ارائه راهکار دو مرحله‌ای جهت شناسایی الگوهای مصرف برق

مریم آموزگار[1] ، مربی

۱- گروه پژوهشی کامپیوتر و فناوری اطلاعات –پژوهشگاه علوم و تکنولوژی پیشرفته وعلوم محیطی-دانشگاه تحصیلات تکمیلی و فناوری پیشرفته- کرمان- ایران
amoozegar@kgut.ac.ir

چکیده: با در نظر گرفتن هزینه‌های بسیار بالای تولید انرژی، شناسایی رفتار و الگوی مصرف مشترکین نیروی برق نقش بسیار موثری را در اتخاذ سیاستهای مدیریت انرژی ایفا می‌کند. همچنین با رقابتی شدن فعالیتهای شرکتهای توزیع این امر به تعریف نظامهای تعرفه ای صحیح نیز کمک می‌کند. از مهمترین مسائل این حوزه ارائه راهکاریست که علاوه بر میزان مصرف، شکل الگوی مصرف مشترکین را نیز به خوبی تشخیص دهد. این مقاله راهکاری مبتنی بر شبکه‌های خودسازمانده ارائه کرده است که در قدم اول ضمن طراحی، بهینه سازی و اجرای شبکه، نگاشتی از الگوهای ورودی تهیه می‌کند. سپس فرآیند دو مرحله‌ای خوشه‌بندی را انجام می‌دهد. در مرحله اول خوشه‌بندی نگاشت را با تمرکز بر شکل نمودارهای مصرف انجام داده سپس در مرحله دوم، هر یک از خوشه‌ها را با بر اساس میزان مصرفشان به زیرخوشه‌هایی تقسیم می‌کند. نماینده هر زیرخوشه، معرف یکی از الگوهای رفتاری مشترکین خواهد بود. علاوه بر شناسایی الگوها، تحلیل و تفسیر خروجیهای شهودی شبکه دانش مفیدی را در اختیار مدیران قرار می‌دهد. راهکار پیشنهادی بر روی داده‌های مصرف نیروی برق در سال ۹۳ در شهرستان عنبرآباد استان کرمان اعمال و نتایج با راهکارهای قبلی مورد مقایسه قرار گرفته است.

واژه‌های کلیدی: شبکه‌های خودسازمانده، الگوی مصرف مشترکین، خوشه بندی، مدیریت انرژی،

تاریخ ارسال مقاله : ۱۳۹۴ /۰۳/۲۳

تاریخ پذیرش مقاله : ۱۳۹۴/۰۷/۱۴

نام نویسنده‌ی مسئول : مریم آموزگار

نشانی نویسنده‌ی مسئول : کرمان – انتهای جاده هفت باغ- دانشگاه تحصیلات تکمیلی و فناوری پیشرفته

[DOR: 20.1001.1.23222344.1395.5.1.4.4]

۱- مقدمه

شناسایی، تحلیل و تفسیر الگوهای مصرف مشترکین نیروی برق به منظور ارزیابی رفتار آنان نقش بسیار موثری در طراحی و مدیریت صحیح شبکه‌های توزیع و همچنین مباحث مدیریت مصرف انرژی دارد. علاوه بر این، با توجه به مسائل مطرح در بازار برق و رقابتی شدن سیستمهای توزیع، مدیریت صحیح تعرفه‌ها و ارائه خدمات بهتر به مشترکین، به درک رفتار و الگوی مصرف آنان بستگی دارد.

در کشور ما، تنها یک بخش بندی کلی، تحت عناوین مشترکین خانگی، صنعتی، تجاری و کشاورزی وجود دارد. این در حالیست که مشترکین هریک از این بخش‌ها، خصوصا مشترکین خانگی که محدوده وسیعی را تشکیل می‌دهند، رفتارهای متفاوتی از نظر زمان، نوع و میزان مصرف دارند. لذا شناخت بیشتر و بخش بندی شده الگوهای مصرف مشترکین ضروری و مفید می‌باشد.

مهمترین مراحل دستیابی به الگوهای مصرف عبارتند از جمع آوری، پالایش و پیش پردازش داده‌های خام اولیه، خوشه‌بندی داده‌ها و نهایتا تعیین نماینده هر خوشه به عنوان الگوی نهایی. برای اجرای هر مرحله تکنیکها و روشهای بسیاری وجود دارد که حصول نتیجه مطلوب در گروه انتخاب راهکار مناسب است.

عمده تحقیقات انجام شده در این حوزه، به بررسی الگوریتمهای خوشه‌بندی و ارزیابی عملکرد آنها پرداخته‌اند[۱]، از جمله مهمترین روشهای مورد استفاده، می‌توان به K-means[2, 3], self-organizing Map (SOM)[4-6] و Fuzzy C-means (FCM)[7] اشاره کرد. علاوه بر این روشهایی مثل شبکه‌های عصبی هاپفیلد[۸]، الگوریتم ISODATA[9] و Support Vector Clustering (SVC) [10] به میزان کمتر مورد استفاده قرار گرفته‌اند. همچنین به منظور ارزیابی کارایی الگوریتمهای خوشه‌بندی در تفکیک و تشخیص الگوها، از معیارهای مهمی همچون شاخص دیویس-بولدین[۱۱]، سیلهوت[۱۲] و دان[۱۳] استفاده شده است.

به عنوان نمونه در [۱۴] یک راهکار دو مرحله‌ای ارائه شده است که در مرحله اول حجم بالای نمودارهای مصرف هر مشترک را خوشه‌بندی کرده تا به یک الگوی نمونه از مصرف وی دست یابد. در مرحله دوم الگوهای کلیه مشترکین را خوشه‌بندی کرده است. هدف عمده این تحقیق بررسی روشهای مختلف خوشه بندی Kmeans, Fuzzy Kmeans و AVQ بر اساس معیارهای ارزیابی مختلف بوده است.

راهکارهای فوق عمدتا به دنبال بهبود معیارهای کارایی و یاقتن تعداد بهینه خوشه‌ها هستند. بررسی دقیق و پیاده سازی راهکارهای فوق بر روی داده‌های این تحقیق، ضعف آنها را در خوشه‌بندی صحیح نمودارهای مصرف نشان داد. این تکنیکها به طور همزمان قادر به تشخیص شکل نمودار و میزان مصرف نبودند. بدین ترتیب این پژوهش بر یافتن راهکاری برای رفع مشکل مذکور متمرکز شد.

در [۱۵] ضمن مرور تحقیقات انجام شده و بررسی تکنیکهای مختلف مورد استفاده، نویسنده اذعان کرده است که استفاده از فاصله اقلیدسی به عنوان معیار تشابه نتایج مطلوبی را فراهم نمی‌آورد. فاصله اقلیدسی نقاط اوج بار و کم باری و یا بعبارتی شکل نمودار را منعکس نمی‌کند.

برای رفع مشکل مذکور دو سیاست عمده در تحقیقات اخیر مشاهده شد. سیاست اول استفاده از توابع تشابه دیگر مثل EEUPC است. در مرجع [۱۶] این تابع شباهت معرفی شده است که بر گرفته از تکنیکی با همین نام در مباحث بازیابی صوت است که برای مقایسه شباهت دو سیگنال صوتی مورد استفاده قرار می‌گیرد.

سیاست دیگر در مرجع [۱۷] ارائه شده است که خوشه‌بندی مشترکین را با استفاده از روش Kmeans به تعداد چندین برابر (۴۰ خوشه) انجام می‌دهد. بدین ترتیب در تنوع ایجاد شده، شکل نمودارها نیز تا حدی بروز می‌کند. سپس با ادغام خوشه‌های مشابه توسط کارشناس خبره، خوشه‌های نهایی بدست می‌آیند.

سیاست اول بار محاسباتی ایجاد می‌کند و سیاست دوم نیز در واقع یک راهکار نیمه‌خودکار و خسته‌کننده محسوب می‌شود. بدین ترتیب این تحقیق به منظور بهبود روشهای موجود، دو هدف را دنبال کرده است: هدف اول، توجه همزمان به هر دو مقوله شکل و میزان (دامنه) مصرف در الگوهای شناسایی شده و ارائه راهکاریست که مشکل راهکارهای ذکر شده را نداشته باشد. هدف دوم، ارائه راهکاریست که بیشترین بصری سازی و تجرید اطلاعات را داشته باشد تا امکان تحلیل و تفسیرهای ساده، صریح و کامل را بدون نیاز به دانش تخصصی برای مدیران و کارشناسان فراهم کند.

از آنجا که در بسیاری از مناطق، تعداد مشترکین بیشر از صد هزار نفر بوده که خوشه‌بندی را زمانبر می‌سازد و همچنین به منظور تحقق هدف دوم، ابتدا نگاشتی دو بعدی از داده‌ها توسط الگوریتم نگاشت خودسازمانده[۱] (SOM) بدست می‌آید که ضمن کوچک کردن فضای بررسی، خروجیهای تولید شده تحلیل و تفسیرهای بسیار خوبی را در اختیار قرار می‌دهند.

به منظور تحقق هدف اول، شناسایی الگوها به صورت سلسله مراتبی از روی نگاشت تولید شده، در دو مرحله انجام می‌شود. در مرحله اول، تنها شکل نمودار مورد توجه قرار گرفته و مشترکین خوشه‌بندی می‌شوند. در مرحله دوم با تمرکز بر میزان مصرف، هر یک از خوشه‌های بدست آمده به تعداد محدودی زیرخوشه تقسیم می‌شوند. نکته قابل توجه اینست که چون نگاشت حاصل تعداد بسیار محدودی گره دارد فرآیند خوشه‌بندی در زمان بسیار اندکی انجام می‌شود.

راهکار پیشنهادی بر روی داده‌های مصرف مشترکین خانگی شهرستان عنبرآباد استان کرمان، تحت پوشش شرکت توزیع نیروی برق جنوب استان کرمان اعمال شده است. ضمن استخراج الگوهای مصرف، خروجیهای بدست آمده با خروجی روشهای قبلی مورد مقایسه قرار گرفته است. راهکار پیشنهادی به صورت کاملا شهودی، ضمن تامین اهداف تحقیق، نتایج بهتری را بهمراه داشته است. الگوهای استخراج شده ضمن تفکیک مشترکین و نمایش رفتار آنها، تعریف صحیح تعرفه ها و سیاستگذاریهای مدیریت مصرف را امکان پذیر می‌سازند. همچنین با تعیین ماههای کم مصرف در هر الگو، می توان با کمترین اختلال برای اصلاح و یا توسعه شبکه برنامه‌ریزی کرد.

در ادامه در بخش دوم نگاشت خودسازمانده معرفی می‌شود، سپس در بخش سوم راهکار پیشنهادی ارائه و در بخش چهارم کلیه مراحل تا شناسایی و تفسر الگوهای مصرف مشترکین شهرستان عنبرآباد تشریح می‌شود. بخش انتهایی نیز به جمع بندی و نتیجه گیری می‌پردازد.

۲- نگاشت خودسازمانده

در این بخش به اختصار، نگاشت خودسازمانده به عنوان پایه و اساس تحقیق پیش رو معرفی می‌گردد.

نگاشت خود سازمانده (SOM) نوعی شبکه عصبی منظم و دوبعدی از نرون‌ها می‌باشد که هر نرون یا گره، مبین مدلی از مشاهدات می‌باشد. این گره‌ها در فرآیند یادگیری رقابتی نسبت به الگوهای ورودی منظم می‌شوند. محل واحدهای تنظیم شده در شبکه به گونه ای نظم می‌یابد که برای ویژگیهای ورودی، یک مختصات معنی دار روی شبکه ایجاد می‌شود. لذا SOM، یک نگاشت توپوگرافیک از الگوهای ورودی را تشکیل می‌دهد که در آن محل قرارگرفتن واحدها متناظر با ویژگیهای ذاتی الگوهای ورودی است.

مراحل کار در الگوریتم SOM به شرح زیر است:

الف- مرحله طراحی نگاشت از لحاظ ساختار و اندازه: طراحی شبکه با تعیین موارد زیر انجام می‌شود. مورد اول ساختار شبکه بندی محلی است که شکل ارتباطی نرونها یا گره‌ها با یکدیگر را مشخص می‌کند و می‌تواند در قالب مربعی و یا شش ضلعی باشد. مورد دوم تعیین شکل سراسری شبکه است که نحوه قرارگیری کل گره‌ها در شبکه را نشان می‌دهد و می‌تواند به صورت صفحه تخت، استوانه ای یا توری شکل باشد. مورد سوم نیز اندازه شبکه است که تعداد گره‌ها و ابعاد آنرا تعیین می‌کند.

ب- مرحله آموزش شبکه و تولید نگاشت خودسازمانده: آموزش شبکه، از نوع یادگیری رقابتی است. مراحل آموزش و ساخت این شبکه به شرح زیر می‌باشد:

۱- در ابتدای امر اوزان مربوط به هر یک از گره‌ها مقداردهی اولیه می‌شود.

۲- برداری از داده‌های آموزشی به طور تصادفی انتخاب شده و به شبکه داده می‌شود.

۳- فاصله اقلیدسی نمونه ورودی تا بردار وزن تمام گره‌های شبکه محاسبه می‌شود. گره ای که بردار وزنش بیشترین شباهت را به بردار ورودی داشته باشد برنده است. این گره بهترین واحد تطبیق یافته (BMU) نامیده می‌شود.

۴- شعاع همسایگی محاسبه می‌شود. مقدار این شعاع در ابتدا بزرگ و معمولا برابر شعاع نگاشت است ولی با هر گام زمانی کوچکتر می‌شود. هر گره داخل این شعاع به عنوان همسایه BMU در نظر گرفته می‌شود.

۵- اوزان هر یک از گره‌های همسایه که در گام قبل مشخص شده‌اند، طبق رابطه (۱) بروزرسانی می‌شوند تا میزان تمایل و گرایش آنها به همسایگان مشابه شان محاسبه و تاثیر آن ثبت شود. طبیعتا همسایگان نزدیکتر تغییرات بیشتری خواهند داشت.

$$W_i(t+1)=W_i(t)+\alpha(t)h_{ci}(t)\big(X(t)-W_i(t)\big) \qquad (۱)$$

در این رابطه $W_i(t)$ و $W_i(t+1)$ وزن i امین نرون برنده در تکرارهای t و t+1 هستند و $\alpha(t)$ ، $h_{ci}(t)$ و $X(t)$ به ترتیب، نرخ یادگیری، همسایگی اطراف نرون c و مشخصات بردار ورودی در تکرار t هستند.

۶- مرحله ۲ تا ۵ تکرار می‌شوند تا شرط خاتمه برقرار شود. شرط خاتمه می‌تواند رسیدن به تعداد تکرارهای مشخص یا دقت لازم برای شبکه باشد.

به طور کلی دو شیوه آموزش تعریف شده است: شیوه آموزش مرحله‌ای[۲]، که به روند آن در بالا اشاره شد. روش دیگر که یکجا[۳] خوانده می‌شود به جای ورود تک به تک بردارهای ورودی کلیه محاسبات و تغییرات به صورت یکجا اعمال می‌گردد.

به طور کلی آموزش شبکه چه در روش مرحله‌ای و چه روش یکجا، در دو فاز انجام می‌شود.

- فاز خودسازمان‌دهی[۴]: با تغییرات اساسی، وزنها در این فاز مرتب می‌شوند، اندازه همسایگی ابتدا به گونه‌ای انتخاب می‌شود که تقریبا همه واحدها در همسایگی باشند سپس تدریجا کوچک می‌شود تا در پایان فاز مرتب شدن به واحد برنده و یا چند همسایه اش برسد.
- فاز همگرائی[۵] : این فاز برای تنظیم دقیق نگاشت است. همسایگی در این فاز در ابتدا شامل واحد برنده و چند همسایه آن می‌شود که سپس به تدریج به خود واحد برنده می‌رسد.

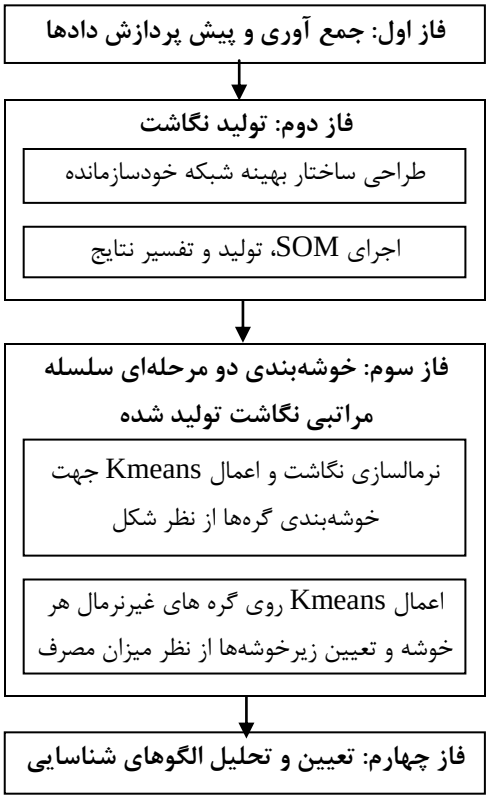
ج- بصری سازی نگاشت بدست آمده: در انتها می‌توان به تولید نمودارهای U-Matrix، ماتریس ویژگیها و بسیاری از خروجیهای قابل درک پرداخت که تحلیل، تفسیر و درک نگاشت تولید شده در SOM را بسیار ساده کرده و اطلاعات بسیار مفیدی در اختیار قرار می‌دهد.

۳- راهکار پیشنهادی شناسایی الگوهای مصرف

استخراج و شناسایی الگو از بین حجم انبوه داده‌ها یک مساله غیرنظارتی است که با استفاده از روشهای خوشه‌بندی قابل اجراست. در این راستا راهکارهای متعددی ارائه شده است که با استفاده از الگوریتمهای خوشه بندی، داده‌های مشابه را در یک خوشه قرار داده و الگویی را به عنوان نماینده ارائه می‌کنند. در واقع این نماینده بیانگر رفتار داده‌های آن خوشه می‌باشد.

این مساله در بانک اطلاعاتی مشترکین برق به منظور کشف الگوهای مصرف، قابل طرح و نتایج آن در امر مدیریت انرژی بسیار کارگشا می‌باشد. این مقاله به دنبال راهکاری است که به شکلی ساده، الگوهای مصرف را با تمرکز بر هر دو معیار شکل و میزان مصرف استخراج کند. و همچنین به صورت مصور و ساده نتایج قابل تفسیری در اختیار مدیران قرار دهد.

شکل (۱) مراحل راهکار پیشنهادی را نشان می‌دهد. در بخش بعد ضمن اعمال راهکار پیشنهادی بر روی نمونه مطالعاتی، جزئیات کامل هر مرحله تشریح خواهد شد.



شکل (۱): مراحل راهکار پیشنهادی

۴- پیاده سازی راهکار پیشنهادی

مراحل اجرای راهکار پیشنهادی، در این بخش تشریح خواهد شد.

۴-۱- جمع آوری، درک و پیش پردازش د ا د ه ها

با بررسی مناطق مختلف تحت پوشش شرکت توزیع نیروی برق جنوب استان کرمان و مذاکره با کارشناسان خبره این مجموعه، شهرستان عنبرآباد به دلیل تنوع بیشتر الگوی مصرف، به عنوان نمونه مطالعاتی انتخاب شد. این شهرستان در منطقه گرم استان کرمان واقع شده است و حدود بیست هزار مشترک خانگی در سال ۹۳ (۹۳ درصد کل مشترکین) دارد. از آنجا که محدوده و الگوی مصرف مشترکین صنعتی و کشاورزی بسیار متفاوت با مشترکین خانگی است و همچنین تعداد آنها نیز اندک است کنار گذاشته شدند. همچنین شهرستان عنبرآباد از پتانسیل بالای کشاورزی برخوردار است. لذا استخراج الگوهای مصرف مشترکین خانگی می‌تواند منجر به تشخیص مشترکین غیرمجازی شود که از تعرفه خانگی برای کشاورزی استفاده می‌کنند.

بدین ترتیب از بانک اطلاعات نرم افزار مشترکین، شش دوره مصرف سال ۹۳ مشترکین خانگی استخراج شد. دوره اول مربوط به دو ماهه اول سال و به همین ترتیب دوره ششم مربوط به دو ماه آخر سال می‌باشد. با توجه به قرارگیری این شش دوره در شرایط فصلی و آب و هوایی متفاوت، این اطلاعات رفتار مشترکین را به خوبی پوشش می‌دهد. به منظور پالایش داده‌ها مراحل زیر انجام شد:

در مرحله اول بررسی نظم در قرائتها یا همان طول دوره‌های قرائت مشترکین انجام شد. به طور معمول و متداول طول دوره‌های قرائت مصرف مشترکین بین ۵۵ تا ۶۵ روز می‌باشد. طبیعتا هر چه پراکندگی مربوط به طول دوره‌ها کمتر باشد تحلیلهای قابل اتکاتری انجام پذیر است. بدین منظور مشترکینی که قرائت آنها خارج از این بازه انجام شده است حذف شدند.

مورد بعدی که اثر بسیار زیادی بر کیفیت تحلیلها دارد نقص در قرائتها می‌باشد. دو نوع نقص عمده در قرائتها به چشم میخورد. مورد اول، موارد عدم قرائتهاست که مقدار خالی برای دوره ذخیره شده است و نقص دوم مربوط به مواردی است که به دلایل فنی مقدار مصرف، صفر ثبت شده است. با توجه به طولانی بودن طول دوره‌ها امکان پیش بینی این مقادیر وجود ندارد لذا مشترکین دارای این نواقص، حذف شدند.

بعد از انجام پالایشهای فوق، پردازشهای زیر بر روی ۹۴۸۲ مشترک، انجام شد:

یکسان سازی طول دوره‌ها: از آنجا که طول دوره‌ها برای مشترکین متفاوت است و این امر مقایسه را دچار مشکل می‌کند برای هر مشترک، میزان مصرف روزانه از تقسیم میزان مصرف دوره بر طول دوره محاسبه و در عدد ۶۰ ضرب شد. بدین ترتیب برای کلیه مشترکین طول دوره‌ها یکسان سازی شد.

نرمالسازی: یکی از مهمترین پارامترهای موثر در کیفیت خوشه‌بندی، انتخاب روش نرمالسازی مناسب است. به طور مرسوم نرمالسازی خطی و یا واریانس در این حوزه استفاده می‌شود. اما نرمالسازی بر اساس واریانس و میانگین، شرایط حاکم بر مصرف در دوره‌های مختلف را حذف می‌کند و میزان مصرف کلیه دوره‌ها را یکنواخت می‌کند. از آنجا

[DOR: 20.1001.1.23222344.1395.5.1.4.4]

که این تحقیق بر حفظ شکل الگوی مصرف تمرکز کرده است، نرمالسازی خطی انتخاب شد. با فرض اینکه مجموعه $D=\{C_i, i=1,2...N\}$ بانک اطلاعاتی N مشترک باشد و $C_i=\{C_{i1}, C_{i2},...C_{id}, d=1,2...6\}$، شش دوره مصرف مشترک iام، نرمالسازی هر دوره، طبق رابطه (۲) انجام شد.

$$\hat{C}_{ij}=\left(C_{ij}-C_{\min}\right)/\left(C_{\max}-C_{\min}\right) \qquad (۲)$$

در این رابطه C_{min} و C_{max} به ترتیب کمترین و بیشترین مقدار مصرف در دوره jام هستند. نهایتا در انتهای این فاز داده‌های شش دوره مصرف مشترکین آماده سازی شد.

۴-۲- طراحی شبکه خودسازمانده بهینه

گرچه استفاده از مقادیر پیش فرض برای طراحی و اجرای شبکه خودسازمانده می‌تواند پاسخگو باشد اما حصول نتایج مطلوب با قابلیت اطمینان بیشتر، به تنظیم صحیح پارامترهای آن بستگی دارد. لذا مرحله طراحی اهمیت زیادی دارد. در ابتدا مهمترین معیارهای سنجش صحت و دقت شبکه که در این تحقیق مورد استفاده قرار گرفته‌اند، معرفی می‌شوند:

خطای تدریج[۶](QE): میانگین فاصله هر نمونه تا مرکز گره مربوطه در نگاشت را محاسبه می‌کند. هر چه مقدار فاصله کمتر باشد خطای تدریج کمتر خواهد بود.

خطای توپوگرافیک[۷](TE): برای محاسبه این خطا، برای هر رکورد داده، اولین و دومین گره مربوطه در شبکه مشخص می‌شود. در صورتی که این دو گره مجاور یکدیگر نباشند یک واحد خطا لحاظ می‌شود. جمع نرمال شده این خطا، در بازه [۱و۰] ارائه می‌شود. هر چه مقدار این خطا به صفر نزدیکتر باشد خطای کمتری رخ داده است.

دو معیار فوق ساختار و دقت شبکه را به خوبی مورد ارزیابی قرار می‌دهند. مبنای ارزیابی عملکرد شبکه جمع این دو خطا طبق رابطه (۳) است:

$$I_s=QE+TE \qquad (۳)$$

در ادامه مراحل طراحی و تنظیم شبکه تشریح خواهد شد. کلیه مراحل پیاده سازی در ابزار Matlab انجام شده است.

- مرحله اول - تعریف ساختار و اندازه شبکه

ساختار شبکه بندی شش وجهی و شکل صفحه تخت در نظر گرفته شد. تعداد گره‌های شبکه بر اساس قاعده $5*\sqrt{N}$ برابر ۴۹۰ گره بدست آمد. ابعاد شبکه بر اساس دستورالعمل مرجع [۱۸] به صورت ۱۴*۳۵ تعیین شد.

- مرحله دوم - تعیین شیوه آموزش

در ابتدا شیوه آموزش مرحله‌ای با سه تابع یادگیری Power, Linear و Inv در مقایسه با شیوه آموزش یکجا با تابع همسایگی Gaussian و تعداد ۱۲۰۰ تکرار مورد بررسی قرار گرفت. جدول(۱) میزان نهایی خطای محاسبه شده طبق رابطه (۲) را نشان می‌دهد.

جدول(۱): خطای شبکه در آموزش یکجا و مرحله‌ای

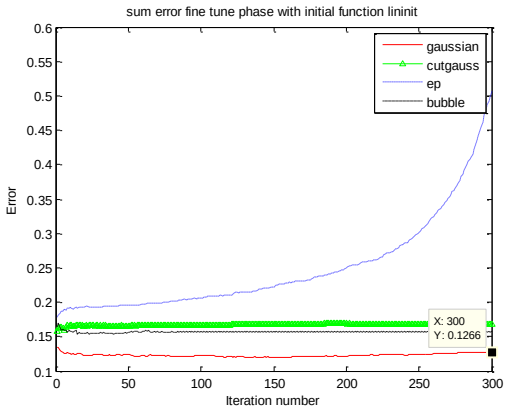
شیوه آموزش	یکجا	مرحله‌ای با نرخ آموزش		
		Power	Linear	Inv
مقدار I_s	۰.۱۱۹	۰.۱۰۹	۰.۱۲۳	۰.۱۲۱

با توجه به اینکه شیوه مرحله‌ای چندین برابر کندتر از روش یکجا عمل می‌کند و کیفیت بهتری را ایجاد نکرده است، شیوه آموزشی یکجا انتخاب می‌شود.

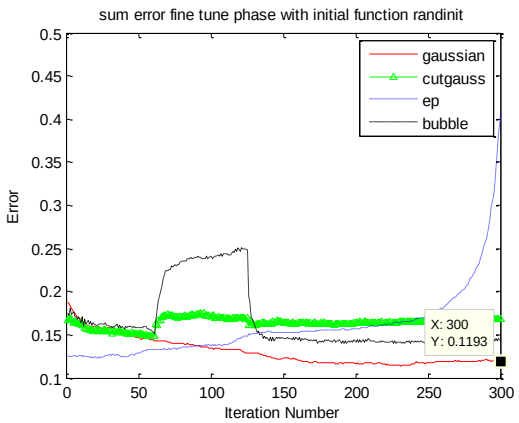
- مرحله سوم: تعیین روش مقداردهی اولیه و تابع همسایگی بهینه

در این مرحله این دو پارامتر مهم و تاثیرگذار مشخص می‌شوند. برای روش مقداردهی اولیه دو حالت خطی و تصادفی در نظر گرفته شد و در هر حالت نیز چهار تابع همسایگی شامل توابع Gaussian, Cutgauss, bubble, و ep مورد بررسی قرار گرفت.

شکل (۲) و شکل (۳) نمودار تغییرات خطای شبکه در ۳۰۰ تکرار فاز همگرایی را نشان می‌دهد. همانطور که از شکل مشخص است مقداردهی تصادفی و تابع همسایگی Gaussian وضعیت مطلوبتری را بوجود آورده‌اند. همچنین از نمودار مشخص است که ۳۰۰ تکرار برای حصول نتیجه مطلوب کفایت می‌کند.



شکل (۲): نمودار خطای شبکه در مقداردهی خطی



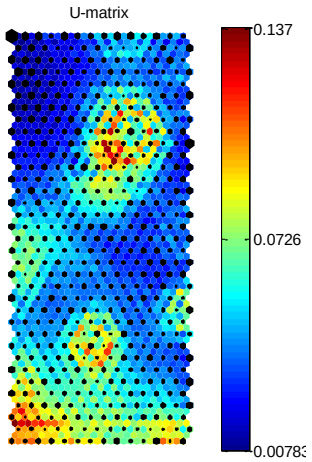
شکل (۳): نمودار خطای آموزش شبکه با مقداردهی تصادفی

۴-۳- اجرای شبکه و تفسیر نتایج

در این بخش شبکه، با تنظیمات بهینه بدست آمده در بخش قبل اجرا و خروجیهای حاصله مورد تحلیل و بررسی قرار خواهند گرفت.

یکی از مهمترین خروجیها، U-Matrix است که در شکل (۴) نمایش داده شده است.

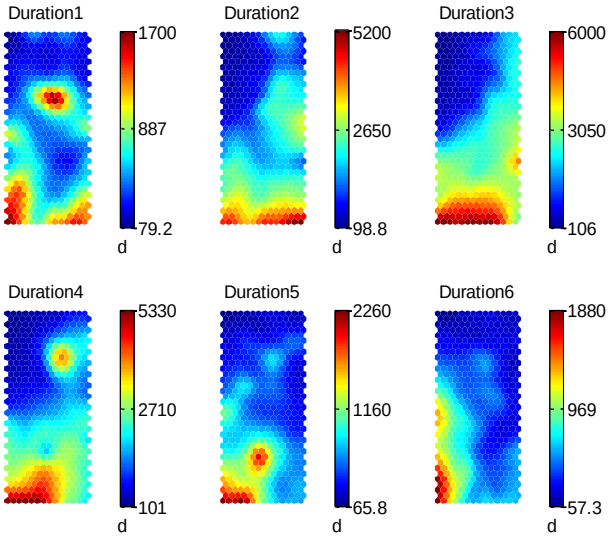
این ماتریس فاصله بین گره‌ها در شبکه را نشان می‌دهد. طبق این ماتریس گره‌های واقع در نوار پایینی و بخشهای مرکزی رو به بالا فاصله زیادی از یکدیگر دارند. اما در بخشهای دیگر گره‌ها به یکدیگر نزدیکترند. این ماتریس به گونه‌ای نیست که مرزبندی مشخصی از گره‌ها ارائه کند.



شکل (۴): U-matrix حاصل از اجرای SOM

تحلیل مفید دیگر، توزیع تعلق مشترکین به گره‌ها است که در واقع چگالی هر گره را نشان می‌دهد. چگالی گره‌ها در شکل (۴) و در قالب نقاط مشکی نمایش داده شده است. جز در تعداد اندکی از گره‌ها در قسمت بالای سمت چپ نگاشت که چگالی بیشتر است در بقیه نواحی تقریبا توزیع یکنواختی مشاهده می‌شود.

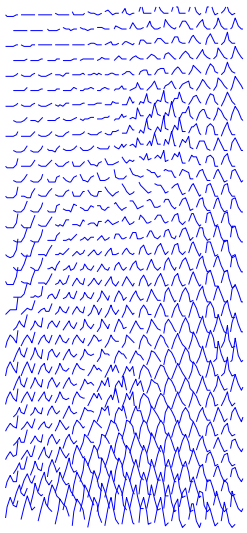
یکی دیگر از خروجیها، نگاشت‌های ویژگی هستند که توزیع مقادیر هریک از ویژگیها را بر روی نگاشت نشان می‌دهند. همچنین می‌توان از آنها برای بررسی همبستگی بین ویژگیها استفاده کرد. شکل (۵) نگاشت شش دوره مصرف سالیانه مشترکین را با مقدار واقعی و غیر نرمال نشان می‌دهد.



شکل (۵): ماتریس ویژگیهای حاصل از اجرای SOM

همانطور که از شکل استنباط می‌شود تعداد گره‌های با طیف رنگی بالا در دوره دوم و سوم و چهارم (خصوصا دوره سوم) بسیار بیشتر است. این دوره‌ها مربوط به ماههای خرداد تا شهریور می‌باشند. در واقع شهرستان عنبرآباد در منطقه گرم استان کرمان واقع شده و دوره گرمای طولانی دارد و همانطور که مشخص است مصرف این سه دوره، به طور مشخصی بیشتر است.

همانطور که قبلا هم گفته شد، الگوریتم SOM مدل‌ها را محاسبه می‌کند، به قسمی که فضای مشاهدات را بصورت بهینه توصیف کند. یکی از بهترین نمودارهایی که برای بصری سازی اطلاعات بسیار مفید است نمودار مصرف (مقادیر مصرف در شش دوره) هر گره در نگاشت است. شکل (۶) نمودار مذکور را نشان می‌دهد. دلیل به هم ریختگی و بلند و کوتاه بودن نمودارها، تنوع میزان مصرف مشترکین در گره‌های مختلف است. (**Error! Reference source not found.** که به منظور خوشه‌بندی در بخشهای بعدی تشریح و مورد استفاده قرار خواهد گرفت نظم و ترتیب بیشتری را نشان می‌دهد. اطلاعات مربوط به ماتریسهای شش دوره در این نمودار خلاصه شده است. (به عنوان نمونه گره های پرمصرف واقع در نوار پایین ماتریس دوره سوم (شکل (۵)).



شکل (۶): مصرف شش دوره هر گره از نگاشت تولید شده

۴-۴- خوشه‌بندی

همانطور که شکل (۶) نشان می‌دهد، نگاشت بدست آمده از SOM، الگوهای مشابه را به یکدیگر نزدیک کرده است. با این تعبیر می‌توان SOM را یک گراف تشابه یا یک دیاگرام خوشه‌یابی تلقی نمود. لذا در اولین قدم طبق روشهای مرسوم، نگاشت تولید شده خوشه‌بندی و سپس در ادامه، راهکار پیشنهادی اعمال و نتایج حاصله مورد تحلیل و مقایسه قرار خواهند گرفت.

یکی از مسائل اثرگذار در خوشه‌بندی و تشخیص صحیح الگوها، تعیین تعداد بهینه خوشه‌هاست. بدین منظور از شاخص دیویس –

بولدین استفاده می‌شود. طبق رابطه (۴) این شاخص تابعی از نسبت پراش درون خوشه به فاصله بین خوشه‌هاست.

$$DB = \frac{1}{n}\sum_{\substack{i=1 \\ i\neq j}}^{n} \max\left\{\frac{S_n(Q_i)+S_n(Q_j)}{S(Q_i,Q_j)}\right\} \qquad (۴)$$

در این رابطه، S_n میانگین فاصله داده‌های خوشه از مرکز آن و $S(Q_i,Q_j)$ فاصله بین مراکز خوشه‌هاست. هنگامی که درون خوشه به هم نزدیک و خوشه‌ها از یکدیگر دور باشند این نسبت کوچک شده و خوشه‌بندی، معتبرتر خواهد بود.

نمودار **Error! Reference source not found.**) تغییرات این شاخص را در افزایش خوشه‌ها از ۲ تا ۸ نشان می‌دهد. برای هر یک از تعداد خوشه‌ها، الگوریتم Kmeans، ۲۰ بار اجرا و نتیجه با کمترین خطا ثبت شده است. با ارجاع به نمودار مشخص است که تعداد ۳ خوشه حالت بهینه می‌باشد. **Error! Reference source not found.**) نگاشت خوشه‌بندی شده را با سه خوشه نشان می‌دهد. یک مقایسه ساده بین خوشه‌بندی بدست آمده در **Error! Reference source not found.**) و نمودارهای مصرف گره‌ها در شکل (۶) و (۸) نشان می‌دهد که تشابه شکل نمودارهای مصرف، مورد توجه قرار نگرفته است. در این خوشه بندی، نگاشت به ۳ بخش تقسیم شده و عمدتا میزان مصرف تعیین کننده بوده است. به عنوان مثال مشترکین بالای نگاشت، فارغ از شکل الگوی مصرفشان همگی در یک خوشه قرار گرفته‌اند، ولی سه شکل یا الگوی متمایز دارند. این در حالیست که در شناسایی الگوهای مصرف، شکل نمودار از اهمیت زیادی برخوردار است.

در بسیاری از کاربردهای SOM ، خوشه‌بندی U-matrix انجام می‌شود. اما در داده‌های این تحقیق، به دلیل یکنواختی U-matrix پاسخ مطلوبی بدست نمی آید.

در ادامه، فرآیند دو مرحله‌ای پیشنهادی برای خوشه‌بندی نگاشت تولید شده تشریح می‌شود.

در مرحله اول، تمرکز بر خوشه‌بندی گره‌هایی است که شکل نمودار مصرف آنها مشابه است. بدین منظور، بدون هیچگونه پردازش پیچیده و زمانبری، مقادیر شش دوره مصرف هر گره در نگاشت حاصله، بین صفر و یک نرمال می‌شوند. دقت شود نرمال سازی مرحله پیش پردازش به ازای هر دوره کلیه مشترکین انجام شد اما در اینجا شش دوره مصرف هر گره، نرمال خطی می‌شوند. نگاشت نرمال شده در **Error! Reference source not found.**) نمایش داده شد.

سپس خوشه‌بندی نگاشت حاصل با الگوریتم Kmeans برای ۲ تا ۸ خوشه انجام و مقادیر شاخص دیویس بولدین ثبت گردید. مطابق نمودار **Error! Reference source not found.**) تعداد شش خوشه نتیجه مطلوبتری دارد. نگاشت خوشه‌بندی شده در سمت راست این شکل نشان داده شده است.

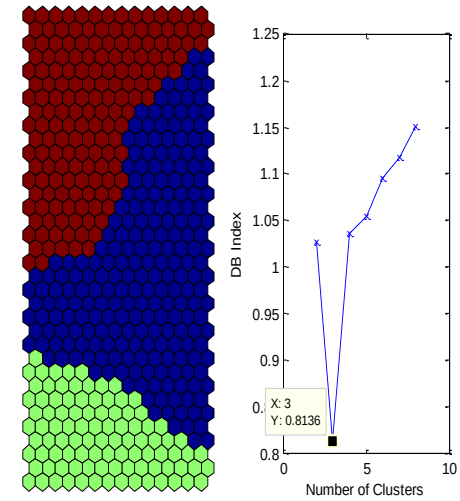
با یک مقایسه ساده بین خوشه‌های بدست آمده و نمودار مصرف گره‌ها می‌توان به راحتی عملکرد بهتر خوشه‌بندی انجام شده را مشاهده کرد. در نیمه بالایی نگاشت، نمودارهای V یا U شکل از نمودارهای معکوس یا سینوسی شکل جدا شده‌اند و سه خوشه متمایز از نظر شکل، به خوبی از یکدیگر تفکیک شده‌اند. حتی ناحیه دایره مانند وسط نگاشت، به خوبی از بقیه نواحی جدا شده است.

در اینجا لازم است دقت شود کار خوشه‌بندی به پایان نرسیده است، چرا که نرمالسازی مصارف یک گره به منظور انعکاس شکل نمودار، اثر تنوع مصرف را حذف کرده است. این در حالیست که مشترکین هر خوشه ممکن است محدوده‌های مصرف متفاوتی داشته باشند. لذا در مرحله دوم، تمرکز بر میزان مصرف قرار می‌گیرد.

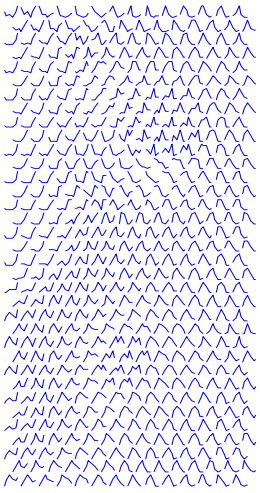
برای لحاظ کردن میزان مصرف، الگوریتم Kmeans بر روی هر یک از خوشه‌های بدست آمده از مرحله قبل اعمال می‌شود (نه بر روی کل نگاشت) تا به دو زیرخوشه تقسیم شوند. از آنجا که میزان مصرف ملاک است، برای گره های هر خوشه، به نگاشت اولیه تولید شده، یعنی شکل (۶) رجوع می‌شود.

شکل (۱۰) خوشه‌بندی نهایی را نشان می‌دهد. در این شکل خوشه‌ها با رنگهای متمایز مشخص شده‌اند. در هر گره رقم اول، شماره خوشه و رقم دوم شماره زیرخوشه را مشخص می‌کند. به عنوان نمونه عدد ۴۱ نشان دهنده تعلق گره به خوشه چهار و زیرخوشه یک است.

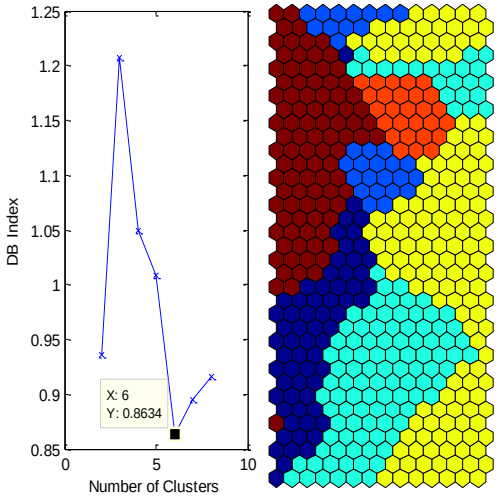
[DOR: 20.1001.1.23222344.1395.5.1.4.4]



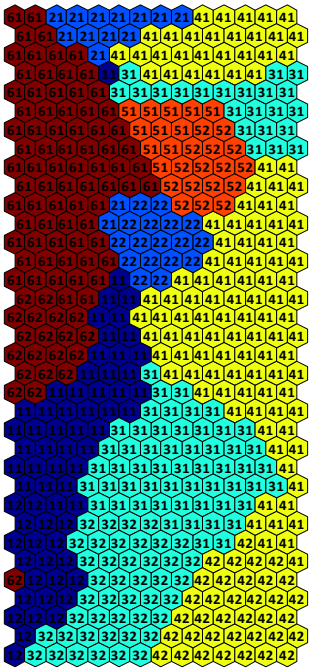
شکل (۷): خوشه‌بندی نگاشت و نمودارشاخصDB



شکل (۸): نگاشت نرمال شده



شکل (۹): خوشه‌بندی نگاشت نرمال شده و شاخصDB



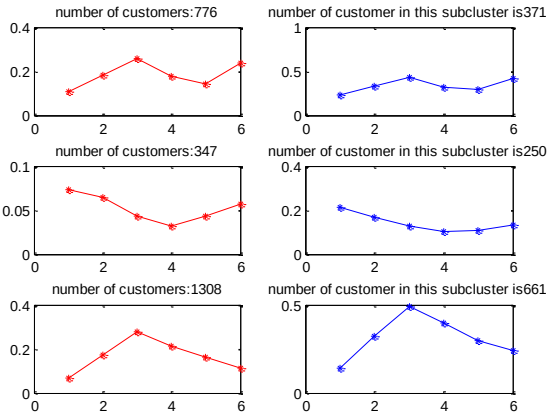
شکل (۱۰): نگاشت بهمراه خوشه‌ها و زیرخوشه‌ها

مقایسه شکل (۱۰) و **Error! Reference source not found.**) نشان می‌دهد که گره‌ها، هم از نظر شکل نمودار مصرف و هم میزان مصرف به خوبی از هم تفکیک شده‌اند. به عنوان نمونه گره‌های ۳۱ و ۳۲ همشکل، اما از نظر میزان مصرف به ترتیب کم مصرف و پرمصرف هستند.

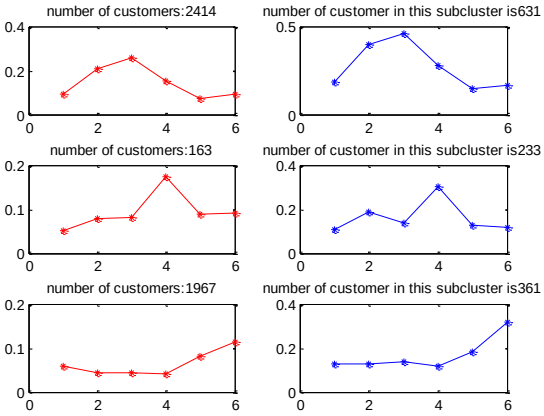
البته لزوما تعداد دو زیر خوشه مد نظر نیست بلکه می‌توان با تعیین بیشینه، کمینه و محدوده مصرف هر خوشه، بر اساس تنوع میزان مصرف، تعداد زیرخوشه‌ها را بین یک تا ۳ تنظیم نمود.

۴-۵- شناسایی و تفسیر الگوهای نهایی

پس از مرحله خوشه‌بندی لازم است، نمایندگان ۱۲ زیرخوشه تعیین شده به عنوان الگوهای نهایی مشخص گردند. برای تعیین نماینده در هر زیرخوشه کافیست به ازای هر دوره، میانگین مصرف گره‌های عضو آن زیرخوشه محاسبه شود. در شکل (۱۱) کلیه الگوهای شناسایی شده به ترتیب کد خوشه و زیر خوشه نمایش داده شده‌اند. (خوشه‌ها از ردیف ۱تا ۶ و در هر ردیف نمودار قرمز، سمت چپ زیرخوشه ۱ و نمودار آبی در سمت راست زیرخوشه ۲ می‌باشد)



[DOR: 20.1001.1.23222344.1395.5.1.4.4]



شکل (۱۱): نمودار الگوهای شناسایی شده

در بالای هر نمودار تعداد کل مشترکین در هر الگو درج شده است. همچنین جدول ۲) درصد تعلق مشترکین به هر یک از زیرخوشه‌ها (الگوهای متمایز از نظر شکل و محدوده مصرف) و همچنین کل خوشه (الگوهای با شکل متمایز) را نشان می‌دهد. الگوی متعلق به خوشه ۴ با ۳۲ درصد مشترکین، الگوی غالب است که ۲۵.۴ درصد آنان در دسته مصرف پایین و متوسط و ۶.۶ درصد در دسته پرمصرف هستند.

به طور کلی الگوهای ردیف ۳ و ۴ که موید مصرف بیشتر در دوره‌های گرم سال هستند جمعا ۵۲ درصد مشترکین را پوشش می‌دهند. نکته جالب توجه حضور ۲۴ درصد مشترکین در الگوی ردیف ۶ است که دوره ۶ پرمصرفی دارند. آماری از این دست کمکهای شایانی به کارشناسان و مدیران، جهت کنترل اوضاع و یا برنامه ریزیهای آتی می‌کند.

جدول ۲: درصد تعلق مشترکین در هر الگو یا زیرخوشه

	زیر خوشه ۱	زیرخوشه ۲	کل
خوشه ۱	۸.۱۸	۳.۹۱	۱۲.۱۰
خوشه ۲	۳.۶۶	۲.۶۴	۶.۳۰
خوشه ۳	۱۳.۷۹	۶.۹۷	۲۰.۷۷
خوشه ۴	۲۵.۴۶	۶.۶۵	۳۲.۱۱
خوشه ۵	۱.۷۲	۲.۴۶	۴.۱۸
خوشه ۶	۲۰.۷۴	۳.۸۱	۲۴.۵۵

۵- نتیجه

در این مقاله به منظور شناسایی الگوهای مصرف برق، راهکاری ارائه شد که علاوه بر میزان مصرف، شکل الگوی مصرف مشترکین را نیز در نظر دارد. بدین منظور از شبکه‌های خودسازمانده برای تولید نگاشت دوبعدی از داده‌های مصرف مشترکین استفاده شد تا ضمن تولید خروجیهای شهودی و قابل تفسیر، امکان خوشه‌بندی ساده و کم هزینه داده های مصرف فراهم گردد.

مراحل کار بدین صورت بود که به منظور تولید پاسخ مطمئن و بهینه، در ابتدای امر، طراحی شبکه انجام و بهینه سازی شد. سپس با اجرای SOM تحت شرایط بهینه تعیین شده، خروجیهای مهمی همچون، ماتریس ویژگیها، U-Matrix و نگاشت شهودی در قالب نمودارهای مصرف استخراج شد.

ابتدا با نرمالسازی داده‌های مصرف در هر گره از نگاشت تولید شده، شکل الگوی مصرف در اولویت قرار گرفت. با خوشه‌بندی این نگاشت، مشترکین دارای الگوی مشابه در خوشه‌های مجزا قرار گرفتند. سپس هر یک از خوشه‌ها به منظور رده‌بندی میزان مصرف، به زیر خوشه‌هایی تقسیم شدند. در انتها نماینده، با میانگین‌گیری از مصرف مشترکین هر زیرخوشه، مشخص و به عنوان الگوی نهایی ارائه شد. بدین ترتیب راهکار پیشنهادی توانست هم شکل الگوی مصرف و هم میزان مصرف مشترکین را برای خوشه‌بندی لحاظ کرده، در عین حال با ارائه خروجیهای ساده و شهودی اطلاعات و دانش مفیدی را در اختیار مدیران قرار دهد. کلیه مراحل فوق بر روی داده‌های مصرف برق سال ۹۳ شهرستان عنبرآباد انجام و خروجیها نمایش داده شد.

به منظور تحقیقات تکمیلی می‌توان ابتدا تحلیل مصرف را در هر یک از خوشه‌های بدست آمده در مرحله اول انجام و تعداد صحیح زیرخوشه‌ها را تعیین نمود. علاوه بر این می‌توان برای خوشه‌بندی نگاشت، از الگوریتمهای دیگری غیر از Kmeans استفاده نمود. همچنین استفاده از شاخصهای ارزیابی دیگر در کنار شاخص دیویس بولدین به ارزیابی جامعتر و بهتر الگوریتم خوشه‌بندی و تقویت این تحقیق کمک می‌کند.

سپاسگزاری

این طرح با حمایت شرکت توزیع نیروی برق جنوب استان کرمان اجرا شده و در اینجا از زحمات کلیه همکاران این مجموعه، به ویژه سرکار خانم امیرتیموری کارشناس بخش مشترکین قدردانی می‌گردد.

مراجع

[1] I. P. Panapakidis, M. C. Alexiadis, and G. K. Papagiannis, "Deriving the optimal number of clusters in the electricity consumer segmentation procedure," in *European Energy Market (EEM), 2013 10th International Conference on the*, 2013, pp. 1-8.

[2] D. Arthur and S. Vassilvitskii, "k-means++: The advantages of careful seeding," in *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, 2007, pp. 1027-1035.

[3] I. Benítez, A. Quijano, J.-L. Díez, and I. Delgado, "Dynamic clustering segmentation applied to load profiles of energy consumption from Spanish customers," *International Journal of Electrical Power & Energy Systems*, vol. 55, pp. 437-448, 2014.

[4] G. Tsekouras, P. Kotoulas, C. Tsirekis, E. Dialynas, and N. Hatziargyriou, "A pattern recognition methodology for evaluation of load profiles and typical days of large electricity customers," *Electric Power Systems Research*, vol. 78, pp. 1494-1510, 2008.

[5] T. Räsänen, D. Voukantsis, H. Niska, K. Karatzas, and M. Kolehmainen, "Data-based method for creating electricity use load profiles using large amount of

[DOR: 20.1001.1.23222344.1395.5.1.4.4]

customer-specific hourly measured electricity use data," *Applied Energy*, vol. 87, pp. 3538-3545, 2010.

[6] A. Seret, T. Verbraken, S. Versailles, and B. Baesens, "A new SOM-based method for profile generation: Theory and an application in direct marketing," *European Journal of Operational Research*, vol. 220, pp. 199-209, 2012.

[7] M. J. Hossain, A. Kabir, M. M. Rahman, B. Kabir, and M. R. Islam, "Determination of typical load profile of consumers using fuzzy c-means clustering algorithm," *International Journal of Soft Computing and Engineering (IJSCE)*, pp. 2231-2307, 2011.

[8] J. J. López, J. A. Aguado, F. Martín, F. Munoz, A. Rodríguez, and J. E. Ruiz, "Hopfield–K-Means clustering algorithm: A proposal for the segmentation of electricity customers," *Electric Power Systems Research*, vol. 81, pp. 716-724, 2011.

[9] A. Mutanen, M. Ruska, S. Repo, and P. Järventausta, "Customer classification and load profiling method for distribution systems," *Power Delivery, IEEE Transactions on*, vol. 26, pp. 1755-1763, 2011.

[10] G. Chicco and I.-S. Ilie, "Support vector clustering of electrical load pattern data," *Power Systems, IEEE Transactions on*, vol. 24, pp. 1619-1628, 2009.

[11] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, pp. 224-227, 1979.

[12] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics*, vol. 20, pp. 53-65, 1987.

[13] J. C. Dunn†, "Well-separated clusters and optimal fuzzy partitions," *Journal of cybernetics*, vol. 4, pp. 95-104, 1974.

[14] G. J. Tsekouras, N. D. Hatziargyriou, and E. N. Dialynas, "Two-stage pattern recognition of load curves for classification of electricity customers," *Power Systems, IEEE Transactions on*, vol. 22, pp. 1120-1128, 2007.

[15] I. P. Panapakidis, T. A. Papadopoulos, G. C. Christoforidis, and G. K. Papagiannis, "Pattern recognition algorithms for electricity load curve analysis of buildings," *Energy and Buildings*, vol. 73, pp. 137-145, 2014.

[16] Y. Wang, X. Wang, Y. Qian, H. Luo, F. Ge, Y. Yang, *et al.*, "Residential Load Pattern Analysis for Smart Grid Applications Based on Audio Feature EEUPC," *Global Applications of Pervasive and Ubiquitous Computing*, p. 107, 2012.

[17] B. A. Smith, J. Wong, and R. Rajagopal, "A simple way to use interval data to segment residential customers for energy efficiency and demand response program targeting," in *ACEEE Summer Study on Energy Efficiency in Buildings*, 2012.

[18] A. A. Akinduko and E. M. Mirkes, "Initialization of Self-Organizing Maps: Principal Components Versus Random Initialization. A Case Study," *arXiv preprint arXiv:1210.5873*, 2012.

فهرست علائم

$W_i(t)$	وزن نرون i در تکرار t
$\alpha(t)$	نرخ یادگیری در تکرار t
$h_{ci}(t)$	همسایگی اطراف نرون c در تکرار t
$X(t)$	مشخصات بردار ورودی در تکرار t
N	تعداد مشترکین در بانک اطلاعات
C_{ij}	میزان مصرف دوره jام مشترک i
C_{min}	کمترین مقدار مصرف در دوره jام
C_{max}	بیشترین مقدار مصرف در دوره jام
QE	خطای تدریج
TE	خطای توپوگرافیک
S (Q)	میانگین فاصله داده‌های خوشه Q از مرکز آن
$S(Q_i,Q_j)$	فاصله بین مراکز خوشه Q_i و Q_j

زیرنویس‌ها

۱ Self-Organizing Map
۲ Sequencing
۳ Batch
۴ Ordering
۵ Tuning
۶ Quantization Error
۷ Topographic Error

[DOR: 20.1001.1.23222344.1395.5.1.4.4]