ارائه روشی برای رگرسیون بر مبنای استخراج ویژگی و مجموعه های فازی مردد

مهلا مختیا[۱]، مهدی افتخاری[۲]، فرید صابری موحد[۳]

[۱]دانشجوی کارشناسی ارشد هوش مصنوعی، بخش مهندسی کامپیوتر، دانشگاه شهید باهنر کرمان

[۲]دانشیار، بخش مهندسی کامپیوتر، دانشگاه شهیدباهنر کرمان، m.eftekhari@uk.ac.ir

[۳]استادیار، گروه ریاضی کاربردی، دانشکده علوم و فناوری‌های نوین، دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته، کرمان

چکیده

در این مقاله روشی کارا برای رگرسیون ارائه شده است که در آن از انواع روش های خوشه بندی فازی و مفاهیم مجموعه های فازی مردد استفاده می شود. در ابتدا الگوریتم خوشه بندی فازی روی داده ها به کار رفته و بعد ازتصویر کردن تابع عضویت خوشه ها روی ویژگی های مختلف، به تعداد خوشه‌ها مجموعه های فازی روی هربعد (یا ویژگی) بدست می‌آید. سپس این مجموعه های فازی را به صورت یک مجموعه فازی مردد روی هرویژگی در نظر گرفته و ماتریس ضریب همبستگی فازی مردد را برای ویژگی ها به دست می آوریم. در ادامه یک نگاشت غیرخطی براساس تجزیه مولفه های اصلی این ماتریس برای تبدیل ویژگی های مجموعه داده به ویژگی های جدید استفاده شده است. در پایان، ویژگی‌های استخراج شده جدید را به الگوریتم خوشه‌بندی فازی داده ویک سیستم فازی سوگنو به منظور رگرسیون برازش شده است. روش پیشنهادی با چندین روش دیگر روی چندین مجموعه داده رگرسیون مقایسه شده است. نتایج آزمایش‌ها نشان دهنده موفق بودن روش پیشنهادی در استخراج وکاهش ویژگی‌ها و همچنین افزایش دقت رگرسیون است. همچنین تعداد قوانین مدل رگرسیون فازی در روش پیشنهادی در حد قابل قبولی کم است .

کلیدواژه

فازی، خوشه بندی فازی، مجموعه فازی مردد، همبستگی فازی مردد

مقدمه

سیستم‌های مبتنی بر قوانین فازی روش شناخته شده‌ای در محاسبات نرم برای رسیدگی به مسائل پیچیده دنیای واقعی هستند. آن‌ها روش قدرتمندی در مقابله با مسائل مختلف از جمله عدم اطمینان، نادقیق بودن و غیرخطی بودن بوده‌اند که در زمینه‌های مختلف علمی و مهندسی مانند داده‌کاوی، مهندسی کنترل، بازشناسی الگو و رباتیک استفاده می‌شوند [۱]. در این سیستم‌ها، روابط بین متغیرها با استفاده از قوانین اگر-آنگاه بیان می‌شود. سیستم‌های مبتنی بر قوانین فازی همچنین به عنوان سیستم‌های استنتاج فازی نیز شناخته شده‌اند . اگر این سیستم‌ها برای موارد خاصی استفاده شوند نیز نام‌های خاصی به آن‌ها تعلق می‌گیرد مانند: کنترل‌کننده‌های فازی [۱]. یکی از مدل‌های فازی مبتنی بر قانون مدل فازی (*FRBS*[۱])، مدل تاکاگی سوگنو[۲]است. در این مدل، قسمت مقدم به صورت گزاره فازی است ولی قسمت تالی آن یک تابع از ورودی‌ها است.

قوانین فازی در ابتدا توسط افراد خبره با استفاده از فرآیندهای مهندسی‌دانش بدست می‌آمد. با این وجود این رویکرد ممکن است در مواجه با کارهای پیچیده یا زمانی که افراد خبره در دسترس نباشد، امکان پذیر باشد. یک جایگزین موثر برای تولید مدل *FRBS* به صورت خودکار از داده‌ها با استفاده از روش‌های یادگیری است. روش‌های یادگیری بسیار زیادی مانند افرازفضا [۲]، روش‌های‌ابتکاری [۳]، تکنیک‌های‌فازی‌عصبی [۴، ۵]، رویکردهای‌خوشه بندی [۶]، الگوریتم ژنتیک [۱، ۶، ۷].

در این مقاله ما قوانین فازی را براساس روش یادگیری با الگوریتم‌های مختلف خوشه‌بندی روی چندین مجموعه داده رگرسیون بدست می‌آوریم و برای تبدیل ویژگی‌ها به ویژگی‌های جدید و همچنین انتخاب بهترین آن‌ها، از روش‌های استخراج ویژگی استفاده می‌کنیم. در استخراج ویژگی هدف از تصویرکردن، نگاشت الگوهایی از فضای p بعدی به q بعد است، درحالی که ساختارداده‌ها حفظ شده است. در این مقاله از یکی از روش‌های استخراج ویژگی به نام روش تحلیل مولفه‌ی اصلی[۳] استفاده شده است و سعی کرده‌ایم با استفاده از مفاهیم فازی و همبستگی روشی جدید ارائه دهیم.

روش تحلیل مولفه‌ی اصلی یک روش استخراج ویژگی رایج در علوم‌داده‌ها است. در عمل *PCA* بردارهای ویژه ماتریس کواریانس را با توجه به بیشترین مقادیر ویژه آن‌ها پیدا می‌کند و سپس از آن ها برای تصویر کردن داده‌ها در زیرفضای جدیدی با تعداد

[۱] Fuzzy Rule Base System

[۲] Takagi-Sugeno

[۳] Principal Component Analysis

تاریخ دریافت: ۱۳۹۷/۱۰/۲۶ تاریخ پذیرش: ۱۳۹۸/۱۰/۰۸

ابعاد یکسان یا کمتر استفاده می‌کند [۸، ۹، ۱۰]. همچنین می‌توان در روش‌های فازی به جای استفاده از ماتریس کواریانس، از ماتریس همبستگی فازی نیز استفاده کرد و داده ها را به فضای دیگری نگاشت کرد. در سال‌های اخیر مجموعه‌های فازی مردد در یادگیری ماشین کاربردهای جالبی داشته‌اند [۱۱]. از این رو در این مقاله یک روش برای بدست آوردن ماتریس همبستگی فازی مردد روی ویژگی‌ها ارائه می‌شود که از آن برای استخراج ویژگی استفاده می‌کنیم. سپس با توجه به الگوریتم‌های خوشه-بندی و روش‌های مختلف استخراج ویژگی و کاهش بعد، میانگین خطای مجذور مربعات را برای تعدادی مجموعه داده رگرسیون بدست آورده و با هم مقایسه کردیم.

با توجه به هدف این مقاله در بخش دوم ابتدا مروری بر کارهای مرتبط با تحقیق صورت می‌گیرد. در بخش سوم روش‌های خوشه‌بندی و رگرسیون فازی و در بخش چهارم همبستگی فازی مردد توضیح داده می‌شود. بخش ۵ روش‌های مختلف استخراج ویژگی و کاهش بعد داده‌ها مطرح خواهد شد. در بخش ۶ روش پیشنهادی معرفی می‌شود. در بخش هفتم نتایج بدست‌آمده از الگوریتم‌ها با یکدیگر مقایسه خواهند شد و در نهایت بخش هشتم نتایج و اهمیت به کارگیری این روش را بیان می‌کند.

مروری بر کارهای انجام شده

مجموعه‌های فازی مردد نسخه ای از مجموعه‌های فازی هستند که عدم اطمینان ناشی از تردید را زمانی که لازم است درجه عضویت یک عنصر را به یک مجموعه فازی اختصاص دهیم نشان می‌دهند [۱۲].

در سال‌های اخیر کارهای زیادی به منظور کاهش بعد و استخراج ویژگی بر اساس مجموعه‌های فازی و فازی مردد در کنار مفاهیمی مانند همبستگی فازی مردد و تحلیل مولفه‌های اصلی فازی انجام گرفته‌اند [۸، ۱۱]. در روش حسین زاده، از مفهوم تحلیل مولفه‌های اصلی که به کمک خوشه‌بندی فازی به صورت فازی تغییر یافته است برای استخراج ویژگی و سپس طبقه‌بندی داده‌های نامتوازن استفاده شده است [۸]. ابراهیم‌پور یک روش انتخاب ویژگی بر اساس انرژی اطلاعاتی مجموعه‌های فازی مردد[۴] و مفهوم همبستگی فازی مردد ارائه کرده است [۱۱]. که روش پیشنهادی روی مسائل انتخاب ویژگی از نظر دقت طبقه بندی و کاهش ابعاد در مجموعه داده‌هایی با ابعاد بسیار بالا کارا بوده است [۱۱].

نسخه‌های دیگر از مجموعه‌های فازی مانند مجموعه‌های فازی شهودی وجود دارند و همچنین نسخه‌های توسعه یافته آنها هم معرفی شده‌اند [۱۲]. تحلیل مولفه‌های اصلی بر اساس مجموعه-های فازی شهودی بازه ای مقدار از جمله کارهای دیگری بوده است که در سال‌های اخیر به منظور کاهش بعد و تصمیم گیری پیشنهاد شده است [۱۳]. همچنین تحلیل مولفه ای اصلی بر مبنای مجموعه‌های فازی نوع دوم اخیرا برای تشخیص سرطان به کار گرفته شده است [۱۴].

روش های خوشه بندی و رگرسیون فازی

خوشه بندی یکی از شاخه‌های یادگیری بدون نظارت می‌باشد و فرآیندی است که در آن الگوهای مشابه به یک دسته اختصاص داده می‌شوند که به این دسته‌ها خوشه[۵] می‌گویند. بنابراین در هر خوشه مجموعه‌ای از نمونه‌ها با بیشترین شباهت به یکدیگر و کمترین شباهت به دیگر خوشه‌ها قرار دارند. اساس استفاده از هر تکنیک خوشه بندی، محاسبه اندازه‌گیری شباهت یا فاصله بین الگوهای مربوطه است که برای بررسی میزان شباهت می‌توان معیارهای مختلفی را در نظرگرفت ; مثلا می‌توان معیار فاصله را برای خوشه‌بندی مورد استفاده قرار داد و اشیائی را که به یکدیگر نزدیکتر هستند را به عنوان یک خوشه در نظر گرفت که به این نوع خوشه‌بندی، خوشه‌بندی مبتنی بر فاصله[۶] می‌گویند. همچنین می‌توان از دیگر روش‌ها مانند ضرب نقطه‌ای[۷] و تانی موتو[۸] برای بدست آوردن معیار شباهت استفاده‌کرد. در خوشه-بندی کلاسیک، هر داده فقط متعلق به یک خوشه است و نمی-تواند عضو خوشه‌ای دیگر باشد. الگوریتم k میانگین یک الگوریتم خوشه بندی کلاسیک است. این الگوریتم برخلاف سادگی آن یک روش پایه در بسیاری از مسائل خوشه بندی است و کاربرد فراوانی‌دارد.

مراحل الگوریتم k میانگین عبارت انداز [۱۵]:

۱) k داده را به صورت تصادفی به عنوان مرکز خوشه انتخاب می‌کنیم.

۲) ماتریس عضویت u را که مقادیر مولفه‌های آن صفر و یک است را با استفاده از فرمول زیر تعیین می‌کنیم:

$$u_{ij} = \begin{cases} 1 & \text{اگر } \| x_j - c_k \| \le \| x_j - c_i \| \text{ برای هر } i \ne k \\ \cdot & \text{درغیراینصورت} \end{cases} \qquad (۱)$$

۳) تابع هزینه J را با استفاده از فرمول زیر محاسبه می‌کنیم:

$$J(C) = \sum_{i=1}^{k} \sum_{j=1}^{N} \| x_j - c_i \|_2^2 \; u_{ij}. \qquad (۲)$$

[۴] Information of Energy of Hesitant Fuzzy Set

[۵] Cluster

[۶] Distance-based Clustering

[۷] Dot Product

[۸] Tanimoto

که مقدار آن اگر از یک حدی کمتر شد ویا پیشرفت آن نسبت به حالت قبل کمتر از مقدار مشخصی باشد، الگوریتم متوقف می‌شود در غیر این صورت ادامه می-دهیم و به مرحله بعد می‌رویم.

۴) مراکز خوشه را به روزرسانی می‌کنیم و به مرحله ۲ می‌رویم.

در بعضی از حالات ممکن است میزان شباهت داده‌ای در یک خوشه با خوشه‌های دیگر یکسان باشد در این حالت می‌توان از خوشه‌بندی فازی استفاده‌کرد. خوشه بندی فازی یک روش جایگزین برای الگوریتم‌های خوشه‌بندی متعارف یا سخت است که درجه عضویت داده‌ها به خوشه‌ها در آن به صورت فازی تعیین می‌شود. خوشه بندی فازی با خوشه بندی سخت توسط ماهیت غیرخطی آن و انعطاف پذیری در گروه بندی داده‌های عظیم در مقابله است. این خوشه‌بندی راه‌حل‌های دقیق تر و نزدیک به ماهیتی را برای خوشه‌ها فراهم می‌کند [۶، ۷، ۱۵].

در این روش می‌توان میزان تعلق هر نمونه به هرخوشه را با استفاده از درجه عضویت مشخص‌کرد.

الگوریتم C-میانگین فازی [۹]

الگوریتم C-میانگین یک الگوریتم برای خوشه‌بندی داده‌ها است. FCM یک روش بدون نظارت است و یکی از مزایای اصلی آن، مدیریت داده‌های پرت به صورت کارا و موثر است [۱۵]. هدف این الگوریتم حداکثرسازی فشردگی درون کلاس و جداسازی بیرون کلاسی است. با این حال الگوریتم FCM به انتخاب مقادیر اولیه بسیار حساس است و در مورد مجموعه داده‌های بزرگ نیازمند زمان همگرایی طولانی است [۱۵].

فرض کنید هدف برنامه یک مجموعه داده N تایی است که $X=\{x_k\}^{N}_{k=1}$ نمایش داده می‌شود. هر داده x_k و مرکز خوشه v_i در فضای R^l قرار دارند که در آن l بعد داده است. الگوریتم خوشه بندی FCM یک الگوریتم کلاسیک از روش‌های خوشه‌بندی براساس توابع هدف است. در این روش برای مشخص کردن میزان تعلق هر داده به هر کلاس، عضویت فازی را اعمال می‌کند و سپس نتایج خوشه نهایی را از طریق ماتریس عضویت به دست می‌آورد.

در واقع تابع هدف به حداقل می‌رسد و هر داده به هر خوشه‌ای که توسط درجه عضویت تعیین می‌شود، اختصاص‌داده‌می‌شود. تابع هدف نشان دهنده‌ی فاصله‌ی بین یک نمونه داده شده و یک مرکز خوشه است [۱۵]. تابع هدف مینیمم الگوریتم FCM به صورت زیر بیان می‌شود:

$$\min_{v_i} J=\sum_{i=1}^{c}\sum_{j=1}^{n}u_{ij}^{m}d^2=\sum_{i=1}^{c}\sum_{j=1}^{n}u_{ij}^{m}\parallel x_j-v_i\parallel^2. \quad (۳)$$

در فرمول فوق x_j داده j ام و v_i مرکز خوشه i ام است که میزان تشابه را نشان می‌دهد. m یک عدد حقیقی بزرگتر از یک است. با استفاده از u_{ij} می‌توان ماتریسی با c سطر و n ستون به دست آورد که هر یک از مولفه‌های آن عددی بین صفر و یک است. در این ماتریس مجموع مقادیر هر ستون برابر یک است.

بااستفاده از شرط فوق و تابع هدفی که گفته شد، می‌توان u_{ij} و v_i را با استفاده از فرمول‌های زیر بدست‌آورد:

$$v_i=\frac{\sum_{j=1}^{n}u_{ij}^{m}x_j}{\sum_{j=1}^{n}u_{ij}^{m}}, \quad (۴)$$

$$u_{ij}=\frac{1}{\sum_{k=1}^{c}(\frac{d_{ij}}{d_{kj}})^{2/(m-1)}}. \quad (۵)$$

مراحل اجرای الگوریتم به صورت زیر است [۱۵]:

۱. ایجاد ماتریس تعلق با مقادیر تصادفی بین صفر و یک.
۲. محاسبه مراکز خوشه‌ها با استفاده از v_i.
۳. محاسبه تابع هزینه با استفاده از فرمول J.
مقدار آن اگر از یک حدی کمتر شد و یا پیشرفت آن نسبت به حالت قبل کمتر از مقدار مشخصی باشد، الگوریتم متوقف می‌شود در غیر این صورت ادامه می-دهیم و به مرحله بعد می‌رویم.

۴. محاسبه u جدید با استفاده از فرمول u_{ij} و رفتن به مرحله ۲.

الگوریتم خوشه بندی کاهشی [۱۰]

الگوریتم دیگری به نام الگوریتم خوشه‌بندی کاهشی پیشنهاد شد که یک الگوریتم کاربردی است و در آن برای انتخاب مراکز خوشه‌ها از خود داده‌ها استفاده می‌کند. از آنجا که هر نقطه داده یک نامزد برای مراکز خوشه است، معیار چگالی برای هر داده x_i به صورت زیر محاسبه می‌شود [۶، ۷، ۱۵].

داده ای که بیشترین مقدار چگالی را دارا باشد به عنوان مرکز خوشه انتخاب می‌شود.

$$D_i=\sum_{j=1}^{n}\exp\left(-\frac{\|x_i-x_j\|^2}{(r_a/2)^2}\right). \quad (۶)$$

۹ Fuzzy C-means Algorithm

۱۰ Subtractive Clustering Algorithm

یک ثابت معین است که نشان دهنده‌ی شعاع همسایگی r_a است. سپس برای بدست آوردن بقیه‌ی مراکز خوشه‌ها، از فرمول اصلاح شده چگالی استفاده می‌کنیم [۶، ۷، ۱۴]:

$$D_i = D_i - D_{c1}\sum_{j=1}^{n} \exp\left(-\frac{\|x_i - x_{c1}\|^2}{(r_b/2)^2}\right), \qquad (۷)$$

x_{c1} مرکز خوشه انتخاب شده براساس رابطه (۵) است و D_{c1}مقدار این رابطه به ازای $x_i = x_{c1}$ است r_b شعاع کاهش چگالی خوشه‌های قبلی است و باعث می‌شود که تابع چگالی در حوالی x_{c1} با شعاع r_b کاهش پیدا کند. معمولا r_b برابربا ۱/۵ برابر r_a است. در این فرمول میزان چگالی مراکزخوشه‌های قبلی به صفر کاهش پیدا می‌کند.

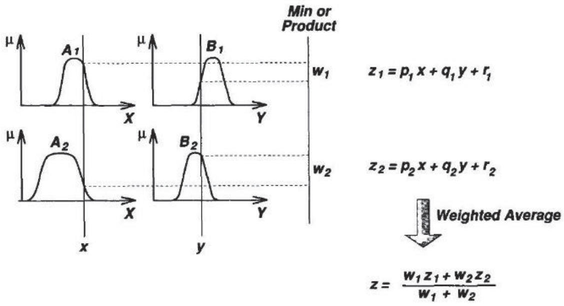
مدل های رگرسیون فازی براساس خوشه بندی

مدل فازی سوگنو توسط تاکاگی، سوگنو و کانگ معرفی‌شد تا یک رویکرد سیستماتیک برای تولید قوانین فازی از یک مجموعه داده ورودی-خروجی ایجاد کند. یک مدل فازی سوگنو به فرم زیر می‌باشد [۶، ۷، ۱۵]:

$$\text{if } x \text{ is } A \text{ and } y \text{ is } B, \text{ then } z = f(x,y), \qquad (۸)$$

که A و B مجموعه‌های‌فازی در قسمت مقدم است در حالی‌که z یک تابع دقیق در قسمت تالی است که می‌تواند یک مقدار ثابت یا یک ترکیبی از x وy باشد. اگر $f(x,y)$ به صورت یک چندجمله ای مرتبه اول باشد به آن مدل فازی سوگنو مرتبه اول[۱۱] می‌گویند و اگر f یک ثابت باشد به آن مدل فازی سوگنو مرتبه صفر[۱۲]می‌گویند.

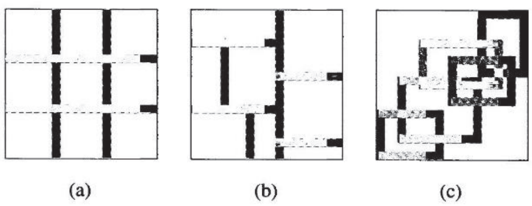
شکل (۱) روش استدلال فازی برای یک مدل فازی سوگنو مرتبه اول را نشان‌می‌دهد. چون هر قانون خروجی دقیقی دارد، خروجی کلی از طریق میانگین وزنی [۱۳] به دست می‌آید در نتیجه نیازی به مرحله فازی‌زدایی ندارد.



شکل ۱. مدل فازی سوگنو [۱۵]

ایده سیستم‌های استنتاج فازی شبیه ایده "تقسیم مسئله[۱۴] " می‌باشد. مقدم یک قاعده فازی در واقع یک ناحیه فازی محلی تعریف کرده و تالی آن قاعده رفتار محلی سیستم در آن ناحیه را مشخص‌می‌کند. تالی می‌تواند یک تابع عضویت (مدل ممدانی[۱۵] و تیسوکاموتو [۱۶]) ، یک مقدار ثابت (مدل سوگنو مرتبه صفر) ، معادله خطی (مدل سوگنو مرتبه یک) و یا معادله غیر خطی (مدل سوگنو مرتبه بالاتر) باشد. تفاوت در تالی‌ها منجر به سیستم‌های استنتاج فازی مختلف می‌شود، اما در هر حال مقدم ها همچنان یکسان خواهند بود. بنابراین بخش‌بندی فضای ورودی برای هر سه نوع سیستم استنتاج فازی یکسان است [۱۵].

در مدل فازی، برای افراز فضای ورودی چندین روش مختلف مانند بخش‌بندی شبکه‌ای[۱۷]، بخش بندی درختی[۱۸] و بخش بندی پراکندگی[۱۹] وجود دارد. شکل(۲) روش‌های مختلف بخش‌بندی فضای ورودی را نشان می‌دهد.



شکل ۲. انواع روش‌های مختلف افراز فضای ورودی. *(a)* افراز شبکه،*(b)* افراز درختی ، *(c)* افراز پراکندگی [۱۵].

معمولا در خوشه‌بندی برای افراز فضای ورودی از روش افراز براساس پراکندگی استفاده می‌شود.در این روش افراز براساس داده‌ها است. برای مثال اگر یک خوشه‌بندی به صورت فازی با توجه به شکل (۳) داشته باشیم که توسط الگوریتم فازی -Cمیانگین ایجاد شده باشد، با توجه به داده‌ها دو خوشه بدست می‌آید . در این روش هر ناحیه یک قانون فازی تولید می‌کند.

---

[۱۱] First-order Sugeno Fuzzy Model

[۱۲] Zero-order Sugeno Fuzzy Model

[۱۳] Weighted Average

[۱۴] Divide and conquer

[۱۵] Mamdani

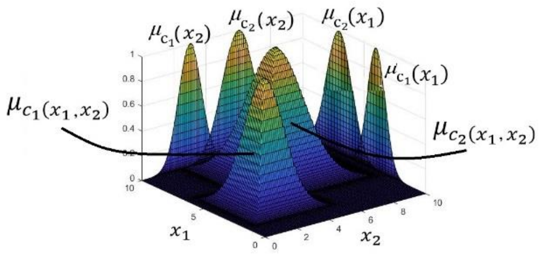
[۱۶] Tsukamoto

[۱۷] Grid Partitioning

[۱۸] Tree Partitioning

[۱۹] Scatter Partitioning

بعد از تصویر کردن خوشه‌های فازی روی محورها، توابع عضویت بدست می‌آیند. در افراز به صورت پراکندگی خوشه‌هایی که در فضا ایجاد می‌شوند با یکدیگر تداخل[20] دارند و همچنین فضای توابع عضویت بر هم عمود نیستند. در این مقاله برای تولید مدل‌های فازی سوگنو از روش خوشه‌بندی فازی استفاده می‌کنیم.



شکل ۳. افراز فضای ورودی با کمک افراز پراکندگی برای خوشه‌بندی فازی با دو ورودی و دو خوشه.

همبستگی فازی مردد[21]

تا به حال سیستم‌های استنتاج فازی با مجموعه‌های فازی که توسط زاده معرفی شد مدل می‌شدند. بعد از بیان این موضوع بسط‌های مختلفی از مجموعه فازی خالص مانند مجموعه فازی شهودی [۱۳]، نوع ۲ [۱۴، ۱۶] و مجموعه فازی مردد [۱۲، ۱۷] بوجود آمد.

در مجموعه‌های فازی مردد که توسط تورا معرفی‌شد، او استدلال کرد که درجه عضویت‌ها در یک مجموعه فازی به جای اینکه فقط یک مقدار داشته باشند می‌توانند به صورت بردار باشند [۱۲، ۱۷]. HFS[22]ها بسطی از مجموعه‌های فازی هستند که عدم اطمینان ناشی از تردید را زمانی که لازم است درجه عضویت یک عنصر را به یک مجموعه فازی اختصاص‌دهیم نشان‌می‌دهند.

در این بخش برخی اطلاعات اساسی که در این روش پیشنهادی استفاده شده است گفته خواهد شد.

تعریف ۱: فرض کنید X یک مجموعه مرجع باشد. یک مجموعه فازی مردد $A(HFS)$ روی X به صورت تابع $h_A(x)$ تعریف می‌شود، و زمانی که روی X اعمال می‌شود، یک زیر مجموعه متناهی از [۰،۱] برمی‌گرداند. یعنی،

$$A = \{\langle x, h_A(x)\rangle \mid x \in X\}, \qquad (۹)$$

که $h_A(x)$ مجموعه مقادیر ممکن بین بازه صفر و یک است. در واقع $h_A(x)$ عناصر فازی مردد[23] (HFE) نام‌گذاری می‌شود.

تعریف ۲: برای دو مجموعه HFS معمولی، همبستگی بین دو HFS به صورت زیر تعریف می‌شود [۱۱،۱۲،۱۷]:

$$C_{HFS}(A,B) = \sum_{i=1}^{n}\left(\frac{1}{l_i}\sum_{j=1}^{l_i} h_{A\sigma(j)}(x_i)h_{B\sigma(j)}(x_i)\right), \qquad (۱۰)$$

که در آن n اندازه‌ی مجموعه مرجع، l_i تعداد افراد خبره و $h_{A\sigma(j)}(x_i)$، j امین عنصر از iامین عضو مجموعه مرجع در HFS مربوطه است.

تعریف ۳ : برای دو HFS معمولی A و B، ضریب همبستگی بین آن‌ها از فرمول زیر بدست می‌آید [۱۱،۱۲،۱۷]:

$$\rho_{HFS}(A,B) = \frac{C_{HFS}(A,B)}{[C_{HFS}(A,A)]^{(1/2)}.[C_{HFS}(B,B)]^{(1/2)}} \qquad (۱۱)$$

$$= \frac{\sum_{i=1}^{n}\left(\frac{1}{l_i}\sum_{j=1}^{l_i} h_{A\sigma(j)}(x_i)h_{B\sigma(j)}(x_i)\right)}{[\sum_{i=1}^{n}(\frac{1}{l_i}\sum_{j=1}^{l_i} {h_{A\sigma(j)}}^2(x_i)]^{(1/2)}[\sum_{i=1}^{n}(\frac{1}{l_i}\sum_{j=1}^{l_i} {h_{B\sigma(j)}}^2(x_i)]^{(1/2)}}.$$

ضریب همبستگی باید شرایط زیر را برآورده کند:

$$\rho_{HFS}(A,B) = \rho_{HFS}(B,A), \qquad (۱۲)$$

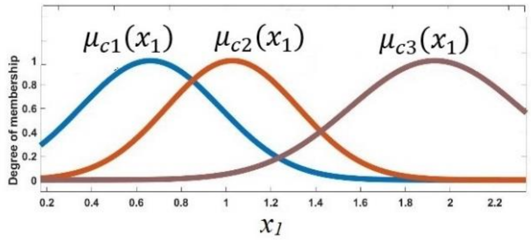
$$0 \le \rho_{HFS}(B,A) \le 1, \qquad (۱۳)$$

$$\rho_{HFS}(A,B) = 1 \quad if\ A = B. \qquad (۱۴)$$

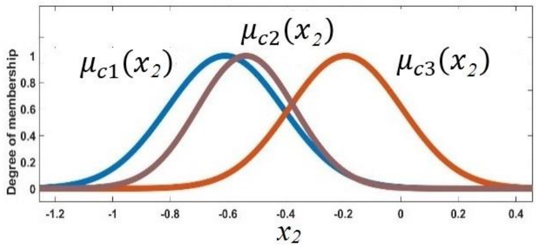
اولین شرط نشان‌می‌دهد که ماتریس ضریب همبستگی باید متقارن باشد. شرط دوم نشان می‌دهد که تمام ضرایب همبستگی در بازه [۰،۱] قراردارند. در نهایت شرط سوم نشان‌می‌دهد که حداکثرمیزان ضریب همبستگی، اگر هر دو HFS شبیه هم باشند، بدست می‌آید [۱۱].

برای بدست آوردن مجموعه‌های فازی مردد روی هر ویژگی ابتدا باید به تعداد مشخصی تابع عضویت تعیین‌کرد که در این مقاله ما بااستفاده از الگوریتم‌های‌خوشه‌بندی فازی روی هر ویژگی ۳ تابع عضویت مشخص‌کردیم. شکل‌های (۴) و (۵) به ترتیب دو ویژگی اول و دوم از یک مجموعه‌داده را با توابع عضویت سه تایی که توسط الگوریتم خوشه‌بندی C-میانگین ایجاد شده است، نشان‌می‌دهند. با توجه به توابع عضویت موجود می‌توان برای هر یک از نمونه داده‌ها در هر ویژگی یک مجموعه فازی مردد تعریف کرد. بنابراین برای هریک از ویژگی‌ها در مجموعه‌داده‌ها ماتریسی به ابعاد تعداد داده‌ها و تعداد توابع عضویت داریم که با استفاده از آن ها می‌توانیم ماتریس ضریب همبستگی مردد را به کمک فرمول (۱۰) بدست‌آوریم.

[20] Overlap
[21] Hesitant Fuzzy Correlation
[22] Hesitant Fuzzy Set
[23] Hesitant Fuzzy Element



شکل ۴. توابع عضویت روی ویژگی اول مجموعه داده Housing



شکل ۵. توابع عضویت روی ویژگی دوم مجموعه داده Housing

با توجه به شکل های بالا دو مجموعه‌ی فازی مردد با توجه به توابع عضویت هر یک داریم:

$$A = \{< x_1, h_A(x_1) > | x_1 \in X\}$$

$$B = \{< x_2, h_B(x_2) > | x_2 \in X\}$$

$$h_A(x_1) = \{\mu_{c_1}(x_1), \mu_{c_2}(x_1), \mu_{c_3}(x_1)\}$$

$$h_B(x_2) = \{\mu_{c_1}(x_2), \mu_{c_2}(x_2), \mu_{c_3}(x_2)\}$$

که $\mu_{c_1}(x_1)$ ، $\mu_{c_2}(x_1)$ و $\mu_{c_3}(x_1)$ به ترتیب نشان دهنده‌ی توابع عضویت فازی حاصل از تصویر کردن خوشه اول، دوم و سوم روی ویژگی اول (X_1) مجموعه داده Housing است. همچنین، $\mu_{c_1}(x_1)$ ، $\mu_{c_2}(x_1)$ و $\mu_{c_3}(x_1)$، به ترتیب، نشان دهنده‌ی توابع عضویت فازی حاصل از تصویر کردن خوشه اول، دوم و سوم روی ویژگی دوم (X_2) مجموعه داده Housing است.

استخراج ویژگی و کاهش بعد داده ها

برای استخراج ویژگی و کاهش بعد داده‌ها می‌توان از روش‌های مختلفی استفاده‌کرد. در این راستا، روش PCA، یکی از مرسوم‌ترین روش‌های کاهش بعد داده‌ها بوده که برای استخراج ویژگی بکار برده می‌شود. در روش PCA، ابتدا توزیع داده به صورت توزیع نرمال تبدیل می‌شود. سپس، ماتریس کواریانس داده‌های آموزشی نرمال شده محاسبه‌گردیده و مقادیر ویژه و بردارهای ویژه مربوط به ماتریس کواریانس، استخراج می‌شوند. در مرحله بعد، به تعداد m بردار ویژه (که $m \leq d$) متناظر با m مقدار ویژه بزرگ ماتریس کواریانس انتخاب می‌شوند. در نهایت، داده‌های آموزشی و تست، بر روی بردارهای ویژه انتخاب شده تصویر شده و بعد داده‌ها کاهش پیدا می‌کند.

در بررسی آماری ساختار داده‌ها، علاوه بر ماتریس کواریانس، ماتریس همبستگی از اهمیت ویژه‌ای برخوردار است. در ادامه، سعی بر این است که مقایسه ای بین ماتریس کواریانس و ماتریس همبستگی در روش PCA انجام دهیم .

۱) یک مشکل بزرگ روش PCA براساس ماتریس کواریانس، حساسیت مولفه‌های اصلی به واحد اندازه‌گیری مورد استفاده برای هر عنصر متعلق بـه X است. اگر تفاوت زیادی بین واریانس عناصر X وجود داشته باشد، آن دسته از متغیرهایی که بزرگترین واریانس را دارند تمایل به تسلط بر چندین مولفه‌ی اصلی اول را دارند. از طرف دیگر، اگر تمام عناصر X در واحدهای مشابه اندازه‌گیری شوند، در این صورت ماتریس کواریانس درست عمل می‌کند. اما در ماتریس همبستگی، استانداردسازی متغیرها به عنوان یک تلاش برای حذف مشکل وابستگی به واحد اندازه‌گیری یا معیار از PCA انجام می‌شود. بنابراین، نتایج حاصل از تجزیه و تحلیل براساس ماتریس همبستگی برای مجموعه‌های مختلف متغیرهای‌تصادفی، به طور مستقیم قابل مقایسه‌تر از آن نتایجی است که براساس ماتریس کواریانس استفاده می‌شوند [۱۰].

۲) توانایی ماتریس کواریانس برای مقایسه نتایج حاصل از تجزیه و تحلیل‌های متفاوت، بسیار کمتر از ماتریس همبستگی است. اندازه واریانس‌های مولفه‌های اصلی، مفهوم مشابهی برای ماتریس‌های وابستگی مختلف در بعدهای یکسان دارند. اما برای ماتریس‌های کواریانس مختلف، این طور نیست [۱۰].

۳) الگوهای‌ضرایب در مولفه‌های اصلی را می‌توان به راحتی برای ماتریس‌های همبستگی مختلف مقایسه کرد و دید که آیا دو ماتریس همبستگی، مولفه‌های اصلی مشابهی ارائه می‌دهند یا خیر؟ در حالی که مقایسه‌ها برای ماتریس کواریانس بسیار پیچیده هستند. با این حال، روش‌های دقیقی برای مقایسه مولفه‌های اصلی ماتریس کواریانس در دسترس هستند [۱۰].

در این مقاله، با توجه به استدلال های اشاره شده در بالا، تصمیم داریم، بجای استفاده از ماتریس کواریانس، تاثیر استفاده از ماتریس همبستگی را در روش PCA و نسخه‌ی فازی آن بررسی کنیم.

فرض کنید، ماتریس کواریانس با نماد Σ نشان‌داده شود. سازوکار روش PCA به صورت زیر فرمول‌بندی می‌شود :

$$\text{PCA:}\quad \max \quad u^T \Sigma u \qquad (۱۵)$$
$$\text{subject to}\quad u^T u = 1.$$

فرض کنید D ماتریس قطری حاصل از درایه‌های قطری ماتریس Σ باشد، یعنی

$$D = \text{diag}(\Sigma) \qquad (۱۶)$$

در این صورت ماتریس D *ماتریسی است که تمام درایه های آن به جز درایه‌های قطر اصلی برابر با صفر و درایه‌های قطر اصلی آن برابر با درایه‌های قطر اصلی ماتریس کواریانس است.*

حال فرض کنید قید $u^T u=1$ به صورت زیر تغییر یابد:

$$u^T D u = 1 \qquad (۱۷)$$

از طرف دیگر تابع هدف $u^T \Sigma u$ به صورت زیر می‌تواند بازنویسی شود:

$$u^T \Sigma u = u^T D^{1/2} D^{-1/2} \Sigma D^{-1/2} D^{1/2} u. \qquad (۱۸)$$

بنابراین مسئله‌ی PCA به صورت زیر می‌تواند نوشته شود:

$$\max \quad u^T D^{1/2} D^{-1/2} \Sigma D^{-1/2} D^{1/2} u \qquad (۱۹)$$
$$\text{subject to}\quad u^T D^{1/2} D^{1/2} u = 1.$$

حال اگر فرض شود $y=D^{(1/2)}u$، در این صورت با توجه به این که ماتریس همبستگی R در رابطه زیر صدق می‌کند:

$$R = D^{-1/2} \Sigma D^{-1/2},$$

آنگاه فرم دیگری از روش PCA براساس ماتریس همبستگی ، به صورت زیر خواهد بود:

$$\max \quad y^T R y$$
$$\text{subject to}\quad y^T y = 1. \qquad (۲۰)$$

اکنون، با توجه به اینکه ماتریس همبستگی برابر است با

$$R = D^{-1/2} \Sigma D^{-1/2}$$

بنابراین، داریم:

$$\max \quad y^T R y$$
$$\text{subject to}\quad y^T y = 1 \qquad (۲۱)$$

اکنون، مشابه با فرآیندی که در ارتباط با استفاده از ماتریس همبستگی در روش PCA گفته شد، می‌توان فرآیندی مشابه را برای نسخه فازی روش PCA استخراج کرد. برای این منظور،

۱) ابتدا، با استفاده از یکی از الگوریتم‌های خوشه‌بندی فازی، بردارهای درجه عضویت برای هر ورودی بدست آورده می‌شوند. سپس در مرحله بعد، با استفاده از این بردارها، ماتریس همبستگی مردد فازی بین ویژگی‌ها ساخته می‌شود.

۲) سپس، مقادیرویژه و بردارهای ویژه مربوط به ماتریس همبستگی مردد فازی، استخراج می‌شوند. در مرحله بعد، به تعداد m بردار ویژه متناظر با m مقدار ویژه بزرگ ماتریس همبستگی مردد فازی انتخاب می‌شوند. در نهایت، داده‌های‌آموزشی بر روی بردارهای ویژه انتخاب شده تصویر شده و بعد داده‌ها کاهش پیدا می‌کند.

روش پیشنهادی

در این بخش نسخه‌ی فازی جدیدی برای استخراج ویژگی در مدل‌های رگرسیون پیشنهاد شده است. برای این منظور، با استفاده از مفهوم مجموعه‌های فازی مردد بیان شده در بخش‌های قبل، یک رویکرد جدید برای محاسبه‌ی همبستگی فازی مردد ارائه شده است تا با توجه به آن با کمک روش PCA تعداد m بردار ویژه متناظر با m مقدار ویژه بزرگ ماتریس همبستگی مردد فازی انتخاب شوند. سپس با کمک ویژگی‌های انتخاب شده، انواع مدل‌های خوشه بندی فازی ایجاد می‌شوند. در نهایت برای ارزیابی هر یک از مدل‌ها، میزان خطای میانگین مربعات و همچنین تعداد قوانین بدست آمده در هر روش محاسبه می‌شوند.

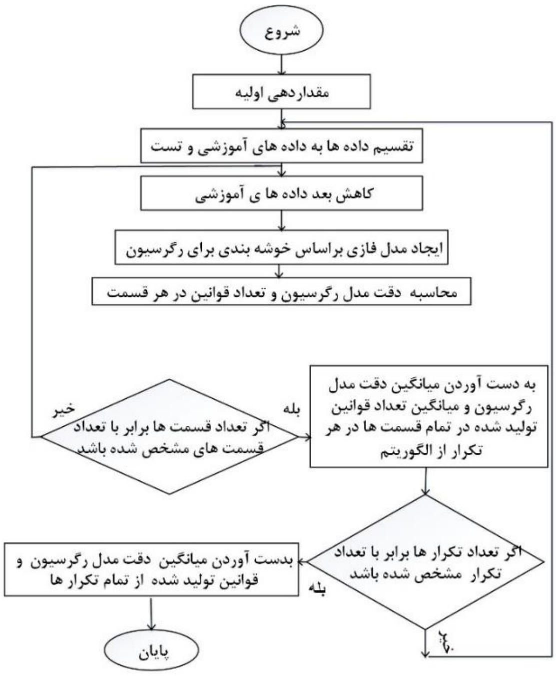
این روش شامل چند بخش است که در شکل(۶) روند نمایی آن نشان داده شده است. حال به توضیح مراحل الگوریتم از روی شکل ۶ می پردازیم:

مقداردهی اولیه :

در مرحله اول مانند تمام الگوریتم‌های دیگر پارامترهایی مانند تعداد تکرار الگوریتم، تعداد قسمت‌های مورد نیاز برای روش اعتبار سنجی متقابل و همچنین تعداد ویژگی‌هایی که توسط کاربر مشخص می‌شود، تعیین می‌شود.

تقسیم داده به داده‌های آموزشی و داده‌های تست با روش اعتبار سنجی متقابل:

در این مرحله داده‌ها با استفاده از روش اعتبارسنجی متقابل به k زیر مجموعه افراز می‌شود تا داده‌های آموزشی و تست بدست آیند. روش کار به این صورت است که هر بار از k زیرمجموعه یک زیرمجموعه برای اعتبارسنجی و k-1 زیر مجموعه دیگر برای آموزش بکارمی‌رود. این روال k بار تکرار می‌شود و همه داده‌ها دقیقا یک بار برای آموزش و یک بار برای اعتبارسنجی بکارمی‌روند. در هر قسمت دقت داده‌های تست بدست می‌آید و بعد از k بار تکرار میانگین k دقت را محاسبه می‌کنیم. در این الگوریتم k برابر با ۵ است.



شکل ۶. روند نمای الگوریتم

استخراج ویژگی وکاهش بعد داده ها:

مشابه با روش PCA کلاسیک، داده‌های آموزشی و تست، بر روی بردارهای ویژه انتخاب شده تصویر شده و بعد داده‌ها کاهش پیدا می‌کند. سپس، m تا از بهترین ویژگی‌ها انتخاب می‌شوند.

ایجاد مدل فازی براساس خوشه بندی برای رگرسیون:

در هر قسمت روی داده های آموزشی کاهش بعد یافته می‌توان از الگوریتم های خوشه بندی مانند الگوریتم خوشه بندی C-میانگین فازی و الگوریتم خوشه بندی کاهشی استفاده‌کرد و یک مدل خوشه بندی فازی ایجاد کرد.

ارزیابی داده های تست با استفاده از مدل ایجاد شده:

با توجه به مدل ایجاد شده در مرحله قبل می‌توان میزان خطای میانگین مربعات بر روی داده‌های تست و همچنین تعداد قوانین ایجاد شده توسط مدل را محاسبه کرد. در نهایت میانگین خطا در ۳۰ بار تکرار لحاظ می‌شود.

نتایج و مقایسه‌ها:

در این مقاله میزان خطای‌میانگین مربعات و تعداد قوانین تولید شده از الگوریتم پیشنهادی توسط هر یک از روش‌های کاهش بعد ذکر شده و الگوریتم‌های خوشه بندی موجود ، ارزیابی و با یکدیگر مقایسه می‌شود.

در جدول (۱) مجموعه داده‌هایی که در این مقاله مورد استفاده قرار گرفته، آورده شده است.حال با توجه به آنچه که گفته‌شد، میزان خطای میانگین مربعات[۲۴] (MSE) محاسبه می‌شود سپس دقت هر روش در هر مجموعه داده با توجه به فرمول $Accuracy=1-MSE$ بدست می‌آید. نتایج بدست‌آمده در جدول (۲) نشان داده‌شده است. همچنین تعداد قوانین در هر روش برای هر مجموعه داده محاسبه می‌شود. سپس برای نرمال سازی آن از فرمول زیر استفاده شد :

$$NNR = 1 - \frac{\text{تعداد قوانین}}{\text{حداکثر تعداد قوانین}} \qquad (۲۲)$$

که تعداد قوانین میانگین تعداد قوانین تولید شده در هر روش روی هر یک از مجموعه داده‌ها است و حداکثر تعداد قوانین بیشترین میانگین تعداد قوانین موجود در هر مجموعه داده است. نتایج حاصل از فرمول (۲۱) در جدول (۳) نشان داده شده است.

این الگوریتم را می‌توان بر اساس انواع روش‌های خوشه‌بندی و همچنین روش‌هایی که برای بدست آوردن بردارها و مقادیر ویژه بیان شد، توصیف کرد:

M۱) الگوریتم خوشه‌بندی کاهشی با استخراج ویژگی با روش آنالیز مولفه‌ی اصلی.

M۲) الگوریتم خوشه‌بندی کاهشی روی مجموعه فازی مردد و استخراج ویژگی توسط روش آنالیز مولفه‌ی اصلی با این تفاوت که به جای استفاده از ماتریس کواریانس از ماتریس ضریب همبستگی مردد ایجاد شده الگوریتم خوشه بندی C-میانگین فازی استفاده می‌شود.

M۳) الگوریتم خوشه بندی کاهشی روی مجموعه فازی مردد و استخراج ویژگی توسط روش آنالیز مولفه‌ی اصلی با این تفاوت که به جای استفاده ازماتریس کواریانس از ماتریس ضریب همبستگی مردد ایجادشده توسط روش خوشه‌بندی کاهشی استفاده می‌شود.

M۴) الگوریتم خوشه بندی C-میانگین فازی با استخراج ویژگی با روش آنالیز مولفه‌ی اصلی.

M۵) الگوریتم خوشه بندی C-میانگین فازی روی مجموعه فازی مردد و استخراج ویژگی توسط روش آنالیز مولفه‌ی اصلی با این تفاوت که به جای استفاده از ماتریس کواریانس از ماتریس ضریب همبستگی مردد ایجادشده توسط روش C-میانگین فازی استفاده می‌شود.

[۲۴] Mean Square Error

M۶) الگوریتم خوشه‌بندی *C*-میانگین فازی با استخراج ویژگی توسط روش آنالیز مولفه‌ی اصلی با این تفاوت که به جای استفاده از ماتریس کواریانس از ماتریس همبستگی استفاده می‌شود.

جدول ۱: مجموعه داده های استفاده شده درمقاله.

مجموعه داده ها	*اسم مختصر*	تعداد ویژگی	تعداد نمونه ها
Residential	D۱	۱۰۸	۳۷۲
Parkinson	D۲	۲۱	۵۸۷۵
Foe	D۳	۱۳	۲۲۷
stock	D۴	۹	۹۵۰
Wizmir	D۵	۹	۱۴۶۱
Default Feature _1059_tracks	D۶	۶۹	۱۰۵۶
Housing	D۷	۱۳	۵۰۶

جدول ۲. میانگین دقت در ۳۰ بار اجرا برای مجموعه داده های مختلف.

روش ها	D1	D۲	D۳	D۴	D۵	D۶	D۷
M۱	۰,۹۷۴۶	۰,۹۹۴۹	۰,۹۵۳۲	۰,۹۹۹۶	۰,۹۸۸۰	۰,۹۶۱۷	۰,۸۶۱۸
M۲	۰,۵۳۶۱	۰,۹۹۴۳	۰,۹۴۷۹	۰,۹۹۸۴	۰,۹۹۰۱	۰,۹۵۸۰	۰,۹۸۳۷
M۳	۰,۲۳۷۵	۰,۹۹۶۶	.۰,۹۲۰۸	۰,۹۹۸۴	۰,۹۸۰۹	۰,۹۶۱۷	۰,۹۸۷۷
M۴	۰,۹۳۶۴	۰,۹۹۱۰	۰,۸۹۷۱	۰,۹۹۲۳	۰,۹۸۸۶	۰,۹۵۲۵	۰,۹۹۸۴
M۵	۰,۹۵۴۷	۰,۹۹۳۰	۰,۸۹۴۰	۰,۹۹۴۰	۰,۹۹۹۷	۰,۹۵۲۹	۰,۹۶۸۸
M۶	۰,۹۵۱۳	۰,۹۹۰۱	۰,۸۹۷۶	۰,۹۹۶۰	۰,۹۹۴۳	۰,۹۵۲۵	۰,۹۶۵۹

جدول ۳. میانگین تعداد قوانین نرمال شده در ۳۰ بار اجرا.

روش ها	D1	D۲	D۳	D۴	D۵	D۶	D۷
M۱	۰	۰,۱۰۳۰	۰,۵۸۰۰	۰,۷۰۳۱	۰,۳۶۶۶	۰,۹۸۳۱	۰,۹۵۲۱
M۲	۰,۳۸۴۱	۰	۰,۵۷۸۰	۰,۵۸۱۷	۰	۰	۰,۹۵۰۵
M۳	۰,۳۶۹۶	۰,۶۰۱۱	۰	۰	۰,۵۴۱۶	۰,۹۸۳۹	۰
M۴	۰,۹۹۰۰	۰,۷۱۱۸	۰,۷۵۴۰	۰,۹۰۹۳	۰,۴۷۹۱	۰,۹۷۹۴	۰,۰۸۹۴
M۵	۰,۹۸۳۴	۰,۷۶۵۲	۰,۸۳۸۰	۰,۹۱۴۴	۰,۴۹۱۶	۰,۹۸۱۰	۰,۹۴۴۰
M۶	۰,۹۸۹۸	۰,۵۲۶۷	۰,۷۶۴۰	۰,۹۰۹۳	۰,۴۷۹۱	۰,۹۸۰۰	۰,۹۴۵۰

براساس جدول (۲) و (۳) می توان میانگین هندسی بین دقت و تعداد قوانین نرمال شده را با استفاده از فرمول زیر بدست آورد:

$$G_{mean} = \sqrt{Accuracy * NNR} \tag{۲۳}$$

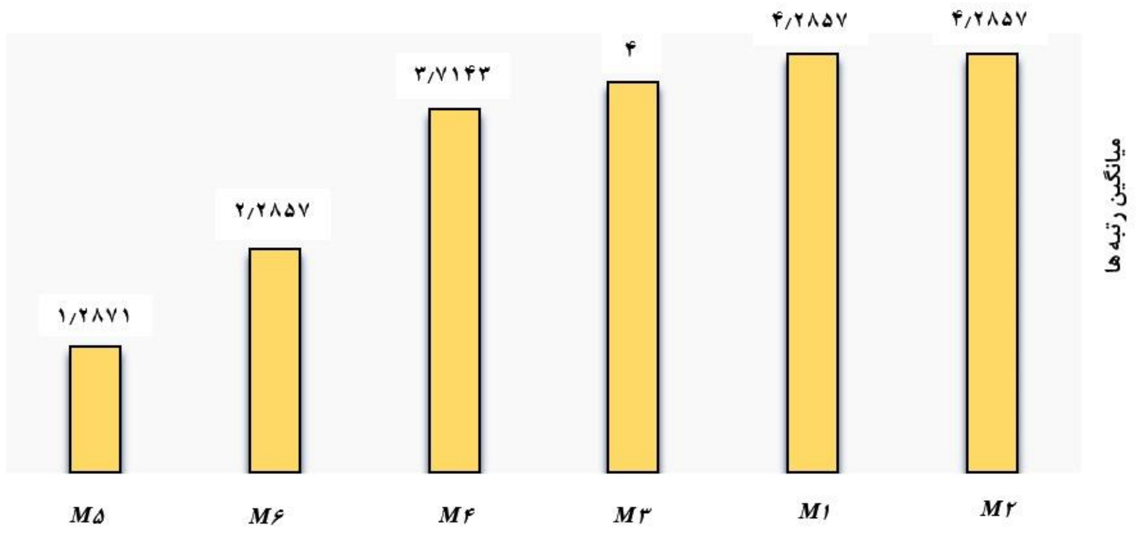
در نتیجه الگوریتمی بهتر است که با توجه به دومعیار (تعداد قوانین کمتر و دقت بیشتر)، بهتر از سایر الگوریتم‌ها عمل کرده باشد.

در جدول (۴) میانگین هندسی برای هر روش در هر مجموعه داده نشان داده شده است.

جدول۴. میانگین هندسی بین دو معیار مذکور.

روش ها	D1	D۲	D۳	D۴	D۵	D۶	D۷
M۱	۰	۰,۳۲۰۱	۰,۷۴۳۰	۰,۸۳۸۳	۰,۶۰۱۸	۰,۹۷۲۳	۰,۹۰۵۸
M۲	۰,۴۵۳۷	۰	۰,۷۴۰۱۲	۰,۷۶۲۰	۰	۰	۰,۹۶۶۹
M۳	۰,۲۹۶۱	۰,۷۷۳۹	۰	۰	۰,۷۲۸۸	۰,۹۷۲۷	۰
M۴	۰,۹۶۲۸	۰,۸۳۹۸	۰,۸۲۲۴	۰,۹۵۰۲۵	۰,۶۸۸۱	۰,۹۶۵۸	۰,۲۹۸۶
M۵	۰,۹۶۸۹	۰,۸۷۱۶	۰,۸۶۵۵	۰,۹۵۳۳	۰,۷۰۰۰	۰,۹۶۶۷	۰,۹۵۶۲
M۶	۰,۹۷۰۳	۰,۷۲۲۰	۰,۸۲۸۰	۰,۹۵۱۶	۰,۶۹۰۱	۰,۹۶۶۱	۰,۹۵۵۳

با توجه به نتایج بدست آمده از جدول(۴)، میانگین هندسی در روش $M5$ که بردارهای ویژه و مقادیر ویژه با روش آنالیز مولفه‌ی اصلی توسط ماتریس همبستگی فازی مردد که با الگوریتم خوشه‌بندی C-میانگین فازی بدست آمده، بیشتر از روش‌های دیگر است. برای مقایسه الگوریتم‌ها ، می‌توان از آزمون‌های آماری مانند آزمون ناپارامتریک فریدمن نیز بهره برد. نتایج این آزمون در نمودار میله ای شکل (۵) نمایش داده شده است.



شکل ۵. نمودار میله ای میانگین رتبه های روش های مختلف روی مجموعه داده های بیان شده.

براساس رتبه‌های موجود در شکل (۵) الگوریتم $M5$ را می‌توان به عنوان روش کنترلی در نظرگرفت چون میانگین رتبه آن از بقیه کمتر(بهتر) شده است.

همچنین مقدار آماره‌ی مربوط به این آزمون نشان می‌دهد که اختلاف این روش‌ها معناداراست . بنابراین برای این که بتوان معنادار بودن اختلاف بین هریک از الگوریتم‌ها با روش کنترلی

M۵ را بررسی‌کرد از آزمون‌های تعقیبی یا پس‌آزمون[۲۵] استفاده شد. بر اساس جدول (۵)، حاصل از بکارگیری آزمون تعقیبی *Li*، فرضیه‌های صفر (*H0*) در مواردی که *P-value* کمتر از ۰٫۰۴۲۱۸۲ باشد، رد می‌شوند. از آنجا که فرضیه صفر به معنی یکسان بودن روش کنترلی با روش موجود در جدول می‌باشد، بنابراین می‌توان‌گفت که روش M۵ با روش‌های M۳ و M۲ و M۱ و M۴ اختلاف معنی‌داری دارد و تنها با روش M۶ اختلاف معنی‌داری ندارد.

جدول ۵. جدول مقایسه *Post hoc*

Li	P	z	Algorithm	i
۰٫۰۴۲۸۲	۰٫۰۱۲۴۱۹	۲٫۵۰۰۰۰	M۳	۵
۰٫۰۴۲۸۲	۰٫۰۱۸۴۱۶	۲٫۳۵۷۱۷۳	M۲	۴
۰٫۰۴۲۸۲	۰٫۰۳۲۱۲۵	۲٫۱۴۲۸۵۷	M۱	۳
۰٫۰۴۲۸۲	۰٫۰۱۱۶۰۸۳	۱٫۵۷۱۴۲۹	M۴	۲
۰/۰۵	۰/۱۹۸۵۴۳	۱/۲۸۵۷۱۴	M۶	۱

نتیجه گیری:

در این تحقیق با استفاده از روش‌های خوشه‌بندی فازی و مفهوم مجموعه‌های فازی مردد و همبستگی فازی مردد، یک روش کارا برای رگرسیون ارائه داده شد که همزمان به کاهش ویژگی‌ها هم می‌پردازد. نتایج آزمایش‌ها نشان‌داد که به کارگیری ماتریس همبستگی و تحلیل مولفه‌های اصلی آن باعث استخراج ویژگی‌های مناسب و کاهش ابعاد و در نتیجه بهبود کارایی مدل‌های رگرسیون می‌شود. همچنین می‌توان این روش را برای انتخاب و استخراج ویژگی در مدل‌های طبقه‌بندی نیز استفاده کرد. می‌توان گفت که این اولین باری نیست که مفاهیم مجموعه‌های فازی مردد در یادگیری ماشین مورد استفاده قرارگرفته‌اند. در روش پیشنهادی این تحقیق، از روش‌های خوشه‌بندی نیز برای ایجاد مجموعه‌های مردد فازی استفاده شده است که نشان می‌دهد همیشه نیاز به شخص خبره انسانی برای ایجاد مجموعه‌های فازی مردد نیست و الگوریتم‌های داده‌کاوی هم می‌توانند چنین نقشی را ایفا کنند. استفاده از ماتریس همبستگی فازی مردد به جای ماتریس کوواریانس و فرمول‌بندی مساله به شکلی که بتوان از ماتریس همبستگی استفاده کرد، از جمله نوآوری‌های اصلی این تحقیق بوده است. نتایج حاصل از ماتریس همبستگی فازی مردد روی داده های مختلف نشان دهنده‌ی کارایی و قابل قبول بودن روش‌های مبتنی بر همبستگی فازی مردد در کاهش بعد و استخراج ویژگی است.

مراجع

[1] F. Aghaeipoor, and M. Eftekhari, "EEFR-R: extracting effective fuzzy rules for regression problems through the cooperation of association rule mining concepts and evolutionary algorithms," Soft Comput., vol. 23, no. 22, pp. 11737–11757, 2019.

[2] L.X. Wang, and J.M. Mendel, "Generating fuzzy rules by learning from examples," IEEE Trans. Syst., pp. 1414-1427, 1992.

[3] H. A. Bazoobandi, and M. Eftekhari, "A differential evolution and spatial distribution based local search for training fuzzy Wavelet neural network," Int. J. Eng., 27, 8, 1185-1194, 2014.

[4] J. JSR, "ANFIS: Adaptive-NetworkBased Fuzzy Inference System," I IEEE Trans. Syst., vol. 23, no. 3, pp. 665-685, 1993.

[۲۵] Post hoc

[5] M. Eftekhari, M. Eftekhari, and H. Karimpour, “Neuro-fuzzy adaptive control of a revolute stewart 2 platform carrying payloads of unknown inertia,” Robotica, vol. 33, no.9, 2001-2024, 2015.

[6] M. Mahdizadeh, and M. Eftekhari, “Generating fuzzy rule base classifier for highly imbalanced datasets using a hybrid of evolutionary algorithms and subtractive clustering,” Journal of Intelligent and Fuzzy Systems, 27, 6, 3033-3046, 2014.

[7] M. Eftekhari, and M. Mahdizadeh, “A novel cost sensitive imbalanced classification method based on new hybrid fuzzy Cost assigning approaches, fuzzy clustering and evolutionary algorithms,” Int. J. Eng. Trans. B, vol. 28, no. 8, pp. 1160-1168, 2015.

[8] M. Hosseinzadeh, and M. Eftekhari, “Using fuzzy under-sampling and fuzzy PCA to improve imbalanced classification through Rotation Forest algorithm,” in: Comput. Sci. Softw. Eng. (CSSE), 2015 Int. Symp., IEEE, pp. 1–7, 2015.

[9] M. Khamar, and M. Eftekhari, “Multi-manifold based rotation forest for classification,” Appl. Soft Comput., vol. 68, pp. 626-635, 2018.

[10] I. Jolliffe, “Principal Component Analysis,” Second Edition ed., Springer, 2002.

[11] M.K. Ebrahimpour, and M. Eftekhari, “Distributed feature selection: A hesitant fuzzy correlation concept for microarray high-dimensional datasets,” Chemom. Intell. Lab. Syst., vol. 173, no. 1, pp. 51-64, 2018.

[12] V. Torra, “Hesitant fuzzy sets,” Int. J. Intell. Syst, vol. 25, pp. 529-539, 2010.

[13] L., Bingsheng et al., “An interval-valued intuitionistic fuzzy principal component analysis model-based method for complex multi-attribute large-group decision-making,” Eur. J. Oper. Res., vol. 245, no. 1, pp. 209-225, 2015.

[14] V. Singh, N. K. Verma and Y. Cui, “Type-2 fuzzy PCA approach in extracting salient features for molecular cancer diagnostics and prognostics,” IEEE Trans. Nanobiosci., vol. 18, no. 3, pp. 482-489, 2019.

[15] C.-T. S. Jyh-Shing Roger Jang, “Neuro-fuzzy and soft computing: A computational approach to learning and machine intelligence,” Upper Saddle River, NJ, USA: Prentice-Hall, Inc, 1997.

[16] J. Mendel, “Uncertain rule-based fuzzy systems,” Springer, pp. 259–306, 2017.

[17] R.M. Rodríguez et al., “Hesitant fuzzy sets: state of the art and future directions,” Int. J. Intell. Syst., vol. 29, no. 6, pp. 495–524, 2014.