

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

مهلا مختیا^۱، مهدی افتخاری^۲، فرید صابری موحد^۳

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

^۲دانشیار، بخش مهندسی کامپیوتر، دانشگاه شهید باهنر کرمان، m.eftkhari@uk.ac.ir

^۳استادیار، گروه ریاضی کاربردی، دانشکده علوم و فناوری های 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 از تصویر کردن تابع عضویت خوشه ها روی ویژگی های مختلف، به تعداد خوشه ها مجموعه های فازی روی هر بعد (یا ویژگی) بدست می آید. سپس این مجموعه های فازی را به صورت یک مجموعه فازی مردد روی هر ویژگی در نظر گرفته و ماتریس ضریب همبستگی فازی مردد را برای ویژگی ها به دست می آوریم. در ادامه یک نگاشت غیرخطی بر اساس تجزیه مولفه های اصلی این ماتریس برای تبدیل ویژگی های مجموعه داده به ویژگی های جدید استفاده شده است. در پایان، ویژگی های استخراج شده جدید را به الگوریتم خوشه بندی فازی داده و یک سیستم فازی سوگنو به منظور رگرسیون برازش شده است. روش پیشنهادی با چندین روش دیگر روی چندین مجموعه داده رگرسیون مقایسه شده است. نتایج آزمایش ها نشان دهنده موفق بودن روش پیشنهادی در استخراج و کاهش ویژگی ها 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

دسترس نباشد، امکان پذیر باشد. یک جایگزین موثر برای تولید مدل $FRBS$ به صورت خودکار از داده ها با استفاده از روش های یادگیری است. روش های یادگیری بسیار زیادی مانند افزایش [۲]، روش های ابتکاری [۳]، تکنیک های فازی عصبی [۴، ۵]، رویکردهای خوشه بندی [۶]، الگوریتم ژنتیک [۱، ۶، ۷]. در این مقاله ما قوانین فازی را بر اساس روش یادگیری با الگوریتم های مختلف خوشه بندی روی چندین مجموعه داده رگرسیون بدست می آوریم و برای تبدیل ویژگی ها به ویژگی های جدید و همچنین انتخاب بهترین آن ها، از روش های استخراج ویژگی استفاده می کنیم. در استخراج ویژگی هدف از تصویر کردن، نگاشت الگوهایی از فضای p بعدی به q بعد است، در حالی که ساختار داده ها حفظ شده است. در این مقاله از یکی از روش های استخراج ویژگی به نام روش تحلیل مولفه اصلی^۳ استفاده شده است و سعی کرده ایم با استفاده از مفاهیم فازی و همبستگی روشی جدید ارائه دهیم.

روش تحلیل مولفه اصلی یک روش استخراج ویژگی رایج در علوم داده ها است. در عمل PCA بردارهای ویژه ماتریس کواریانس را با توجه به بیشترین مقادیر ویژه آن ها پیدا می کند و سپس از آن ها برای تصویر کردن داده ها در زیر فضای جدیدی با تعداد

سیستم های مبتنی بر قوانین فازی روش شناخته شده ای در محاسبات نرم برای رسیدگی به مسائل پیچیده دنیای واقعی هستند. آن ها روش قدرتمندی در مقابله با مسائل مختلف از جمله عدم اطمینان، نادقیق بودن و غیرخطی بودن بوده اند که در زمینه های مختلف علمی و مهندسی مانند داده کاوی، مهندسی کنترل، بازشناسی الگو و رباتیک استفاده می شوند [۱]. در این سیستم ها، روابط بین متغیرها با استفاده از قوانین اگر-آنگاه بیان می شود. سیستم های مبتنی بر قوانین فازی همچنین به عنوان سیستم های استنتاج فازی نیز شناخته شده اند. اگر این سیستم ها برای موارد خاصی استفاده شوند نیز نام های خاصی به آن ها تعلق می گیرد مانند: کنترل کننده های فازی [۱]. یکی از مدل های فازی مبتنی بر قانون مدل فازی ($FRBS$)، مدل تاکاگی سوگنو است. در این مدل، قسمت مقدم به صورت گزاره فازی است ولی قسمت تالی آن یک تابع از ورودی ها است.

قوانین فازی در ابتدا توسط افراد خبره با استفاده از فرآیندهای مهندسی دانش بدست می آمد. با این وجود این رویکرد ممکن است در مواجهه با کارهای پیچیده یا زمانی که افراد خبره در

^۳ Principal Component Analysis

^۱ Fuzzy Rule Base System

^۲ Takagi-Sugeno

معرفی شده‌اند [۱۲]. تحلیل مولفه‌های اصلی بر اساس مجموعه‌های فازی شهودی بازه ای مقدار از جمله کارهای دیگری بوده است که در سال‌های اخیر به منظور کاهش بعد و تصمیم‌گیری 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 نظر گرفت؛ 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 استفاده کرد. در خوشه‌بندی کلاسیک، هر داده فقط متعلق به یک خوشه است و نمی‌تواند 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 فراوانی دارد.

مراحل الگوریتم k میانگین عبارت انداز [۱۵]:

(۱) 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:

$$u_{ij} = \begin{cases} 1 & \text{اگر } \|x_j - c_i\| \leq \|x_j - c_k\| \text{ برای هر } k \neq i \\ 0 & \text{در غیر این صورت} \end{cases} \quad (1)$$

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

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

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 داده‌ها 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 خوشه‌بندی 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 بدست آمده 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

^۷ Dot Product
^۸ Tanimoto

^۴ Information of Energy of Hesitant Fuzzy Set
^۵ Cluster
^۶ Distance-based Clustering

$$\min_{v_i} J = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m d^2 = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m \|x_j - v_i\|^2. \quad (3)$$

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

$$v_i = \frac{\sum_{j=1}^n u_{ij}^m x_j}{\sum_{j=1}^n u_{ij}^m}, \quad (4)$$

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{kj}}\right)^{2/(m-1)}}. \quad (5)$$

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 می‌دهیم 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.

$$D_i = \sum_{j=1}^n \exp\left(-\frac{\|x_i - x_j\|^2}{(r_a/2)^2}\right). \quad (6)$$

که 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 می‌دهیم 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 استفاده کرد. 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 هر خوشه را با استفاده از درجه عضویت مشخص کرد.

الگوریتم C-میانگین فازی^۹

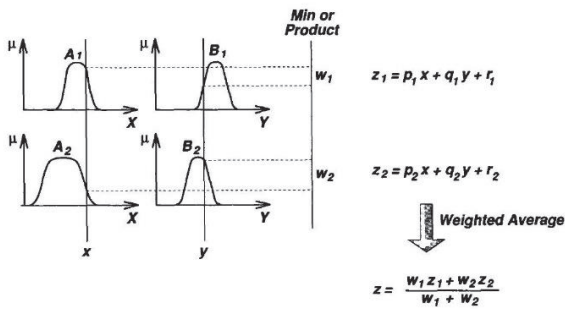
الگوریتم C-میانگین یک الگوریتم برای خوشه‌بندی داده‌ها است. FCM یک روش بدون نظارت است و یکی از مزایای اصلی آن، مدیریت داده‌های پرت به صورت کارا و موثر است [۱۵]. هدف این الگوریتم حداکثرسازی فشردگی درون کلاس و جداسازی بیرون کلاسی است. با این حال الگوریتم FCM به انتخاب مقادیر اولیه بسیار حساس است و در مورد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

فرض کنید هدف برنامه یک مجموعه داده N تایی است که $X = \{x_k\}_{k=1}^N$ 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، اختصاص داده می‌شود. 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:

^{۱۰} Subtractive Clustering Algorithm

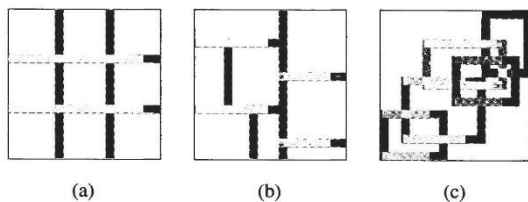
^۹ Fuzzy C-means Algorithm



شکل ۱. مدل فازی سوگنو [۱۵]

ایده سیستم های استنتاج فازی شبیه ایده "تقسیم مسئله" ۱۴ می باشد. مقدم یک قاعده فازی در واقع یک ناحیه فازی محلی تعریف کرده و تالی آن قاعده رفتار محلی سیستم در آن ناحیه را مشخص می کند. تالی می تواند یک تابع عضویت (مدل ممدانی^{۱۵} و تیسو کاموتو^{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 تالی ها منجر به سیستم های استنتاج فازی مختلف می شود، اما در هر حال مقدم ها همچنان یکسان خواهند بود. بنابراین بخش بندی فضای ورودی برای هر سه نوع سیستم استنتاج فازی یکسان است [۱۵].

در مدل فازی، برای افزایش فضای ورودی چندین روش مختلف مانند بخش بندی شبکه ای^{۱۷}، بخش بندی درختی^{۱۸} و بخش بندی پراکندگی^{۱۹} وجود دارد. شکل (۲) روش های مختلف بخش بندی فضای ورودی را نشان می دهد.



شکل ۲. انواع روش های مختلف افزایش فضای ورودی. (a) افزایش شبکه، (b) افزایش درختی، (c) افزایش پراکندگی [۱۵].

معمولاً در خوشه بندی برای افزایش فضای ورودی از روش افزایش براساس پراکندگی استفاده می شود. در این روش افزایش براساس داده ها است. برای مثال اگر یک خوشه بندی به صورت فازی با توجه به شکل (۳) داشته باشیم که توسط الگوریتم فازی - C میانگین ایجاد شده باشد، با توجه به داده ها دو خوشه بدست می آید. در این روش هر ناحیه یک قانون فازی تولید می کند.

یک ثابت معین است که نشان دهنده شعاع همسایگی r_a است. سپس برای بدست آوردن بقیه ی مراکز خوشه ها، از فرمول اصلاح شده چگالی استفاده می کنیم [۶، ۷، ۱۴]:

$$D_i = D_i - D_{c1} \sum_{j=1}^n \exp\left(-\frac{\|x_i - x_{c1}\|^2}{(r_b/2)^2}\right), \quad (7)$$

x_{c1} مرکز خوشه انتخاب شده براساس رابطه (۵) است و D_{c1} مقدار این رابطه به ازای $x_i = x_{c1}$ است شعاع کاهش چگالی خوشه های قبلی است و باعث می شود که تابع چگالی در حوالی x_{c1} با شعاع r_b کاهش پیدا کند. معمولاً r_b برابر با $1/5$ برابر r_a است. در این فرمول میزان چگالی مراکز خوشه های قبلی به صفر کاهش پیدا می کند.

مدل های رگرسیون فازی براساس خوشه بندی

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

$$\text{if } x \text{ is } A \text{ and } y \text{ is } B, \text{ then } z = f(x, y), \quad (8)$$

که A و B مجموعه های فازی در قسمت مقدم است در حالی که z یک تابع دقیق در قسمت تالی است که می تواند یک مقدار ثابت یا یک ترکیبی از x و y باشد. اگر $f(x, y)$ به صورت یک چندجمله ای مرتبه اول باشد به آن مدل فازی سوگنو مرتبه اول^{۱۱} می گویند و اگر f یک ثابت باشد به آن مدل فازی سوگنو مرتبه صفر^{۱۲} می گویند.

شکل (۱) روش استدلال فازی برای یک مدل فازی سوگنو مرتبه اول را نشان می دهد. چون هر قانون خروجی دقیقی دارد، خروجی کلی از طریق میانگین وزنی^{۱۳} به دست می آید در نتیجه نیازی به مرحله فازی زدایی ندارد.

^{۱۶} Tsukamoto
^{۱۷} Grid Partitioning
^{۱۸} Tree Partitioning
^{۱۹} Scatter Partitioning

^{۱۱} First-order Sugeno Fuzzy Model
^{۱۲} Zero-order Sugeno Fuzzy Model
^{۱۳} Weighted Average
^{۱۴} Divide and conquer
^{۱۵} Mamdani

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

$$C_{HFS}(A, B) = \sum_{i=1}^n \left(\frac{1}{l_i} \sum_{j=1}^{l_i} h_{A\sigma(j)}(x_i) h_{B\sigma(j)}(x_i) \right), \quad (10)$$

که در آن n اندازه‌ی مجموعه مرجع، l_i تعداد افراد خبره و $h_{A\sigma(j)}(x_i)$ j امین عنصر از i امین 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 [۱۷، ۱۲، ۱۱]:

$$\rho_{HFS}(A, B) = \frac{C_{HFS}(A, B)}{[C_{HFS}(A, A)]^{(1/2)} [C_{HFS}(B, B)]^{(1/2)}} \quad (11)$$

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

ضریب همبستگی باید شرایط زیر را برآورده کند:

$$\rho_{HFS}(A, B) = \rho_{HFS}(B, A), \quad (12)$$

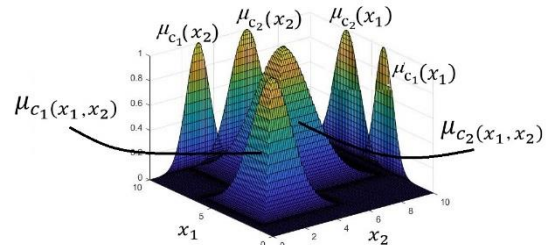
$$0 \leq \rho_{HFS}(B, A) \leq 1, \quad (13)$$

$$\rho_{HFS}(A, B) = 1 \quad \text{if } A = B. \quad (14)$$

اولین شرط نشان می‌دهد که ماتریس ضریب همبستگی باید متقارن باشد. شرط دوم نشان می‌دهد که تمام ضرایب همبستگی در بازه $[0, 1]$ قرار دارند. در نهایت شرط سوم نشان می‌دهد که حداکثر میزان 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 تعیین کرد که در این مقاله ما با استفاده از الگوریتم‌های خوشه‌بندی فازی روی هر ویژگی ۳ تابع عضویت مشخص کردیم. شکل‌های (۴) و (۵) به ترتیب دو ویژگی اول و دوم از یک مجموعه داده را با توابع عضویت سه تایی که توسط الگوریتم خوشه‌بندی C -میانگین ایجاد شده است، نشان می‌دهند. با توجه به توابع عضویت موجود می‌توان برای هر یک از نمونه داده‌ها در هر ویژگی یک مجموعه فازی مردد تعریف کرد. بنابراین برای هر یک از ویژگی‌ها در مجموعه داده‌ها ماتریسی به ابعاد تعداد داده‌ها و تعداد توابع عضویت داریم که با استفاده از آن‌ها می‌توانیم ماتریس ضریب همبستگی مردد را به کمک فرمول (۱۰) بدست آوریم.

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^{۲۰} 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^{۲۱}

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 [۱۷، ۱۲] بوجود آمد.

در مجموعه‌های فازی مردد که توسط تورا معرفی شد، 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^{TT} 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 اختصاص دهیم نشان می‌دهند. 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 $[0, 1]$ برمی‌گرداند. یعنی،

$$A = \{ \langle x, h_A(x) \rangle \mid x \in X \}, \quad (9)$$

که $h_A(x)$ مجموعه مقادیر ممکن بین بازه صفر و یک است. در واقع $h_A(x)$ عناصر فازی مردد $(HFE)^{TT}$ نام‌گذاری می‌شود.

مقدار ویژه بزرگ ماتریس کواریانس انتخاب می شوند. در نهایت، داده های آموزشی و تست، بر روی بردارهای ویژه انتخاب شده تصویر شده و بعد داده ها کاهش پیدا می کند.

در بررسی آماری ساختار داده ها، علاوه بر ماتریس کواریانس، ماتریس همبستگی از اهمیت ویژه ای 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 مولفه های اصلی به واحد اندازه گیری مورد استفاده برای هر عنصر متعلق به 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 مولفه ای اصلی اول را دارند. از طرف دیگر، اگر تمام عناصر X

در واحدهای مشابه اندازه گیری شوند، در این صورت ماتریس کواریانس درست عمل می کند. اما در ماتریس همبستگی، استانداردسازی متغیرها به عنوان یک تلاش برای حذف مشکل وابستگی به واحد اندازه گیری یا معیار از *PCA* انجام می شود. بنابراین، نتایج حاصل از تجزیه و تحلیل براساس ماتریس همبستگی برای مجموعه های مختلف متغیرهای تصادفی، به طور مستقیم قابل مقایسه تر از آن نتایجی است که براساس ماتریس کواریانس استفاده می شوند [۱۰].

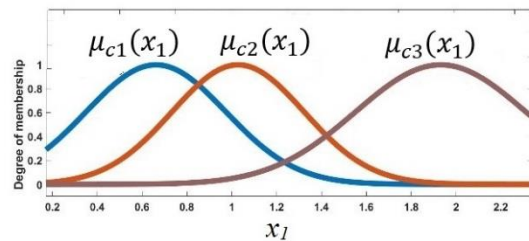
(۲) توانایی ماتریس کواریانس برای مقایسه نتایج حاصل

از تجزیه و تحلیل های متفاوت، بسیار کمتر از ماتریس همبستگی است. اندازه واریانس های مولفه های اصلی، مفهوم مشابهی برای ماتریس های وابستگی مختلف در بعدهای یکسان دارند. اما برای ماتریس های کواریانس مختلف، این طور نیست [۱۰].

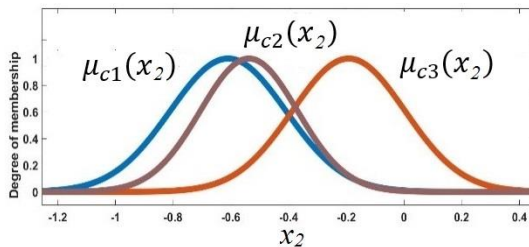
(۳) الگوهای ضرایب در مولفه های اصلی را می توان به

راحتی برای ماتریس های همبستگی مختلف مقایسه کرد و دید که آیا دو ماتریس همبستگی، مولفه های اصلی مشابهی ارائه می دهند یا خیر؟ در حالی که مقایسه ها برای ماتریس کواریانس بسیار پیچیده هستند. با این حال، روش های دقیقی برای مقایسه مولفه های 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 نسخه ی فازی آن بررسی کنیم.



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 مجموعه ی فازی مردد با توجه به توابع عضویت هر یک داریم:

$$A = \{ \langle x_1, h_A(x_1) \rangle \mid x_1 \in X \}$$

$$B = \{ \langle x_2, h_B(x_2) \rangle \mid x_2 \in X \}$$

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

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

که $\mu_{c_1}(x_1)$ ، $\mu_{c_2}(x_1)$ و $\mu_{c_3}(x_1)$ به ترتیب نشان دهنده ی 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 دهنده ی 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 داده ها می توان از روش های مختلفی استفاده کرد. در این راستا، روش *PCA*، یکی از مرسوم ترین روش های کاهش بعد داده ها بوده که برای استخراج ویژگی بکار برده می شود. در روش *PCA*، ابتدا توزیع داده به صورت توزیع نرمال تبدیل می شود. سپس، ماتریس کواریانس داده های آموزشی نرمال شده محاسبه گردیده و مقادیر ویژه و بردارهای ویژه مربوط به ماتریس کواریانس، استخراج می شوند. در مرحله بعد، به تعداد m بردار ویژه (که $m \leq d$) متناظر با m

بعد، به تعداد m بردار ویژه متناظر با m مقدار ویژه بزرگ ماتریس همبستگی مردد فازی انتخاب می‌شوند. در نهایت، داده‌های آموزشی 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 می‌پردازیم:

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 $k-1$ زیرمجموعه دیگر برای آموزش بکار می‌رود. این روال k بار تکرار می‌شود و همه داده‌ها دقیقاً یک بار برای آموزش و یک بار برای اعتبارسنجی بکار می‌روند. 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.

فرض کنید، ماتریس کواریانس با نماد Σ نشان داده شود. سازوکار روش PCA به صورت زیر فرمول بندی می‌شود :

$$\begin{aligned} PCA: \max \quad & u^T \Sigma u \\ \text{subject to} \quad & u^T u = 1. \end{aligned} \quad (15)$$

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

$$D = \text{diag}(\Sigma) \quad (16)$$

در این صورت ماتریس D ماتریسی است که تمام درایه‌های 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:

$$u^T D u = 1 \quad (17)$$

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

$$u^T \Sigma u = u^T D^{1/2} D^{-1/2} \Sigma D^{-1/2} D^{1/2} u. \quad (18)$$

بنابراین مسئله‌ی PCA به صورت زیر می‌تواند نوشته شود:

$$\max \quad u^T D^{1/2} D^{-1/2} \Sigma D^{-1/2} D^{1/2} u \quad (19)$$

$$\text{subject to} \quad u^T D^{1/2} D^{1/2} u = 1.$$

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:

$$\begin{aligned} \max \quad & y^T R y \\ \text{subject to} \quad & y^T y = 1. \end{aligned} \quad (20)$$

اکنون، با توجه به اینکه ماتریس همبستگی برابر است با

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

بنابراین، داریم:

$$\begin{aligned} \max \quad & y^T R y \\ \text{subject to} \quad & y^T y = 1 \end{aligned} \quad (21)$$

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 خوشه بندی 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، مقادیر ویژه و بردارهای ویژه مربوط به ماتریس همبستگی مردد فازی، استخراج می‌شوند. در مرحله

بعد ذکر شده و الگوریتم های خوشه بندی موجود ، ارزیابی و با یکدیگر مقایسه می شود.

در جدول (۱) مجموعه داده هایی که در این مقاله مورد استفاده قرار گرفته، آورده شده است. حال با توجه به آنچه که گفته شد، میزان خطای میانگین مربعات MSE محاسبه می شود سپس دقت هر روش در هر مجموعه داده با توجه به فرمول $Accuracy = 1 - MSE$ بدست می آید. نتایج بدست آمده در جدول (۲) نشان داده شده است. همچنین تعداد قوانین در هر روش برای هر مجموعه داده محاسبه می شود. سپس برای نرمال سازی آن از فرمول زیر استفاده شد :

$$NNR = 1 - \frac{\text{تعداد قوانین}}{\text{حداکثر تعداد قوانین}} \quad (22)$$

که تعداد قوانین میانگین تعداد قوانین تولید شده در هر روش روی هر یک از مجموعه داده ها 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 می توان بر اساس انواع روش های خوشه بندی و همچنین روش هایی که برای بدست آوردن بردارها و مقادیر ویژه بیان شد، توصیف کرد:

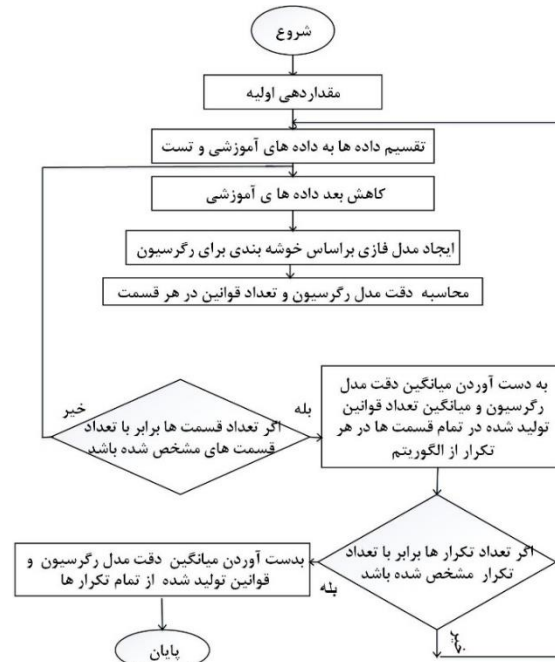
$M1$ الگوریتم خوشه بندی کاهشی با استخراج ویژگی با روش آنالیز مولفه ی اصلی.

$M2$ الگوریتم خوشه بندی کاهشی روی مجموعه فازی مردد و استخراج ویژگی توسط روش آنالیز مولفه ی 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 می شود.

$M3$ الگوریتم خوشه بندی کاهشی روی مجموعه فازی مردد و استخراج ویژگی توسط روش آنالیز مولفه ی اصلی با این تفاوت که به جای استفاده از ماتریس کواریانس از ماتریس ضریب همبستگی مردد ایجاد شده توسط روش خوشه بندی کاهشی استفاده می شود.

$M4$ الگوریتم خوشه بندی C-میانگین فازی با استخراج ویژگی با روش آنالیز مولفه ی اصلی.

$M5$ الگوریتم خوشه بندی C-میانگین فازی روی مجموعه فازی مردد و استخراج ویژگی توسط روش آنالیز مولفه ی اصلی با این تفاوت که به جای استفاده از ماتریس کواریانس از ماتریس ضریب همبستگی مردد ایجاد شده توسط روش C-میانگین فازی استفاده می شود.



شکل ۶. روند نمای الگوریتم

استخراج ویژگی و کاهش بعد داده ها:

مشابه با روش PCA کلاسیک، داده های آموزشی و تست، بر روی بردارهای ویژه انتخاب شده تصویر شده و بعد داده ها کاهش پیدا می کند. سپس، m تا از بهترین ویژگی ها انتخاب می شوند.

ایجاد مدل فازی براساس خوشه بندی برای رگرسیون:

در هر قسمت روی داده های آموزشی کاهش بعد یافته می توان از الگوریتم های خوشه بندی مانند الگوریتم خوشه بندی C- میانگین فازی و الگوریتم خوشه بندی کاهشی استفاده کرد 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 می توان میزان خطای میانگین مربعات بر روی داده های تست و همچنین تعداد قوانین ایجاد شده توسط مدل را محاسبه کرد. در نهایت میانگین خطا در ۳۰ بار تکرار لحاظ می شود.

نتایج و مقایسه ها:

در این مقاله میزان خطای میانگین مربعات و تعداد قوانین تولید شده از الگوریتم پیشنهادی توسط هر یک از روش های کاهش

^{۲۴} Mean Square Error

($M\epsilon$) 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 در مقاله.

مجموعه داده ها	اسم مختصر	تعداد ویژگی	تعداد نمونه ها
Residential	D ^۱	۱۰۸	۳۷۲
Parkinson	D ^۲	۲۱	۵۸۷۵
Foe	D ^۳	۱۳	۲۲۷
stock	D ^۴	۹	۹۵۰
Wizmir	D ^۵	۹	۱۴۶۱
Default Feature_1059_tracks	D ^۶	۶۹	۱۰۵۶
Housing	D ^۷	۱۳	۵۰۶

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

روش ها	D ^۱	D ^۲	D ^۳	D ^۴	D ^۵	D ^۶	D ^۷
M ^۱	۰.۹۷۴۶	۰.۹۹۴۹	۰.۹۵۳۲	۰.۹۹۹۶	۰.۹۸۸۰	۰.۹۶۱۷	۰.۸۶۱۸
M ^۲	۰.۵۳۶۱	۰.۹۹۴۳	۰.۹۴۷۹	۰.۹۹۸۴	۰.۹۹۰۱	۰.۹۵۸۰	۰.۹۸۳۷
M ^۳	۰.۲۳۷۵	۰.۹۹۶۶	۰.۹۲۰۸	۰.۹۹۸۴	۰.۹۸۰۹	۰.۹۶۱۷	۰.۹۸۷۷
M ^۴	۰.۹۳۶۴	۰.۹۹۱۰	۰.۸۹۷۱	۰.۹۹۲۳	۰.۹۸۸۶	۰.۹۵۲۵	۰.۹۹۸۴
M ^۵	۰.۹۵۴۷	۰.۹۹۳۰	۰.۸۹۴۰	۰.۹۹۴۰	۰.۹۹۹۷	۰.۹۵۲۹	۰.۹۶۸۸
M ^۶	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	D ^۱	D ^۲	D ^۳	D ^۴	D ^۵	D ^۶	D ^۷
M ^۱	۰	۰.۱۰۳۰	۰.۵۸۰۰	۰.۷۰۳۱	۰.۳۶۶۶	۰.۹۸۳۱	۰.۹۵۲۱
M ^۲	۰.۳۸۴۱	۰	۰.۵۷۸۰	۰.۵۸۱۷	۰	۰	۰.۹۵۰۵
M ^۳	۰.۳۶۹۶	۰.۶۰۱۱	۰	۰	۰.۵۴۱۶	۰.۹۸۳۹	۰
M ^۴	۰.۹۹۰۰	۰.۷۱۱۸	۰.۷۵۴۰	۰.۹۰۹۳	۰.۴۷۹۱	۰.۹۷۹۴	۰.۸۹۴
M ^۵	۰.۹۸۳۴	۰.۷۶۵۲	۰.۸۳۸۰	۰.۹۱۴۴	۰.۴۹۱۶	۰.۹۸۱۰	۰.۹۴۴۰
M ^۶	۰.۹۸۹۸	۰.۵۲۶۷	۰.۷۶۴۰	۰.۹۰۹۳	۰.۴۷۹۱	۰.۹۸۰۰	۰.۹۴۵۰

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

(۲۳)

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

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

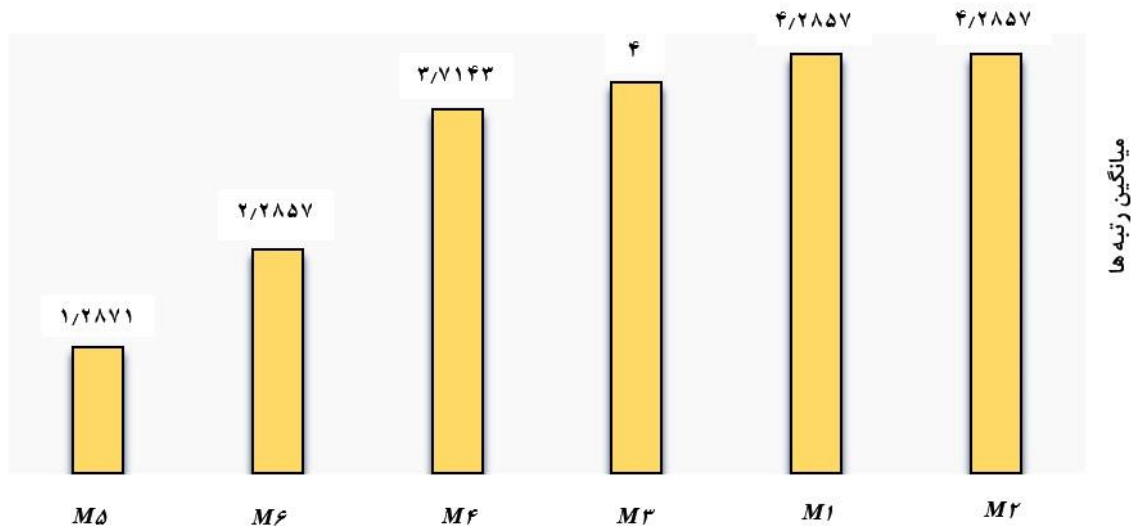
در جدول (۴) میانگین هندسی برای هر روش در هر مجموعه داده نشان داده شده است.

جدول ۴. میانگین هندسی بین دو معیار مذکور.

روش ها	D1	D2	D3	D4	D5	D6	D7
M1	۰	۰,۳۲۰۱	۰,۷۴۳۰	۰,۸۳۸۳	۰,۶۰۱۸	۰,۹۷۲۳	۰,۹۰۵۸
M2	۰,۴۵۳۷	۰	۰,۷۴۰۱۲	۰,۷۶۲۰	۰	۰	۰,۹۶۶۹
M3	۰,۲۹۶۱	۰,۷۷۳۹	۰	۰	۰,۷۲۸۸	۰,۹۷۲۷	۰
M4	۰,۹۶۲۸	۰,۸۳۹۸	۰,۸۲۲۴	۰,۹۵۰۲۵	۰,۶۸۸۱	۰,۹۶۵۸	۰,۲۹۸۶
M5	۰,۹۶۸۹	۰,۸۷۱۶	۰,۸۶۵۵	۰,۹۵۳۳	۰,۷۰۰۰	۰,۹۶۶۷	۰,۹۵۶۲
M6	۰,۹۷۰۳	۰,۷۲۲۰	۰,۸۲۸۰	۰,۹۵۱۶	۰,۶۹۰۱	۰,۹۶۶۱	۰,۹۵۵۳

دیگر است. برای مقایسه الگوریتمها، می توان از آزمونهای آماری مانند آزمون ناپارامتریک فریدمن نیز بهره برد. نتایج این آزمون در نمودار میله ای شکل (۵) نمایش داده شده است.

با توجه به نتایج بدست آمده از جدول (۴)، میانگین هندسی در روش M5 که بردارهای ویژه و مقادیر ویژه با روش آنالیز مولفه اصلی توسط ماتریس همبستگی فازی مردد که با الگوریتم خوشه بندی C-میانگین فازی بدست آمده، بیشتر از روشهای



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

همچنین مقدار آماری مربوط به این آزمون نشان می دهد که اختلاف این روشها معنادار است. بنابراین برای این که بتوان معنادار بودن اختلاف بین هر یک از الگوریتمها با روش کنترلی

براساس رتبه های موجود در شکل (۵) الگوریتم M5 را می توان به عنوان روش کنترلی در نظر گرفت چون میانگین رتبه آن از بقیه کمتر (بهتر) شده است.

یکسان بودن روش کنترلی با روش موجود در جدول می‌باشد، بنابراین می‌توان گفت که روش M_5 با روش‌های M_2 و M_3 و M_1 و M_4 اختلاف معنی‌داری دارد و تنها با روش M_6 اختلاف معنی‌داری ندارد.

M_5 را بررسی کرد از آزمون‌های تعقیبی یا پس‌آزمون^{۲۵} استفاده شد. بر اساس جدول (۵)، حاصل از بکارگیری آزمون تعقیبی Li فرضیه‌های صفر (H_0) در مواردی که P -value کمتر از $0,042182$ باشد، رد می‌شوند. از آنجا که فرضیه صفر به معنی

جدول ۵. جدول مقایسه *Post hoc*

Li	P	z	Algorithm	i
0,04282	0,12419	2,50000	M_3	۵
0,04282	0,18416	2,357173	M_2	۴
0,04282	0,32125	2,142857	M_1	۳
0,04282	0,116083	1,571429	M_4	۲
0,105	0,198543	1,285714	M_6	۱

قرار گرفته‌اند. 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 داده‌های مختلف نشان دهنده‌ی کارایی و قابل قبول بودن روش‌های مبتنی بر همبستگی فازی مردد در کاهش بعد و استخراج ویژگی است.

نتیجه‌گیری:

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 نشان داد 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

مراجع

- [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.

- [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.

- 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.
- [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