمدی [۱] ، دکتر رضا زارعی [۲]

[۱] استادیار گروه آمار، دانشکده علوم ریاضی، دانشگاه گیلان

[۲] کارشناس ارشد گروه آمار، دانشکده علوم ریاضی، دانشگاه گیلان

چکیده: با گسترش روزافزون داده‌ها در حوزه‌های مختلف علمی، نیاز به روش‌های مؤثر برای تحلیل و استخراج الگوهای پنهان در داده‌ها بیش از پیش احساس می‌شود. خوشه‌بندی یکی از روش‌های مهم در داده‌کاوی غیرنظارتی است که با هدف شناسایی ساختارهای پنهان و تقسیم داده‌ها به گروه‌های مشابه به‌کار می‌رود. در این میان، خوشه‌بندی فازی به‌دلیل انعطاف‌پذیری بالا در برخورد با داده‌های نادقیق، جایگاه ویژه‌ای یافته است. در این مقاله ضمن معرفی رویکرد خوشه‌بندی فازی دوطرفه، یک الگوریتم جدید در این حوزه مورد بررسی قرار می‌گیرد. جزئیات محاسباتی این الگوریتم با استفاده از مجموعه داده‌های واقعی در حوزه زیست‌پزشکی تشریح شده و عملکرد آن با چند الگوریتم مشابه مقایسه شده است. نتایج نشان‌دهنده برتری نسبی الگوریتم مذکور از نظر دقت، تفسیرپذیری و معیارهای ارزیابی خوشه‌بندی است.

واژه‌های کلیدی: خوشه‌بندی فازی، خوشه‌بندی دوطرفه، تابع هدف، توابع عضویت

کد موضوع‌بندی ریاضی (۲۰۲۰): 62H30، 03B52، 68T09.

---

۱ مقدمه

تحلیل و دسته‌بندی داده‌های حجیم و پیچیده یکی از چالش‌های اساسی در علوم داده و آمار است. خوشه‌بندی[۱] به عنوان یکی از روش‌های پایه در داده‌کاوی، نقش مؤثری در کشف ساختار درون‌داده‌ها دارد. هدف اصلی خوشه‌بندی، تقسیم داده‌ها به گروه‌هایی است که درون هر گروه، داده‌ها بیشترین شباهت و بین گروه‌ها بیشترین تفاوت را داشته باشند. در روش‌های سنتی خوشه‌بندی یا خوشه‌بندی سخت[۲]، هر داده تنها به یک خوشه اختصاص می‌یابد. با این حال، بسیاری از داده‌های دنیای واقعی دارای مرزهای نامشخص یا هم‌پوشانی هستند.

[۱] سخنران، ahmadizahra19.77@gmail.com

[1] Clustering
[2] Hard Clustering

در چنین شرایطی، خوشه‌بندی فازی[۳] گزینه‌ای مناسب‌تر است زیرا به هر داده اجازه می‌دهد تا به چند خوشه با درجات عضویت متفاوت تعلق گیرد.

در بسیاری از کاربردها، مانند تحلیل بیان ژن، خوشه‌بندی همزمان نمونه‌ها (ردیف‌ها) و ویژگی‌ها (ستون‌ها) اهمیت فزاینده‌ای پیدا کرده است. این رویکرد که به خوشه‌بندی دوطرفه موسوم است، قادر است زیرگروه‌هایی از نمونه‌ها را همراه با زیرگروه‌هایی از ویژگی‌ها که رفتاری مشابه از خود نشان می‌دهند، شناسایی کند. علی‌رغم پیشرفت‌های قابل توجه در خوشه‌بندی فازی دوطرفه، همچنان چالش‌هایی اساسی در مواجهه با داده‌های بسیار پیچیده و نویزی و به ویژه در استخراج خوشه‌های با مرزهای هم‌پوشان و نتایج با قابلیت تفسیرپذیری بالا وجود دارد. روش‌های موجود اغلب در مدل‌سازی دقیق وابستگی‌های پیچیده میان نمونه‌ها و ویژگی‌ها و همچنین مدیریت بهینه ابهام و عدم قطعیت در داده‌ها با محدودیت‌هایی مواجه‌اند. در این راستا، این مقاله به معرفی و بررسی یک الگوریتم جدید مبتنی بر اطلاعات می‌پردازد. نوآوری اصلی این الگوریتم در استفاده از یک تابع فاصله اطلاعاتی و ترکیب دو عبارت آنتروپی مجزا برای کنترل پراکندگی عضویت‌ها در سطرها و ستون‌ها است که همگرایی پایدارتر، مدیریت بهینه عدم قطعیت و در نتیجه نتایج دقیق‌تر و قابل تفسیرتری ارائه می‌دهد.

باقیمانده این مقاله به شرح زیر سازماندهی شده است: در بخش دوم، مبانی نظری خوشه‌بندی دوطرفه فازی به کوتاهی مرور می‌شود. بخش سوم به معرفی جامع و جزئیات محاسباتی الگوریتم پیشنهادی اختصاص دارد. در بخش چهارم، نتایج آزمایش‌های تجربی روی مجموعه داده‌های واقعی و مقایسه عملکرد الگوریتم تحت بررسی با الگوریتم‌های موجود در این زمینه ارائه می‌شود. در نهایت، بخش پنجم شامل بحث و نتیجه‌گیری کلی از یافته‌های پژوهش است.

۲ خوشه‌بندی دوطرفه

خوشه‌بندی کلاسیک معمولاً تنها بر سطرها (نمونه‌ها) یا ستون‌ها (ویژگی‌ها) انجام می‌شود. اما داده‌های پیچیده‌تر، مانند ماتریس‌های با ابعاد بزرگ و دارای ساختارهای پنهان، نیازمند روشی هستند که بتواند همزمان خوشه‌بندی روی نمونه‌ها و ویژگی‌ها انجام دهد. این روش با عنوان خوشه‌بندی دوطرفه[۴] شناخته می‌شود که اولین بار توسط بوریس میرکین مطرح شد. در این روش، هدف شناسایی زیرماتریس‌هایی از داده‌هاست که در آن‌ها گروهی از ویژگی‌ها با گروهی از نمونه‌ها به‌صورت هم‌زمان الگوهای مشترکی دارند. خوشه‌بندی دوطرفه کاربردهای گسترده‌ای در متن‌کاوی، زیست‌فناوری، داده‌های پزشکی و سیستم‌های توصیه‌گر دارد و نسبت به روش‌های سنتی، نتایج دقیق‌تر و تفسیرپذیرتری ارائه می‌دهد. تفاوت این روش با خوشه‌بندی کلاسیک در این است که خوشه‌بندی دوطرفه هم‌زمان روی ردیف‌ها و ستون‌ها کار می‌کند و خروجی آن زیرماتریس‌هایی با الگوهای مشترک است و انعطاف‌پذیری زیادی دارد یعنی هر آیتم می‌تواند در چند خوشه باشد این در حالی است که در خوشه‌بندی کلاسیک خوشه‌بندی فقط روی ردیف یا ستون انجام می‌شود و خروجی آن گروهی از ردیف‌ها یا ستون‌ها است و انعطاف‌پذیری کمی دارد.فرارو.ام.بی. (۲۰۲۴)

فرض کنید X یک ماتریس داده‌گان $n \times p$ باشد. هدف الگوریتم خوشه‌بندی دوطرفه تقسیم‌بندی همزمان n سطر به k خوشه و p ستون به h خوشه است. k-میانگین دوطرفه توسط ویچي و همکارانش (۲۰۰۱) مطرح شد. نسخه فازی آن توسط فرارو و همکارانش (۲۰۱۵) معرفی شد و سپس در سال ۲۰۲۱ عمیقا مورد بررسی قرار گرفت.

[3] Fuzzy Clustering
[4] Co-Clustering

تابع هدف الگوریتم k-میانگین دوطرفه فازی به صورت زیر مدل‌بندی می‌شود

$$J(\mathbf{X},U,C)=\sum_{j=۱}^{k}\sum_{i=۱}^{n}\sum_{f=۱}^{h}\sum_{g=۱}^{p}(x_i-c_j)^{۲}(u_{ij})^m(v_{gf})^l$$

که در آن

$$\sum_{j=۱}^{k}u_{ij}=۱\quad\sum_{f=۱}^{h}v_{gf}=۱,\ u_{ij},v_{gf}\in[۰,۱].$$

همچنین U ماتریسی $n\times k$ حاوی درجات عضویت سطرها و V ماتریسی $p\times h$ حاوی درجات عضویت ستون‌ها می‌باشند. m و l مشخص‌کننده درجات فازی و اعداد حقیقی و بزرگتر از یك هستند.

۳ الگوریتم خوشه‌بندی دوطرفه فازی مبتنی بر اطلاعات

الگوریتم خوشه‌بندی دوطرفه فازی مبتنی بر اطلاعات[۵] (به کوتاهی $ibFCC$) یکی از روش‌های نوین در زمینه خوشه‌بندی فازی دوطرفه است که با هدف بهبود دقت خوشه‌بندی و تفسیر بهتر خوشه‌ها طراحی شده است. این الگوریتم از تابع هدفی بهره می‌برد که شامل سه جزء اصلی است: (۱) تابع فاصله اطلاعاتی بین داده‌ها و مراکز خوشه‌ها، (۲) آنتروپی توزیع عضویت نمونه‌ها در خوشه‌ها، و (۳) آنتروپی توزیع عضویت ویژگی‌ها در خوشه‌ها.

در این مدل، دو ماتریس عضویت به نام‌های U و V تعریف می‌شوند که به ترتیب درجات عضویت نمونه‌ها و ویژگی‌ها را در خوشه‌های مربوطه تعیین می‌کنند. پارامترهای Tu و Tv برای تنظیم میزان پراکندگی عضویت‌ها استفاده می‌شوند. مزیت اصلی ibFCC نسبت به روش‌های مشابه در استفاده از فاصله اطلاعاتی به‌جای فاصله اقلیدسی است که باعث می‌شود ارتباط میان ویژگی‌ها و نمونه‌ها بهتر مدل‌سازی شود.

هدف $ibFCC$ به حداقل رساندن تابع هدف معادله (۱.۳) با توجه به محدودیت‌های معادله زیر است

$$J_{ibFCC}=\sum_{c=۱}^{C}\sum_{i=۱}^{n}\sum_{j=۱}^{k}u_{ci}v_{cj}d_{cij}+T_u\sum_{c=۱}^{C}\sum_{i=۱}^{n}u_{ci}lnu_{ci}+T_v\sum_{c=۱}^{C}\sum_{j=۱}^{k}v_{cj}lnv_{cj}. \qquad (۱.۳)$$

$$\sum_{c=۱}^{C}u_{ci}=۱\quad و\quad\sum_{j=۱}^{k}v_{cj}=۱\qquad i=۱,\dots,n\quad و\quad c=۱,\dots,C$$

عبارت اول معادله (۱.۳) درجه تطابق است که باید طی فرایند خوشه‌بندی دوطرفه به حداقل برسد تا افراد و ویژگی‌های بسیار مرتبط را باهم خوشه‌بندی دوطرفه کند. u_{ci} و v_{cj} دو تابع عضویت هستند که به ترتیب عضویت اسناد و ویژگی‌ها را نشان می‌دهند. عبارت دوم و سوم عوامل تنظیم آنتروپی هستند که u_{ci} و v_{cj} را جداگانه ترکیب می‌کنند. T_u و T_v پارامترهای وزنی هستند که درجه فازی را در خوشه‌های نهایی کنترل می‌کنند. معادلات به‌روزرسانی برای عضویت سند و ویژگی که d_{cij} فاصله بین نقطه داده ویژگی و مرکز خوشه ویژگی است، به صورت زیر بدست می‌آید

$$u_{ci}=\frac{\exp\left(-\frac{\sum_{j=۱}^{k}v_{cj}d_{cij}}{T_u}\right)}{\sum_{c=۱}^{C}\exp\left(-\frac{\sum_{j=۱}^{k}v_{cj}d_{cij}}{T_u}\right)}\quad و\quad v_{cj}=\frac{\exp\left(-\frac{\sum_{i=۱}^{n}u_{ci}d_{cij}}{T_v}\right)}{\sum_{j=۱}^{k}\exp\left(-\frac{\sum_{i=۱}^{n}u_{ci}d_{cij}}{T_v}\right)}. \qquad (۲.۳)$$

[۵] Information-based Fuzzy Co-Clustering

برای محاسبه اطلاعات فاصله مجموعه دادگان از رابطه زیر استفاده می‌کنیم

$$d_{cij} = \frac{۱}{n} x_{ij} log(\frac{x_{ij}}{t_{cij}}) + \frac{|c|}{n} p_{cj} log(\frac{p_{cj}}{t_{cij}}), \quad (۳.۳)$$

که $t_{cij} = \frac{(x_{ij}+|c|*p_{cj})}{(۱+|c|)}$، $|c|$ تعداد اسناد در cامین خوشه است. تعریف مقدار $|c|$ در خوشه‌بندی فازی نسبت به خوشه‌بندی سخت کمی پیچیده‌تر است زیرا باید عملیات فازی‌سازی روی ماتریس عضویت انجام شود. پس از فازی‌سازی به راحتی می‌توان مقدار $|c|$ را در خوشه‌بندی سخت بدست آورد.

در الگوریتم $ibFCC$ بدست آوردن p_{cj} به طور صریح دشوار است حتی اگر مقدار p_{cj} از نظر تئوری به صورت u_{ci} و v_{cj} محاسبه شود ممکن است فرایند محاسباتی بالایی داشته باشد. بنابراین یک رویکرد جایگزین انتخاب می‌شود که از روش میانگین وزنی استفاده می‌کند. در خوشه‌بندی فازی مرکز یک خوشه میانگین تمام نقاط است که براساس درجه تعلق آن‌ها به خوشه وزن می‌شود. درنتیجه معادله به‌روزرسانی نرمال شده p_{cj} به صورت زیر بدست می آید

$$p_{cj} = \frac{\sum_{i=۱}^{n} u_{ci} x_{ij}}{\sum_{i=۱}^{Nn} u_{ci}}. \quad (۴.۳)$$

۱.۳ مراحل اجرای الگوریتم $ibFCC$

مرحله	شرح عملیات
۱	مقداردهی اولیه به پارامترها: تعداد خوشه‌ها C، بیشینه تکرارها $\tau_{\max}$، آستانه همگرایی ε، ضرایب تنظیم آنتروپی T_u و T_v.
۲	تنظیم $\tau = ۱$.
۳	مقداردهی اولیه به درجات عضویت u_{ci} و v_{cj} به‌صورت تصادفی.
۴	شروع تکرارها
۵	محاسبه مراکز خوشه ویژگی‌ها با استفاده از رابطه (۴.۳).
۶	محاسبه فاصله اطلاعاتی بین اسناد و ویژگی‌ها با استفاده از رابطه (۳.۳).
۷	به‌روزرسانی درجات عضویت با استفاده از رابطه (۲.۳).
۸	افزایش شمارنده تکرار: $\tau \leftarrow \tau + ۱$.
۹	بررسی شرط توقف: $\tau = \tau_{\max}$ یا $\max \lvert u_{ci}^{(\tau)} - u_{ci}^{(\tau-۱)} \rvert \leq \varepsilon$.

یکی از تفاوت‌های کلیدی بین الگوریتم $ibFCC$ و الگوریتم‌های فازی دوطرفه کلاسیک، ساختار تابع هدف آن‌هاست. در روش‌های سنتی مانند فازی k-میانگین دوطرفه، تابع هدف بر اساس فاصله اقلیدسی و ترکیب درجه‌های عضویت سطرها و ستون‌ها تعریف می‌شود و تمرکز آن صرفاً بر کمینه‌سازی فاصله بین داده‌ها و مراکز خوشه است. اما این رویکرد در مواجهه با داده‌های پیچیده، نویزی یا دارای هم‌پوشانی بالا، کارایی محدودی دارد. در مقابل، تابع هدف الگوریتم $ibFCC$ با بهره‌گیری از یک تابع فاصله اطلاعاتی (اطلاعات متقابل بین ویژگی و سند) و نیز دو عبارت آنتروپی مجزا برای سطرها و ستون‌ها، رویکردی غنی‌تر و منعطف‌تر ارائه می‌دهد. پارامترهای کنترلی T_u و T_v امکان تنظیم میزان پراکندگی عضویت‌ها را فراهم کرده و از همگرایی سریع و نامناسب جلوگیری می‌کنند. این طراحی باعث می‌شود الگوریتم $ibFCC$ بتواند ساختارهای پنهان و الگوهای معنادارتر را در داده‌ها شناسایی کرده و خوشه‌بندی دقیق‌تری ارائه دهد. به‌طور خلاصه، تفاوت‌های موجود در تابع هدف موجب افزایش تطبیق‌پذیری، دقت، و تفسیرپذیری خروجی الگوریتم $ibFCC$ نسبت به روش‌های سنتی شده‌اند، که در بخش بعدی از طریق ارزیابی عددی در داده‌های واقعی به آن پرداخته می‌شود.

۴ نتایج عددی

برای ارزیابی عملکرد الگوریتم ،ibFCC آزمایش‌هایی روی پنج مجموعه داده واقعی در حوزه زیست‌پزشکی انجام شد. ویژگی‌های این مجموعه داده‌ها در جدول زیر آمده است

جدول ۱: جزئیات مجموعه داده‌های مورد بررسی

نام مجموعه داده	تعداد نمونه	تعداد ویژگی	تعداد دسته‌بندی‌ها	توضیحات
Ohsumed	۲۰۰۰۰ (سنداز ۵۰۲۱۶)	۵۰۰	۱۵	۲۳ دسته بیماری‌های موضوعی پزشکی (MeSH).
سرطان ریه (LC)	۲۷	۵۶	۳	برای شناخت انواع پاتولوژیک سرطان ریه.
بافت سینه (BT)	۱۰۶	۹	۶	یش‌بینی طبقه‌بندی با ادغام کلاس‌های فیبروآدنوم، ماستوپاتی و غده‌ای.
کاردیوتوکوگرافی (CTG)	۲۱۲۶	۲۱	۱۰	۲۱۲۶ کاردیوتوکوگرام جنین، ویژگی‌های تشخیصی اندازه‌گیری شده.
بیان پروتئین موش (MPE)	۱۰۷۶	۶۸	۸	۱۰۷۶ اندازه‌گیری در هر پروتئین، با ۸ دسته از موش‌ها.

هر ۵ الگوریتم به صورت تصادفی مقداردهی و ده بار اجرا شدند تا تاثیر بهینه‌سازی‌های موضعی کاهش یابد. از بین پنج الگوریتم، FCM یک الگوریتم خوشه‌بندی فازی استاندارد است و بقیه، الگوریتم‌های خوشه‌بندی دوطرفه فازی هستند. در ادامه، الگوریتم $ibFCC$ با چهار الگوریتم دیگر شامل FCM، $FCCM$، $RFCC$ و $FCCI$ مقایسه شد. معیارهای ارزیابی شامل آنتروپی، شاخص F و $Log(CS)$ بودند و نتایج نشان داد که الگوریتم $ibFCC$ در اکثر موارد عملکرد بهتری داشته و به‌ویژه در داده‌های BT و MPE توانسته دقت خوشه‌بندی را به‌طور چشم‌گیری افزایش دهد. بعد از $ibFCC$ الگوریتم $FCCI$ عملکرد بهتری از FCM، $FCCM$ و $RFCC$ دارد.

جدول ۲: مقایسه الگوریتم‌ها بر اساس معیارهای F، آنتروپی و $\log(CS)$

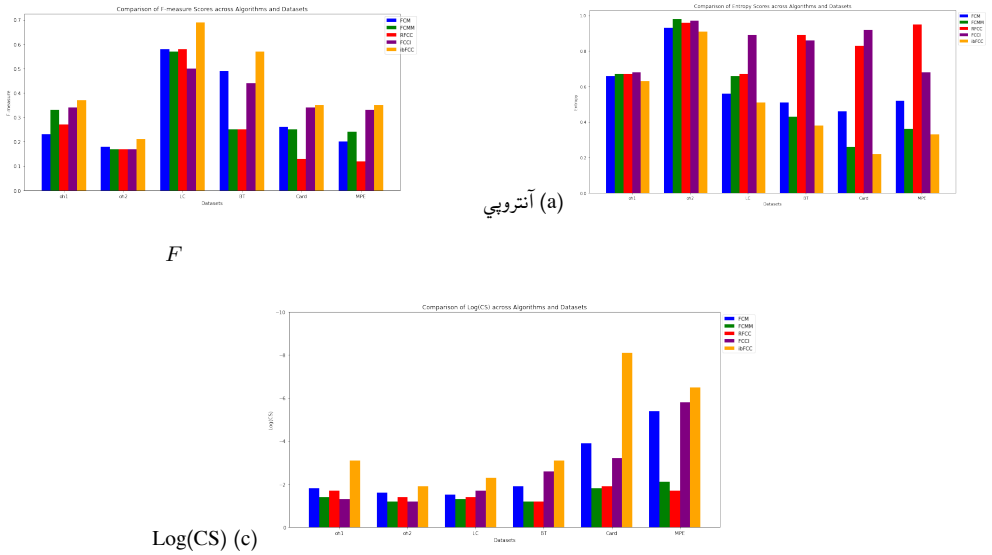
الگوریتم	نویسندگان	داده‌ها	F	آنتروپی	$\log(CS)$
FCM	بزدك (۱۹۷۳)	$Oh۱$	۰/۲۳	۰/۶۶	−۱/۸
		$Oh۲$	۰/۱۸	۰/۹۳	−۱/۶
		LC	۰/۵۸	۰/۵۶	−۱/۵
		BT	۰/۴۹	۰/۵۱	−۱/۹
		$Card$	۰/۲۶	۰/۴۲	−۳/۹
		MPE	۰/۳۵	۰/۲۲	−۵/۱
$FCCM$	او،سی.اچ و همکاران (۲۰۰۱)	$Oh۱$	۰/۳۳	۰/۶۷	−۱/۴
		$Oh۲$	۰/۲۷	۰/۹۶	−۱/۳
		LC	۰/۵۵	۰/۶۶	−۱/۵
		BT	۰/۴۶	۰/۶۰	−۱/۸
		$Card$	۰/۲۵	۰/۳۶	−۲/۱
		MPE	۰/۳۴	۰/۲۶	−۳/۹
$RFCC$	تیجی،دبلیو.سی. و همکاران (۲۰۰۷)	$Oh۱$	۰/۲۷	۰/۶۷	−۱/۷
		$Oh۲$	۰/۲۰	۰/۹۲	−۱/۴
		LC	۰/۵۸	۰/۹۶	−۱/۳
		BT	۰/۵۲	۰/۸۹	−۱/۲
		$Card$	۰/۲۵	۰/۸۵	−۱/۹
		MPE	۰/۳۴	۰/۹۵	−۳/۷
$FCCI$	هانماندلوام. و همکاران (۲۰۱۳)	$Oh۱$	۰/۳۴	۰/۶۸	−۱/۳
		$Oh۲$	۰/۲۸	۰/۹۰	−۱/۲
		LC	۰/۵۷	۰/۸۵	−۱/۲
		BT	۰/۴۹	۰/۸۰	−۱/۴
		$Card$	۰/۲۷	۰/۷۸	−۳/۵
		MPE	۰/۳۶	۰/۶۵	−۵/۸
$ibFCC$	لیو و همکاران (۲۰۱۷)	$Oh۱$	۰/۳۵	۰/۷۰	−۱/۳
		$Oh۲$	۰/۲۹	۰/۹۱	−۱/۲
		LC	۰/۵۷	۰/۸۵	−۱/۲
		BT	۰/۴۹	۰/۸۰	−۱/۴
		$Card$	۰/۳۵	۰/۲۲	−۶/۵
		MPE	۰/۳۵	۰/۳۳	−۵/۳

بحث و نتیجه‌گیری

در این مطالعه الگوریتمibFCC به عنوان یک روش جدید برای خوشه‌بندی فازی دوطرفه مورد مطالعه و بررسی قرار گرفت. این الگوریتم با در نظر گرفتن اطلاعات متقابل بین ویژگی‌ها و نمونه‌ها، نسبت به روش‌های موجود که عمدتاً بر فاصله اقلیدسی تمرکز دارند، برتری محسوسی دارد. علاوه بر این، ibFCC با استفاده از دو عبارت آنتروپی مجزا برای سطرها (نمونه‌ها) و ستون‌ها (ویژگی‌ها)، انعطاف‌پذیری بیشتری در مدل‌سازی ساختارهای پنهان داده‌ها ارائه می‌دهد. وجود پارامترهای کنترلی در این الگوریتم، امکان تنظیم میزان پراکندگی عضویت‌ها را فراهم کرده و به همگرایی سریع و جلوگیری از همگرایی نامناسب کمک کرده است.

نتایج ارزیابی بر روی پنج مجموعه داده واقعی نشان داد که الگوریتم ibFCC در بیشتر موارد عملکرد بهتری داشته و دقت خوشه‌بندی را به‌طور چشمگیری افزایش داده است. این برتری عملکردی ،ibFCC به دلیل توانایی آن در مدل‌سازی بهتر ارتباطات پیچیده میان ویژگی‌ها و نمونه‌ها، و همچنین انعطاف‌پذیری بالای آن در مواجهه با داده‌های نادقیق و هم‌پوشان است.

در نهایت، با توجه به نتایج به‌دست‌آمده از ارزیابی‌های تجربی، می‌توان نتیجه گرفت که الگوریتم ibFCC به‌عنوان یک ابزار مؤثر و کارآمد برای تحلیل داده‌های پیچیده و دارای ساختار پنهان محسوب می‌شود. قابلیت‌های پیشرفته این الگوریتم در کشف ساختارهای



(a) آنتروپي (b) معیار F Log(CS) (c)

شکل ۱: مقایسه سه معیار F، آنتروپي و $Log(CS)$ برای الگوریتم‌هاي جدول نتایج روی داده‌ها

نهفته و ارائه نتایج دقیق‌تر و تفسیرپذیرتر، آن را به گزینه‌ای مناسب برای کاربردهای مختلف در علوم داده، به‌ویژه در حوزه‌هایی مانند زیست‌پزشکی که با داده‌های حجیم و پیچیده سروکار دارند، تبدیل می‌کند.

مراجع

طبیبی،م (۱۴۰۱)، "رویکرد بیزی در خوشه‌بندی تحت شرایط نادقیق"، پایان نامه کارشناسی ارشد آمار ریاضی، دانشگاه گیلان.

, *Bezdek, J. C. (1973). "Fuzzy Mathematics in Pattern Classification" , John Wiley Sons, Hoboken-New Jersey.*

,Ferraro, M.B, Vichi, M., Grzegorzewski, P., Gagolewski, M., Hryniewicz, O., Gil, M., (2015) Fuzzy double clustering: a robust proposal. In: Advances in Intelligent systems and computing 315. Springer, Cham., pp 225232.

,Ferraro,M.B.(2024) ,Fuzzy k-means:history and applications.Elsevier; 2024 . pp.110-123.

,Hanmandlu M, Verma OP, Susan S, Madasu VK (2013) Color segmentation by fuzzy co-clustering of chrominance color features. Neurocomputing 120: 235–249.

,Liu Y, Wu S, Liu Z, Chao H (2017) A fuzzy co-clustering algorithm for biomedical data. PLoS ONE 12(4): e0176536

,Oh CH, Honda K, Ichihashi H. Fuzzy clustering for categorical multivariate data; 2001. pp. 2154–2159 vol. 2154.

,Tjhi WC, Chen L. Robust fuzzy Co-clustering algorithm; 2007. pp. 1–5.

, Vichi, M., Borra, S., Rocci, R., Vichi, M., Schader, M., (2001) Doubl kmeans clustering for simultaneous classification of objects and variables. In: Advances in classification and data analysis. studies in classification, data analysis, and knowledge organization. Springer, Berlin, Heidelberg.

Fuzzy Co-Clustering using Information Distance

Zahra Ahmadi[1], **Reza Zarei**[2]

[1]Department of Statistics, University of Guilan

[2]Department of Statistics, University of Guilan

Abstract: With the rapid growth of data in various scientific domains, the need for effective methods to analyze and extract hidden patterns has become increasingly important. Among clustering techniques, fuzzy clustering has gained significant attention due to its high flexibility in handling ambiguous and overlapping data. This paper introduces and investigates an advanced algorithm called interval-based bidirectional fuzzy co-clustering (ibFCC) designed for imprecise environments. Using real-world biomedical datasets, the performance of this algorithm is compared with several related methods. The results indicate its relative superiority in terms of accuracy, interpretability, and clustering evaluation metrics.

Keywords: Fuzzy clustering, Co-clustering, Membership functions, Objective function

Mathematics Subject Classification (2020): 62H30,۰۳B۵۲,۶۸T۰۹.
