

مدلسازی نرخ خطای گذرا در شبکه‌های توزیع برق مبتنی بر داده‌کاوی در داده‌های میدانی

محمود رضا حقی فام^{۱*}، الهام اخوان رضایی^۲، علیرضا فریدونیان^۳

۱- دانشکده مهندسی برق و کامپیوتر دانشگاه تربیت مدرس،

۲- دانشکده مهندسی برق و کامپیوتر دانشگاه واترلو، کانادا،

۳- بخش مهندسی برق دانشگاه صنعت آب و برق (شهید عباسپور)

haghifam@modares.ac.ir

(دریافت مقاله: ۸۸)

چکیده- ارزیابی قابلیت اطمینان در شبکه‌های توزیع برق تصویر ارزشمندی از عملکرد سیستم در مقابل بهره‌برداران و صاحبان شرکت‌های توزیع قرار می‌دهد. به این ترتیب آنها قادر می‌شوند استراتژی‌های عملیاتی جهت مدیریت دارایی در شبکه خود را بر اساس آن استوار سازند. مدلسازی نرخ خطا از شیوه‌های سودمند جهت حرکت از ارزیابی قابلیت اطمینان به سمت اخذ استراتژی‌های مدیریت دارایی بهینه و توسعه فعالیت‌های نگهداری و تعمیرات قابلیت اطمینان محور می‌باشد. این مقاله روشی مبتنی بر داده‌کاوی جهت مدلسازی نرخ خطای گذرا در خطوط هوایی شبکه توزیع ارائه می‌دهد، که در آن داده‌های خاموشی شبکه توزیع انرژی الکتریکی شهر تهران بزرگ برای ایجاد مدلسازی مورد استفاده قرار می‌گیرد. سه روش مدلسازی شامل رگرسیون چندمتغیره خطی به عنوان یک روش پارامتریک، روش درخت تصمیم و شبکه‌های عصبی مصنوعی به عنوان روش‌های غیر پارامتریک به داده‌های مسئله برازش می‌شوند. سپس، نتایج و کارایی سه مدل نسبت به یکدیگر مورد بحث و بررسی قرار می‌گیرد. نتایج حاصل از مدلسازی‌های این مطالعه بر کارایی بهتر مدل شبکه عصبی نسبت به رگرسیون چندمتغیره خطی و درخت تصمیم در تخمین نرخ خطای گذرا، تاکید دارد.

کلید واژگان: ارزیابی قابلیت اطمینان، تحلیل آماری خطا، داده‌کاوی، مدلسازی نرخ خطا، شبکه‌های توزیع انرژی الکتریکی.

۱- مقدمه

ارزیابی و بهبود قابلیت اطمینان از نیازمندی‌های اصلی در بهره‌برداری و طراحی توسعه شبکه‌های قدرت است. این نیازمندی، خصوصاً در فضای جدید بازار و مقررات‌زدایی،

شرکت‌های برق را با چالش‌های اساسی مواجه می‌کند چراکه متضمن بقای آنها در عرصه رقابت خواهد بود. از این رو، شرکت‌های برق همواره درصدد توسعه روشها و ابزارهای ارزیابی قابلیت اطمینان، به عنوان زیرمجموعه

مدلسازی و تخمین به تعیین رفتار متغیرهایی چون شاخص‌های قابلیت اطمینان منجر می‌شود.

در مطالعات قابلیت اطمینان شبکه توزیع، اهمیت شاخص‌های نقطه بار در کنار شاخص‌های مشترک-محور و در نقش مکمل یکدیگر قابل تامل است. اگرچه عموم شرکت‌ها به شاخص‌های مشترک-محور نظیر SAIFI و SAIDI بیشتر توجه دارند، ماهیت این شاخص‌ها تنها تصویری از میانگین وضعیت قابلیت اطمینان شبکه به دست می‌دهد [۲]. شاخص نرخ خطا، به‌عنوان یکی از شاخص‌های نقطه بار، در شبکه توزیع از جمله شاخص‌های متداولی است که با مطالعه رفتار آن دانش وسیعی از ماهیت خطاهای شبکه به دست می‌آوریم.

نرخ خطا در تجهیزات شبکه توزیع مقدار ثابتی نیست و به فاکتورهای متنوع بستگی دارد، شامل: فاکتورهای ذاتی شبکه (نظیر سن تجهیزات، نقص در ساخت تجهیزات، اندازه هادی‌ها و ...)، فاکتورهای محیطی (نظیر درختان، رطوبت، باد، یخبندان، رعد و برق، حیوانات و ...)، و فاکتورهای زیرمجموعه اشتباه‌های انسانی (نظیر تصادف اتومبیل‌ها با شبکه توزیع، خطای کارکنان شبکه، خشونت‌های خیابانی و ...) [۳]. از آنجا که اعمال نرخ متوسط خطا در مدل‌های قابلیت اطمینان سیستم ممکن است گمراه کننده باشد [۴، ۵]، در سال‌های اخیر تلاش‌هایی صورت گرفته است تا با گذر از فرض ساده انگارانه نرخ متوسط خطا، آن را به شکل تابعی از پارامترهای مختلف مدلسازی کنند. در پاره‌ای از این پژوهش‌ها مدلسازی بر اساس خطاهای متأثر از فاکتورهای ذاتی شبکه صورت گرفته است [۶ و ۷]، حال آن که دسته دیگر به مدلسازی خطاهای متأثر از فاکتورهای محیطی پرداخته‌اند [۵، ۸، ۹، ۱۰، ۱۱]. در برخی مطالعات نیز ترکیبی از

ابزارهای مدیریت دارایی^۱ هستند، تا با در دست داشتن تصویر واقعی از وضعیت قابلیت اطمینان شبکه خود، بسترهای بهبود آن را فراهم نمایند.

شبکه توزیع انرژی الکتریکی که حلقه نهایی تحویل انرژی الکتریکی به مصرف‌کننده است، به علت ویژگی‌هایی نظیر گستردگی، تجهیزات زیاد و متنوع، نزدیکی به مصرف‌کننده، و ... اهمیت بالایی دارد. در نتیجه، مساله حفظ و بهبود قابلیت اطمینان آن بسیار حائز اهمیت است. از آنجا که ارزیابی قابلیت اطمینان مبتنی است بر اطلاعات خاموشی گذشته، مطالعات آماری اتفاقات ثبت شده در سیستم مدیریت خاموشی^۲ به عنوان اولین و اساسی‌ترین حلقه از زنجیره ارزیابی تا بهبود قابلیت اطمینان، اجتناب ناپذیر می‌شود. با توسعه ابزارهای آماری و تحلیل داده، در قالب زیرمجموعه‌های دانش نوین داده-کاوی^۳، عرصه جدیدی در علوم مختلف توسعه یافته است که به استخراج دانش^۴ نهفته در پس داده‌های خامی که در انبارهای داده انباشته شده‌اند، می‌انجامد. از دیدگاه ارزیابی قابلیت اطمینان شبکه توزیع، مطالعه آماری خطاهای شبکه، به شناسایی نقاط حادثه‌خیز و دلایل اصلی رخداد خطاها کمک می‌کند. اهمیت این مطالعات در تصمیم‌گیری جهت تدوین برنامه‌های نگهداری و تعمیرات عیان خواهد شد، چنانکه روی بیلینتون در [۱] بیان می‌کند، شناسایی بخشهایی از سیستم که با صرف هزینه موجب بهبود قابلیت اطمینان می‌شود، با ارزش‌تر است از محاسبه وضعیت کنونی قابلیت اطمینان در سیستم. به علاوه، ادغام این مطالعات با روش‌های

1. Asset Management
2. Outage Management Systems (OMS)
3. Data Mining (DM)
4. Knowledge Discovery (KD)

از قوانین فازی تدوین شده توسط خبرگان شبکه، جهت تخمین نرخ خرابی بهره می‌گیرد.

اگرچه در مطالعات پیشین خاموشی‌های ماندگار بیشتر مورد توجه بودند، رشد بارهای حساس نظیر سرویس‌دهنده‌های رایانه‌ای موجب ضرورت پرداختن به خطاهای گذرا و شاخص‌های کیفیت توان در ارزیابی‌های قابلیت اطمینان شبکه توزیع شده است [۱۴، ۱۶]. از این‌رو نیاز به مطالعات بیشتر در زمینه خطاهای گذرا ضرورت پیدا می‌کند. این مقاله مدلسازی خطای گذرا در شبکه‌های توزیع هوایی را مورد مطالعه قرار می‌دهد و با بهره‌گیری از ابزارهای داده‌کاوی شامل تحلیل آماری اطلاعات خاموشی و دانش خبرگان شبکه توزیع شهری تهران بزرگ، به شناسایی عوامل موثر در رخداد خطای گذرا و ارائه مدل مناسب در تخمین رفتار آن می‌پردازد.

به دلیل پیچیدگی‌های مساله، کارایی سه روش مدلسازی در تخمین نرخ خطای گذرا مورد بررسی قرار می‌گیرد. روشی مبتنی بر درخت تصمیم، به نام روش کارت، و شبکه‌های عصبی به عنوان دو روش غیرپارامتریک، و رگرسیون چند متغیره خطی به عنوان یک روش پارامتریک جهت مدلسازی انتخاب شده است. اگرچه روش کارت در رشته‌های گوناگون به کار رفته است، اما در مطالعات قابلیت اطمینان به ندرت اعمال شده است [۱۷]، و بر اساس اطلاعات ما این نخستین باری است که در مطالعات قابلیت اطمینان شبکه توزیع مورد استفاده قرار می‌گیرد.

فاکتورهای هر دسته در مدلسازی اعمال شده است [۱۳ و ۱۲].

مرجع [۶]، روشی عملیاتی ارائه داده است، که داده‌های گردآوری شده از بازرسی تجهیزات شبکه را به نرخ خرابی آنها نگاشت می‌دهد. در [۷] نرخ خرابی کابل‌ها در شبکه توزیع زمینی بصورت تابعی از سن کابل مدلسازی می‌شود. مرجع [۱۲] همچنین، با مطالعه آماری جامع (۵۰ سال) روی اطلاعات خرابی تجهیزات شبکه‌های توزیع، مدل احتمالاتی برای پیری این تجهیزات به دست آورده است.

از مجموعه پژوهش‌ها روی اثر فاکتورهای محیطی بر خرابی‌ها، عموماً به مطالعه اثر درختان در مجاورت شبکه و یا شرایط جوی بر رخداد خطا در خطوط هوایی توجه شده است. [۸، ۹]، با انجام تحلیل‌های آماری روی اطلاعات گذشته خاموشی یک شبکه توزیع، و با استفاده از روش رگرسیون لاجستیک مدلی برای وقوع یا عدم وقوع خطاهای ناشی از درختان پیشنهاد داده است. [۱۰]، با اعمال داده‌های گذشته خاموشی به مدل‌های رگرسیون خطی، رگرسیون نمایی و شبکه عصبی، توانایی آن‌ها را در تخمین خرابی‌های متاثر از درختان در شبکه توزیع مورد ارزیابی قرار داده است.

از آنجا که خرابی‌ها متاثر از فاکتورهای گوناگون هستند، مطالعه فاکتورهای متنوع‌تر باعث ارتقاء مسیر پژوهش خواهد شد. در مقالات [۵، ۱۱]، ترکیبی از پارامترهای جوی و فعالیت‌های شاخه‌زنی به عنوان عوامل اثرگذار بر رخداد خرابی‌ها پیشنهاد شده است. مدلسازی [۵] بر روش استنتاج فازی استوار است. به این ترتیب که

1. Classification and Regression Tree (CART)

.....

نتایج حاصل از مدلسازی‌های این مطالعه بر کارایی بهتر مدل شبکه عصبی نسبت به رگرسیون چندمتغیره خطی و کارت در تخمین نرخ خطای گذرا، تاکید دارد. هرچند جزئیات کامل ارزیابی‌ها در قالب بحث‌هایی پیرامون نتایج بیان می‌شود.

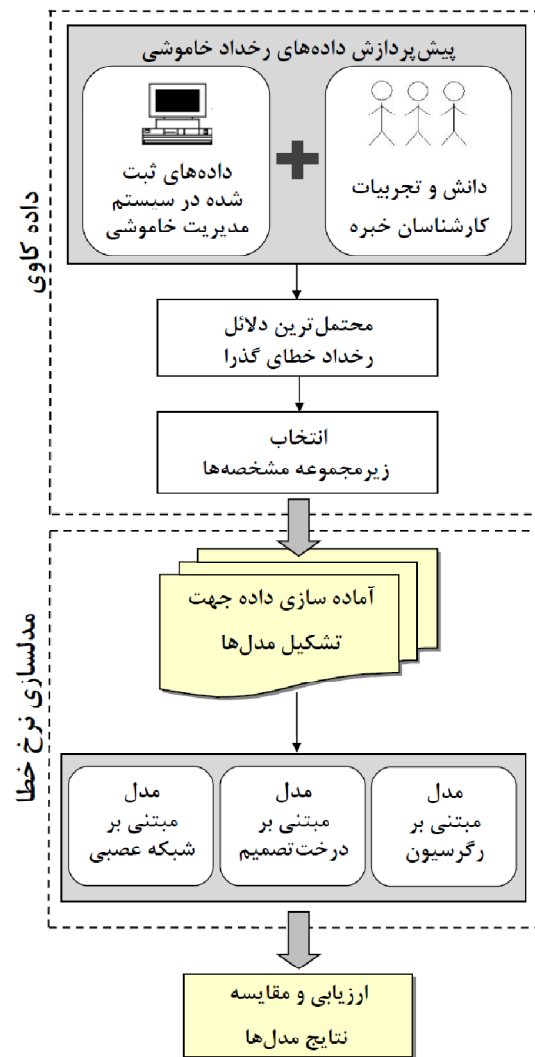
مدلسازی نرخ خطای گذرا ارائه می‌شود. فرآیند داده‌کاوی شامل جمع‌آوری، پیش‌پردازش داده و چگونگی استخراج عوامل اثرگذار بر خطای گذرا در بخش (۳) شرح داده می‌شود. متعاقباً، نتایج مدلسازی‌ها، همچنین مقایسه و بحث روی نتایج بخش (۴) را در برمی‌گیرد، و با جمع‌بندی و ارائه پیشنهادات در بخش (۵) خاتمه می‌یابد.

۲- روش پیشنهادی

شکل (۱) چهارچوب روش پیشنهادی مدلسازی نرخ خطای گذرا در این مطالعه را به تصویر می‌کشد. این روش از دو مرحله تشکیل شده است: مرحله داده‌کاوی و مرحله مدلسازی.

در مرحله داده‌کاوی، ابتدا داده‌های خاموشی گذشته شبکه پیش‌پردازش می‌شوند. این قسمت شامل تحلیل آماری رخدادهای ثبت‌شده در سیستم مدیریت خاموشی و جمع‌آوری نظرات و تجربیات خبرگان شبکه در مورد دلایل وقوع خطای گذرا است. به این ترتیب محتمل‌ترین دلایل اثرگذار بر وقوع خطای گذرا در قالب یک مجموعه داده به‌دست می‌آید. در ادامه این مرحله، یک الگوریتم انتخاب مشخصه‌ها^۱ به مجموعه داده اعمال می‌شود تا متغیرهای اصلی موثر در رخداد خطای گذرا از میان تمام متغیرهای مجموعه داده استخراج شود. به این ترتیب هم فضای جستجوی مساله کوچک‌تر شده و متعاقباً حجم محاسبات در ادامه مطالعه کاهش می‌یابد، و هم دید واقع‌بینانه‌تری از پارامترهای اثرگذار بر خطای گذرا به‌دست خواهد آمد.

سپس، مرحله مدلسازی نرخ خطای گذرا با آماده‌سازی نمونه‌ی داده‌ها برای مدلسازی آغاز می‌شود. سه



شکل ۱ چهارچوب روش پیشنهادی جهت مدلسازی خطای گذرا

ساختار ادامه مقاله به این ترتیب سازمان‌دهی شده است: در بخش (۲) مراحل روش پیشنهادی جهت

1. Feature Selection Algorithm

موقعیت جغرافیایی زیر-حوزه‌ها نام‌گذاری شده است که عبارتند از: شمال غرب، شمال شرق، جنوب غرب، و جنوب شرق.

۲-۳- پیش‌پردازش داده‌های خاموشی و رخداد خطا

تحلیل خاموشی‌های بی‌برنامه گذشته سیستم نخستین گام در مطالعه خطاها است. از این رو، اطلاعات مربوط به رخداد خطاها در تهران بزرگ، در بازه سال‌های ۱۳۸۶-۱۳۸۸ از پایگاه داده سیستم ثبت اطلاعات خاموشی شرکت توزیع نیروی برق تهران بزرگ (ENOX) استخراج شد. بیش از ۴۹۰۰ خاموشی بی‌برنامه مربوط به شبکه فشار متوسط هوایی (۲۰KV)، مربوط به بازه سال‌های ۱۳۸۶-۱۳۸۸، برای مطالعه در نظر گرفته شد. این رخدادها شامل کلاس‌های مختلف خطاهای شبکه است که به شکل از پیش تعریف‌شده در پایگاه داده ثبت می‌شوند. تعداد خطاهای شبکه فشار متوسط هوایی به تفکیک نوع و انرژی توزیع نشده ناشی از هر نوع خطا در شکل‌های (۲) و (۳) به تصویر کشیده شده است. همانطور که در شکل‌های (۲) و (۳) مشاهده می‌شود، خطای گذرا نوع غالب خطاهای شبکه فشار متوسط هوایی است و ۵۷/۴٪ کل تعداد خطاها در بازه زمانی مورد مطالعه را به خود اختصاص می‌دهد (یعنی، ۲۸۵۸ خطا از ۴۹۷۹ کل خطاها) که این منجر به ۳۱/۵٪ انرژی توزیع نشده در شبکه هوایی شده است (۱۱۱۷ MWh).

مطابق این بررسی، خطای گذرا از هر دو منظر فراوانی رخداد و انرژی توزیع‌نشده قابل‌تامل است. از طرف دیگر به دلیل تجهیز نبودن همه بخش‌های شبکه مورد مطالعه به ریکلوزر، خطاهای گذرا گاهی بین ۱۵ تا ۲۰ دقیقه هم

روش مدلسازی جهت تخمین نرخ خطای گذرا انتخاب شده است. روش رگرسیون چند متغیره خطی به‌عنوان یک روش پارامتریک، روش کارت مبتنی بر درخت تصمیم و شبکه‌های عصبی به‌عنوان دو روش غیر پارامتریک (فاقد مدل) انتخاب شدند. در نهایت، مسیر مدلسازی با ارزیابی کارایی مدل‌ها و نتایج پایان می‌رسد.

۳- فرآیند داده‌کاوی

۳-۱- حوزه مطالعه

به دلیل وابستگی این مطالعه به داده‌های واقعی شبکه توزیع انرژی الکتریکی، شبکه توزیع کلان‌شهر تهران بزرگ برای مطالعه و جمع‌آوری داده انتخاب شد. تهران بزرگ نظیر سایر کلان‌شهرها دارای ویژگی‌هایی است که آن را از نقطه‌نظر ملاحظات قابلیت اطمینان حائز اهمیت می‌کند. تنوع الگوی مصرف، حجم مصرف‌کنندگان حساس اعم از صنعتی، تجاری و دولتی، چگالی بالای بار از یک سو، و اثرات منفی ترافیک شدید ساعات تردد، ناکافی بودن اتوماسیون در شبکه توزیع و سیستم‌های مدیریت خاموشی مکانیزه از سوی دیگر، مطالعات جامع-تر جهت ارزیابی‌های قابلیت اطمینان و شناسایی بخشهای ضعیف شبکه را ضروری می‌کند.

این شبکه زیر پوشش شرکت توزیع نیروی برق تهران بزرگ اداره می‌شود. به دلیل گستردگی، منطقه تحت پوشش این شرکت به چهار زیر-حوزه و با مدیریت مجزا تقسیم شده است، که تحت یک نظارت واحد سیاست-گزاری شده و فعالیت می‌کنند. این تقسیم بندی بر اساس

1. Model-free

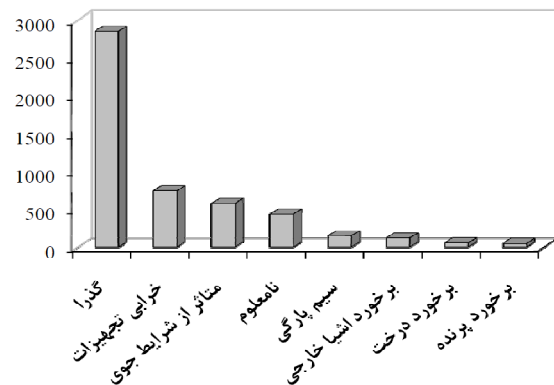
فرض اولیه بر اثرپذیری خطای گذرا از شرایط جوی استوار شد. شرایط جوی که به عنوان عوامل اثرگذار کاندیدا شدند، عبارتند از: میانگین رطوبت و دمای روزانه، میزان بارش روزانه، میانگین و ماکزیموم شدت باد روزانه. اطلاعات مورد نیاز از شرایط جوی شهر تهران بزرگ از وبسایت سازمان هواشناسی جمع‌آوری شد [۱۸]. به دلیل شعاع محدود پوشش دستگاه‌های اندازه‌گیری در ایستگاه‌های هواشناسی، دو منطقه از میان مناطق توزیع انرژی الکتریکی تهران بزرگ که بیشترین تطابق بر ایستگاه‌های هواشناسی، و بیشترین تمرکز فیدرهای هوایی فشار متوسط را داشتند، برای مطالعه انتخاب شدند؛ یعنی مناطق برق شمیران و آزادی.

به طور خلاصه، تلفیق اطلاعات روزانه وقوع خطای گذرا در تک‌تک فیدرهای هوایی هر منطقه به عنوان متغیر وابسته، به همراه اطلاعات ایستگاه هواشناسی متناظر به عنوان متغیرهای مستقل، ساختار کلی پایگاه داده برای بررسی اثر شرایط جوی را تشکیل داد. جهت تحقیق اثرپذیری خطای گذرا از شرایط جوی، از تحلیل همبستگی^۳ بهره گرفتیم. تحلیل همبستگی، ابزاری آماری است برای تعیین نوع و درجه رابطه یک متغیر با متغیر دیگر، که با رابطه (۱) بیان می‌شود

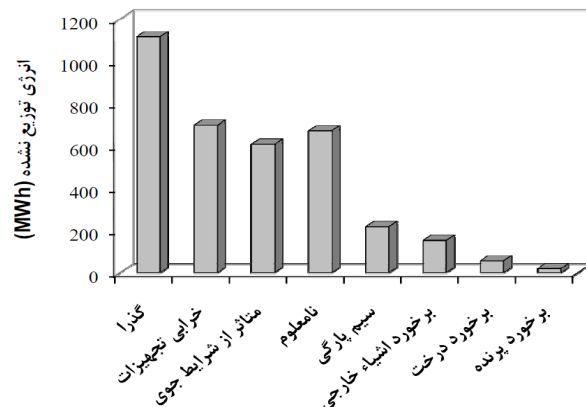
$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)} \sqrt{E(Y^2) - E^2(Y)}} \quad (1)$$

که در آن $cov(X,Y)$ کواریانس دو متغیر X و Y ، σ انحراف معیار، و E امید ریاضی است. وجود یا عدم وجود همبستگی بین متغیرها از طریق ایجاد فرضیه

ادامه می‌یابد^۱ به‌علاوه، از آنجاکه خطاهای گذرا طبیعت ناشناخته‌تری در مقایسه با دیگر حوادث دارند، چرا که فیدر پس از حادثه گذرا، بدون نیاز به عملیات بازیابی قابل وصل جریان است^۲، از این‌رو در این مقاله، مدلسازی خطای گذرا مورد توجه قرار گرفت.



شکل ۲ فراوانی خطاها در شبکه



شکل ۳ انرژی توزیع نشده ناشی از رخداد هر خطا

الف) تحقیق اثرپذیری خطای گذرا از شرایط جوی

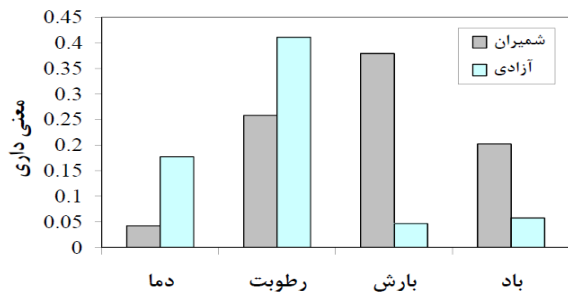
۱. حدود دو تا سه برابر مدت زمان تعیین شده برای خطای گذرا در

استاندارد قابلیت اطمینان IEEE 1366.

۲. مطابق تعریف خطای گذرا در استاندارد IEEE 1366.

3. Correlation Analysis

اشاره شد، اعداد معنی دار در مرز قرار گرفته‌اند (نزدیک به عدد ۰/۰۵ که پیشینه معنی داری است)، این وضعیت در شکل (۴) به تصویر کشیده شده است.



شکل ۴ باند معنی داری همبسته بودن خطای گذرا و پارامترهای جوی

جدول ۱ ضرایب همبستگی بین خطای گذرا و شرایط جوی و تحقیق معنی داری ضرایب

		دما	رطوبت	بارش	باد
منطقه شمیران	ضریب همبستگی	۰/۱۱۳	-۰/۰۳۵	۰/۰۱۷	۰/۰۴۵
	معنی داری	۰/۰۴۴	۰/۲۵۸	۰/۳۷۹	۰/۲۰۲
منطقه آزادی	ضریب همبستگی	۰۴۹/۰	۰۱۲/۰	۱۰۱/۰	۰۸۴/۰
	معنی داری	۱۷۷/۰	۴۱۱/۰	۰۴۷/۰	۰۵۷/۰

با این ملاحظه در می‌یابیم که اثر دو پارامتر دما و بارش به صورت بسیار جزئی و محلی است، نه فراگیر. بر همین اساس می‌توانیم بیان کنیم که در منطقه آزادی هم به دلیل اختلاف (افزونی) جزئی معنی داری پارامتر باد با پیشینه معنی داری، همبستگی بسیار کمی بین خطای گذرا و باد در آن منطقه برقرار است.

به طور کلی، در این فاز مطالعه درمی‌یابیم که اثر شرایط جوی شامل دما، باد و بارش بر رخداد خطای گذرا

آماري^۱ همبستگی قابل تحقیق می‌شود. با فرض اینکه ρ ضریب همبستگی جامعه آماری باشد، فرضیه آماری مطابق رابطه (۲) بیان می‌کنیم:

$$H_0: \rho = 0 \text{ ندارد وجود ندارد}$$

$$H_1: \rho \neq 0 \text{ همبستگی معنی دار وجود ندارد}$$

(۲)

که در آن H_0 و H_1 فرضیه‌های آماری ایجاد شده برای بررسی آزمون آماری است. از نظر آماری تحقیق معنی-داری^۲ فرضیه معادل محاسبه مقدار- p فرضیه است. به این ترتیب که مقدار- p با مقدار خطایی که معمولاً ۰/۰۵ در نظر گرفته می‌شود و با α نمایش می‌دهیم، مقایسه می‌شود [۱۹، ۲۰]. رابطه (۳) را داریم:

(۳)

(وجود همبستگی فرض H_0 رد می‌شود $\rho\text{-value} < \alpha$)
 (عدم همبستگی) فرض H_1 پذیرفته می‌شود $\rho\text{-value} > \alpha$
 نتایج انجام تحلیل همبستگی در جدول (۱) خلاصه شده است، که شامل ضریب همبستگی میان خطای گذرا و هریک از پارامترهای شرایط جوی، و همچنین مقدار معنی داری متناظر آن، به تفکیک منطقه، است. همان‌گونه که در جدول (۱) مشاهده می‌شود، براساس سطح معنی-داری، خطای گذرا در منطقه شمیران تنها با پارامتر دما همبسته است، در حالی که در منطقه آزادی، همبستگی تنها با پارامتر بارش برقرار است. به علاوه، از دقت در ارقام معنی داری در جدول (۱) مشاهده می‌کنیم، حتی در مواردی که در بالا به عنوان پارامترهای همبسته به آن

1. Hypothesis Testing
2. Significance
3. P-value

افراد فوق به دست آمد، به عنوان پارامترهای برگزیده اثرگذار بر رخداد خطای گذرا تلقی شد. به طور کلی عوامل موثر در چهار کلاس اصلی دسته بندی شدند که عبارتند از:

- ضعف در طراحی و اجرای شبکه،
- کاستی در فعالیت‌های بهره برداری،
- شرایط جوی، و
- مداخله عوامل محیطی (خارجی).

هریک از این کلاس‌ها نیز به زیرمجموعه‌هایی تقسیم می‌شوند که در جدول (۲) قابل مشاهده است. در مرحله بعدی باید درستی نتایج به دست آمده تحقیق شود. در این راستا، مجدداً اطلاعات ۱۶۰ فیدر هوایی، این بار براساس وضعیت پارامترهای جدید (جدول (۲)) به عنوان متغیرهای مستقل، جمع‌آوری و در نتیجه پایگاه داده برآمده از آن شکل گرفت.

۳-۳- الگوریتم انتخاب زیرمجموعه مشخصه‌ها

ایده اصلی در انتخاب مشخصه‌ها بر ایجاد و انتخاب زیرمجموعه‌ای از متغیرهایی استوار است که پیش‌بینی کننده‌های قدرتمند متعیر هدف باشند. این ایده نظیر مسیر تحقیق بخش (۳-۲-الف)، از آزمون‌های آماری معنی‌داری در تشخیص مناسب بودن یا نبودن مشخصه‌ها بهره می‌گیرد.

اگر به جای بررسی جداگانه مشخصه‌ها، آنها را به طور جمعی بررسی نماییم، اطلاعات بیشتری بدست می‌آوریم. معمولاً هنگامی که مشخصه‌ها را جداگانه بررسی می‌کنیم، ممکن است برخی از ستونهای جدول داده به اشتباه حذف شوند، و یا ممکن است برخی مشخصه‌ها وقتی جداگانه ملاحظه می‌شوند مفید به نظر آید ولی از نظر قدرت پیش‌بینی متغیر هدف ضعیف باشند [۲۲].

ناچیز است و لذا منشاءهای موثرتر را باید در میان عوامل دیگر جستجو کرد. لازم به ذکر است، علی‌رغم اثر غیر قابل اجتناب شرایط جوی بر رخداد خطاها که در متون و ادبیات قابلیت اطمینان به آن پرداخته‌اند، این نتایج از دو دیدگاه قابل توجیه است: اول، به دلیل وجود آب و هوای معتدل در شهر تهران بزرگ، اثر شرایط جوی در مقایسه با مناطق دیگر که در معرض شرایط جوی سخت‌تر و تغییرات زیاد قرار دارند، کم‌رنگ‌تر است. دیگر اینکه، وجود شرایط آب و هوایی نامناسب بیشتر موجب وقوع خطاهای ماندگار می‌شود و نه گذرا (شکل (۲)).

ب- شناسایی عوامل موثرتر بر خطای گذرا

از آنجا که مطالعه جامع در شرکت توزیع تهران بزرگ در زمینه خطاهای گذرا انجام نشده بود و گزارشهای تحلیلی به شکل مکتوبی موجود نبود، ادامه راه مطالعه با یک چالش مواجه شد و آن یافتن پارامترهای اثرگذارتر جهت مدلسازی خطای گذرا بود. از این رو برای رسیدن به پاسخ، روش پیشنهادی ما بر مبنای یک رویکرد تجربی متشکل از مصاحبه با خبرگان شبکه و نتیجه‌گیری از اجماع نظرات آنها استوار شد. روش قضاوت خبره ۱ پیش از این بعنوان راه حلی کارا، در شرایط فقدان دانش مکتوب کافی، در مطالعات روی شبکه توزیع تهران بزرگ اعمال شده است [۲۱].

از این‌رو، مساله "شناسایی پارامترهای اثرگذار بر خطای گذرا در شهر تهران" در میان خبرگان بخشهای مختلف شبکه اعم از بهره‌برداران شبکه، اعضای مراکز دیسپاچینگ، اپراتورهای اسکادا و مامورهای مانور به بحث گذاشته شد. در نهایت آنچه که از اجماع نظرات

1. Expert Judgment

جدول ۲ عوامل برگزیده اثرگذار بر رخداد خطای گذرا

محتمل ترین دلایل رخداد خطای گذرا	
ضعف در طراحی و بهره‌برداری شبکه	طول نامناسب اسپن‌ها در فیدر
	وجود فلش‌های نامناسب در طول فیدر
	چگالی بار بالا در فیدر
کاستی در فعالیتهای بهره‌برداری	کاستی‌های کمی و کیفی در انجام تعمیرات پیشگیرانه
	شاخه‌زنی‌های ناکافی
	پیری طبیعی در فیدر
	قرارگرفتن در معرض تنش‌های الکتریکی - مکانیکی
شرایط بومی	تغییرات دما
	بارش
مداخله عوامل خارجی	قرارگرفتن در معرض برخورد پرنده‌گان
	پرتاب اشیاء خارجی

جدول ۳ شرایط حذف متغیر یا نمونه داده در مرحله غربال

شرایط حذف متغیر	شرایط حذف نمونه داده
متغیرهایی که مقادیر مفقود- شده دارند.	نمونه‌ای که مقدار هدف (متغیر وابسته) در آن مفقود شده است.
متغیرهایی با مقدار تکراری در مجموعه داده‌ها.	نمونه‌ای که مقدار متغیرهای مستقل آن مفقود شده است.
متغیرهایی که شماره شناسه نمونه‌ها را تعیین می‌کنند.	_____

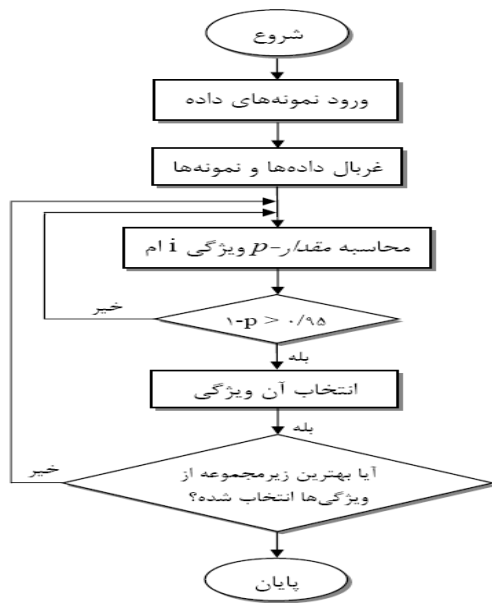
الگوریتم انتخاب زیرمجموعه از مشخصه‌ها شامل سه مرحله است:

- غربال^۱: شامل حذف متغیرهای کم‌اهمیت (متغیرهایی که اطلاعات مفید برای تخمین به دست نمی‌دهند) و یا نمونه داده‌های مشکل‌دار. جدول (۳) شرایطی که موجب حذف متغیر یا نمونه داده می‌شود را نشان می‌دهد.
 - رتبه‌بندی: شامل مرتب‌کردن متغیرهای باقی‌مانده و تخصیص رتبه به آنها،
 - انتخاب: شناسایی زیرمجموعه‌های مهم از ویژگی‌ها برای بکاربردن در مدل.
- رتبه‌بندی مشخصه‌ها بر اساس محاسبه مقدار اهمیت^۲ هر متغیر انجام می‌شود، که برابر است با $1-p$ که در آن p در واقع مقدار p -بدست آمده از انجام آزمون فرض آماری است. اگر متغیر هدف و متغیرهای مستقل (ویژگی‌ها) هر دو از نوع کمی باشند، مقدار p -مبتنی است بر توزیع- t از تبدیل t روی ضریب همبستگی پیرسن r . طبق رابطه (۴) ضریب همبستگی پیرسن r عبارتست از:

$$r = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}) / (N-1)}{\sqrt{S(x)^2 S(y)^2}} \quad (4)$$

که در آن x و y ، به ترتیب، متغیر مستقل و هدف، \bar{x} و \bar{y} میانگین متغیرها و هدف، و $S(x)^2$ و $S(y)^2$ واریانس متغیرها و هدف است.

1. Screening
2. Importance Value
3. t -Distribution



شکل ۵ شمای کلی الگوریتم انتخاب زیرمجموعه مشخصه‌ها

نتیجه الگوریتم در غالب ویژگی‌های استخراج شده جالب توجه است، زیرا همه ویژگی‌های موثر از میان دو کلاس اول یعنی ضعف در طراحی و نصب شبکه، و کاستی در فعالیت-های بهره‌برداری انتخاب شده‌اند. معتقدیم که این نتیجه برای تصمیم‌گیری‌ها در جهت فعالیتهای نگهداری و تعمیرات قابلیت اطمینان محور^۲ سرنخ مفیدی خواهد بود [۲۳].

جدول ۴ نتایج مقدار اهمیت هرمتغیر در خروجی الگوریتم انتخاب مشخصه‌ها

مقدار اهمیت	مشخصه‌ها
۱	چگالی بار بالا در فیدر
۱	وجود فلش‌های نامناسب در طول فیدر
۱	طول نامناسب اسپن‌ها در فیدر
۰/۹۹۵	کاستی‌های کمی و کیفی در انجام تعمیرات
۰/۹۹۴	بشکندانه پیری طبیعی در فیدر

به این ترتیب تبدیل t روی r از رابطه (۵) به دست می‌آید:

$$t = r \sqrt{\frac{N-2}{1-r^2}} \quad (5)$$

در نتیجه تحت فرض صفر که ضریب همبستگی جامعه برابر صفر است ($\rho=0$)، مقدار p از رابطه (۶) محاسبه می‌شود:

$$p\text{-value} = \begin{cases} 0 & \text{اگر } r^2 = 1 \\ 2 \text{ prob} \{T > |t|\} & \text{سایر موارد} \end{cases} \quad (6)$$

که در آن T متغیر تصادفی توزیع t با $N-2$ درجه آزادی است. شمای کلی الگوریتم انتخاب مشخصه‌ها در شکل (۵) قابل مشاهده است.

جهت اعمال الگوریتم بر مجموعه داده گردآوری شده، همان‌گونه که پیش‌تر بیان شد، متغیرهای به‌دست‌آمده در جدول (۲) بعنوان متغیرهای مستقل (مشخصه‌ها)، و متوسط نرخ خطای گذرا بعنوان هدف مساله در نظر گرفته شده‌اند.

با در نظر گرفتن مقدار $0/05$ برای مقدار p به عنوان معیار قطع^۱ در آزمون فرض، نتایج مقدار اهمیت بدست-آمده برای هر مشخصه در جدول (۴) ارائه شده است. مطابق جدول (۴)، شش متغیر مقدار اهمیت بالاتر از حداقل دارند و عبارتند از: چگالی بار فیدر، فلش نامناسب خط، نامناسب بودن اسپن‌ها، ناکافی بودن تعمیرات پیشگیرانه، پیری فیدر، و ناکافی بودن شاخه‌زنی.

1. Cutoff

2. Reliability-centred Maintenance

۴-۲- مدل دوم: درخت تصمیم

درخت تصمیم یکی از ابزارهای قوی و متداول جهت دسته‌بندی و پیش‌بینی است. در این روش، دسته‌بندی در قالب قواعد و یافتن پاسخ برای پرسش‌هایی انجام می‌شود و فرآیند پرسش و پاسخ تا تکمیل دسته‌بندی و بدست آمدن درخت ادامه خواهد داشت. از دسته‌بندی حاصل، سپس، به منظور پیش‌بینی در مدل استفاده می‌شود. به طور کلی، ایجاد درخت تصمیم از دو مرحله تشکیل شده است:

- مرحله رشد و ایجاد درخت: شامل پیوند^۲ و انشعاب^۳. پیوند عبارتست از تخصیص متغیرهای مستقل کم‌معنا^۴ به گروه‌های معنی‌دار، بر اساس متغیر هدف (وابسته). انشعاب شامل انتخاب نقطه شکست از طریق انتخاب متغیری که بتواند بهترین شاخه‌های جدید را در مرحله بعدی انتخاب کند.
- مرحله توقف و هرس^۵ درخت با هدف حداقل کردن خطای پیش‌بینی. توقف قواعدی است که میزان توسعه‌ی انشعاب در گره‌ها را محاسبه می‌کند، و هرس شامل حذف شاخه‌هایی است که اثر ناچیزی در مقادیر تخمینی حاصل از مدل نهایی دارد.

انتخاب نقطه شکست و ایجاد نقطه انشعاب در درخت از اهمیت خاصی برخوردار است. روش‌های مختلف برای انتخاب نقطه شکست وجود دارد. روشی که

2. Merging
3. Splitting
4. Non-significant
5. Pruning

۰/۹۸۲	شاخه‌زنی‌های ناکافی
۰/۹۰۸	بودن در معرض تنش‌های الکتریکی-مکانیکی
۰/۸۳۰	قرار گرفتن در معرض برخورد پرندگان
۰/۴۹۴	پرتاب اشیاء خارجی

۴- مدل‌سازی نرخ خطای گذرا

در این قسمت سه روش مدل‌سازی، جهت تخمین نرخ خطای گذرا اعمال می‌شود. با بررسی‌های فراوان از میان انواع روش‌های موجود، سه روش شامل رگرسیون چندمتغیره خطی به عنوان یک روش پارامتریک، هم‌چنین روشی مبتنی بر درخت تصمیم و شبکه‌های عصبی مصنوعی به عنوان روش‌های غیرپارامتریک جهت مدل‌سازی نرخ خطای گذرا انتخاب شدند.

۴-۱- مدل اول: رگرسیون چندمتغیره خطی

رگرسیون چندمتغیره خطی یکی از ساده‌ترین روش‌های رگرسیون است. اگر متغیرهای مستقل و نرخ خطای گذرا (متغیر وابسته) را، به ترتیب با X_i و λ_m نمایش دهیم، مدل رگرسیون چندمتغیره خطی را می‌توان با رابطه (۷) توصیف کرد:

$$\lambda_m = \alpha_0 + \sum_{i=1}^6 \alpha_i X_i \quad (7)$$

که در آن α_i ضرایب متناظر با هر متغیر در رگرسیون است. حل این معادله بر روش کمترین مربعات استوار است. به این ترتیب که بهترین مدل برازش شده بر مجموعه داده، مدلی خواهد بود که در آن مجموع مربع مانده‌ها^۱ یعنی اختلاف بین داده واقعی و مقداری که از مدل به دست می‌آید، کمینه باشد.

1. Residual

جلوگیری از وقوع بیش‌برازش^۱، فرآیند هرس کردن باعث از بین رفتن شاخه‌هایی که کمترین قابلیت اعتماد را دارند می‌شود و منجر به بهبود توانایی درخت در دسته‌بندی و پیش‌بینی داده‌های آزمون خواهد شد. یعنی، به عبارتی، با افزایش نرخ خطا در آموزش، نرخ خطا در آزمون را کاهش می‌دهد و در واقع مدلی با قابلیت تعمیم بهتر ایجاد می‌کند.

۴-۳- مدل سوم: شبکه‌های عصبی مصنوعی

عملکرد موفق شبکه‌های عصبی مصنوعی در تخمین و پیش‌بینی، قابلیت آنها در مواجهه با داده‌های مغشوش و کمبود دانش در مورد فضای مساله، موجب وسعت کاربرد آنها در مطالعات شبکه قدرت شده است.

شبکه پیش‌خور^۲ یکی از متداول‌ترین شبکه‌ها است که با الگوریتم یادگیری پس‌انتشار^۳ خطا آموزش داده می‌شود. این الگوریتم بر اساس کاهش خطا و به صورت نظارتی^۴ شکل می‌گیرد؛ به این ترتیب که بر اساس وزن‌های تصادفی یک پاسخ توسط شبکه تولید می‌شود و در یک فرآیند تکراری میزان خطای میان خروجی شبکه با مقادیر واقعی بر اساس تغییر وزن‌ها کاهش می‌یابد. با رسیدن خطا به حداقل ممکن، شبکه توسط داده‌ها آموزش داده شده و قادر است برای داده‌های جدید همان الگو را ارائه دهد [۲۴].

با بررسی‌های گسترده، یک شبکه سه لایه‌ای (با یک لایه میانی) و تابع فعال‌سازی باینری سیگموید^۵ برای مدلسازی این مساله انتخاب شد. در شکل (۶) ساختار شبکه عصبی شامل ورودی‌ها و نرخ خطای گذرا به عنوان

ما در این مطالعه انتخاب کردیم روش کارت نام دارد. در الگوریتم پیاده‌سازی کارت شاخه‌ها بصورت دوتایی و تنها بر اساس یک متغیر مستقل (فیلد) انشعاب می‌زنند. هر گره والد بر اساس شباهت بین اعضای آن، به دو زیرگره فرزند تفکیک می‌شود، در نتیجه اعضای هر گره از والد خود خالص‌تر خواهند بود. روش‌های مختلف برای محاسبه سطوح ناخالصی وجود دارد. از آنجاکه در این مطالعه متغیرها همگی کمی هستند، روش کمترین مربعات انحراف برای اندازه‌گیری ناخالصی انتخاب شد. به این ترتیب رابطه بیانگر انشعاب گره والد t به دو گره فرزند راست و چپ (t_L و t_R) برابر است با [۱۷]:

$$\Phi(t) = R^2(t) - p_L \times R^2(t_L) - p_R \times R^2(t_R)$$

$$= \frac{1}{N(t)} \sum_{i \in t} [y_i - \bar{y}(t)]^2 - p_L \frac{1}{N(t_L)} \sum_{i \in t_L} [y_i - \bar{y}(t_L)]^2 - p_R \frac{1}{N(t_R)} \sum_{i \in t_R} [y_i - \bar{y}(t_R)]^2 \quad (۸)$$

که در آن p_L و p_R نسبت اعضا در گره والد است که به ترتیب، در گره‌های فرزند سمت چپ و راست قرار می‌گیرند، y_i مقدار متغیر هدف در نمونه i -ام، $\bar{y}(t)$ مقدار متوسط در گره والد t -ام، $\bar{y}(t_x)$ متوسط مقدار گره فرزند t_x ، $N(t_x)$ تعداد نمونه‌های تخصیص یافته به گره فرزند t_x ، و $R_2(t_x)$ انحراف معیار وزندار مرتبط با گره فرزند t_x . بهترین انشعاب آن است که تابع $\Phi(t)$ را بیشینه کند و این معادل کاهش ناخالصی حین ایجاد درخت است.

این روند بعد از هر شاخه زنی برای گره‌های پائین‌تر ایجاد شده، تکرار می‌شود و تا آنجا پیش می‌رود که به گره‌ای برسیم که ایجاد شاخه جدید، گوناگونی را تغییر ندهد. برای

1. Overfitting
2. Feed-forward
3. Back-propagation
4. Supervised
5. Sigmoid Binary

به صورت مجزا، به ارزیابی و مقایسه آن‌ها خواهیم پرداخت.

۱-۵- روش رگرسیون چندمتغیره خطی

با برازش مساله به مدل رگرسیون چندمتغیره خطی، مدلی مطابق رابطه (۹) خواهیم داشت:

$$Y = -0.01887 + \begin{pmatrix} 0.1422 \\ 0.3104 \\ 0.4806 \\ 0.1553 \\ 0.1431 \\ 0.05001 \end{pmatrix} \cdot \bar{X}, X = x_1: x_6 \quad (9)$$

که در آن \bar{X} بردار متغیرهای ورودی مساله است، شامل شش متغیر اول جدول (۴).

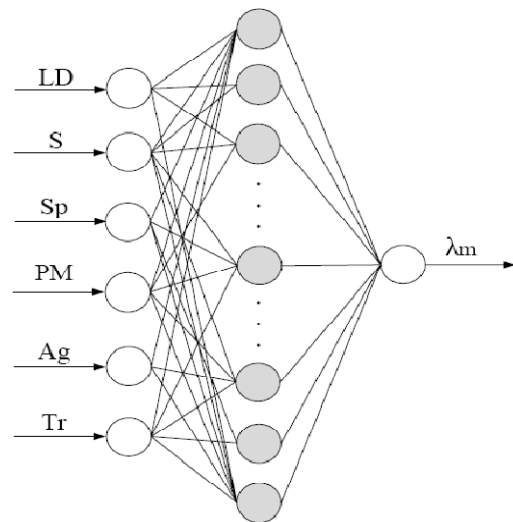
۲-۵- روش کارت

شکل (۷) ساختار درخت هرس شده مساله را نشان می‌دهد. چهار نمونه از قوانین به دست آمده از درخت در جدول (۶) ارائه شده است. این قوانین در قالب "اگر-آنگاه" هستند، به این ترتیب که هر مسیر طولی از ابتدا تا انتها یک قاعده برای تخمین متغیر هدف ایجاد می‌کند. جزء-قاعده‌های طول مسیر با ترکیب عطفی ("و") بخش مقدم^۱ قاعده (یعنی بخش "اگر")، و گره نهایی بخش تالی^۲ (بخش "آنگاه") را تشکیل می‌دهد.

خروجی مدل نشان داده شده است. نام هر ورودی و علامت اختصاری آن در جدول (۵) ارائه شده است.

جدول ۵ علامت اختصاری متغیرهای ورودی شبکه

علامت اختصاری	متغیر ورودی
LD	چگالی بار
S	فلش خط
Sp	اسپن خط
PM	تعمیرات پیشگیرانه
Ag	پیری طبیعی
Tr	شاخه زنی



شکل ۶ شمای کلی شبکه عصبی مساله تخمین خطای گذرا

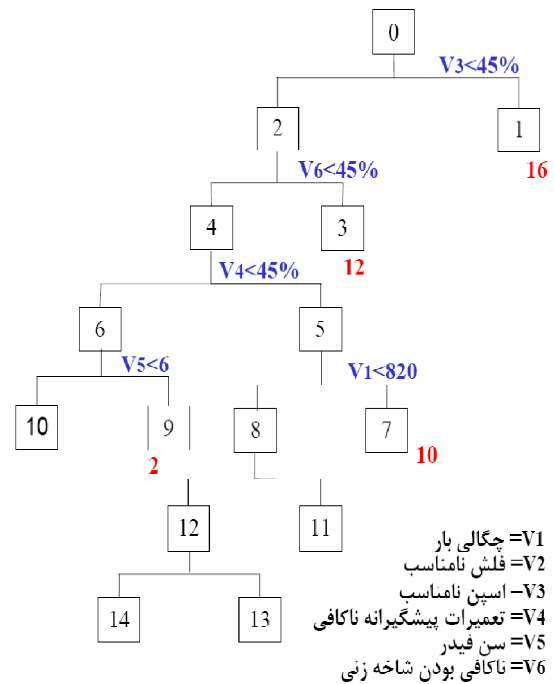
۵- نتایج پیاده‌سازی

هر سه مدل تخمین نرخ خطای گذرا به مجموعه داده به دست آمده در فصل سوم (اطلاعات خاموشی و وضعیت ۱۶۰ فیدر هوایی نسبت به شش متغیر ورودی مساله) برازش شدند. در ادامه پس از ارائه جزئیاتی نتایج مدل‌ها

1. Antecedent
2. Consequent

جدول ۶ نمونه‌هایی از قوانین به دست آمده از درخت

<p>گره ۱</p> <p>اگر ($< 45\%$ اسپن‌های نامناسب)</p> <p>آنگاه وضعیت گره ۱ اتفاق می‌افتد:</p> <p>نرخ خطای گذرا ۱۶ بار در سال</p>
<p>گره ۳</p> <p>اگر ($> 45\%$ اسپن‌های نامناسب) و ($< 45\%$ شاخه‌زنی نامناسب)</p> <p>آنگاه وضعیت گره ۳ اتفاق می‌افتد:</p> <p>نرخ خطای گذرا ۱۲ بار در سال</p>
<p>گره ۷</p> <p>اگر ($> 45\%$ اسپن‌های نامناسب) و ($> 45\%$ شاخه‌زنی نامناسب) و ($< 45\%$ ناکافی بودن تعمیرات پیشگیرانه) و (820 KVA/Km)</p> <p>$<$ چگالی بار)</p> <p>آنگاه وضعیت گره ۷ اتفاق می‌افتد:</p> <p>نرخ خطای گذرا ۱۰ بار در سال</p>
<p>گره ۹</p> <p>اگر ($> 45\%$ اسپن‌های نامناسب) و ($> 45\%$ شاخه‌زنی نامناسب) و ($> 45\%$ ناکافی بودن تعمیرات پیشگیرانه) و (6 سال $<$ سن فیدر)</p> <p>آنگاه وضعیت گره ۹ اتفاق می‌افتد:</p> <p>نرخ خطای گذرا ۲ بار در سال</p>



شکل ۷ درخت حاصل از مدل کارت

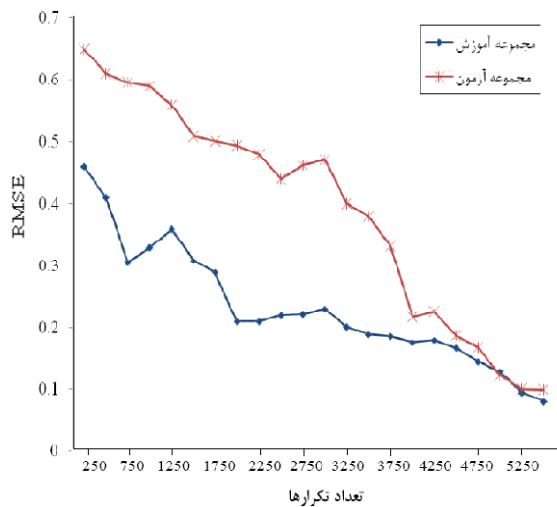
۵-۳- شبکه عصبی

هفتاد درصد مجموعه داده بصورت تصادفی برای آموزش شبکه عصبی انتخاب شد و سی درصد باقی مانده برای ارزیابی قدرت تخمین و تعمیم‌پذیری مدل در قالب مجموعه آزمون مورد استفاده قرار گرفت. طی فرآیند آزمون و خطا جهت رسیدن به شبکه‌ای با پارامترهای مناسب، یک شبکه سه لایه‌ای با نرخ یادگیری $\eta = 0.2$ ، شتاب $\alpha = 0.05$ و 5250 تکرار به جواب مناسب منجر شد. شکل (۸) خطای یادگیری و آزمون شبکه را در تکرارهای مختلف نشان می‌دهد.

۵-۴- ارزیابی، مقایسه و بحث روی نتایج مدل‌ها

پس از ارائه نتایج اولیه هر مدل، در این قسمت تلاش می‌کنیم توان‌مندی مدل‌ها را در مساله تخمین خطای گذرا مورد ارزیابی و مقایسه قرار دهیم. دو شاخص برآورد خطا جهت ارائه تصویر واحد از عملکرد مدل‌ها انتخاب شده است: شاخص مجذور متوسط مربعات خطا (RMSE) برای بیان خطای مطلق تخمین، و شاخص مجذور نسبی مربعات خطا (RRSE) برای محاسبه

1. Root Relative Squared Error



شکل ۸ روند خطای شبکه عصبی در مجموعه‌های آموزش و آزمون به ازای تکرارهای مختلف

جدول ۷ خطاهای مدل‌های پیشنهادی در تخمین خطای گذرا

نام مدل	RMSE (مجموعه آموزش)	RMSE (مجموعه آزمون)	RRSE (مجموعه آموزش)	RRSE (مجموعه آزمون)	قدرت تعمیم
رگرسیون	۰/۱۹۸۸	۰/۱۶۸۲	۰/۸۵۴۷	۰/۹۶۹۵	۰/۸۴۶
کارت	۰/۱۳۱۹	۰/۲۰۵۵	۰/۵۶۷۴	۱/۱۲۳۸	۱/۵۵۸
شبکه عصبی	۰/۰۷۹	۰/۰۹۸	۰/۳۹۳۸	۰/۴۷۸۴	۱/۲۴۰۵

خطای نسبی تک تک مقادیر نسبت به میانگین (روابط (۱۰) و (۱۱)).

$$RMSE = \sqrt{\frac{\sum n(\lambda_{ma} - \lambda_{me})^2}{n}} \quad (10)$$

$$RRSE = \sqrt{\frac{\sum n(\lambda_{ma} - \lambda_{me})^2}{\sum n(\lambda_{ma} - \bar{\lambda})^2}} \quad (11)$$

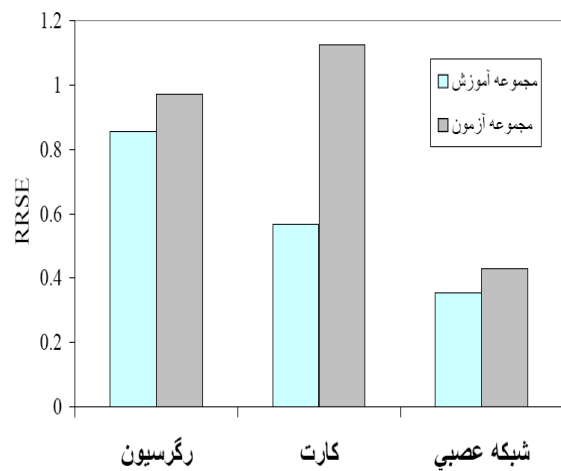
که در آن λ_{me} و λ_{ma} به ترتیب، نرخ واقعی و نرخ تخمینی خطای گذرا، $\bar{\lambda}$ میانگین نرخ واقعی خطای گذرا، و n تعداد نمونه داده است. مقادیر به دست آمده از هر دو ترم خطا برای مجموعه داده‌های هر مدل در جدول (۷) خلاصه شده است. همچنین، از معیاری به نام قدرت تعمیم^۱ مدل برای بیان توانایی مدل نسبت به تخمین مقادیر آزمون - یعنی مقادیر غیر از مقادیر آموزش - استفاده می‌کنیم. این معیار برابر نسبت خطای مطلق (RMSE) مجموعه آزمون به مجموعه آموزش، در هر مدل، است (ستون آخر جدول ۷)، و هرچه مقدار آن کمتر باشد نشان‌دهنده بهتر بودن قدرت تعمیم مدل خواهد بود. همان‌گونه که از نتایج جدول (۷) برمی‌آید، مقدار خطای مدلسازی شبکه عصبی کمتر از دو مدل دیگر بوده و بعد از آن، به ترتیب، خطای حاصل از مدل رگرسیون چندمتغیره و کارت قرار دارند. زیرا اگرچه مقدار خطای مدل کارت در مجموعه آموزش نسبت به مدل رگرسیون کمتر است، اما توانایی تخمین مدل را باید براساس خطای آزمون ارزیابی کرد. برای امکان مقایسه تصویری نتایج، مقدار خطاها در شکل‌های (۹) و (۱۰) ترسیم شده است.

1. Generalization Capability

است. همانگونه که مشاهده می‌کنیم، به طور کلی تخمین شبکه عصبی بهتر از دو مدل دیگر است (نزدیکی مقادیر شبکه عصبی به خطی که نشان‌دهنده برابری مقدار تخمینی با مقدار واقعی است). اما با دقت بیشتر در مقادیر شکل (۱۱) درمی‌یابیم که توانمندی هر سه مدل در تخمین خطاهای کوچک به خوبی خطاهای بزرگ نیست (پراکندگی بیشتر نسبت به خط مذکور). از دلایل آن می‌توان به این نکته اشاره کرد که به دلیل ماهیت تصادفی و ناشناخته‌تر خطای گذرا نسبت به خطای ماندگار، امکان تاثیر پذیری آن از عوامل دیگری که در این مطالعه وارد نشده است، وجود دارد. به عبارت دیگر، از این نتایج می‌توان ادعا کرد که رخداد خطای گذرا به دفعات زیاد (سالانه) در یک فیدر، معلول پارامترهایی است که در مطالعه لحاظ شده است. اما داده‌های نمونه تهیه شده بر اساس همین پارامترها قادر به ارائه تصویر درستی از فیدرهایی که تعداد رخداد خطای گذرای سالانه روی آنها کم است، نبودند. از این رو لازم است در آینده، روی این وضعیت مطالعه و کار بیشتر انجام شود.

علاوه بر این مشاهده می‌کنیم که مدل کارت نه تنها در تخمین خطاهای کوچک، بلکه در مورد خطاهای خیلی بزرگ نیز ضعف دارد، و به عبارتی تنها قادر به ارائه تخمین‌های قابل قبول از مقادیر میانه است. این وضعیت به طور حسی و سرانگشتی در قوانین به‌دست آمده از درخت (جدول (۶)) هم مستتر است.

شکل ۹ مقدار خطای مطلق (RMSE) در مدل‌های پیشنهادی



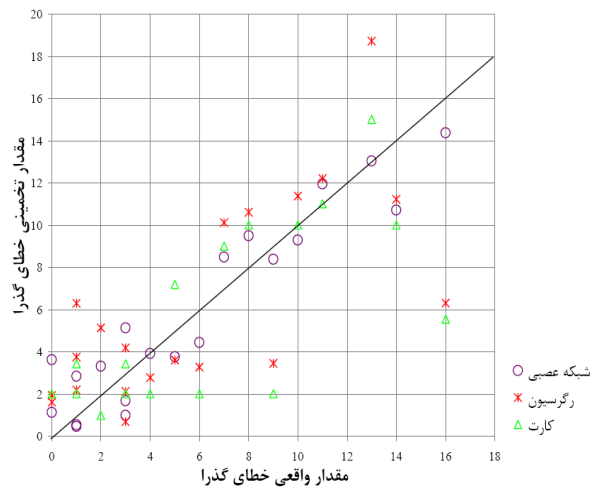
شکل ۱۰ مقدار خطای نسبی (RRSE) در مدل‌های پیشنهادی

اگرچه عملکرد شبکه عصبی از دیدگاه خطای تخمین بهتر از دو مدل دیگر بود، اما از نظر معیار قدرت تعمیم (ستون آخر جدول (۷)) در جایگاه دوم قرار می‌گیرد (بعد از رگرسیون چندمتغیره). در واقع می‌توان گفت که با وجود خطای تخمین بالا در مدل رگرسیون، قدرت این مدل در رویارویی با مقادیر جدید بهتر بوده است (در محدوده مقدار خطای برازش خود (آموزش)). مدل کارت از نقطه نظر قدرت تعمیم نیز در جایگاه آخر قرار می‌گیرد. علت اصلی این وضعیت به تعداد نمونه‌داده و میزان پیچیدگی مدل بستگی دارد. به این مفهوم که با پیچیده‌تر شدن روش مدلسازی به تعداد بیشتری نمونه‌داده برای یادگیری بهتر مدل لازم است.

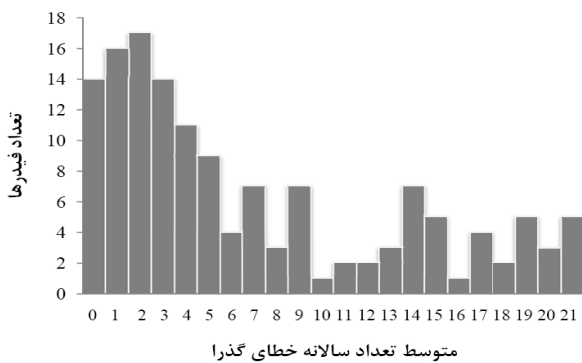
حال عملکرد مدل‌ها در ارتباط با چگونگی تخمین هر نمونه را بررسی می‌کنیم. در شکل (۱۱) مقدار تخمین زده‌شده خطای گذرا نسبت به مقدار واقعی آن، برای ۲۰ فیدر از مجموعه آزمون، و برای هر سه مدل ترسیم شده

جدول ۸ میزان بهبود در خطای تخمین مدل‌های پیشنهادی به ازای حذف برخی داده‌های پرنحرف

نام مدل	RMSE	RMSE	درصد بهبود خطای مدل
	(مجموعه آزمون در حضور کل نمونه‌ها)	(مجموعه آزمون با حذف ۲۱٪ نمونه‌ها)	
رگرسیون	۰/۱۶۸۲	۰/۱۶۱۹	٪ ۰/۶۳
کارت	۰/۲۰۵۵	۰/۱۷۶	٪ ۲/۹۵
شبکه عصبی	۰/۰۹۸	۰/۰۹۴	٪ ۰/۴



شکل ۱۱ مقدار تخمینی خطای گذرا نسبت به مقدار واقعی.



شکل ۱۲ توزیع فراوانی تعداد فیدرها براساس متوسط تعداد سالانه خطای گذرا.

۶- نتیجه‌گیری

در این مقاله سه روش، شامل رگرسیون چندمتغیره خطی، درخت تصمیم و شبکه‌های عصبی، برای مدلسازی نرخ خطای گذرا در خطوط فشارمتوسط هوایی پیشنهاد شد.

ورودی مدل‌ها از فرآیند داده‌کاوی در اطلاعات مربوط به خطاهای شبکه فشارمتوسط، اطلاعات هواشناسی و

در پایان برای نشان دادن اثر نمونه داده‌های (فیدرها) با تعداد کوچک خطای گذرا، بار دیگر از ترم خطای مطلق (RMSE) بهره می‌گیریم. جدول (۸) خطای مطلق (RMSE) مجموعه آزمون را در دو حالت نشان می‌دهد؛ حالت اول در حضور نمونه داده‌های با تعداد کوچک میانگین خطای گذرا، مشابه جدول (۷)، و حالت دوم بدون حضور تعدادی از نمونه داده‌ها (فیدرها) که تعداد متوسط رخداد خطای گذرای سالانه در آنها کمتر از ۲ بار در سال بوده، محاسبه شده است. حذف این تعداد نمونه داده معادل حذف حدود ۲۰٪ از کل مجموعه داده است (مطابق توزیع فراوانی تعداد فیدرها براساس متوسط تعداد سالانه خطای گذرا (شکل (۱۲))، و نشان‌دهنده کاهش در خطای تخمین مدل‌ها است، این بهبود بخصوص در مدل کارت بارزتر است. با ارزیابی نتایج به میزان توانمندی مدل‌ها در برابر مساله تخمین نرخ خطای گذرا پی می‌بریم. در جدول (۹) مزایا و معایب هر روش نیز به اختصار قابل مشاهده است.

به علاوه، تخمین دقیق‌تر از نرخ خطا موجب برآورد صحیح‌تر از شاخص‌های قابلیت اطمینان مشترک-محور نظیر SAIFI و SAIDI خواهد شد.

۷- سپاسگزاری

این مطالعه با همکاری شرکت توزیع نیروی برق تهران بزرگ انجام شده است؛ که در اینجا از همکاری و مساعدت این همکاران قدردانی می‌شود: آقایان ح. ترابی (قائم‌مقام مدیرعامل)، پ. خوارزمی (کارشناس مسئول)، الف. کاظم‌بخشی (معاونت فنی)، و همچنین مدیران مراکز دیسپاچینگ: آقایان ح. مظفری، خ. ترنگان، م. بهرامی، و ح. سیوانی.

۸- منابع

- [1] Billinton, R., Allan, R., "Reliability Evaluation of Power Systems", Plenum Press. New York and London, 1996.
- [2] Brown, R. E., 2002, "Electric Power Distribution Reliability", Marcel Dekker Inc.
- [3] Pahwa, A., Gupta, Sh., R.-E., Brown, "Data Needs for Reliability Assessment of Distribution Systems", in Proc. IEEE Power Engineering Society Summer meeting, 2002.
- [4] Bowles, J.-B., "Commentary—Caution: Constant Failure-Rate Models May Be Hazardous to Your Design", IEEE Trans. on Reliability, Volume 51, Issue 3, pp. 375-377, 2002.
- [5] Gupta, S., Pahwa, A., Brown, R.E., Zhou, Y., Das, S., "An Adaptive Fuzzy Model for Failure Rates of Overhead Distribution Feeders", Electric Power Components and Systems, pp. 1175-1190, 2005.
- [6] Brown, R.E., Frimpong, G., Willis, H.L., "Failure Rate Modeling Using Equipment Inspection Data", IEEE Trans. Power systems, Volume 19, Issue 2, Page(s): 782 - 787, 2004.
- [7] Zhou, Y., Brown, R.-E., "A Practical Method for Cable Failure Rate Modeling", in Proc. IEEE Transmission and Distribution Conference and Exhibition, 2006.

بهره‌گیری از قضاوت خبره‌ها، با استفاده از الگوریتم انتخاب زیرمجموعه مشخصه‌ها استخراج شد. متعاقب آن کارایی سه مدل به‌دست آمده براساس نتایج هریک مورد ارزیابی قرار گرفت که با برشمردن مزایا و معایب هر مدل به پایان رسید.

جدول ۹ مقایسه مزایا و معایب مدل‌های پیشنهادی

مدل	مزایا	معایب
چندمتغیره خطی رگرسیون	سادگی پیاده‌سازی.	عدم مقابله با داده‌های پرت و مغشوش. لزوم خطی بودن روابط.
مبتنی بر درخت تصمیم رگرسیون	سادگی پیاده‌سازی. امکان بررسی نتایج در قالب یک سری قواعد. مدل حاصل به راحتی قابل درک است.	افزایش نرخ خطا در صورت کم‌بودن نمونه‌ها در دسته‌های ایجادشده.
شبکه عصبی	دقت بالا در تخمین متغیر هدف. مقاوم در برابر داده‌های مغشوش.	حساس بودن آموزش شبکه. مانند جعبه سیاه عمل می‌کند. زمان‌بر بودن فرآیند یادگیری.

نتایج این مدلسازی از چند دیدگاه مفید است. از دیدگاه اول که به شناسایی نقاط ضعف شبکه کمک می‌کند و تصویری واقع‌بینانه از آن برای ما ترسیم می‌کند. اما از دیدگاه دوم از این مدل می‌توان جهت توسعه روش‌هایی برای نگهداری و تعمیرات پیشگیرانه و قابلیت اطمینان-محور استفاده کرد (این دیدگاه توسط نویسندگان مقاله در دست مطالعه است).

- [16] Warren, C. M., "The effect of reducing momentary outages on distribution reliability indices," IEEE Trans. Power Delivery, Volume 7, Issue 3, pp.1610-1615, 1992.
- [17] Bevilacqua, M., Braglia M., Montanari, R., "The Classification and Regression Tree Approach to Pump Failure Rate Analysis", Reliability Engineering and System safety, pp. 59-67, 2003.
- [18] www.irimo.ir/farsi/province/tehran.asp.
[۱۹] آذر، ع.، مومنی، م.، "آمار و کاربرد آن در مدیریت"، سازمان مطالعه و تدوین کتب علوم انسانی دانشگاهها (سمت)، ۱۳۸۰.
- [20] Montgomery, D., C., Runger, G., C., "Applied Statistics and Probability for Engineers", Jon Wiley and Sons Inc, 2003.
- [21] A. Fereidunian, M. A. Zamani, H. Lesani, C. Lucas and M. Lehtonen, "An Expert System Realization of Adaptive Autonomy in Electric Utility Management Automation," J. Applied Sci., Vol.9, no.8, , pp.1524-1530, 2009.
[۲۲] غضنفری، م.، علیزاده، س.، تیمورپور، ب.، "داده-کاوی و کشف دانش"، انتشارات دانشگاه علم و صنعت ایران، ۱۳۸۷.
- [23] Haghifam, M.R., Akhavan-Rezai, E., Fereidunian, A., 2009, "A Data Mining Approach to MV Network Failure Analysis Using Field Recorded Data", to appear in Proceedings of IEEE 9th International Conference on Electric Power and Energy Conversion Systems, Sharjah, Dubai, Nov. 10-12.
[۲۴] جکسون، آر.، (نگارنده)، البرزی، م.، (مترجم) "آشنایی با شبکه‌های عصبی"، انتشارات دانشگاه صنعتی شریف، چاپ دوم ۱۳۸۶.
- [8] Xu, L., Mo-yuen Chow, Taylor, L., S., "Data Mining and Analysis of Tree-Caused Faults in Power Distribution Systems", in Proc. IEEE Power Systems Conference and Expositions, 2006.
- [9] Xu, L., Chow, M.-Y., Taylor, L., S., "Analysis of Tree-Caused Faults in Power Distribution Systems", in Proc. IEEE North Amer. Power Symp, 2003.
- [10] Radmer, D.T., Kuntz, P.A., Christie, R.D., Venkata, S.S.; Fletcher, R.H., "Predicting Vegetation-Related Failure Rates for Overhead Distribution Feeders", IEEE Trans. Power Delivery, Volume 17, Issue 4, pp. 1170 - 1175, 2002.
- [11] Gupta, S., Pahwa, A., Brown, R.E., Zhou, Y., Das, S., "An Adaptive Fuzzy Model for Failure Rates of Overhead Distribution Feeders", Electric Power Components and Systems, pp. 1175- 1190, 2005.
- [12] Zhang, X., Gockenbach, E., "Component Reliability Modeling of Distribution Systems Based on the Evaluation of Failure Statistics", IEEE Trans. Dielectrics and Electrical Insulation, Volume 14, Issue 5, pp. 1183-1191, 2007.
- [13] Lang, B.-P., Pahwa, A., "Power Distribution System Reliability Planning Using a Fuzzy Knowledge-based Approach", IEEE Trans. Power Delivery, volume 15, Issue 1, pp.279-284, 2000.
- [14] Brown, R.-E, Gupta, S., Christie, R.-D., Venkata, S.-S., Fletcher, R., "Distribution System Reliability Assessment: Momentary Interruptions and Storms", IEEE Trans. Power Delivery, Volume 12, Issue 4, pp. 1569-1575, 1997.
- [15] Brown, R.-E., Ochoa, J.-R., "Distribution System Reliability: Default Data and Model Validation", IEEE Trans. Power Systems, Volume 13, Issue 2, pp. 704-709, 1998.