تولید قواعد مشارکت فازی با هدف تشخیص فِراد[۱] مشترکین شرکت توزیع نیروی برق مرکز خراسان رضوی

شهرام کیخائی[۲]
شرکت توزیع نیروی بر خراسان رضوی

سید حسین محدث[۳]
شرکت توزیع نیروی برق خراسان رضوی

تجدید ساختار در صنعت برق از جمله مباحثی است که کشور ایران نیز همپای کشورهای توسعه یافته در حال تجربه آن می باشد. در محیط تجدید ساختار یافته یکی از اصول ادامه حیات شرکتهای توریع حضور و رقابت در بازار برق خواهد بود. از سوی دیگر شبکه های توزیع دارای تلفات انرژی هستند لذا این شرکتها میبایست در آینده تلفات خود را کاهش دهند تا حضوری موثر در بازار داشته باشند. از جمله تلفاتی که میبایست مورد بررسی قرار گیرند خطای ناشی از تجهیزات اندازه گیری خواهد بود، این خطا یا به واسطه اتمام عمر مفید لوازم اندازه گیری رخ میدهد و یا توسط مشترک با هدف عدم پرداخت بهای معادل انرژی مصرفی ایجاد می گردد(فِراد).

در این تحقیق با کاوش قواعد مشارکت فازی که یکی از تکنیک های داده کاوی اطلاعات میباشد قواعدی که دارای پشتیبان و اطمینان قابل توجهی هستند، از پایگاه داده مشترکین شرکت توزیع خراسان رضوی تولید میگردد. هدف از تولید این قواعد فراهم نمودن قواعد مورد نیاز در پروژه طراحی سیستم خبره نرو-فازی است که وظیفه پیش بینی مصارف و تشخیص مصارف مشکوک را خواهد داشت.

واژه های کلیدی:داده کاوی اطلاعات،قواعد مشارکت فازی، اطمینان، پشتیبان، فِراد

۱ مقدمه

کاوش قواعد در پایگاههای بزرگ یکی از حوزه های وسیع تحقیق در داده کاوی اطلاعات میباشد. در داده کاوی اطلاعات، قواعد انجمنی اغلب برای نمایش و شناسائی ارتباطات مابین خصیصه ها در پایگاه داده بکار میرود. ایده اصلی به دهه ۹۰ میلادی باز میگردد، زمانیکه کاربرد آنالیز سبد بازار رواج پیدا کرده بود. در اغلب کاربردهای امروزی، پایگاه داده شامل تعدادی از مقادیر کمی است. یکی از راههای بررسی خصیصه های کمی جایگزینی آن با تعدادی خصیصه های کریسپ[۴] میباشد که این خصیصه ها ناشی از پارتیشن بندی خود خصیصه کمی هستند. برای مثال در کاربردهای عملی ممکن است تصمیم بگیریم خصیصه سن کارمندان را در یک پایگاه داده با خصیصه هایی چون جوان، میانسال و پیر معادل فواصل)۰،۳۵) و)۳۵،۶۰)و)۶۵،۱۰۰) نامگذاری نمائیم. از جمله مشکلات این نوع تقسیم بندی آنستکه خواص اعداد مرزی لحاظ نخواهد شد. معرفی تئوری مجموعه فازی به علت عدم درک مدلهای مفهومی مبهمی چون جوان و میانسال در مجموعه های کریسپ میباشد. درست به این دلیل که فردی که سنی کمتر از ۳۵ دارد، دقیقاً زمانیکه جشن تولد ۳۵ سالگی خود را میگیرد فردی میانسال در مفهوم کریسپ مطرح میگردد. در دنیای حقیقی تغییر مابین جوان و غیر جوان به صورتی آنی نیست بلکه تدریجی است. این دلیل خوبی برای مدلسازی مفاهیم مبهم با مجموعه های فازی در عوض مجموعه های کریسپ است.

---

[۱] بکار بردن نیرنگ و خدعه برای بهره گیری از سرویسها و منابع با هدف عدم پرداخت بهای معادل

[۲] کارشناس مسئول مکانیزاسیون اتوماسیون Email:nikpco@gmail.com

[۳] مدیر دفتر خدمات مشترکین

[۴] crisp

امروز از این استدلال استفاده شده و قواعد انجمنی فازی معرفی گردید [۱], [۵]. در این فرآیند یک پایگاه داده شامل خصیصه های کمی با مقادیر مابین [۰،۱] جایگزین میگردند. مسئله اصلی پارتیشن بندی مقادیر خصیصه های کمی و ساخت توابع عضویت میباشد. اولین راه حل به کمک فرد خبره (به صورت دستی بازه ها و توابع عضویت تعریف میگردد) و راه حل دوم بر اساس داده در دسترس (به صورت اتوماتیک بازه ها و توابع عضویت تولید میگردند) میباشد. در این تحقیق بر اساس روش دوم بازه ها و توابع فازی از ماهیت داده در دسترس تولید گردیده و بر این اساس پشتیبان و اطمینان قواعد مشارکت فازی اهمیت آنها بررسی میگردد.

۲ معرفی کاوش قواعد مشارکت فازی

مسئله کاوش قواعد انجمنی فازی امری است که به تازگی رایج شده است. مجموعه ای از تراکنشها را در نظر بگیرید، در حالیکه هر تراکنش مجموعه ای از آیتمهاست، یک قاعده انجمنی عبارتی به فرم $X \Rightarrow Y$ می باشد در صورتیکه X وY مجموعه هایی از آیتمها میباشند. یک مثال از یک قاعده انجمنی بدینصورت می باشد:"۳۰٪ از تراکنشهایی که شامل قهوه هستند شامل کیک هم میباشند"، "۲٪ از کلیه تراکنشها شامل هر ۲ آیتم هستند"، در اینجا ۳۰٪ به عنوان اطمینان قاعده و ۲٪ هم به عنوان پشتیبان قاعده نامیده میشود. مسئله پیدا نمودن قواعد انجمنی است که قیود مینی مم پشتیبان و مینی مم اطمینان تعیین شده توسط کاربر را ارضاء مینماید [۱].

همانطوریکه میدانیم مجموعه فازی A بر روی یک مجموعه جهانی X به صورت نگاشت $X \longrightarrow [0,1]$ تعریف و همچنین تابع عضویت A نیز نامیده میشود. برای هر x در X، A(x) درجه عضویت هر عضو مجموعه فازی A را نمایش میدهد. اگر تابع عضویت صرفاً اعضاء را به سمت مقادیر {0,1} ببرد مفهوم مجموعه های crisp را خواهد داشت. اجازه دهید X مجموعه تمامی تراکنشها (در مثال ما $|X| = 5$ میباشد). هدف ما مطالعه قاعده به فرم $A \Rightarrow B$ است در حالیکه A وB خصیصه های متفاوتی هستند. از A(x) برای بیان مقدار خصیصه A برای تراکنش x استفاده میگردد. در این راه، A یا زیرمجموعه crisp از X است (A(x)=۰ به مفهوم عدم عضویت x به A و A(x)=۱ بیانگر عضویت x به A است) یا مجموعه فازی در X (A(x) بیانگر درجه عضویت x در A است). همانطوریکه میدانیم پشتیبان از یک قاعده انجمنی $A \Rightarrow B$ معادل $\sup(A \Rightarrow B) = \frac{|A \cap B|}{|X|}$ در حالیکه $|X|$ معادل تعداد رکوردهای و یا تراکنشهای مورد مطالعه میباشد. همچنین اطمینان قاعده $A \Rightarrow B$ به صورت $Conf(A \Rightarrow B) = \frac{|A \cap B|}{|A|}$ تعریف میگردد. پشتیبان بیانگر اهمیت آماری قاعده انجمنی است در حالیکه اطمینان قاعده بیانگر قدرت قاعده انجمنی میباشد. ما هم از تعریف تقاطع و کاردینالیتی آقای عسگری زاده برای مجموعه های A و B استفاده مینمائیم.

$$(A \cap B)(x) = \min(A(x), B(x))$$
$$|A| = \sum_{x \in X} A(x)$$

بنا بر تعاریف فوق در صورتیکه قاعده انجمنی $A \Rightarrow B$ را به صورت A مصرف دوره قبل و B مصرف دوره بعدی بگیریم میتوانیم بر اساس تعاریف فوق پشتیبان دوره های مختلف را بدست آوریم [۳], [۲], [۱].

۲_۱ استخراج قواعد فازی مشارکتی

لازم به ذکرست در این تحقیق برای پیاده سازی الگوریتم کاوش قواعد مشارکت فازی از نرم افزار متلب استفاده شده است. لذا بر اساس مشخصات مشترکینی که به عنوان فِراد توسط بازرسی شهرستان نیشابور شناسائی و کسب اطلاعات مصارف انرژی همان مشترکین در دوره های دو و سه و چهار در سال ۱۳۸۵ این تحقیق صورت پذیرفت. با عنایت به اینکه خصیصه های (مصارف دوره ها) منتخب کمی بودند در ابتدا میبایست آنها را دسته بندی میکردیم. در این گام دو روش در پیش روی ما قرار داشت.

۱.بکارگیری از روشهای کلاسیک کریسپ

۲.بکارگیری از روشهای نرم

در روشهای کلاسیک کریسپ توسط کاربر توابع عضویت تعریف میگردد و بر این اساس مراحل استخراج قواعد مشارکتی ادامه می یابد. در روشهای نرم با استفاده از روشهای خودکار تولید تابع عضویت سعی میگردد بر اساس ماهیت داده توابع عضویت تولید گردد [۴].در این تحقیق با بکارگیری c-mean و با k=۳ سعی گردید توابع عضویت به صورت خودکار و بر اساس ماهیت داده در اختیار ایجاد گردد.(جدول ۱)

۲_۲ معرفی خوشه بندی فازی c-mean

در خوشه بندی فازی, هر نقطه دارای یک درجه تعلق به خوشه میباشد به جای آنکه به طور کامل متعلق به یک خوشه باشد. بنابراین نقاط در لبه یک کلاستر دارای درجه تعلق کمتری نسبت به نقاطی خواهند بود که در مرکز خوشه قرار میگیرند. برای هر نقطه x ضریبی خواهیم داشت که میزان استقرار آنرا در خوشه k نمایش میدهد.

$$\forall x \quad \sum_{k=1}^{num_cluster} \mu_k(x) = 1$$

با فازی c-mean مرکز ثقل یک خوشه میانگین کلیه نقاط بوده که با درجه عضویت خود وزندار شده اند.

$$center_k = \frac{\sum_x \mu_k(x)^m x}{\sum_x \mu_k(x)^m}$$

درجه تعلق متناسب با معکوس فاصله تا خوشه میباشد.

$$\mu_k(x) = \frac{1}{d(center_k, x)}$$

بنابراین ضرایب نرمال و با پارامتر حقیقی m>۱ به گونه ای فازی سازی شده که مجموع آن یک خواهد بود. پس:

$$\mu_k(x) = \frac{1}{\sum \left(\frac{d(center_k, x)}{d(center_k, x)}\right)^{2/(m-1)}}$$

برای m=۲ این معادل آنستکه ضرایب به صورت خطی نرمال شده تا مجموع آنها یک گردد. زمانیکه m به یک نزدیک گردد مرکز خوشه به نقطه ای نزدیک خواهد شد که دارای وزن بیشتری از دیگران باشد و مشابه الگوریتم k-mean می باشد.

۲_۳ الگوریتم فازی c-mean

۱.انتخاب تعداد خوشه ها

۲.انتصاب ضرایبی رندم برای تعلق به خوشه به هر نقطه .

۳.این امر را تا زمان همگرایی الگوریتم ادامه میدهیم.

۴.نقطه ثقل هر خوشه محاسبه شده (با استفاده از فرمول بالا).

۵.برای هر نقطه ضرایب تعلق به خوشه ها با استفاده از فرمول بالا محاسبه میگردد.

الگوریتم واریانس مابین خوشه ها را به میزانی مناسب کاهش میدهد اما مسائل مشابهی با k-means را داشته ومینی مم مطرح شده مینی مم محلی میباشد.

۳ بکارگیری قواعد مشارکت فازی در تشخیص کسر مصرف مشترکین

در شبکه های توزیع هر چه طول شبکه توزیع بیشتر گردد تلفات نیز افزایش خواهد یافت، البته بررسی اینگونه از تلفات موضوع این مقاله نمیباشد. آنچه که با محاسبات ریاضی و قوانین مهندسی برق قابلیت محاسبه ندارد تلفات غیر تکنیکال هستند. منشاء تمامی این تلفات نقص در تجهیزات اندازه گیری مشترک میباشد. البته این خرابی میتواند ناشی از قدمت تجهیزات منصوبه بوده و یا به واسطه دستکاری در تجهیزات اندازه گیری مشترکین و با هدف عدم پرداخت بهای معادل برق مصرفی صورت پذیرد. در حال حاضر در شرکتهای توزیع بررسی و تشخیص این گونه از موارد به واسطه بازدیدهای سالیانه و یا تشخیص عدم تطابق مصرف با نوع کاربری توسط مامور قرائت و یا واحد خدمات مشترکین صورت می پذیرد. طبیعی است که این امر غیر مکانیزه صرفاً به نیروی انسانی وابسته بوده و نمیتواند نیازهای شرکتهای توزیع در شرایط آینده را برآورده نماید[۵],[۴]. در شرکتهای توزیع اطلاعات مصارف ماهیانه انرژی مشترکین در پایگاه های داده جمع آوری میگردد. بدیهی است کاوش و جداسازی اطلاعات مفید توسط انسان از این اطلاعات گسترده میسر نمی باشد.

هدف از تعریف این تحقیق بکار گیری داده کاوی اطلاعات در استخراج دانش و قواعد انجمنی از بطن اطلاعات در دسترس می باشد. کاربرد قواعد انجمنی مستخرج در یک سیستم خبره نرو فاری خواهد بود. از مزایای این روش آنستکه قواعد به صورت خودکار بوده و مزایای مجموعه های فازی را دارد. سیستم خبره نرو- فاری مورد نظر سیستمی است که میتواند بنا بر مصارف گذشته مشترکین، مشترکینی را که مصارف (به دلیل نقص لوازم اندازه گیری) نمایش یافته شان کمتر از میزان مصرف واقعی است، تشخیص دهد. لازم به توضیح است که در این مقاله صرفاً قواعد انجمنی قابل استفاده در تشخیص تراکنشهای مشکوک تولید میگردد و ارائه یک سیستم خبره نرو- فازی از برنامه های تحقیقاتی آتی نویسندگان مقاله میباشد.

برای بیان ساده مسئله و راه حل پیشنهای مصارف چندین مشترک (جدول ۱۶) را که دستکاری در کنتور ایشان در سال ۱۳۸۵ محرز بوده بررسی نموده و سعی گردید قواعدی استحصال گردد که براساس این قواعد بتوان سیستم خبره ای طراحی نمود که از آن به بعد به صورت خودکار، بر

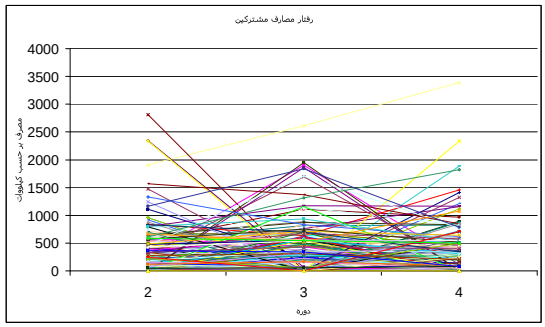
اساس قواعد استحصالی مشترکین با رفتار مشابه و مشکوک را تشخیص دهد.

۳_۱ بیان مسئله

مطلوبست اعمال کاوش قواعد مشارکت فازی بر اطلاعات جدول ۱۵ و ارائه قواعدی که مینی مم پشتیبان ۲۰٪ ومینی مم اطمینان ۵۰٪ باشد.

۳_۲ شرح نتایج آزمونهای اعمال شده

جدول ۱۶ را در نظر گرفته و اطلاعات مصارف مشترکین را در دوره های دو تا چهار سال ۸۵ استحصال نمودیم. از جمله نتایج جالبی را که در نگاه اول از شکل ۲ میتوان برداشت نمود آنستکه چگالی مصارف این نوع از مشترکین در بازه سه دوره اکثراً از ۱۰۰۰ کیلووات کمتر بوده و دستکاری به گونه ای انجام شده که مورد ظن شرکت برق قرار نگیرد. کما اینکه شرکت برق برخی از اینگونه موارد را با بازرسیهای سالیانه و حتی موردی تشخیص میدهد.



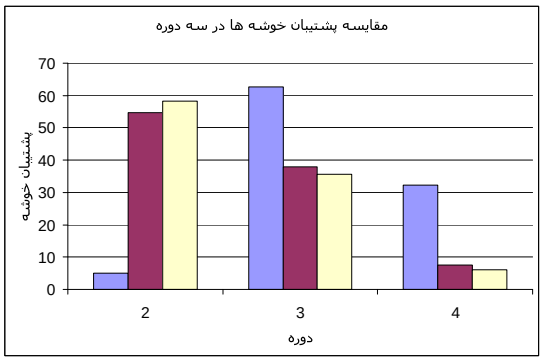
شکل ۱. مصارف مشترکینی که در سال ۱۳۸۵ کسر مصرف داشته اند

در این تحقیق با بکارگیری c-mean و با k=۳ سعی گردید توابع عضویت به صورت خودکار و بر اساس ماهیت داده در اختیار ایجاد گردد.(جدول ۱)

پشتیبان	خوشه		
۵,۱	خوشه ۱	۲۰۵۲,۳	دوره ۲ سال ۸۵
۶۲,۷		۷۹,۱	
۳۲,۲		۶۴۴,۲	
۵۴,۶	خوشه ۲	۷۰,۶	دوره ۳ سال ۸۵
۳۷,۹		۶۲۵,۴	
۷,۵		۱۷۶۹,۱	
۵۸,۳	خوشه ۳	۶۳,۱	دوره ۴ سال ۸۵
۳۵,۶		۵۷۸,۲	
۶,۱		۱۷۹۰,۱	

جدول ۱. اعمال c-mean و پشتیبان هر یک از توابع عضویت

همانطوریکه ملاحظه میگردد در رنجهای متوسط پشتیبان بیشتر میباشد.



شکل ۲. مقایسه پشتیبان هر تابع عضویت در سه خوشه

هدف از تحقیق فوق ارائه قواعدی به فرم A=>B است در ادامه به کاوش قواعدی به فرم شرطی می پردازیم. با تعریف رنجهای تابع عضویت به فرم:

رنجهای تابع عضویت فازی			
A۱	۲۰۵۲,۳	خوشه ۱	دوره ۲ سال ۸۴
A۲	۷۹,۱		
A۳	۶۴۴,۲		
B۱	۷۰,۶	خوشه ۲	دوره ۳ سال ۸۴
B۲	۶۲۵,۴		
B۳	۱۷۶۹,۱		
C۱	۶۳,۱	خوشه ۳	دوره ۴ سال ۸۴
C۲	۵۷۸,۲		
C۳	۱۷۹۰,۱		

جدول۲. تعریف مجموعه های فازی

در صورتیکه مقادیر مجموعه های مصارف دورهای دو و سه سال ۸۵ را در هر پارتیشن با دوره های سه و چهار ترکیب نمائیم قواعدی به فرم جدول۶ بدست خواهد آمد. با اعمال الگوریتم فازی c-mean مراکز خوشه ها به صورت جدول ۵ حاصل گردید و با استفاده از نرم افزار متلب و ترکیب درجه عضویت هر خوشه با خوشه های دوره های دیگر قواعدی حاصل گردید که میزان پشتیبان و اطمینان دوره های ترکیبی در جدول ۳ ملاحظه میگردد.

	B۱	B۲	B۳	C۱	C۲	C۳
A۱	٤	۱	۱	۲	۳	۲
A۲	٤۳	۲۰	٤	۵۱	۱۵	۲
A۳	۱۲	۲۱	۳	۱۱	۲۳	٤
B۱	۰	۰	۰	۳۹	۱۷	٤
B۲	۰	۰	۰	۲۰	۲۱	۲
B۳	۰	۰	۰	٤	۳	۲

جدول ۳.ترکیب توابع عضویت و محاسبه پشتیبان بر هر یک از قواعد انجمنی

	B۱	B۲	B۳	C۱	C۲	C۳
A۱	۷۰	۲۸	۲۹	۳۶	۶۰	۷۰
A۲	۶۹	۳۱	۶	۸۱	۲۴	۶۹
A۳	۳۷	۶۶	۱۱	۳۳	۷۰	۳۷
B۱	۰	۰	۰	۷۱	۳۱	۰
B۲	۰	۰	۰	۵۴	۵۵	۰
B۳	۰	۰	۰	۵۷	۴۵	۰

جدول ۴.ترکیب توابع عضویت و محاسبه اطمینان هر یک از قواعد انجمنی

همانطوریکه در ابتدا بحث گردید مطابق بند ۳_۱ میبایست قواعدی را با اطمینان و پشتیبان مشخص شده تعیین نمود لذا بر این اساس:

خوشه ۳	خوشه ۲	خوشه ۱
۶۳,۱	۷۰,۶	۲۰۵۲,۳
۵۷۸,۲	۶۲۵,۴	۷۹,۱
۱۷۹۰,۱	۱۷۶۹,۱	۶۴۴,۲

جدول ۵ مرکز خوشه های الگوریتم c-mean

جدول ۶ شروط حاصله با یک مقدم و یک تالی را نمایش میدهد.

قاعده مشارکت A=>B

اطمینان	پشتیبان	تالی	مقدم
۶۹	۴۳	B۱-۷۱	A۲-۷۹
۸۱	۵۱	C۱-۶۳	A۲-۷۹
۶۶	۲۱	B۲-۶۲۵	A۳-۶۴۴
۷۰	۲۳	C۲-۵۷۸	A۳-۶۴۴
۷۱	۳۹	C۱-۶۳	B۱-۷۱
۵۴	۲۰	C۱-۶۳	B۲-۶۲۵
۵۵	۲۱	C۲-۵۷۸	B۲-۶۲۵

جدول ۶

طبیعی است از این شروط میتوان به عنوان یک طبقه بندی کننده در یک سیستم استنتاج فازی استفاده نمود. حال تعداد خصیصه ها در پیش شرط را به ۲ رسانده و قاعده انجمنی با ۳ عضو به صورت $A \cap B => C$ رابررسی مینمائیم. منتجه این عملیات در جدول ۹ نمایش داده شده است.

	مقدمA	مقدمB	تالیC	پشتیبان	اطمینان
۱	A۲-۷۹	B۲-۷۱	C۱-۶۳	۳۶	۸۳

جدول ۹ قواعد استحصالی با دو مقدم و یک تالی (توضیح اینکه نحوه نمایش هر خوشه به صورت نام خوشه + مرکز خوشه میباشد)

اطمینان قواعد انجمنی با دو مورد تابع و یک مورد تالی به صورت جدول ۹ و در ستون conf خواهد بود. با اعمال پیش شرط های مینی مم پشتیبان ۲۰ درصد و اطمینان پنجاه درصد و اعمال این شروط در نرم افزار متلب نتایج حداول ۸و۹ بدست آمد. لازم به ذکرست پس از کسب قواعد میبایست توسط افراد خبره شروط استحصالی مجدداً پالایش شده و قواعد مفید استخراج گردد. کاربرد این قواعد میتواند در سیستمهایی با خاصیت بازخورد باشد در این سیستمها بر اساس رفتار گذشته رفتار آینده تصحیح میگردد. ناگفته پیداست با توجه به اینکه مجموعه دادگان مورد مطالعه ما (جدول ۱۶) مجموعه ای کوچک بوده قواعد اندکی استحصال گردید، البته هدف از این مطالعه بررسی امکان ایجاد قواعد مشارکتی بوده است. با توجه به اینکه برخی از موارد دستکاری موجب از کارافتادگی کامل لوازم اندازه گیری میشوند مصارف زیادی در سطح صفر در جدول ۱۶ مشاهده میگردد.

۳_۴ افزایش تعداد پارتیشن های توابع عضوی

با افزایش تعداد پارتیشن ها از ۳ مورد به چهار مورد، قواعدی با یک تالی و یک مقدم به شرح جدول ۱۰ حاصل گردید. لازم به ذکرست مراکز خوشه ها در جدول ۱۳ نمایش داده شده است.

	مقدم A	تالی B	پشتیبان	اطمینان
۱	A۲-۵۱	B۲-۲۲	۳۲	۵۷
۲	A۲-۵۱	C۴-۳۷	۴۴	۷۸
۳	A۴-۵۰۲	C۲-۴۵۶	۲۰	۶۳
۴	B۲-۲۲	C۴-۳۷	۲۷	۶۴

جدول ۱۰

با کاوش قواعد مشارکت فازی بر روی اطلاعات جدول ۱۶ قواعدی با دو مقدم و یک تالی به صورت جدول ۱۱ حاصل گردید. همانطوریکه ملاحظه میگردد قواعد استحصالی دارای مراکز کوچکتری بوده و پشتیبان قاعده استحصالی تضعیف شده است.

	Aمقدم	Bمقدم	Cتالی	پشتیبان	اطمینان
۱	A۲-۵۱	B۲-۲۲	C٤-۳۷	۲۵	۸۰

جدول ۱۱

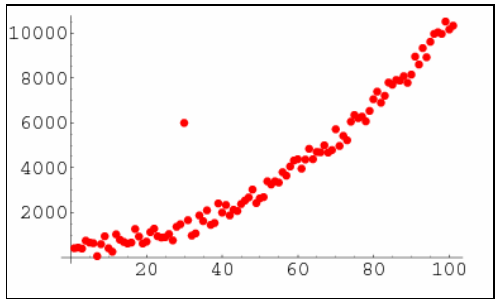
خوشه ۱	خوشه ۲	خوشه ۳
۱۱۳۸,۸	۷٦۹,٦	۱۱۱۱,۲
۵۰,٦	۲۲	٤۵۵,۸
۲۳۸۸,۳	۱۸٤۰,٤	۲٦۹۲,۸
۵۰۲,٤	۳۸٤,٦	۳۷,۵

جدول ۱۲

۴ داده های اوت لایر

۴_۱ داده اوت لایر

یک اوت لایر داده ای است که خارج از الگوی توزیع کلی داده ها بوده و برخی از مشکلات همانند کاهش پشتیبان و قاده برای اطلاعات ماحصل در یک خوشه را در پی خواهد داشت. طبق تعریف نقطه اوت لایر نقطه ایست که دارای ۱،۵ برابر فاصله از دامنه میانی بوده و بالای یک چارک و یا تحت یک چارک قرار میگیرد.



شکل ۳.نمایش اوت لایر در یک الگو

۴_۲بررسی اثرات اوت لایر در داده های مورد مطالعه

جهت بررسی اثرات این نوع داده در جدول ۱۶ مجدداً این نوع داده را با اعمال الگوریتم فازی c-mean مطالعه مینمائیم.

	خوشه	پشتیبان
دوره ۲	۷۹	٦۳
	٦٤٤	۳۲
	۲۰۵۲	۵
دوره ۳	٦۲۵	۳۸
	۷۱	۵۵
	۱۷٦۹	۷
دوره ۴	٦۳	۵۸
	۵۷۸	۳٦
	۱۷۹۰	٦

جدول ۱۱

در برخی از خوشه های بدست آمده پشتیبان ناچیز است (کمتر از ۲۰ درصد) این احتمال میرود که داده های موجود در این خوشه ها اوت لایر باشند. در این قسمت در راستای بررسی حساسیت الگوریتم به این نوع از داده ها برخی از خوشه ها که دارای پشتیبان کمتری هستند (کمتر از ۷ درصد) از مجموعه داده ها حذف گردید. به عنوان مثال داده هایی که در خوشه سوم از دوره ۲ و۳و۴ قرار گرفته اند را حذف نموده و مجدداً قواعد مشارکت فازی را کاوش مینمایئم.

	خوشه	پشتیبان
دوره ۲	۲٦	۵۲
	۷۵۱	۱۸
	۳۵۵	۳۰
دوره ۳	۷۹٦	۲۱
	٤۰۱	۳۳
	۲۵	٤۵
دوره ۴	٤۲۰	۳٦
	۹۵۹	۱۱
	۳۰	۵۳

جدول ۱۳

جدول ۱۴ اطلاعات مربوط به قواعدی با یک تالی و یک مقدم را پس از حذف خوشه های ضعیف از مجموعه داده ها و اجرای مجدد روال تشخیص قواعد مشارکت فازی در متلب نمایش میدهد. قواعد منتجه نشان دهنده کاهش تعدادی قواعد حاصله با یک تالی و یک مقدم میباشد.

مقدم	تالی	پشتیبان	اطمینان
A۱-۲۹	B۳-۲۵	۳۳	٦۳
A۱-۲۹	C۳-۳۰	٤۳	۸۳
A۳-۳۵۵	C۱-٤۲۰	۲۰	٦۷
B۳-۲۵	C۳-۳۰	۳۰	٦۷

جدول ۱۴

همچنین قواعد کسب شده با دو مقدم و یک تالی نیز دارای پشتیبان و اطمینان با تفاوت زیادی نسبت به جدول ۹ نمیباشد. نتیجه این امر بیانگر عدم حساسیت الگوریتم به اوت لایرها بوده و نشان میدهد الگوریتم با وجود اوت لایرها باز هم قادرست قواعدی مناسب را تولید نماید.

	Aمقدم	Bمقدم	Cتالی	پشتیبان	اطمینان
۱	A۱-۲۹	B۳-۲۵	C۳-۳۰	۲۸	۸٤

جدول ۱۵

نتیجه گیری و کارهای آینده

داده های مصارف انرژی هم مانند دیگر پدیدهای حقیقی ارتباط مستقیمی با زندگی روزمره انسانها دارد، خصصیه اصلی این داده ها آنستکه به صورت کمی هستند. همانطوریکه در اثنای مقاله اشاره گردید هدف از این تحقیق ارائه قواعد مشارکتی فازی بود که بتوان از آنها در تصحیح رفتار مشترکین در یک سیستم نرو فازی استفاده نمود.

البته علت بهره گیری از سیستم فازی حذف اثرات مرزی در مطالعه قواعد مشارکتی است. قواعد خروجی این تحقیق نشان داد که میتوان از داده در دسترس مربوط به مشترکینی که به نوعی در سیستم اندازه گیری خود دستکاری نموده اند قواعدی کارآمدی تولید نمود که در شناسائی افراد مشابه استفاده گردد. همچنین با حذف خوشه هایی با پشتیبان کم روشن گردید قواعد تولید شده بایک مقدم و تالی کاهش یافته و قواعد منتجه با دو مقدم و تالی نیر تفاوتی از دیدگاه پشتیبان و اطمینان با حالت اولیه ندارند. میتوان نتیجه گیری نمود الگوریتم حساسیت کمی به خوشه های ضعیف داشته و قواعد منتجه قابل اعتماد هستند.

از جمله کارهای آتی نویسندگان مقاله طراحی سیستمی با خاصیت پیشگویی مصارف مشترکین میباشد. در این سیستم داده های تولید شده در خروجی سیستم پیشگو به واسطه سیستم فازی تصحیح گردیده و مصارف مشترکینی که الگوی مصارفشان مشابه با قواعد استحصالی در این مقاله باشد به عنوان مشترکین مشکوک نشان گذاری شده و مابقی به عنوان مشترک نرمال نشانه گذاری میشود.

مراجع:

[۱] Fuzzy Versus Quantitative Association Rules: A Fair Data-Driven Comparison, Hannes Verlinde, Martine De Cock, and Raymond Boute , IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, VOL. ٣٦, NO. ٣, JUNE ٢٠٠٦

[۲] **Mining Fuzzy Association Rules in a Bank-Account Database**, Wai-Ho Au and Keith C. C. Chan, IEEE TRANSACTIONS ON FUZZY SYSTEMS, VOL. ١١, NO. ٢, APRIL ٢٠٠٣

[۳] Fraud Detection in Electrical Energy Consumers Using Rough Sets ,JosC E. Cabral, Edgar M. , ٢٠٠٤ IEEE International Conference on Systems, Man and Cybernetics

[٤]An Electric Energy Consumer characterization Framework Based on Data Mining Techniques, Vera Figueiredo, F?tima Rodrigues, Zita Vale, *Member, IEEE*, and Joaquim Borges Gouveia, IEEE TRANSACTIONS ON POWER SYSTEMS, VOL. ٢٠, NO. ٢, MAY ٢٠٠٥

[۵] Mining Quantitative Association Rules in Large Relational Tables, Ramakrishnan Srikant" Rakesh Agrawal

ردیف	۱۳۸۵_۲	۱۳۸۵_۳	۱۳۸۵_٤	ردیف	۱۳۸۵_۲	۱۳۸۵_۳	۱۳۸۵_٤	ردیف	۱۳۸۵_۲	۱۳۸۵_۳	۱۳۸۵_٤
۱	۳۳۱	۱۹۸	۱٤۱	۵۰	۲۱۸	۲۵۷	۳۰۸	۹۹	۸۰۲	۱۱۷۰	۱۱٦٦
۲	۲۹۳	۱۹۰۱	۱۸۸	۵۱	۷۵۱	۱۱۳۹	۸۲۱	۱۰۰	۰	٤۸۳	۰
۳	۲۲۷	۲۹۹	۲۵۵	۵۲	۲۳٤٤	٦٦	۰	۱۰۱	۱٦٤	۰	٤۰
٤	٦٦٤	٦٤۱	۵۰۱	۵۳	۵۲۷	۷۰۷	۲۷	۱۰۲	۰	۷۳۲	۰
۵	۱۷۱	۳۰۸	۲۲۱	۵٤	۲۲٤	۳۲۳	۳۰۵	۱۰۳	۱۲٤۳	۰	٤٦۰
٦	۵۵۰	۱۱۱۱	۹٦۵	۵۵	۳۰۵	۳۲۰	۳۲۰	۱۰٤	۳۹٤	٦۷۱	۵۷۹
۷	۵۵۱	۸۱٦	۸۲۱	۵٦	۱۱۰۹	۲٤٤	۳٦۱	۱۰۵	٤۲۲	۵۷۳	٤۳۵
۸	۲۳۷	۲۸۷	٤۵٤	۵۷	۵۹۰	٤۳۸	۲۰	۱۰٦	٤٦۲	۱۷۰۱	۷۷۱
۹	٤۳٦	۵۹۲	٤۸۸	۵۸	۲۳۳۸	۰	۲۳۳۸	۱۰۷	۰	۰	۰
۱۰	۰	۰	۰	۵۹	۰	۳۰۸	۰	۱۰۸	۰	۰	۰
۱۱	٤۲۰	۳٤۷	۲۸۷	٦۰	۰	۱۸٦۳	۰	۱۰۹	۰	۰	۰
۱۲	۱۹۱۱	۲٦۱۲	۳۳۹۲	٦۱	۰	۰	۰	۱۱۰	۰	۰	۰
۱۳	۵۵۲	۷۰٤	۳٤٤	٦۲	۷۰۰	۰	۸۷٦	۱۱۱	٤۷۷	۵۵۳	۵٤۱
۱٤	۲۱۵	۲۲۵	۳۷۷	٦۳	۰	۲٤٦	۰	۱۱۲	۰	۰	۰
۱۵	۵۵۵	۳۷۱	۲۳۲	٦٤	۱۸۷	۰	۷۲۱	۱۱۳	۰	۷	۰
۱٦	۷۱۵	٤۳۹	۰	٦۵	۰	۱۸۱	۰	۱۱٤	۰	۰	۰
۱۷	۹۲۲	۰	۱۲۱۳	٦٦	۳۱۰	۰	۳۲۷	۱۱۵	۱	۰	۰
۱۸	۹۵٦	۰	۱۸۸۳	٦۷	۰	۰	۰	۱۱٦	۱۵٦٦	۱۳۷۲	۸٤۵
۱۹	۹۵۸	۰	۱۱۱٦	٦۸	۰	۲۸٤	۰	۱۱۷	۱۸۸	۲۵۹	۲۷۱
۲۰	۲۵۵	۰	۷۷	٦۹	۰	۳۳	۰	۱۱۸	۱٤٦	٦۳	۸۸
۲۱	۰	۰	۰	۷۰	۰	۲۲۱	۰	۱۱۹	۵۰	۵۵۹	۳۰۵
۲۲	٤۸۳	۰	٤٤۰	۷۱	۰	۸٤۳	۰	۱۲۰	۱۸٦	۳۵۸	۳٤٤
۲۳	۳۹٦	٤٦۵	٤۲۲	۷۲	۰	۳۹	۰	۱۲۱	۱۰۷	۵۵	۲۹۲
۲٤	٦٤٦	۸۸۹	۵۵۷	۷۳	۰	۱۰	۰	۱۲۲	۸٦	۱	۲۳۵
۲۵	۱۰۱	۲۹۵	۳۰۸	۷٤	۰	٤٦۲	۰	۱۲۳	۱۷	۱۵۹	۱۰۵
۲٦	۱۸۱	۵۳٦	۲۸۱	۷۵	۰	۰	۰	۱۲٤	۱٤۲	۱۵۵	٤٤
۲۷	٤۲۷	۰	۸۹٦	۷٦	۰	٤۷٤	۰	۱۲۵	۷۷	۳۱٤	٤۲۹
۲۸	٦٤۰	۸۷۰	۸۳۰	۷۷	۰	۰	۰	۱۲٦	٦۵۵	۵٦۳	٤۲۰
۲۹	۱۱۰	٤۲	٤۳٦	۷۸	۰	۳۸۹	۰	۱۲۷	۱۳۳۰	۸۳۰	٦۱۷
۳۰	۲۸۱	۰	٤٦۵	۷۹	۸۷۸	۰	٤۱۲	۱۲۸	۸۰۲	۹۳۲	۵۳٤
۳۱	۱۹۸	۲۰٦	۲۱۲	۸۰	۰	۰	۰	۱۲۹	۲۲۱	۵۹	٤۹۹
۳۲	۰	۰	۱٤۰	۸۱	۰	۱۹۵	۱۹۵	۱۳۰	٤۸۵	۷۱٦	٦٤۵
۳۳	۱٤۰۱	۳٦۱	۳۹۸	۸۲	۰	۵۸٦	۰	۱۳۱	٦٦۹	۵۲۲	۱۰۸۰
۳٤	۸٤۰	٦۳۲	۱٤۵۵	۸۳	۰	۱۹۵۲	۰	۱۳۲	۱۱۷	۱٦۸	۹۹
۳۵	۲٤٦	٤۷۷	۱۸۹	۸٤	۳۳۱	۰	۱٤۵	۱۳۳	۵۱	۲۷	۵٦۳
۳٦	٤۰۳	۲٤٤	۱۳۸	۸۵	۱٤۷٦	۰	۱۳۲٦	۱۳٤	۱۸	۰	۸٤٤
۳۷	٤۳۳	٦۷۵	٤۷۵	۸٦	۰	٦۲۵	۰	۱۳۵	۸۳۵	۷۰۰	۳۳۹
۳۸	۳٦۹	۵۸٦	٤۷٤	۸۷	۷۹۵	۰	۵۰۹	۱۳٦	٦۰۸	۱۳۱۸	۱۸۲۰
۳۹	۱۹۱	۲۸٤	۲۲۸	۸۸	۰	۱۱۱۸	۰	۱۳۷	٤۹	۵۳	۷۹
٤۰	۲۸۱۲	۰	۷۰۱	۸۹	۰	٦۰۹	۰	۱۳۸	۵٦۷	۷٤۹	۵۱۷
٤۱	٤۹۱	۵۰۹	۰	۹۰	۰	۱۱۵٦	۰	۱۳۹	۲۷۵	۳۱۰	۲۰٦
٤۲	۰	۵۵۵	۱۱٤	۹۱	۰	۰	۰	۱٤۰	۳۳۵	٤۳۷	۳۹۷
٤۳	٦٦۰	٦۹۱	٤۵۷	۹۲	۰	۵٤۹	۰	۱٤۱	۱۱٦۸	۱۸۳٤	۷۸٦
٤٤	۳۲۰	۷۵۸	۱۱۵۳	۹۳	۰	۷۰	۰	۱٤۲	٤	٦	۰
٤۵	۳۳۵	٤۲۵	۳۲۰	۹٤	۰	۰	۰	۱٤۳	٤۳۵	۵٤۲	۵٤٤
٤٦	٦۰۰	۷۲۵	٦۰۰	۹۵	۲۱	۰	۲۲٤	۱٤٤	۲٤۵	۰	۷۱۰
٤۷	۸٤	۷٤	۲۳	۹٦	۰	٦۹۷	۰	۱٤۵	۵۸۸	۵٤۸	۵۰۱
٤۸	۲۳۰	۷٦۵	۰	۹۷	۰	۰	۱٤۱۲	۱٤٦	۳٦٦	۳٤۰	۷۹
٤۹	۲۸٤	۱٦۹۳	۱۱۷	۹۸	۰	۳٦۵	۰	۱٤۷	۰	۰	۰

جدول ۱۶ اطلاعات مصارف فِراد مشترکین ۱۴۷ مشترک در دوره های دو و سه و چهار در سال ۱۳۸۵