ارائه یک مدل دومرحله‌ای جهت تشخیص تقلب در شبکه توزیع به‌وسیله یادگیری عمیق

مهدی عمادالاسلامی[۱] حسن مجیدی[۲] محمودرضا حقی‌فام[۳]

۱- دانشجوی دکتری- دانشکده مهندسی برق و کامپیوتر- دانشگاه تربیت مدرس- تهران- ایران
mahdi.emadaleslami@modares.ac.ir

۲- دانش آموخته کارشناسی ارشد- دانشکده مهندسی برق و کامپیوتر- دانشگاه تربیت مدرس- تهران- ایران
hassan.majidi20@modares.ac.ir

۳- استاد- دانشکده مهندسی برق و کامپیوتر- دانشگاه تربیت مدرس- تهران- ایران
haghifam@modares.ac.ir

چکیده: شرکت‌های برق از دیرباز به دنبال شناسایی و کاهش موارد برق‌دزدی به‌عنوان اصلی‌ترین بخش تلفات غیر فنی بوده‌اند.از طرفی شناسایی این موارد لزوماً از طریق بازرسی مشترکین ممکن است که شرکت‌های برق به دلایلی نظیر هزینه بالا، تعداد مشترکین و ... به دنبال کاهش محدوده بازرسی به موارد با احتمال برق‌دزدی بیشتر هستند. یکی از راهکارهای کاهش محدوده بازرسی، استفاده از روش‌های هوش مصنوعی است، اما چالش مهمی که در این حوزه وجود دارد عدم تعادل در نسبت مشترکین مشکوک به مشترکین عادی است که منجر به عملکرد ضعیف الگوریتم‌ها می‌شود. در این مقاله باهدف غلبه بر این چالش با فرض اینکه بتوان رفتار مشترک مشکوک را به‌صورت تابع ریاضی از رفتار مشترک عادی بیان کرد، در مرحله اول الگوی مصرف مشترکین عادی و مشکوک دسته‌بندی‌شده است؛ سپس یک شبکه عمیق اولیه جهت مدل‌سازی رفتار مشترکین مشکوک آموزش داده‌شده است. در ادامه به کمک شبکه آموزش داده‌شده اولیه، سناریوهای محتمل برق‌دزدی به ازای مشترکین عادی پیش‌بینی‌شده است. درنهایت یک شبکه عمیق ثانویه جهت تفکیک مشترکین عادی و مشکوک آموزش داده‌شده است. بررسی مدل پیشنهادی به ازای سناریوهای مختلف و مقایسه با تحقیقات پیشین بر روی مجموعه داده واقعی با بیش از ۶۰۰۰ مشترک عملکرد بالای آن را نشان می‌دهد.

واژه‌های کلیدی: برق‌دزدی، تشخیص تقلب، دسته‌بندی مشترکین مشکوک، پیش‌بینی الگو مصرف، یادگیری عمیق

نوع مقاله: پژوهشی

DOI: 10.52547/jiaeee.19.1.13

تاریخ ارسال مقاله: ۱۴۰۰/۰۲/۲۳
تاریخ پذیرش مشروط مقاله: ۱۴۰۰/۰۵/۲۵
تاریخ پذیرش مقاله: ۱۴۰۰/۰۶/۱۶
نام نویسنده‌ی مسئول: دکتر محمودرضا حقی‌فام
نشانی نویسنده‌ی مسئول: ایران – تهران – خیابان جلال آل احمد –دانشگاه تربیت مدرس – دانشکده‌ی برق و کامپیوتر

۱- مقدمه

تلفات شبکه توزیع به‌صورت تفاوت بین انرژی الکتریکی ورودی به شبکه توزیع از منابع داخلی، شبکه بالادست یا پایین‌دست و انرژی الکتریکی خروجی از آن به‌صورت درصدی برای بازه زمانی مشخص تعریف می‌شود. این تلفات به‌طور معمول به دو گروه تلفات فنی و غیرفنی تقسیم می‌شوند [۱،۲۱]. امروزه تلفات غیر فنی یکی از معضلات اصلی صنعت برق به شمار می‌رود به‌طوری‌که مشترکین متخلف سعی می‌کنند در عین مصرف زیاد انرژی، اطلاعات ثبت‌شده توسط کنتورها را به کمک روش‌های متعددی دستکاری‌شده کنند و سعی دارند با انتقال موارد پرمصرف و غیرمجاز به زمان‌هایی با تعرفه پایین‌تر، مصرفی کمتر از مصرف واقعی را گزارش کنند. از طرفی شناسایی این موارد لزوماً از طریق بازرسی مشترکین ممکن است اما به دلایل متعددی نظیر هزینه بالا، زیاد بودن تعداد مشترکین، پیچیدگی شبکه و ... امکان بازرسی و صحت‌سنجی اطلاعات مصرف تمامی مشترکین و مشخص کردن صحت مصرف آن‌ها وجود ندارد و شرکت‌های برق عموماً اطلاعات محدودی از مشترکین برق دزد دارند و هموار به دنبال محدود کردن محدوده بازرسی به موارد با احتمال برق‌دزدی بیشتر هستند. ازجمله راهکارهای کاهش محدوده بازرسی، استفاده از روش‌های مبتنی بر تخمین حالت، نظریه بازی[1] و هوش مصنوعی به‌منظور استخراج الگوریتمی جهت شناسایی موارد برق‌دزدی و عادی است [۱]. روش‌های مبتنی بر تخمین حالت توانایی شناسایی موارد برق‌دزدی را ندارد و در بهترین حالت منطقه برق‌دزدی را با احتمال بالا تشخیص می‌دهند علاوه بر این، هزینه تجهیزات و بسترهای مخابراتی آن‌ها زیاد است [۲, ۳]. روش‌های مبتنی بر نظریه بازی به دلایلی نظیر پیچیدگی مدل‌سازی رفتار مشترکین، تحلیل و بررسی اطلاعات مصرف و الگوهای بار مشترکین عملکرد خوبی از خود نشان نمی‌دهند. موارد مذکور به‌عنوان چالش‌های اصلی در تشخیص برق‌دزدی، سبب روی‌آوری به علومی مانند داده‌کاوی و هوش مصنوعی شده است و اهمیت این موضوع درگذر زمان و با پیشرفت تجهیزات سخت‌افزاری نه‌تنها کم‌رنگ‌تر نشده بلکه امروزه بیشتر مطالعات صورت‌گرفته در مسائل تحلیل بار، شناسایی برق‌دزدی و دیگر مسائل حوزه برق با تمرکز بر روش‌های داده‌کاوی و هوش مصنوعی است. اما چالش مهمی که در حوزه هوش مصنوعی وجود دارد، مسئله عدم تعادل در نسبت داده‌های صحیح و غلط است به‌طوری‌که اگر این نسبت در نمونه‌های جمع‌آوری‌شده رعایت نشود، عملکرد الگوریتم‌های آموزش داده‌شده بر اساس نمونه‌های جمع‌آوری‌شده قابل‌اتکا نیست و سبب کشیدگی الگوریتم به سمت داده‌های صحیح می‌شود. همین امر تشخیص برق‌دزدی را که به‌صورت طبیعی یک مسئله نامتعادل از جهت نسبت مشرکین سارق به مشترکین عادی است را دشوار می‌کند و منجر به عملکرد ضعیف الگوریتم‌های آموزش داده‌شده در این حوزه می‌شود.

یکی از روش‌های غلبه بر مسئله عدم تعادل داده‌ها، پیش‌بینی و تخمین رفتار مشترکین مشکوک با فرض اینکه بتوان رفتار یک مشترک مشکوک را به‌صورت توابع ریاضی از رفتار مشترکین عادی بیان کرد، است. در آن صورت می‌توان این انتظار را داشت که با تکیه‌بر این توابع ریاضی، تعداد نمونه‌های موردنیاز جهت آموزش الگوریتم را افزایش داد و بر مسئله عدم تعادل چیره شد. در این تحقیق با تکیه‌بر این فرض که رابطه بین مصرف واقعی و مصرف ثبت‌شده به‌صورت ریاضی قابل‌بیان است، در ابتدا سعی شده الگوی مصرف مشترکین عادی و مشکوک به کمک الگوریتم خوشه‌بندی K-means دسته‌بندی‌شده است. سپس این رابطه به‌صورت ریاضی مدل شده و یک شبکه عصبی عمیق DNN جهت مدل‌سازی رفتار مشترکین مشکوک آموزش داده‌شده است.

به‌طور خلاصه نوآوری‌های این مقاله به شرح زیر است:

- استفاده از اطلاعات مصرف، آماری و معیشتی مشترکین در تحلیل رفتار مشترکین
- استخراج الگوهای مختلف مصرف مشترکین مشکوک و عادی با در نظر گرفتن تعداد صحیح دسته‌ها
- ارائه یک شبکه عمیق جهت مدل‌سازی و تخمین الگوی رفتاری مشترکین مشکوک
- ارائه یک شبکه عمیق ثانویه جهت دسته‌بندی مشترکین
- بررسی سناریوهای مختلف تشخیص برق‌دزدی

۱-۱- پیشینه تحقیق

به‌طور خلاصه روش‌های تشخیص برق‌دزدی را براساس تحقیقات اخیر می‌توان به دو دسته سخت‌افزاری و نرم‌افزاری تقسیم کرد. روش نرم‌افزاری عموماً براساس تخمین حالت ، تئوری بازی یا الگوریتم‌های دسته‌بندی انجام می‌شود [۴]. با توجه به اینکه روش پیشنهادی در این تحقیق یک روش نرم‌افزاری مبتنی بر دسته‌بندی است. بنابراین، در ادامه پژوهش‌هایی که به‌وسیله بهره‌گیری از داده‌های کنتور هوشمند سعی در تشخیص مشترکین متقلب را دارند موردبررسی قرار می‌گیرند. نویسندگان در[۴] ابتدا مجموعه انرژی مصرفی کنتورهای مشترکین هر منطقه و میزان تلفات خط را محاسبه کرده‌اند و با مقایسه انرژی مصرف‌شده کل و میزان انرژی گزارش‌شده توسط کنتور ترانس توزیع هر منطقه، مناطقی که دارای اختلاف زیادی هستند را به‌عنوان مناطق مستعد برق‌دزدی انتخاب کرده‌اند. سپس با فرض این نکته که الگوی رفتار مشترکین مشکوک را می‌توان به‌صورت تابعی از رفتار مشترکین عادی هر منطقه دانست، به کمک تعریف تابع‌های متعددی از الگو مشترکین عادی یک مجموعه داده ساختی ایجاد کرد. درنهایت به کمک دیتاست ساختگی[2] و اطلاعات مصرف مشترکین عادی، یک الگوریتم دسته‌بندی مبتنی بر ماشین بردار پشتیبان[3] (SVM) را آموزش می‌دهد و به کمک آن اگر مشترکی برای تعداد بار مشخصی به‌عنوان دزد توسط سیستم شناسایی شود، سیستم آن را به‌عنوان یک

کاندید برق‌دزدی جهت بازدید حضوری معرفی می‌کند. مرجع [۵] از یک روش بالا به پایین مبتنی بر درخت تصمیم[۴] و SVM جهت شناسایی برق‌دزدی استفاده کرده است. درخت تصمیم در ابتدا میزان مصرف مورد انتظار یک مشترک را بر اساس اطلاعاتی نظیر تعداد تجهیزات خانه، افراد خانه، دمای بیرون و ... تخمین می‌زند و سپس این مقدار به همراه دیگر ویژگی‌های ورودی درخت تصمیم جهت آموزش SVM استفاده می‌شوند. درنهایت SVM تشخیص می‌دهد الگوی مصرف یک مشترک عادی است یا غیرعادی. مرجع [۶] از الگوریتم خوشه‌بندی مسیر بهینه جنگل[۵] جهت دسته‌بندی الگو مصرف مشترکین استفاده کرده است. پس از دسته‌بندی مشترکین به گروه‌های مختلف اگر فاصله هر پروفیل بار از مرکز خوشه از آستانه‌ای مشخص بیشتر باشد، مشترک موردنظر به‌عنوان یک ناهنجاری شناسایی می‌شود. مرجع [۷] به جای دسته‌بندی تمامی پروفیل‌های بار مشترکین به دسته‌بندی پروفیل بار، هر مشترک پرداخت است، روش پیشنهادی این مقاله به این صورت است که برای هر مشترک پروفیل مصرف عادی و غیرعادی را به دست می‌آورد و سپس یک سیستم دسته‌بندی را بر اساس مجموعه‌ای از منحنی بار عادی و غیرعادی مشترکین آموزش می‌دهد. مطالعه موردی صورت‌گرفته در مقاله نشان می‌دهد روش ارائه‌شده دقت بیشتری نسبت به SVM دارد. مرجع [۸] براساس تبدیل فوریه[۶] اطلاعات مصرف مشترکین در یک بازه یک‌ساله را به حوزه فرکانس می‌برد و با الگوبرداری از ویژگی‌های یک بازه زمانی مرجع (به‌عنوان‌مثال ماه اول سال) که در ابتدا فرایند تحلیل توسط کاربر تعیین‌شده است، به بررسی بازه زمانی مورد آزمایش (به‌عنوان‌مثال ماه ششم سال) پرداخته است و تشخیص می‌دهد که الگو مصرف عادی یا غیرعادی است. روش پیشنهادی در این مقاله توانایی اجرا به‌صورت موازی برای چندین هزار مشترک را دارد. نویسندگان در مرجع [۹] با بیان اینکه مسئله برق‌دزدی یک مسئله غیرمتعادل از جهت تعداد مشترکین است، دو روش جهت غلبه بر مشکل غیر متعادل بودن نسبت داده‌ها بیان کرده و معایب هریک را بررسی کرده است. سپس اطلاعات مشترکین غیرعادی را به کمک الگوریتم نزدیک‌ترین همسایه[۷] (KNN) دسته‌بندی می‌کند و از طریق هر دو مشترک که کمترین فاصله را نسبت به یکدیگر دارند، یک مشترک ساختگی بر اساس اختلاف فاصله آن‌ها ضربدر یک عدد تصادفی بین صفر و یک تولید می‌کند. بعد از تولید مجموعه داده ساختگی با توجه به ماهیت داده‌های انرژی که سری زمانی هستند، از ترکیب الگوریتم CNN[۸]، LSTM[۹] جهت ایجاد یک سیستم تشخیص برق‌دزدی استفاده می‌کند. مرجع [۱۰] به‌منظور شناسایی برق‌دزدی در سطح مشترکین صنعتی و تجاری اطلاعات مصرف مشترکین در یک بازه ۱۰ساله به همراه اطلاعات جانبی که توصیف‌کننده ویژگی‌های جغرافیایی و فنی هر مشترک هستند را جمع‌آوری کرده است. سپس به کمک اطلاعات مشترکینی که حداقل یک‌بار بازرسی حضوری شده‌اند الگوریتم‌های متعددی ازجمله KNN، XG Boost و رگرسیون لجستیک[۱۰] آموزش داده‌شده‌اند و بهترین آن‌ها بر اساس دقت بالا در پیش‌بینی داده‌های آزمون و زمان کمتر در مرحله آموزش انتخاب‌شده و بر اساس این الگوریتم به تشخیص برق‌دزدی در مشترکین جدید پرداخته‌شده است. درنهایت یک لیست رده‌بندی‌شده از احتمال وجود برق‌دزدی ارائه‌شده است. نویسندگان در مرجع [۱۱] با بیان اینکه مصرف نادرست عموماً به سه روش دستکاری‌شده، تخریب و هک کنتور باهدف ارسال مصرف کمتر از مقدار واقعی و یا عدم ارسال اطلاعات مصرف انجام می‌شود.به‌منظور شبیه‌سازی الگوهای مخرب در شبکه به ایجاد یک مجموعه داده از نمونه‌های مخرب بر اساس الگوی مصرف مشترکین عادی پرداخته‌اند و به ازای هر مشترک عادی، ۳ نمونه الگو مخرب ایجاد کرده‌اند. درنهایت یک سیستم SVM براساس این اطلاعات جهت دسته‌بندی و شناسایی برق‌دزدی آموزش داده‌شده است. معیارهای دقت[۱۱]، نرخ شناسایی غلط داده صحیح[۱۲] جهت سنجش الگوریتم استفاده‌شده است. در [۱۲] نویسندگان به کمک توابع ریاضی مرجع [۴] و انتخاب الگوی مصرف تعدادی از مشترکین به ایجاد یک مجموعه داده مصنوعی پرداخته‌اند و با تکیه‌بر این نکته که رفتار مصرفی یک مشترک دارای سیکل هفتگی است و صرفاً تفاوت‌های روزانه دارد، به آموزش یک شبکه عمیق به ازای هرروز جهت استخراج ویژگی‌های توصیف‌کننده رفتار مصرفی مشترکین پرداخته‌اند. سپس با تکیه‌بر ویژگی‌های استخراج‌شده به آموزش یک شبکه عمیق جهت شناسایی مشترکین متخلف پرداخته‌اند و به‌وسیله معیارهای FPR، DR الگوریتم را موردسنجش قرار داده‌اند. نوآوری معرفی‌شده در این مقاله استفاده از دو شبکه عمیق به‌منظور استخراج ویژگی‌ها و پیش‌بینی نتایج و همچنین استفاده از آموزش نیمه نظارت یافته است که امکان استفاده از مجموعه داده‌های بدون برچسب در فرایند آموزش شبکه را فراهم می‌کند. مرجع [۱۳] ابتدا الگوریتم CNN را باتوجه به توانایی‌اش در تحلیل تصاویر جهت شناسایی تفاوت‌ها در ساعات مختلف و روزهای مختلف مورداستفاده قرار داده است. سپس الگوریتم جنگل‌های تصادفی[۱۳] به کمک ویژگی‌های استخراج‌شده جهت تشخیص مشترکین متخلف استفاده‌شده است. پارامترهای الگوریتم جنگل‌های تصادفی توسط الگوریتم جست‌وجو شبکه[۱۴] تنظیم‌شده‌اند. معیارهای ماتریس اغتشاش[۱۵]و ROC[۱۶] جهت ارزیابی استفاده‌شده است. مرجع [۱۸] به ارائه یک روش مبتنی بر XG Boost پرداخته است؛ به این منظور به ازای هر مشترک ۶ سناریوی برق دزدی ایجاد کرده است و به‌وسیله سناریوهای ایجاد شده و الگوهای مصرف عادی مشترکین یک مدل آموزش داده است. براساس نتایج روش پیشنهادی در این مقاله نسبت به الگوریتم درخت تصمیم و SVM عملکرد بهتری دارد و در مقابل عدم تعادل داده‌ها مقاوم است. مرجع [۲۰] با ارائه الگوریتمی دومرحله‌ای ابتدا مصرف برق مشترکین را در هر دوره مصرف به مفهومی تحت عنوان رأی تبدیل کرده است. در مرحله اول، الگوی مصرف مشترکین را با توجه به بیشینه و کمینه مصرف برق در هر دوره به سه دسته کلی الگوی

نرمال، غیر نرمال و الگوی پرمصرف تقسیم‌شده و مشترکین غیرعادی را بر اساس میزان رأی آن‌ها جهت بازرسی اولویت‌بندی می‌کند. در مرحله دوم مصرف برق هر مشترک با مشابه آن در سال قبل مقایسه می‌شود و بر اساس درصد کاهش، آن را به چهار دسته تقسیم می‌کند. روش ارائه‌شده به‌وسیله داده‌های مصرف ۳۰۰ مشترک در سه ناحیه مختلف ازنظر سطح اقتصادی و معیشتی موردبررسی و آزمایش قرارگرفته است.

۲- بیان مسئله

به‌طورکلی سارقان انرژی سعی می‌کنند در عین مصرف زیاد انرژی، اطلاعات ثبت‌شده توسط کنتورهای هوشمند را به کمک روش‌های سایبری، فیزیکی و داده‌ای که در جدول ۱ ذکرشده‌اند دستکاری‌شده کنند[۱۴].

از طرفی شناسایی این موارد لزوماً از طریق بازرسی مشترکین ممکن است و شرکت‌های برق همواره به دنبال محدودکردن محدوده بازرسی خود به مواردی با احتمال برق‌دزدی بیشتر هستند. یکی از راهکارهای کاهش محدوده بازرسی، استفاده از روش‌های مبتنی بر هوش مصنوعی است ولی چالش مهمی که در حوزه هوش مصنوعی وجود دارد، مسئله عدم تعادل در نسبت داده‌های مشرکین عادی و مشکوک است که منجر به عملکرد ضعیف الگوریتم‌های آموزش داده‌شده می‌شود.

جدول (۱): روش‌های مختلف برق‌دزدی

نوع	روش‌ها
سایبری	• دستکاری‌شده سیستم‌عامل/حافظه کنتور • تغییر/ مداخله در ارتباطات کنتور • اشغال کردن ram/cpu • سرقت اطلاعات ورود به کنتور • از بین بردن یا مختل کردن شبکه ارتباطی کنتورها و مرکز داده
فیزیکی	• تخریب کنتور • معکوس کردن اتصالات کنتور • قطع کردن کنتور • دور زدن کنتور به‌وسیله سیم
داده	• عدم ارسال اطلاعات مصرف یا ارسال مصرف صفر • تغییر مقادیر انرژی اندازه‌گیری شده به‌صورت نسبتی از کل انرژی مصرفی • گزارش مصرف منفی (عمل کردن به‌صورت تولیدکننده) • حذف انرژی مصرفی تجهیزات پرمصرف از اندازه‌گیری • استفاده غیرمجاز از تعرفه

پیش‌بینی و تخمین رفتار مشترکین برق دزد به‌عنوان یکی از روش‌های قدیمی و قابل‌اتکا در بحث تشخیص برق‌دزدی، همواره سعی در پیش‌بینی رفتارهای برق‌دزدی بر اساس مطالعه رفتارهای اجتماعی مشترکین و تحلیل منحنی مصرف آن‌ها داشته است. به‌طوری‌که درگذشته عموماً متخصصین با توجه به شناخت کلی از رفتارهای اجتماعی مشترکین مجموعه‌ای از معیارهای ساده جهت تفکیک مشترکین برق دزد و عادی را مطرح می‌کردند. اما امروزه با تغییر رفتار مصرفی مشترکین و تنوع زیاد در نوع و زمان مصرف امکان استفاده از معیارهای ساده در تفکیک مشترکین برق دزد و عادی وجود ندارد و لازم است که معیارهای پیچیده‌تری جهت تفکیک مورداستفاده قرار گیرد. یکی از روش‌های شناسایی این معیارها (با فرض اینکه بتوان رفتار یک مشترک برق دزد را به‌صورت تابع ریاضی از رفتار مشترکین عادی بیان کرد)، تخمین و استخراج این توابع ریاضی است که در آن صورت می‌توان انتظار داشت که با تکیه‌بر این توابع ریاضی، تعداد نمونه‌های موردنیاز جهت آموزش الگوریتم افزایش یابد و بر مسئله عدم تعادل چیره شد. امروزه عمده مقالات مطرح‌شده در سطح جهانی تمرکز خود را بر پیش‌بینی توابع ریاضی جهت توصیف رفتار مشترکین برق دزد قرار داده‌اند، اما توابع ارائه‌شده توسط مقالات بعضاً بسیار ساده و صرفاً با تکیه‌بر تجربه بوده است و همین امر سبب می‌شود که توانایی پیش‌بینی تعداد محدودی از موارد برق‌دزدی را داشته باشند. مدل ارائه‌شده در این تحقیق باهدف پیش‌بینی و تخمین رفتار مشترکین مشکوک که به‌وسیله روش‌های سایبری و مبتنی بر داده مقدار مصرف خود را دستکاری‌شده کرده‌اند، طراحی‌شده است.

۳- مدل‌سازی مسئله

به‌طورکلی اگر اطلاعات مصرف واقعی یک مشترک و اطلاعات ارسال‌شده توسط کنتور به شرکت برق در یک بازه زمانی مشخص را در نظر بگیریم. میزان مصرفی واقعی با اختلاف ناچیزی که عموماً این اختلاف ناشی از خطای اندازه‌گیری کنتور است برابر با مقدار ارسال‌شده است اگر و فقط اگر مشترک در مصرف خود صالح و کنتور دستکاری‌شده نشده باشد و برعکس اگر مقدار ارسالی توسط یک مشترک برابر با مقدار مصرف واقعی‌اش نباشد، آنگاه این مشترک می‌تواند کاندید استفاده غیرمجاز باشد و در لیست بازرسی شرکت برق قرار بگیرد.

مسئله حائز اهمیت این است که عموماً اطلاعات مصرف واقعی مشترکین مشکوک که استفاده غیرمجاز آن‌ها توسط شرکت برق تأییدشده است وجود ندارد و تنها اطلاعات دستکاری‌شده آن‌ها موجود است. درنتیجه امکان سنجش و مقایسه این اطلاعات با مقادیر ارسالی وجود ندارد. اما با تکیه به آخرین بازرسی مشترکین مشکوک که در آن بازه اطلاعات ارسالی به‌صورت صحیح بوده است، می‌توان با بهره‌گیری از یک ماژول پیش‌بینی بار تحت عنوان STLF[۱۷] میزان مصرف واقعی مشترک را در بازه مورد مطالعه بازسازی کرد و تخمین زد. این عمل درواقع این اصل را بیان می‌کند که اگر رفتار مشترک همانند گذشته بود، الگوی مصرف به چه صورت می‌شد. با توجه به اینکه مبحث

پیش‌بینی بار یک فضای مطالعاتی متفاوت است، در این تحقیق از پرداختن به آن اجتناب شده و با فرض اینکه یک ماژول STLF با دقت قابل‌قبول در دسترس خواهد بود مقادیر مصرف مشترکین مشکوک در بازه زمانی مورد مطالعه تخمین زده‌شده‌اند. ازاینجا به بعد منظور از مقدار واقعی مصرف، مقدار تخمین زده‌شده توسط ماژول STLF است. درصورتی‌که رابطه مقدار مصرف واقعی و ارسال‌شده را به‌صورت زیر در نظر بگیریم:

$$\vec{Y}_i = h(\vec{X}_i + \varepsilon) \quad (۱)$$

در این فرمول $\vec{X}_i$ بردار مصرف واقعی مشترک آم i، $\vec{Y}_i$ بردار مصرف ارسالی مشترک آمi و ε خطای ناشی از اندازه‌گیری کنتور یا ارسال اطلاعات است که بر اساس کلاس دقت اندازه‌گیری کنتور تعیین می‌شود و عموماً کمتر از ۲٪ خواهد بود. درنتیجه تابع $h(\vec{X}_i + \epsilon)$ بیانگر رفتار مشترک است. به‌طوری‌که اگر مشترک در مصرف خود صالح و کنتور دستکاری‌شده نشده باشد، بیانگر یک رابطه یک‌به‌یک است و مقدار ارسالی برابر مقدار واقعی مصرف بعلاوه مقدار خطای کنتور در نظر گرفته می‌شود و برعکس اگر مقدار ارسالی توسط یک مشترک دستکاری‌شده باشد، یک رابطه غیرخطی است و بیانگر رفتار ریاضی مشترک مشکوک است. بر اساس مطالعه سناریوهای متعدد برق‌دزدی و با تکیه‌بر علم و تجربه مهندسی می‌توان انتظار داشت که حالت‌های ساده آن پیش‌بینی و تخمین زده شود. به‌عنوان‌مثال ساده‌ترین حالت این تابع با صرف‌نظر کردن از خطای کنتور و با توجه به اینکه هدف مشترکین برق دزد گزارش مقدار مصرف کمتر از مقدار واقعی است، به‌صورت رابطه ۲ قابل‌بیان است:

$$h(x) = \alpha x, 0 < \alpha < 1 \quad (۲)$$

که در این رابط ضریب α معادل میزان حریص بودن یک مشترک در گزارش مقدار مصرف است به‌طوری‌که اگر α برابر یک باشد یعنی مشترک صالح و اگر برابر صفر باشد به معنای عدم ارسال مقادیر مصرف است. به‌عنوان‌مثال α برابر ۰/۸ به این معناست که شرکت برق فقط ۸۰٪ مصرف واقعی را دریافت می‌کند و ۲۰٪ مصرف توسط مشترک به روش‌های مختلف ارسال نمی‌شود. دیگر حالت‌های ساده *h(x)* در [۴] قابل‌مشاهده است.

نکته حائز اهمیت این است که تمامی حالات ممکن تابع *h(x)* به‌خصوص حالت‌های پیچیده و غیرخطی آن قابل پیش‌بینی و تخمین با تکیه‌بر تجربه نیست؛ اما با فرض اینکه اطلاعات تعدادی از مشترکین مشکوک در دسترس است آنگاه رابطه *y=h(x)* را می‌توان با کنار هم قرار دادن بردارهای $\vec{X}_i$ و $\vec{Y}_i$ مشترکین به‌صورت رابطه ماتریسی ۳ بیان کرد:

$$\begin{bmatrix} x_{1n} & \dots & x_{nn} \\ \dots & \dots & \dots \\ x_{1m} & \dots & x_{nm} \end{bmatrix} \cdot \begin{bmatrix} \theta_{11} & \dots & \theta_{1n} \\ \dots & \dots & \dots \\ \theta_{n1} & \dots & \theta_{nn} \end{bmatrix} = \begin{bmatrix} y_{1n} & \dots & y_{nn} \\ \dots & \dots & \dots \\ y_{1m} & \dots & y_{nm} \end{bmatrix} \quad (۳)$$

m: تعداد مشترکین مشکوک

n: بازه زمانی ثبت اطلاعات برحسب ساعت

xmn: میزان مصرف واقعی مشترک mآم در ساعت nآم

θnn: ضریب تأثیر ساعت nآم در مصرف ثبت‌شده nآم

ymn: میزان مصرف ثبت‌شده مشترک mآم در زمان nآم

در این رابطه ضرایب θ مجهول و مقادیر x, y معلوم هستند. تعیین مجهولات در این رابطه با بزرگ‌تر شدن n پیچیده می‌شود و با توجه به غیرخطی بودن، حل آن بسیار سخت و بعضاً ناممکن است. یکی از راه‌های تخمین ماتریس ضرایب که بیان‌کننده رابطه بین x, y است، مدل‌سازی رابطه فوق به‌صورت یک مسئله رگرسیون چند متغیره و استفاده از یک شبکه عصبی عمیق جهت تخمین این ضرایب است؛ به‌طوری‌که در این شبکه عصبی مقادیر مصرف واقعی به‌عنوان ورودی و مقادیر ارسال‌شده توسط کنتور به‌عنوان خروجی در نظر گرفته می‌شوند و شبکه عصبی آموزش داده‌شده بر اساس نمونه‌های موجود بیانگر ضرایب مجهول θ است. مرحله مهمی که پیش از آموزش شبکه عصبی لازم است به آن پرداخته شود، تفکیک و بررسی الگوهای رفتاری متفاوت مشترکین عادی است. الگوی رفتاری یک مشترک در یک بازه زمانی لزوماً ثابت نیست و همواره دچار تغییر می‌شود؛ این تغییرات لزوماً بیانگر رفتار غیرعادی و نابهنجار نیستند بلکه ممکن است در اثر تغییر در نحو استفاده از تجهیزات، تغییر در میزان زمان حضور در منزل، تغییر عادات مصرفی و ... ایجادشده باشند. عدم توجه به این تغییرات و تأثیر آن‌ها به‌طور یقین سبب خطای بسیار زیادی در تشخیص برق‌دزدی می‌شود و ممکن است حتی به‌اشتباه موارد صحیح و عادی به‌عنوان برق دزد شناسایی شوند. لذا لازم است در ابتدا الگوهای رفتاری متفاوت این مشترکین شناسایی و دسته‌بندی شوند.

یکی از روش‌های استخراج الگوهای مختلف مصرف با تکیه‌بر ویژگی‌های مختلف، استفاده از الگوریتم‌های خوشه‌بندی نظیر K-means است [۱۵]. مسئله مهمی که در استفاده از این الگوریتم وجود دارد، تعیین تعداد درست خوشه‌ها است. اگر تعداد آن‌ها به‌اندازه کافی بزرگ نباشد یا بیش‌ازحد بزرگ باشد، امکان توصیف کامل و جامع رفتارهای مختلف جامعه مطالعاتی وجود ندارد و همواره لازم است که تعداد صحیح خوشه‌ها جهت توصیف مشترکین تعیین شود. یکی از معیارهای تعیین تعداد صحیح خوشه‌ها معیار Silhouette است. معیار Silhouette یکی از روش‌های ارزیابی خوشه‌بندی است که بر اساس تحلیل میزان پیوستگی و تفکیک‌پذیری خوشه‌ها میزان تعلق هر نمونه به خوشه‌ای که در آن قرار دارد را در مقایسه با خوشه‌های مجاور اندازه می‌گیرد. معیار Silhouette به کمک رابطه (۷) محاسبه می‌شود:

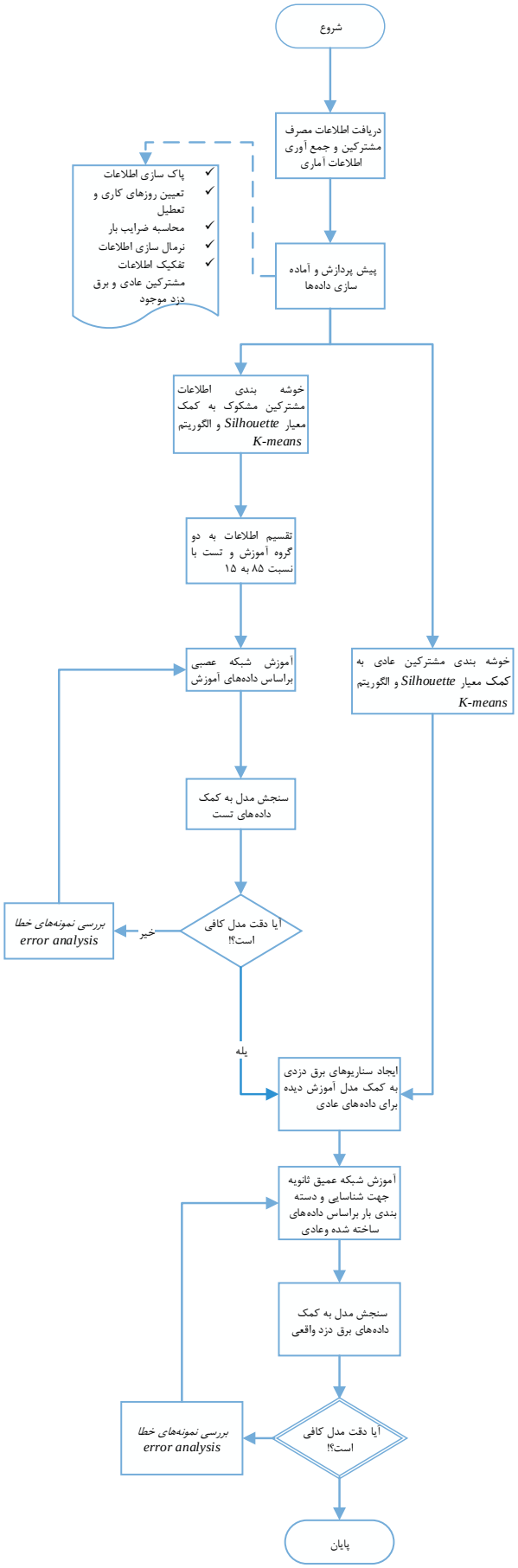
$$s(i) = \frac{b(i) - a(i)}{\max(b(i), a(i))} \quad (۴)$$

در این رابطه *a(i)* بیانگر میانگین فاصله یک نقطه از خوشه با نقاط دیگر آن خوشه است. *b(i)* بیانگر حداقل میانگین فاصله یک نقطه با خوشه‌های دیگر است. با توجه به رابطه بالا مقدار این شاخص بین

[۱،۱-] تغییر می‌کند و مقدار نزدیک به یک بیانگر انطباق خوب بین نقطه و خوشه است.

بر اساس توضیحات ارائه‌شده در قدم اول مرحله اول به آماده‌سازی اطلاعات مشترکین جهت خوشه‌بندی مشترکین پرداخته می‌شود که پس از آماده‌سازی درمجموع ۷۰ ویژگی شامل اطلاعات مصرف ۲۴ ساعته مشترک و اطلاعات آماری و معیشتی نظیر تعداد افراد ساکن در منزل، درآمد خانواده، سیستم گرمایشی و سرمایش، تعداد اتاق و سن افراد می‌شوند. در قدم دوم به‌وسیله معیار Silhouette تعداد مناسب خوشه‌های توصیف‌کننده مشترکین عادی تعیین می‌شوند و الگوریتم K-means جهت تعیین خوشه‌ها استفاده می‌شود. در قدم سوم و نهایی مرحله اول اطلاعات مشترکین مشکوک را به دودسته آموزش و آزمون با نسبت ۸۵ به ۱۵ درصد تقسیم می‌کنیم و به‌وسیله داده‌های آموزشی چند شبکه DNN را با پارامترهای متفاوت ازجمله تعداد لایه‌ها، تعداد نورون‌ها و توابع فعال‌ساز مختلف با تابع هزینه میانگین مربعات خطا[۱۸]، اندازه دسته[۱۹] برابر ۸ و بهینه‌ساز[۲۰] آدام جهت پیش‌بینی مصرف مشترکین مشکوک برای ۲۴ ساعت آینده آموزش داده می‌شوند. سپس به کمک شبکه عمیق آموزش‌دیده در مرحله قبل تعداد نمونه مشکوک بر اساس اطلاعات مشترکین عادی ایجاد می‌کنیم. این نمونه‌ها درواقع شبیه‌سازی سناریوهای محتمل برق‌دزدی و دستکاری‌شده کنتور به ازای مشترکین عادی است و این واقعیت را بیان می‌کند که اگر روزی یک مشترک عادی با تغییر رفتار مصرفی خود به سمت استفاده غیرمجاز یا دستکاری‌شده کنتور برود تأثیر این رفتار در الگوی مصرف به چه صورت است. درنهایت در مرحله دوم بر اساس اطلاعات مشترکین عادی و مشکوک به آموزش یک شبکه عمیق DNN که در لایه ورودی از ۷۰ نورون، در لایه‌های میانی از ترکیب ۱۰۰ و ۵۰ نورون و در لایه پایانی از L نورون تشکیل‌شده است جهت تفکیک مشترکین پرداخته می‌شود. L برابر تعداد خوشه‌های استخراج‌شده در مرحله خوشه‌بندی مشترکین عادی به‌علاوه برچسب مشترکین مشکوک است. تابع فعال‌ساز نورون‌های میانی Relu و تابع فعال‌ساز نورون‌های لایه آخر با توجه به این‌که وظیفه لایه آخر تفکیک موارد عادی و مشکوک است از نوع SoftMax است و به‌تبع آن تابع هزینه categorical_crossentropy جهت آموزش مدل استفاده‌شده است.

پس از آموزش مدل و اطمینان از صحت عملکرد مدل، اطلاعات مصرف روزانه مشترکین را دسته‌بندی می‌کنیم و مشترکین که تعداد دفعات تشخیص آن‌ها توسط مدل به‌عنوان موارد مشکوک بیشتر از یک حد مشخصی باشد به لیست موارد مشکوک شرکت برق جهت بازرسی اضافه می‌شوند. حد مشخص تکرار بر اساس نظر متخصصین شرکت‌های برق برای هریک از مناطق موردمطالعه می‌تواند متفاوت تعیین گردد و این اصل را بیان می‌کند که مناطق مختلف رفتاری مشابه ندارند. درنهایت بر اساس توضیحات داده‌شده روند نمای کلی کار به‌صورت شکل ۱ است.



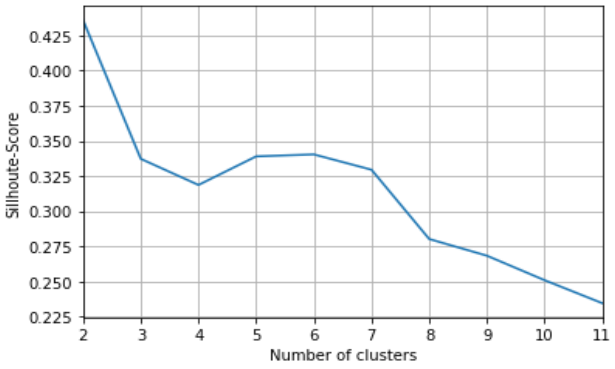
شکل (۱): روند نمای کلی روش پیشنهادی

۴- شبیه‌سازی و نتایج

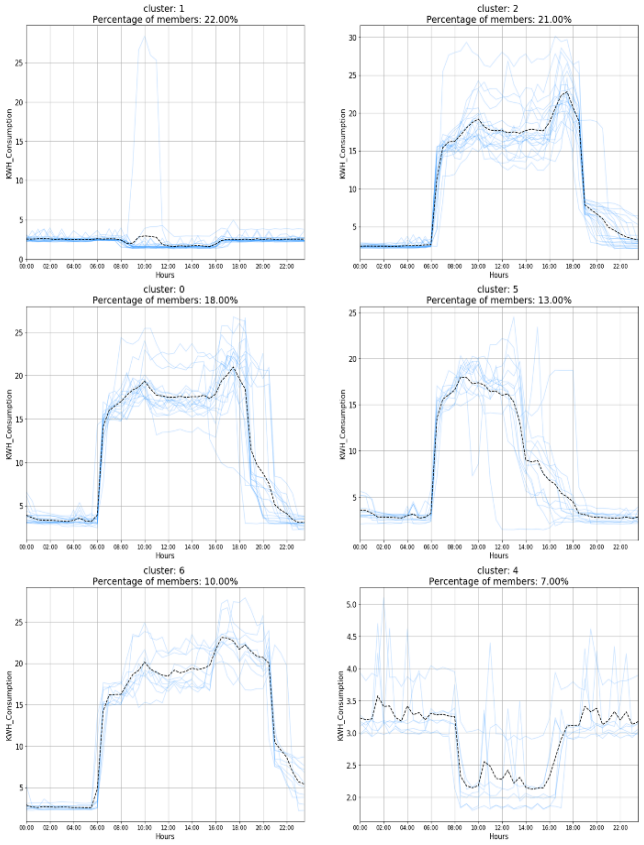
جهت سنجش مدل ارائه‌شده در این تحقیق، از اطلاعات کنتورهای هوشمند مشترکین کشور ایرلند استفاده‌شده است[۱۶]. این مجموعه داده شامل اطلاعات مصرف ۶۴۴۵ مشترک به ازای بازه‌های ۳۰ دقیقه در سال‌های ۲۰۰۹-۲۰۱۰ است و به ازای هرروز ۴۸ مقدار توسط کنتورها برای هر مشترک ثبت‌شده است. در این مجموعه داده علاوه بر اطلاعات مصرف دو پرسشنامه جهت جمع‌آوری اطلاعات آماری و معیشتی مشترکین تهیه و توسط مشترکین تکمیل‌شده است.

در این تحقیق که تلفیقی از داده‌کاوی و هوش مصنوعی است، از کتابخانه‌های pandas,keras,sklearn و زبان برنامه‌نویسی پایتون جهت ارائه نتایج شبیه‌سازی استفاده‌شده است.

در قدم اول، اطلاعات جمع‌آوری‌شده مورد پیش‌پردازش و آماده‌سازی قرارگرفته‌اند. اطلاعات مصرف مشترکین در قالب ۶ فایل با پسوند .txt ارائه‌شده‌اند که در هر خط این فایل‌ها شماره اشتراک مشترک (عدد صحیحی در بازه [۱۰۰۰،۷۴۴۴] تاریخ ثبت مصرف به‌صورت کد شده و مقدار مصرف برحسب kw/h ثبت‌شده است که درنهایت اطلاعات خام اولیه پس از آماده‌سازی به‌صورت جدولی مطابق شکل ۲ درآورده شده‌اند. در ادامه جهت استخراج انواع الگوهای مصرف مشترکین عادی در فصل تابستان بر اساس معیار Silhouette مناسب‌ترین تعداد خوشه ۵ و ۶ است. شکل ۳ بیانگر این موضوع است؛ لذا الگوریتم K-means به ازای تعداد ۶ خوشه با هدف درک الگوهای رفتاری مشترکین مورداستفاده قرار می‌گیرد. با توجه به اینکه الگوی رفتاری یک مشترک همواره دارای نوساناتی است که لزوماً بیانگر رفتار غیرعادی و نابهنجار نیستند؛ ممکن است این نوسانات در اثر تغییر در نحو استفاده از تجهیزات، میزان زمان حضور در منزل، عادات مصرفی و ... ایجادشده باشند. عدم توجه به این نوسانات و تأثیر آن‌ها بر الگو مصرف به‌طور یقین سبب خطای بسیار زیادی در تشخیص برق‌دزدی می‌شود و ممکن است حتی به‌اشتباه مشترکین صالح به‌عنوان برق دزد شناسایی شوند. لذا در این مرحله از تحقیق لازم است که انواع الگوهای مصرف مشترکین شناسایی و تفکیک بشوند. شکل ۴ بیانگر الگوهای مصرف تفکیک‌شده توسط الگوریتم K-means است . این داده‌های تفکیک‌شده در مرحله دوم جهت تشخیص برق دزدی استفاده می‌شوند.



شکل (۳): معیار silhouette به ازای تعداد خوشه‌های متفاوت



شکل (۴): انواع الگوی بار دسته‌بندی‌شده مشترکین عادی به‌وسیله الگوریتم k-means

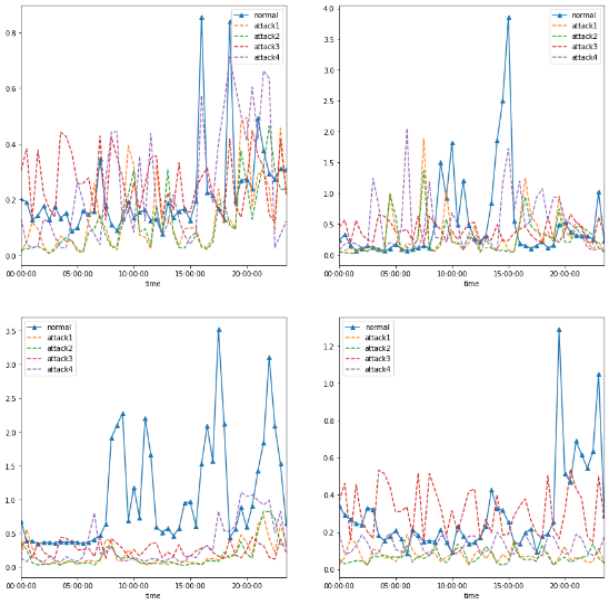
روز کاری	روز تعطیل	انحراف معیار	میانگین مصرف	حداکثر مصرف	حداقل مصرف	۲۳:۳۰:۰۰	۲۳:۰۰:۰۰	۲۲:۳۰:۰۰	...	۰۰:۳۰:۰۰	۰۰:۰۰:۰۰	تاریخ	شماره مشترک
۱	۰	۰.۷۲۰۴۸۴۰۸۷	۰.۲۹۳۶	۳.۴۵۳	۰.۰۲۴	۰.۱۵۲	۰.۲۱۱	۱.۱۳۳	...	۰.۰۵	۰.۱۱۷	۱/۱/۲۰۰۹	۱۰۰۰
۱	۰	۰.۵۴۶۶۹۱۱۵۲	۰.۲۲۳۷۸	۲.۷۶۲	۰.۰۲۴	۰.۱۳۵	۰.۱۸۴	۰.۲۱	...	۰.۰۲۴	۰.۰۶۳	۱/۲/۲۰۰۹	۱۰۰۰
۰	۱	۰.۱۶۱۷۶۱۵۷۳	۰.۱۵۹۵۶	۰.۷۳	۰.۰۲۳	۰.۳۹۳	۰.۱۱۴	۰.۰۴۴	...	۰.۰۵۷	۰.۱۶۱	۱/۳/۲۰۰۹	۱۰۰۰
۰	۱	۰.۶۵۶۴۴۷۴۷۲	۰.۲۸۳۵۸	۲.۸۱۷	۰.۰۲۴	۰.۱۵	۰.۱۰۱	۰.۱۰۳	...	۰.۰۹۸	۰.۲۲۵	۱/۴/۲۰۰۹	۱۰۰۰
۱	۰	۰.۶۴۰۹۷۳۸۰۳	۰.۲۸۱۹۴	۲.۷۶۷	۰.۰۲۴	۰.۰۸۵	۰.۰۲۴	۰.۰۲۴	...	۰.۰۶	۰.۱۶۷	۱/۵/۲۰۰۹	۱۰۰۰

```
1392 19506 0.145
1392 19507 0.145
1392 19501 0.157
1392 19502 0.144
1392 19724 0.128
1392 19725 0.142
1392 19726 0.145
```

شکل (۲): نمونه اطلاعات یک مشترک قبل و بعد از آماده‌سازی

یکی از فرضیات روش پیشنهادی وجود تعداد حداقلی اطلاعات مشترکین مشکوک است. اما با توجه به اینکه در این مجموعه داده هیچ‌گونه اطلاعات در مورد مشترکین مشکوک یا برق دزد وجود ندارد لازم است که یک مجموعه داده به‌عنوان مرجع اطلاعات مشکوک ایجاد شود. در این راستا با استناد به مرجع [۴] تعدادی نمونه برق‌دزدی ایجادشده است. شکل ۵ نمونه الگوی مصرف واقعی و دستکاری‌شده چند مشترک برق دزد را نشان می‌دهد. در این شکل منحنی‌های ممتد (آبی پررنگ) الگوی مصرف واقعی چند هستند که به‌وسیله ماژول STLF، در بخش قبل توضیح داده شده است، پیش‌بینی شده است. درحالی‌که هر یک منحنی‌های نقطه‌چین یک نمونه الگوی مصرف دستکاری و ارسال شده این مشترکین است. بر اساس شکل کاملاً مشهود است که اولاً الگوهای ایجاد شده کاملاً تصادفی و متنوع هستند و ثانیاً سطح زیر الگوی مصرف ارسال‌شده، که بیانگر انرژی مصرفی است، عموماً بسیار کمتر از سطح زیر الگوی مصرف واقعی هستند . این اختلاف انرژی مصرفی بیانگر نوعی استفاده غیرمجاز است که با توجه به معادلات ماتریسی ذکر شده در بخش قبل اگر اطلاعات تعداد کافی از مشترکین برق دزد را در اختیار باشیم، می‌توان به کمک یک شبکه عصبی عمیق رفتار این مشترکین مشکوک را مدل‌سازی کنیم. درنتیجه به کمک مدل ساخته‌شده می‌توان رفتار مشترکین مشکوک را بسط داد و سناریوهای محتمل برق‌دزدی به ازای مشترکین عادی پیش‌بینی‌کرد.

به‌طور خلاصه هدف از مدل‌سازی رفتار مشترکین برق دزد شناسایی و کشف رابطه ریاضی است که سبب تبدیل شدن الگوی مصرف واقعی به هر یک از الگوهای مصرف دستکاری‌شده می‌شود.



شکل (۵): نمونه الگو مصرف واقعی و دستکاری‌شده چند مشترک برق‌دزدی

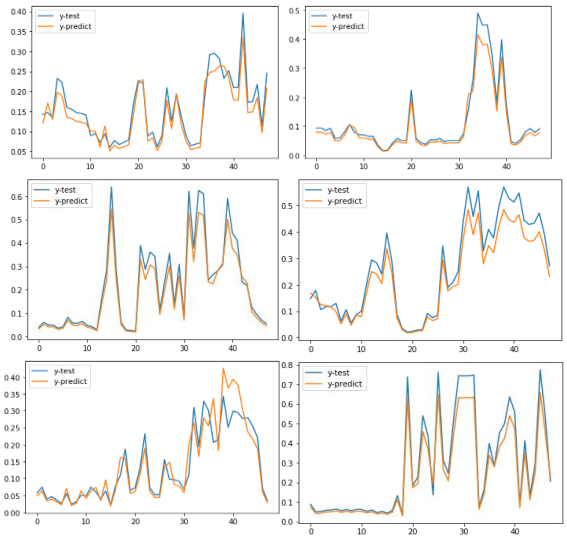
درنهایت با توجه به توضیحات ارائه شده، ۵۰۰۰ نمونه برق دزدی مطابق شکل ۵ جهت آموزش شبکه عصبی اولیه تولیدشده است. نتایج مربوط به پیاده‌سازی مدل‌های استفاده‌شده در روش پیشنهادی با پارامترهای مختلف برحسب معیار MSE در جدول ۲ ارائه‌شده است. همچنین نمودارهای مربوط به خروجی بعضی از مدل‌ها در شکل ۶ به نمایش گذاشته‌شده است. مطابق شکل ۶ منحنی‌های آبی رنگ الگوهای مصرف دستکاری‌شده توسط مشترکین برق دزد را نشان می‌دهند، که در شکل ۵ به صورت نقطه‌چین نمایش داده شده بود؛ منحنی‌های نارنجی رنگ الگوهای مصرفی پیش‌بینی شده توسط مدل‌های آموزش داده شده است. به‌طور خلاصه این مدل‌ها در ورودی منحنی مصرف عادی را دریافت می‌کنند و در خروجی الگوی مصرف دستکاری‌شده احتمالی را پیش‌بینی می‌کنند.

جدول (۲): نتایج مربوط به پیاده‌سازی شبکه DNN

تعداد لایه‌ها	تعداد نورون‌ها	توابع فعال‌ساز	تعداد تکرار	اندازه دسته‌ها	Train-MSE	Test-MSE
۳	(۷۰،۶۰،۴۸)	Relu	۳۰۰	۸	۴/۵۵	۵/۴۳
۴	(۷۰،۶۴،۵۰،۴۸)	Relu	۳۰۰	۱۶	۴/۲۴	۵/۳۱
۳	(۷۰،۶۰،۴۸)	Tanh	۳۰۰	۸	۵/۹۲	۶/۰۱

براساس نتایج جدول ۲ محدوده خطا مدل DNN در حدود ۴-۶ واحد است. براساس شکل ۶ اگرچه میزان خطای شبکه DNN پایین است اما مقایسه نتایج پیش‌بینی الگوی مصرف نشان می‌دهد که مدل توانایی تخمین نقاط اوج را به‌درستی ندارد که می‌توان این ضعف را به دو صورت توجیه کرد؛ اولاً با توجه به اینکه در داده‌های ورودی هیچ‌گونه توجهی به سیر زمانی داده‌ها نشده است درنتیجه مدل توانایی یادگرفتن روند را ندارد؛ دوماً با توجه به تعداد کم داده‌ها به ازای هر یک از گروه‌ها نمی‌توان انتظار عملکرد بسیار دقیق از مدل داشت. پس از اطمینان از دقت مدل‌های آموزش داده شده؛ به کمک این مدل‌ها تعدادی سناریوهای برق‌دزدی برای مشترکین عادی پیش‌بینی می‌شود. منظور از پیش‌بینی سناریوهای برق دزدی این است که از میان اطلاعات مشترکین عادی به صورت تصادفی تعدادی از منحنی مصرف آن‌ها را به مدل می‌هیم تا مدل متناسب با رابطه‌ای که از رفتار مشترکین برق دزد کشف کرده است الگوی مصرف دستکاری‌شده احتمالی آن مشترکین را پیش‌بینی کند.

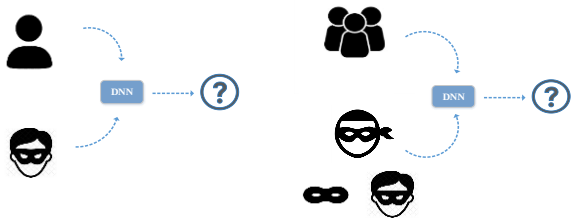
در نهایت به کمک این نمونه‌های پیش‌بینی شده و اطلاعات مشترکین عادی، که در مرحله اول تفکیک و خوشه بندی شده بودند، یک شبکه عصبی ثانویه جهت دسته‌بندی مشترکین عادی و مشکوک از یکدیگر ارائه می‌شود.



شکل (۶): مقایسه نمونه الگوهای خروجی مدل DNN و الگوهای واقعی

به‌منظور سنجش عملکرد روش پیشنهادی می‌توان سناریوهای متفاوتی برای مدل ثانویه مطابق شکل ۷ مطرح کرد:

- سناریوی اول: در این حالت مسئله تشخیص برق‌دزدی به‌صورت یک مسئله دسته‌بندی ۲ تایی۲۱ مطرح می‌شود. به‌طوری‌که در این حالت هدف صرفاً تشخیص برق دزد یا نبودن یک مشترک بدون توجه به الگوهای مختلف برق‌دزدی و عادی که در مراحل خوشه‌بندی استخراج‌شده‌اند است. در این سناریو درصورتی‌که مدل تشخیص دهد نمونه ورودی مشکوک است خروجی یک می‌شود و در غیر این صورت صفر است.
- سناریوی دوم: در این حالت مسئله تشخیص برق‌دزدی به‌صورت یک مسئله دسته‌بندی چندتایی۲۲ از دیدگاه تفکیک مشترکین برق دزد و عادی مطرح می‌شود. در این حالت هدف تشخیص برق دزد یا نبودن یک مشترک در کنار توجه به الگوهای مختلف مشترکین مشکوک و عادی است. به‌طوری‌که علاوه بر تشخیص نوع فعالیت مشترکین شماره گروهی که به آن تعلق دارند نیز تعیین می‌شود. درصورتی‌که مدل تشخیص دهد نمونه ورودی مشکوک است در خروجی شماره گروه مشترک را نمایش می‌دهد و در غیر این صورت شماره گروه مشترک را نمایش می‌دهد.



شکل (۷): سناریوهای مختلف تشخیص برق‌دزدی

نتایج حاصل از مقایسه سناریوی اول، دوم و تحقیقات پیشین در جدول ۳ موردبررسی قرارگرفته است. بر اساس نتایج در سناریوی اول مسئله تشخیص برق‌دزدی به‌صورت دسته‌بندی دوتایی در نظر گرفته‌شده است که درنتیجه آن مدل نمونه‌های بیشتری جهت یادگیری در اختیار داشته و به‌تبع آن بهترین نتیجه ممکن حاصل‌شده است. همچنین قابل‌مشاهده است که با افزایش تعداد گروه‌ها توانایی مدل در تشخیص کاهش می‌یابد که با توجه به کاهش تعداد نمونه‌ها به ازای هر یک از گروه‌ها جهت آموزش مدل کاملاً امری طبیعی است. همچنین قابل مشاهده است که عمده تحقیقات گذشته بر اساس شبکه‌های یادگیری عمیق انجام‌شده است و ثانیاً در عمده تحقیقات از پرداختن به تأثیر عوامل آماری و اجتماعی اجتناب شده است.

جدول (۳): مقایسه روش پیشنهادی و روش‌های موجود

اطلاعات استفاده‌شده	Accuracy(%)	FPR(%)	DR(%)	مرجع
پروفیل بار مشترکین	-	۵/۱۴	۹۵/۰۴	[۱۲]
	-	۵/۵۶	۹۷/۲۷	
پروفیل بار مشترکین و ترانس	-	۱۱	۹۴	[۴]
	-	۹	۹۶	[۱۷]
پروفیل بار مشترکین، تعداد اعضا خانه، تعداد لوازم‌خانگی	۹۲/۵۰	۵/۱۲	-	[۵]
پروفیل بار مشترکین	۹۳/۳۶	۵/۸۴	۹۲/۵۶	[۱۸]
پروفیل بار مشترکین	-	۳/۱۷	۹۷/۵۳	[۱۹]
پروفیل بار مشترکین، تعداد اعضا خانه، نوع سیستم گرمایش و سرمایش، میزان درآمد، سن افراد	۹۵/۱	۵/۱۵	۹۸	سناریو۱
	۹۰/۶	۱۴/۴	۹۲/۳	سناریو۲

۵- نتیجه‌گیری

در این پژوهش ، یک مدل دو مرحله‌ای مبتنی برشبکه‌های عمیق جهت تشخیص برق دزدی در بستر کنتورهای هوشمند ارائه شده است. در مرحله اول ، الگوهای مصرف مشترکین به کمک الگوریتم خوشه‌بندی استخراج شده؛ سپس با تکیه‌بر این فرض که رابطه بین مصرف واقعی و مصرف ثبت‌شده به‌صورت ریاضی قابل‌بیان است، یک شبکه عمیق اولیه جهت پیش‌بینی الگوهای برق دزدی براساس الگوهای مشترکین متقلب توسعه داده شده است. به کمک این شبکه سناریوهای محتمل برق دزدی به ازای مشترکین عادی جهت غلبه بر مسئله عدم تعادل در نسبت مشترکین مشکوک تولید می‌شود.

درنهایت با تکیه‌بر سناریوهای تولیدشده و اطلاعات مشترکین عادی چندین شبکه عمیق ثانویه جهت دسته‌بندی مشترکین ارائه شده است . بررسی مدل پیشنهادی بر روی مجموعه داده واقعی با بیش از ۶۰۰۰ مشترک و مقایسه آن با دیگر روش‌های موجود عملکرد بالای آن را از دیدگاه نرخ شناسایی و دقت نشان می‌دهد. در پژوهش‌های آتی سعی می‌شود از ترکیب شبکه‌های LSTM,CNN که عملکرد بهتری در حوزه رگرسیون دارند و همچنین داده‌های مربوط به داخل کشور استفاده شود.

مراجع

[1] P. Glauner, J. A. Meira, P. Valtchev, R. State, and F. Bettinger, "The Challenge of Non-Technical Loss Detection Using Artificial Intelligence: A Survey," arXiv preprint: 1606.00626, 2016.

[2] J. B. Leite, and J. R. S. Mantovani, "Detecting and locating Non-Technical Losses in Modern Distribution Networks," IEEE Transactions on Smart Grid, vol. 9, no. 2, pp. 1023-1032, 2016.

[3] S.-C. Huang, Y.-L. Lo, and C.-N. Lu, "Non-technical loss Detection Using State Estimation and Analysis of Variance," IEEE Transactions on Power Systems, vol. 28, no. 3, pp. 2959-2966, 2013.

[4] P. Jokar, N. Arianpoo, and V. C. Leung, "Electricity theft Detection in AMI Using Customers' Consumption Patterns," IEEE Transactions on Smart Grid, vol. 7, no. 1, pp. 216-226, 2015.

[5] A. Jindal, A. Dua, K. Kaur, M. Singh, N. Kumar, and S. Mishra, "Decision Tree and SVM-Based Data Analytics for Theft Detection in Smart Grid," IEEE Transactions on Industrial Informatics, vol. 12, no. 3, pp. 1005-1016, 2016.

[6] L. A. P. Júnior, C .C. O. Ramos, D. Rodrigues, D. R. Pereira, A. N. de Souza, K. A. P. da Costa, and J. P. Papa, "Unsupervised Non-Technical Losses Identification Through Optimum-Path Forest," Electric Power Systems Research, vol. 140, pp. 413-423, 2016.

[7] A. Nizar, Z. Dong, and Y. Wang, "Power Utility Nontechnical Loss Analysis with extreme Learning Machine Method," IEEE Transactions on Power Systems, vol. 23, no. 3, pp. 946-955, 2008.

[8] V. Botev, M. Almgren, V. Gulisano, O. Landsiedel, M. Papatriantafilou, and J. van Rooij, "Detecting non-Technical Energy Losses Through Structural Periodic Patterns in AMI Data." pp. 3121-3130, 2016.

[9] M. Hasan, R. N. Toma, A.-A. Nahid, M. Islam, and J.-M. Kim, "Electricity Theft Detection in Smart Grid Systems: a CNN-LSTM Based Approach ",Energies, vol. 12, no. 17, pp. 3310, 2019.

[10] M. M. Buzau, J. Tejedor-Aguilera, P. Cruz-Romero, and A. Gómez-Expósito, "Detection of Non-Technical Losses Using Smart Meter Data and Supervised Learning," IEEE Transactions on Smart Grid, vol. 10, no. 3 ,pp. 2661-2670, 2018.

[11] N. E. I. k. Abdelaziz Amara Korba, "Smart Grid Energy Fraud Detection Using SVM," International Conference on Networking and Advanced Systems (ICNAS), 2019.

[12] T. Hu, Q. Guo, X. Shen, H. Sun, R. Wu, and H. Xi, "Utilizing Unlabeled data to Detect Electricity Fraud in AMI: A Semisupervised Deep Learning Approach," IEEE Transactions on Neural Networks and Learning Systems, vol. 30, no. 11, pp. 3287-3299, 2019.

[13] S. Li, Y. Han, X. Yao, S. Yingchen, J. Wang, and Q. Zhao, "Electricity Theft Detection in Power Grids with Deep Learning and Random Forests," Journal of Electrical and Computer Engineering, vol, 2019.

[14] S. McLaughlin, B. Holbert, A. Fawaz, R. Berthier, and S. Zonouz, "A Multi-Sensor Energy Theft Detection Framework for Advanced Metering Infrastructures," IEEE Journal on Selected Areas in Communications, vol. 31, no. 7, pp. 1319-1330, 2013.

[15] J. A. Hartigan, and M. A. Wong, "AK-Means Clustering Algorithm," Journal of the Royal Statistical Society: Series C (Applied Statistics), vol. 28, no. 1, pp. 100-108, 1979.

[16] "Irish Social Science Data Archive. [Online]. Available: http://www.ucd.ie/issda/data/commissionforenergy

[17] S. Salinas, M. Li, and P. Li, "Privacy-Preserving Energy Theft Detection in Smart Grids: A P2P computing approach," IEEE Journal on Selected Areas in Communications, vol. 31, no. 9, pp. 257-267, 2013.

[18] Ibrahem MI, Nabil M, Fouda MM, Mahmoud MM, Alasmary W, Alsolami F. "Efficient privacy-preserving electricity theft detection with dynamic billing and load monitoring for AMI networks," IEEE Internet of Things Journal, vol.8, no.2, pp.1243-58, 2020.

[19] Yan, Z. and Wen, H, "Electricity theft detection base on extreme gradient boosting in AMI," IEEE Transactions on Instrumentation and Measurement, vol. 70, pp.1-9, 2021.

[۲۰] وحید یوسفی، مصطفی محمودی و شهریار شیخ اویسی، طراحی یک قالب برای تشخیص تقلب در شبکه توزیع برق، سومین کنفرانس بین‌المللی مهندسی برق، مهندسی مکانیک، کامپیوتر و علوم مهندسی، صوفیه - بلغارستان، ۱۳۹۸.

[۲۱] مویدی راد حجت، فلقی حمید، فرشاد محسن. یک الگوریتم ابتکاری برای تجدید آرایش شبکه‌های توزیع به منظور کاهش تلفات اهمی مبتنی بر نظریه‌ی گراف. نشریه مهندسی برق و الکترونیک ایران. ۱۳۹۳; ۱۱ (۱) :۵۹-۷۲

زیرنویس‌ها

[1] Game theory
[2] Syntactic Dataset
[3] Support Vector Machine
[4] Decision Tree
[5] Optimum-Path Forest
[6] Fourier Transform
[7] K-Nearest Neighbor
[8] Convolutional Neural Network
[9] Long-Short Term Memory
[10] Logistic Regression
[11] Precision
[12] False Positive Rate
[13] Random Forest
[14] Grid Search Algorithm
[15] Confusion Matrix
[16] Receiver Operating Characteristic
[17] Short Term Load Forecasting
[18] Mean Score Error
[19] Batch Size
[20] Adam Optimizer
[21] Binary-Class Classification
[22] Multi-Class Classification