دانشگاه صنعتی اراک

هو الصانع

"گواهی پذیرش مقاله"

جناب آقایان حمزه صادقی، روزبه شاد، محمود ققفور مغربی و سرکار خانم مرجان قائمی

بدینوسیله گواهی می شود، مقاله ارزشمند شما با عنوان: "تخمین نرخ شکست خطوط لوله در شبکه آب شهری با استفاده از LS-SVM و خوشه بندی مبتنی بر PSO در محیط GIS" در "چهارمین کنفرانس ملی کاربرد سامانه اطلاعات مکانی (GIS) در صنعت آب و برق"، در تاریخ ۲۲ تا ۲۳ آذر ماه ۱۳۹۶ در شهر اراک، جهت چاپ پذیرفته شده است.

با ابراز خرسندی و تبریک به این مناسبت، از خداوند متعال سلامت و توفیق روز افزون شما را مسئلت می نماییم. اطمینان دارد با استمرار این کوششها در جهت افزایش توانمندی های علمی و پژوهشی کشور گام های موثری برداشته خواهد شد.

4th GIS
چهارمین کنفرانس ملی
سامانه اطلاعات مکانی
در صنعت آب و برق
استان مرکزی - اراک

دکتر محسن نجفی
رئیس کمیته علمی کنفرانس

مهندس فرهاد شبیهی
رئیس کنفرانس

تخمین نرخ شکست خطوط لوله در شبکه آب شهری با استفاده از LS-SVM و خوشه بندی فازی مبتنی بر PSO در محیطGIS

حمزه صادقی[1]*، روزبه شاد[2]،محمود فقفور مغربی[3]، مرجان قائمی[4]

۱- دانشجوی کارشناسی ارشد سیستم اطلاعات مکانی، دانشکده عمران، دانشگاه فردوسی مشهد

Hamzeh.s.d66@gmail.com

۲- استادیار گروه عمران، دانشکده مهندسی عمران، دانشگاه فردوسی مشهد

r.shad@um.ac.ir

۳- استاد گروه عمران، دانشکده مهندسی عمران، دانشگاه فردوسی مشهد

maghrebi@um.ac.ir

۴- استادیار گروه عمران، دانشکده مهندسی عمران، دانشگاه فردوسی مشهد

mghaemi270@gmail.com

چکیده

با افزایش روز افزون جمعیت و گسترش شهرها، شبکه‌های تاسیسات زیرزمینی از جمله شبکه‌های آب، گاز، تلفن و غیره در هم تنیده‌تر و پیچیده‌تر شده‌اند. از آن‌جایی که شبکه توزیع آب در زیر زمین مدفون شده و امکان مشاهده‌ی بخش چندانی از آن وجود ندارد، لذا کنترل و نظارت بر روی شبکه به کاری بسیار سخت و پیچیده تبدیل می‌شود. شبکه توزیع آب متشکل از لوله‌ها و اتصالاتی با ویژگی‌های مختلفی همچون جنس، قطر، طول، سال تولید و سال نصب بوده و در هر نقطه از شبکه نیز عوامل بیرونی متعددی نظیر رطوبت خاک، دما و غیره بر روی آن تأثیر می‌گذارند. این شرایط عمر قطعات در شبکه را متغیر و ناهمگون ساخته و با توجه به نیاز بخش‌های فرسوده و مستهلک شبکه به نوسازی و تعمیر، تشخیص این بخش‌های شبکه (که قابل مشاهده نیستند) بسیار دشوار خواهد بود. لذا در این پژوهش برای تخمین نرخ شکست لوله در شبکه توزیع آب، رویکرد جدیدی از خوشه بندی فازی و روش‌های LS-SVM معرفی شده و از طرف دیگر به شناسایی ارتباط عوامل مؤثر در شکست و چگونگی آن نیز پرداخته می‌شود. بدین منظور از داده‌های شکست خطوط توزیع آب شهر مشهد، مشاهده شده در بین سال‌های ۱۳۸۷ تا ۱۳۹۵ استفاده شده است. در فاز اول با پیاده سازی روش LS-SVM بر روی کل داده‌های منطقه‌ای، نرخ شکست پیش‌بینی می‌شود. در ادامه، برای حساس‌تر شدن و دقیق‌تر شدن نرخ شکست، کل منطقه با روش خوشه‌بندی مبتنی بر PSO به خوشه‌هایی با خصوصیات مشابه تقسیم شده و هر خوشه نیز به یک زیرمنطقه اطلاق می‌شود. در مرحله بعد، بار دیگر با روش LS-SVM، نسبت به پیش‌بینی نرخ شکست برای هر زیرمنطقه اقدام و سپس نتایج حاصل از مجموع زیرمنطقه‌ها با نتایج کل منطقه (که در فاز اول بررسی شد) مقایسه می‌گردد. در نتیجه یک روش ترکیبی بین LS-SVM و خوشه‌بندی PSO به عنوان بهترین روش ترکیبی به منظور پیش بینی نرخ شکست معرفی گردید.

واژه‌های کلیدی: **پیش بینی شکست، خطوط لوله توزیع آب، LS-SVM، خوشه بندی مبتنی بر PSO، خوشه بندی فازی**

3

4

۱- مقدمه

با افرایش روز افزون جمعیت و گسترش شهرها، شبکه‌های تاسیسات زیرزمینی از جمله شبکه‌های آب وگاز و تلفن و غیره در هم تنیده‌تر و پیچیده‌تر شده‌اند. بهره برداری بهینه، اجرا و نظارت بر این شبکه‌های پیچیده تنها با کمک متخصصان و تکنسین‌های کار آزموده امکان پذیر می‌باشد. درکشور ما که از نظر کمبود آب در تنگنا قرار دارد، لذا مسأله تأمین آب اهمیت ویژه‌ای پیدا می‌کند. محدودیت منابع آب موجود از یک سو و رشد سریع صنعتی و توسعه زیر ساختارهای شهری بدون برنامه ریزی لازم از سوی دیگر، لزوم استفاده از یک تفکر سیستمی و جامع در زمینه مدیریت منابع و توسعه زیر ساختارهای آب-رسانی را روز به روز آشکارتر و با اهمیت‌تر می‌سازد. از طرف دیگر، مواجه شدن با حجم عظیمی از داده‌ها و اطلاعات در این گونه زیرساخت‌های مکانی و مطرح شدن زمان به‌عنوان مهمترین سرمایه در مدیریت‌های نوین شهری، استفاده از فناوری‌های نوین جهت دسترسی، جمع‌آوری، پردازش، مدل‌سازی و نهایی نمودن فرآیند تصمیم گیری در کوتاه‌ترین زمان، بالاترین دقت و کمترین هزینه را ناگزیر می‌سازد. طراحی، اجرا و بهره‌برداری از شبکه‌های توزیع آب شهری، وقت و هزینه‌های سنگینی را می‌طلبد. لذا محققین همواره به دنبال توسعه روش‌های بهتر در طراحی، اجرا و بهره‌برداری از شبکه‌های توزیع آب در قالب سیستم اطلاعات مکانی[5] (GIS) بوده‌اند.

یکی از راه‌هایی که بعد از صرفه‌جویی در مصرف آب از مردم انتظار می‌رود، حفظ و نگهداری شبکه تأمین آب است. نگهداری، تعمیرات یا بهسازی به‌موقع خطوط لوله و شریان‌های حیاتی مدفون و سطحی در زمان بهره‌برداری و حین بروز مشکل از اهمیت ویژه‌ای برخوردار است و تداوم بهره‌برداری مناسب از آن‌ها به میزان زیادی بستگی به نحوه مراقبت وضعیت آن‌ها (پایش مداوم) و کیفیت اجرای اولیه طرح و از همه مهم‌تر طرح اولیه دارد. بعضی از نقایص نگهداری و یا مشکلات کوچک پیش‌آمده حین تعمیرات در زمان بروز حوادث طبیعی بزرگ نیز باعث بروز مشکلات متعدد می‌گردند که این مشکلات را به‌راحتی با یک پایش اولیه می‌توان در پیش از رخداد از بین برد. لذا اهمیت بهسازی و نگهداری سیستم از پیش نمایان است.

با گسترش GIS و بهبود حسگرها در شبکه آب‌رسانی، لحظه‌به‌لحظه داده‌های حجیمی نمایش داده و ثبت می‌شوند. با افزوده شدن داده‌ها، تفسیر و نمایش آن‌ها بسیار مشکل می‌شود که این مشکل با استفاده از داده‌کاوی تا حد قابل قبولی مرتفع می‌گردد. خوشه‌بندی یکی از روش‌های داده‌کاوی است که نقش مهمی در این عرصه ایفا می‌کند. ازجمله استفاده گسترده از خوشه‌بندی بر روی داده‌ها می‌توان به طبقه‌بندی داده‌ها در زمینه‌های مختلف مهندسی، گسسته‌سازی داده-های مستمر، تشخیص نقطه خطا، حذف نویز از داده‌ها، تشخیص الگو و پردازش تصویر اشاره نمود.

آمار و اطلاعات نشان می‌دهد کشور ما در معرض خشک‌سالی است. بنابراین با نگاهی به آینده مصرف رو به افزایش آب، نیاز به تعیین معیارهای کاهش تلفات آب در سطح مدیریت کلان ضروری به نظر می‌رسد. به‌طورکلی مقدار هدر رفت آب بیانگر بازدهی شبکه توزیع آب و نحوه نگهداری از آن و به‌طور مستقیم میزان سوددهی مجموعه است. بدین معنی که هزینه‌های تولید، بهره‌برداری و نگهداری از شبکه‌های آب در صورتی قابل توجیه هستند که از وجوه پرداختی مصرف‌کننده کمتر باشند در غیر این صورت مجموعه متضرر می‌گردد. با توجه به اهمیت معیارهایی نظیر بقای آب در اکوسیستم و حفاظت از محیط زیست در دنیا، کنکاش در زمینه کاهش آب به‌حساب نیامده و کنترل شاخص‌های آن روزبه‌روز مورد توجه قرار گرفته و ارزش واقعی خود را در میان مدیران ارشد به‌دست می‌آورد. هدر رفت آب به دو دسته تقسیم-بندی می‌شوند، اول هدر رفت ظاهری که شامل تلفات غیرقابل رؤیت نظیر مصارف غیرمجاز، خطای اندازه‌گیری و دوم هدر رفت واقعی یا فیزیکی که شامل نشت و سرریز از مخازن خطوط انتقال، شبکه توزیع و انشعابات می‌باشند. یکی از راهکارهای ضروری در جهت کاهش هدر رفت فیزیکی آب، اصلاح خطوط انتقال فرسوده، نوسازی انشعابات و بازسازی شبکه‌های فرسوده هست.

طبق گزارش دفتر آمار و فن‌آوری اطلاعات شرکت مهندسی آب و فاضلاب کشور، طول کل شبکه‌های اصلی توزیع آب کشور در پایان سال ۱۳۹۰ برابر با ۱۳۳۱۶۳ کیلومتر بوده است (گزارش عملکرد سالانه صنعت آب و فاضلاب کشور، ۱۳۹۱). از طرفی با توجه به توسعه روزافزون شهرها، نیاز به طراحی و توسعه شبکه‌های موجود و همچنین احداث شبکه‌های جدید است، به طوری که در بودجه سال ۱۳۹۰ شرکت مهندسی آب و فاضلاب کشور پیش‌بینی شده بود ۲۳۷۷ کیلومتر احداث و توسعه شبکه توزیع آب شهری با اعتبار ۱،۱۳۸،۵۹۶،۷۸۵ هزار ریال به انجام برسد. ولی با توجه به نیاز به وجود آمده، مقدار توسعه و احداث شبکه جدید در پایان سال مذکور بیشتر از مقدار پیش‌بینی شده و برابر با ۳۲۹۰ کیلومتر بوده است. از طرف دیگر قسمت اعظم شبکه‌های موجود دارای عمر نسبتاً طولانی است. لذا مساله بازسازی و نوسازی آنها نیز ضرورتی انکارناپذیر بوده و هزینه‌های هنگفتی را به شرکت‌های آب و فاضلاب تحمیل می‌کند، به طوری که طبق پیش‌بینی به عمل امده در بودجه سال ۱۳۹۰ شرکت مهندسی آب و فاضلاب کشور، مقرر شده بود مقدار ۲۸۴۵ کیلومتر از طول شبکه‌های توزیع آب شهری با اعتباری بالغ بر ۱،۵۹۴،۷۰۵،۷۳۸ هزار ریال اصلاح و نوسازی شود. ولی با توجه به ضرورت نوسازی هرچه بیشتر شبکه‌های موجود، مقدار اصلاح و

[5] - Geospatial Information System

نوسازی نیز در پایان سال مذکور بیشتر از مقدار پیش‌بینی شده و برابر ۳۲۵۴ کیلومتر بوده است (گزارش عملکرد سالانه صنعت آب و فاضلاب کشور، ۱۳۹۱).

همان‌طور که از این آمار پیداست، مبالغ هنگفتی صرف توسعه و احداث شبکه‌های جدید توزیع آب و همچنین اصلاح و نوسازی شبکه-های موجود توزیع آب می‌شود. اما نکته قابل تأمل این است که در حال حاضر عملیات بازسازی و نوسازی شبکه‌های توزیع آب کشور به صورت سنتی و غیر علمی صورت می‌گیرد و هیچ توجیهی به بحث هزینه، کارایی و قابلیت اطمینان شبکه‌ها نمی‌شود. همچنین در روش مرسوم طراحی شبکه‌ها، طراحی بر اساس نیازهای آبی سال انتهای طرح صورت می‌گیرد که باعث می‌شود اولاً همه هزینه‌های ساخت در سال ابتدای طرح متمرکز شود و ثانیاً در سال‌های ابتدای طرح با توجه به مقادیر کم مصرف، سرعت جریان در لوله‌ها کم بوده و فشار گرهی افزایش می‌یابد که در نتیجه باعث کاهش کارایی شبکه‌های توزیع آب می‌شود. لذا ضرورت است هرچه سریع‌تر نسبت به تغییر و اصلاح وضعیت موجود اقدام شود چرا که بودجه و منابع مالی شرکت‌های آب و فاضلاب محدود بوده و باید هزینه کردن منابع مالی به صورت بهینه صورت گیرد، به طوری که هم هزینه‌ها کمینه شود و هم قابلیت اطمینان و کارایی شبکه‌ها بیشینه شود. به عبارت دیگر باید طراحی اولیه برنامه‌های بازسازی و نوسازی شبکه‌های توزیع آب به طور هم‌زمان بهینه شده و طرح‌های بهینه ارائه شود تا اولاً از صرف هزینه‌های اضافی غیر ضرری در سال اول طرح جلوگیری به عمل آمده و هزینه‌ها در طول دوره طرح پخش شود و ثانیاً کارایی شبکه در طول دوره طرح افزایش یابد. در این پژوهش احداث یا اجرای شبکه، به عملیات لوله‌گذاری قبل از شروع دوره اطلاق می‌شود. منظور از تعمیر یا بازسازی شبکه توزیع آب، عملیات رفع حوادث و اتفاقات لوله‌های شبکه است. منظور از نوسازی یا اصلاح شبکه نیز تعویض لوله‌های موجود و یا اضافه کردن لوله‌های جدید به موازات لوله‌های موجود (تزویج) است.

سیستم زیرساخت شهری از مهم‌ترین عناصر ساختاری شهر به‌حساب می‌آید، بررسی وضعیت موجود و نظارت بر این زیرساخت‌ها اهمیت بالایی دارد. ازآنجایی‌که شبکه آب‌رسانی مدفون بوده، نظارت و کنترل بسیار دشوار می‌گردد. پس از آگاهی از عوامل دینامیکی آب، نظیر فشار و دبی در داخل شبکه، شناسایی عوامل دیگری مانند ترکیدگی، شکست و نشت لوله و تجهیزات، باعث می‌شود نظارت و کنترل بهتری حاصل شود. به‌منظور مدیریت بهتر شبکه و کاهش تلفات آب، آگاهی از فراوانی نقاط شکست در شبکه (برای تعیین مکان‌هایی که احتمال شکست بیشتری دارد) امری ضروری است[1].

۲- پیشینه تحقیق

تکنیک‌های آماری یکی از قدیمی‌ترین روش‌هایی است که که برای تخمین شکست لوله در شبکه توزیع آب توسط شمیر و هاوارد در سال ۱۹۷۹ میلادی به‌کار گرفته شد؛ رابطه‌ای که این دو به‌دست آوردند نشان می‌دهد نرخ شکست لوله نسبت به سن لوله به صورت نمایی در حال افزایش است.

کوپر با استفاده از روش‌های آماری در سیستم اطلاعات جغرافیایی GIS، روابط رگرسیونی را برای تخمین شکست در خطوط لوله آب شهری پیشنهاد دادند[2]. میکاییل نیشیاما و فیلیون اکثر مدل‌های آماری مربوط به پیش‌بینی و تخمین شکست در خطوط لوله آب شهری طی سال های ۲۰۰۲ تا ۲۰۱۲ را در یک مقاله گنجاندند. تاریخچه خصوصیات آماری داده‌های شکست، از نکات اساسی مدل‌هایی است که در پژوهش‌های اخیر مورد بررسی قرار گرفته اند؛ مدل‌هایی که عمدتا در قالب دسته‌های قطعی و احتمالی و یا دسته‌های گوناگون روش‌های محاسبه نرم‌افزاری تقسیم بندی شده‌اند. کاربرد هر مدل تابعی از اطلاعات و داده‌هایی مورد نیاز آن مدل بوده، وجه تمایز و تفکیک مدل-ها متاثر از داده‌های موردنیاز آنان خواهد بود[3].

شبکه آب شهری با گذشت زمان آسیب دیده و کارایی خود را از دست می دهد. از این رو، تلاش برای حفظ و بهبود کیفیت قابل‌قبول آب و عملکرد مناسب هیدرولیک در شبکه توزیع آب شهری نیازمند بازسازی و جایگزینی قسمت های معیوب و آسیب دیده بوده و این امر اجتناب ناپذیر می نماید. عدم نوسازی و یا تعویض به‌موقع تأسیسات خطوط لوله، شبکه آب شهری را متحمل هزینه‌های سرسام‌آوری در تعمیر شکستگی‌ها نموده و شبکه را با نشتی‌های پی‌درپی مواجه می‌کند. برای نوسازی و جایگزینی قسمتی از شبکه، غالباً از مدل‌های پیش‌بینی نرخ شکست استفاده می‌شود. این مدل‌ها معمولاً با بررسی خصوصیات لوله، مکان دفن و داده‌های شکست که در زمان مختلفی رخ داده نسبت به پیش‌بینی نرخ شکست خط لوله در شبکه اقدام می‌کنند[4].

تعداد مدل‌های ارائه‌شده طی ۳۰ سال گذشته، به صورت آماری و فیزیکی بررسی و توسعه یافته ند:

- مدل‌های آماری توسعه‌یافته با توجه به تاریخچه‌ی نقاط شکست لوله‌ها، الگوی شکست را شناسایی و با درون‌یابی داده‌ها به الگوی تخمین نرخ شکست دست می‌یابند.
- مدل‌های فیزیکی با توجه به شبیه‌سازی در شبکه و دانستن مکانیسم شکست لوله، ظرفیت و مقاومت آن‌ها، شکست را پیش‌بینی می‌کنند.

کلینر و راجانا [5] بررسی جامعی از مدل‌های آماری و فیزیکی در تعین نرخ شکست لوله ارائه داده‌اند. تعدادی از روش‌های جدید بر اساس شبکه‌های عصبی مصنوعی[6] (ANN)، مدل‌های نمایی چند متغیره‌ی تعمیم‌یافته و مجموعه‌های فازی توسعه یافته‌اند و بسیاری از آن‌ها در عمل رواج پیدا کرده‌اند. این مقاله ضمن مرور و نقد مدل‌های ارائه‌شده طی سال‌های ۲۰۰۲ تا ۲۰۱۰، به بررسی چگونگی رشد و توسعه این مدل‌ها و شناسایی حوزه تحقیقات در زمینه پیش‌بینی نرخ شکست خطوط لوله و ارائه راه‌کارهای مفید برای آینده می‌پردازد.

[6] Artificial Neural Network

روش‌های آماری با توجه به کاربرد آن‌ها با سطوح مختلفی از داده‌های ورودی، برای پیش‌بینی شکست در شبکه توزیع آب مناسب‌تر هستند. تمرکز این پژوهش بر مدل‌های آماری شکست لوله است؛ مدل-هایی که خود به دو صورت آماری- قطعی و آماری – احتمالی مشخص می‌شوند. مدل‌های آماری قطعی، شکست لوله را با توجه به مناسب بودن داده‌های شکست لوله در معادلات وابسته به زمان پیش‌بینی می-کند. معادلات وابسته به زمان، تابعی خطی یا نمایی از زمان می‌باشند که شکست لوله را به‌صورت تابعی تجمعی شکست از زمان شروع اولین شکست نمایش می‌دهند. مدل اصلی نمایی زمان برای پیش‌بینی شکست، اولین بار توسط شامیر و هاوارد در سال ۱۹۷۹ پیشنهاد شد[6].

روش‌های مبتنی بر فازیِ تولیدشده، به‌صورت آنالیز فاصله‌ای است که می‌توان اطلاعات غیرقابل‌قبول و اشتباهی از آن استخراج کرد. تکنیک‌های فازی نشان می‌دهد برای مدل زوال‌پذیر زیرساخت‌هایی که اطلاعاتی اغلب نادر و در دانش بی‌دقت و مبهم دارند، مناسب می‌باشند.

مطالعات مرتبط کلاینر و همکاران[7] روشی برای مدل‌سازی زوال‌پذیری زیرساخت‌های مدفون‌شده را نشان می‌دهد که با استفاده از یک قانون ناهمگن فازی فرآیند مارکوف و تفسیر شرایط این فرآیند در یک شکست احتمالی درجه‌بندی می‌شود. یک مدل ناهمگن مارکوف برای شبیه‌سازی فرآیند قواعد مبتنی بر فازی زوال لوله استفاده می‌شود که شامل «جریان» وضعیت شرایط که از بالا به پایین قرار می‌گیرد. نتایج تولیدشده با فرایند مارکوف می‌تواند به امکان شکست تبدیل شود که با نتیجه شکست ترکیب می‌شود تا یک منحنی عمر- خطر را ارائه دهد. این منحنی می‌تواند توسط برنامه ریزان برای ارزیابی برنامه‌های بهسازی خطوط لوله باقابلیت محاسبه تحمل خطر و انتخاب بر اساس این معیارها اعمال شود.

با گسترش GIS و بهبود حسگرها در شبکه آب‌رسانی، لحظه‌به‌لحظه داده‌های حجیمی نمایش و ثبت می‌شوند. با افزوده شدن داده‌ها، تفسیر و نمایش آن‌ها بسیار مشکل شده که با استفاده از داده-کاوی این مشکل به میزان زیادی برطرف می‌گردد. خوشه‌بندی یکی از روش‌های داده‌کاوی است که نقش مهمی در این عرصه ایفا می‌کند. ازجمله استفاده گسترده از خوشه‌بندی بر روی داده‌ها می‌توان به طبقه-بندی داده‌ها در زمینه‌های مختلف مهندسی، گسسته‌سازی داده‌های مستمر، تشخیص نقطه خطا، حذف نویز از داده‌ها، تشخیص الگو و پردازش تصویر اشاره نمود.

اولین بار [8] از الگوریتم‌های خوشه‌بندی مکانی بر روی داده‌های مکانی شکست در خطوط شبکه توزیع آب با هدف تعیین نرخ شکست استفاده نمود. همچنین[9] از این الگوریتم‌ها برای مطالعه تأثیر خرابی شبکه آب‌رسانی بر اثر زلزله به‌کار برد. بااین‌حال این روش‌های خوشه-بندی مکانی چندان مثمر ثمر واقع نشدند تا اینکه [10] بر روی الگوریتم DBSCAN[7] اصلاحاتی انجام تا در خوشه‌بندی‌هایی که با داده‌های گوناگون در بعدهای مختلفی مواجه هستند، نتایج پایدارتر و مستحکمی حاصل گردد.

روش LS-SVM[8] بر اساس روش SVM توسعه و برای اولین بار توسط[11] پیشنهاد شده است. روش‌های جدید شبکه عصبی مصنوعی (ANN) و منطق فازی (FL) در چند سال اخیر برای تجزیه و تحلیل روابط پیچیده شبکه آب و شبیه‌سازی آن‌ها بسیار استفاده شده و موردقبول بوده است. بدین منظور برای ارزیابی عملکرد مدل LS-SVM از دو روش مختلف ANN مانند شبکه عصبی رگرسیون تعمیم شده (GRNN)[9] و شبکه‌های عصبی تغذیه به جلو (FFNN)[10] بر روی منطقه موردنظر استفاده‌شده و تفاوت‌ها و برتری هر یک شرح داده می‌شود .

۳- روش تحقیق

۳-۱- خوشه‌بندی

هدف اصلی از خوشه‌بندی[11] شناسایی ساختارهای مشترک در داده‌ها برای ایجاد خوشه هست. معمولاً استفاده از خوشه‌بندی فازی منجر به ارائه نتایج مطلوبی در اجرای تحلیل‌های مکانی خواهد شد. در خوشه‌بندی فازی، هر خوشه دارای یک مرکز خوشه می‌باشد که نشان‌گر عضو نمونه همان خوشه هست. برتری خوشه‌بندی فازی بر روش‌های معمول خوشه‌بندی آن است که اشیا با درجات مختلف عضویت (Membership Values) به چندین خوشه تعلق داشته دارند. درحالی‌که خوشه‌بندی‌های معمولی و سخت (Hard Clustering) هر پدیده فقط متعلق به یک خوشه است. مقدار عضویت نشان‌گر آن است که هر عضو مجموعه چقدر به مرکز خوشه نزدیک است. اگر مقدار (درجه) عضویت یک شی به یک خوشه بالا باشد (مثلاً ۹۵٪) بدان معنی است که شی مذکور تقریباً منطبق بر مرکز خوشه بوده و در نتیجه آن خوشه قرار می‌گیرد (برچسب آن خوشه به آن داده تعلق می‌گیرد). از طرف دیگر، اگر درجه‌ی عضویت کم باشد (مثلاً ۵٪) شیء موردنظر شباهت کمی به خوشه داشته و دارای فاصله زیادی تا مرکز آن است.

اساس کار اکثر الگوریتم‌های خوشه‌بندی بدین‌صورت است که در گام اول نماینده‌ای اولیه انتخاب کرده، سپس با توجه به تشابه اعضا با نماینده انتخابی، خوشه‌ای که هر شیء به آن تعلق دارد مشخص می‌گردد. در مرحله بعد نماینده‌های جدید هر خوشه (مراکز جدید خوشه) انتخاب‌شده و نمونه‌ها دوباره خوشه‌بندی می‌شوند. این فرآیند تا زمانی که نماینده جدید با نماینده مرحله قبل تفاوت نداشته باشد، ادامه می‌یابد. هدف از انجام هر خوشه‌بندی شناسایی گروه‌ها (خوشه‌ها) مجزا از یکدیگر در داده‌هاست. برای انجام خوشه‌بندی الگوریتم‌های متفاوتی وجود دارند که بر اساس نوع نیاز و کاربرد انتخاب می‌گردند.

[7] Density Based Spatial Clustering of Applications with Noise
[8] Least squares support vector machine
[9] Generalized Regression Neural Networks
[10] Feed Forward Neural Network
[11] Clustering

۳-۲- خوشه‌بندی فازی

همان‌طور که گفته شد در خوشه‌بندی سخت، داده‌ها به خوشه‌های کاملاً مجزایی از یکدیگر تقسیم می‌شوند، اما در خوشه‌بندی فازی (خوشه‌بندی نرم) هر شیء می‌تواند به بیش از یک خوشه تعلق داشته باشد. لذا به هرکدام از داده‌ها یک درجه عضویت اختصاص می‌یابد. در واقع درجه عضویت نشان‌دهنده میزان ارتباط داده‌ها با خوشه مشخص شده است. فرایند خوشه‌بندی فازی به مفهوم یافتن درجات عضویت مذکور و سپس استفاده از آن‌ها برای تخصیص هر داده به یک یا چند خوشه می‌باشد. یکی از معروف‌ترین الگوریتم‌های خوشه‌بندی فازی (Fuzzy Clustering) یا FCM (Fuzzy C-means) است، در FCM تعداد خوشه‌ها (C) از قبل مشخص بوده و تابع هدف آن به صورت رابطه (۱) می‌باشد.

$$j = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m}\, d_{ik}^{2} = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m}\, ||x_k - v_i||^2 \qquad (۱)$$

در این رابطه m (یا فازی‌سازی) عدد حقیقی بزرگ‌تر از ۱ می‌باشد که در اکثر مواقع مقدار آن معادل با ۲ انتخاب می‌شود. در این رابطه تعداد داده‌ها n و تعداد خوشه‌ها c در نظر گرفته شده‌اند. همچنین پارامتر x_k نشان‌دهنده نمونه k ام از میان n نمونه است. v_i نشان دهنده نماینده خوشه i ام از میان c خوشه می‌باشد. علاوه بر این u بیانگر درجه عضویت و u_{ik} میزان تعلق نمونه k ام را به خوشه i ام می‌باشند.

در رابطه (۱) علامت || || نمایانگر فاصله (میزان تشابه) هر نمونه با هر یک از خوشه‌هاست. لازم به ذکر است که برای محاسبه فاصله از تابع نرم می‌توان استفاده نمود. به کمک همه مقادیر u_{ik} یک ماتریس با ابعاد $c \times n$ را می‌توان ایجاد نمود. که هر مؤلفه آن مقداری بین ۰ و ۱ می‌باشد. در این ماتریس مجموع مؤلفه‌های هر ستون برابر با عدد ۱ است.

$$j = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m}\, d_{ik}^{2} = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m}\, ||x_k - v_i||^2 \qquad (۲)$$

$$\sum_{i=1}^{c} u_{ik} = 1, \forall k = 1, \dots, n \qquad (۳)$$

منظور از رابطه (۳) آن است که مجموع تعلقات هر نمونه به تمام خوشه‌ها بایستی معادل با ۱ باشد. لذا با کمک شرط بالا و کمینه نمودن تابع هدف روابط زیر به دست خواهد آمد.

$$u_{ik} = \frac{1}{\sum_{j=1}^{c}\left(\frac{d_{ik}}{d_{jk}}\right)^{\frac{2}{(m-1)}}}$$

و

$$v_i = \frac{\sum_{k=1}^{n} u_{ik}^{m}\, x_k}{\sum_{k=1}^{n} u_{ik}^{m}}$$

در روابط مذکور، u_{ik} تعلقات نهایی هر داده به هر خوشه را نشان می‌دهد. همچنین v_i نشان‌دهنده مرکز و نماینده هریک از خوشه‌ها می‌باشند.

مراحل انجام الگوریتم FCM به شرح زیر است:

۱- مقداردهی اولیه پارامترهای c,m,u° به‌صورت کاملاً حدسی و تصادفی.

۲- محاسبه مرکز خوشه‌ها (v_iها).

۳- تعیین ماتریس تعلقات از روی خوشه‌های تولید شده.

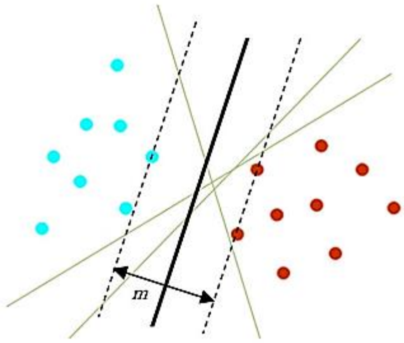
۴- با شرط $||v_{i+1} - v_i|| \le \varepsilon$ الگوریتم به پایان رسیده (در اینجا منظور از v_i مرکز به‌دست‌آمده برای خوشه در تکرار i ام می‌باشد و v_{i+1} مرکز خوشه در تکرار بعدی است) در غیر این صورت از مرحله ۲ الگوریتم تکرار خواهد شد. در اینجا منظور از ε یک مقدار انتخابی برای توقف کار یا همان دقت مورد نظر می‌باشد.

الگوریتم FCM همواره همگرا بوده و به‌صورت نظارت‌نشده (Unsupervised) کار می‌کند؛ اما در مقابل، محاسبات طولانی و پرحجم دارای ضعف بوده و به نویز در داده‌ها حساس می‌باشد. الگوریتم مذکور به حدس‌های اولیه نیز وابسته می‌باشد.

۳-۳-ماشین بردار پشتیبان SVM

روش SVM توسط Vapnik در سال ۱۹۹۵ ارائه گردید و خیلی زود توانست به یکی از ابزارهای قدرتمند در طبقه‌بندی داده‌ها تبدیل شود. این روش برای جداسازی داده‌های دو کلاس از یکدیگر ابداع شد. برای حالت‌های چند کلاسی می‌توان ترکیبی از چند SVM دوکلاسی را استفاده کرد که در ادامه بیشتر مورد بحث قرار خواهد گرفت. از ویژگی‌های مهم این روش می‌توان به پایه‌ی قوی ریاضی آن اشاره کرد. در این روش طبقه‌بند در نتیجه حل یک مسأله‌ی بهینه‌سازی با تابع هدف محدب با تعدادی قید به‌دست می‌آید. تعداد این قیود مسأله، به اندازه‌ی تعداد نمونه‌های یادگیری است. نتیجه‌ی حل این مسأله، یافتن ابرسطح جداکننده‌ی بهینه با حداکثر حاشیه[12] است؛ یعنی تا آنجا که ممکن است طبقه‌بند دور از نقاط داده‌ای هر دو کلاس باشد. همان‌طور که در شکل ۱ مشخص است، SVM به دنبال یافتن بهترین جداکننده است.

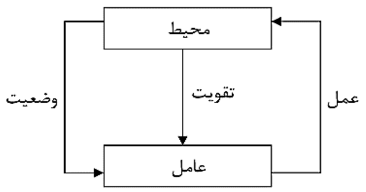
12.Margin



شکل ۱: سطح جداکننده‌ی بهینه با ماکزیمم حاشیه m

ماشین بردار پشتیبان یکی از روش‌های یادگیری بانظارت[13] است. همان‌طور که گفته شد یادگیری ماشین[14]، یکی از زیرشاخه‌های هوش مصنوعی است که به سه دسته تقسیم می‌شود:

۱. یادگیری بدون نظارت که عبارتست از ساختن مدل‌های یادگیری برای کشف یک نظم و سازمان مشخص در داده‌ها؛ برای مثال می‌توان از خوشه‌بندی داده‌ها به‌عنوان نمونه‌ای از شیوه‌ی یادگیری نام برد.
۲. یادگیری با نظارت که می‌توان آن را ساخت مدل-های یادگیری به کمک داده‌های پیشین تعریف کرد. (مانند طبقه‌بندی داده‌ها)
۳. یادگیری با تقویت[15] که عبارت است از ساخت مدل‌های یادگیری با استفاده از بازخورد مدل حال حاضر (مانند یادگیری ربوتیک). شکل ۲بیان‌گر روش یادگیری با تقویت است.



شکل ۲: یادگیری با تقویت

SVM در ابتدا برای تشخیص الگو توسعه یافت و سپس سعی شد از این سامانه به‌منظور رگرسیون استفاده شود. در شکل رگرسیون ماشین بردار پشتیبان[16] (SVR) الگوی ورودی x با استفاده از یک تابع انتقال خطی به فضای مشخصه با ابعاد بالا نگاشت داده می‌شود و درنتیجه یک رگرسیون خطی در این فضا ایجاد خواهد شد.

$$f(x) = \sum_{i=1}^{D} w_i \phi_i(x) + b \qquad (۵)$$

که در آن $\{\phi_i(x)\}_{i=1}^{D}$ مشخصه نامیده شده و b ، $\{w_i\}_{i=1}^{D}$ ضرایبی هستند که با استفاده از داده‌های موجود تعیین می‌شوند. بنابراین رگرسیون غیرخطی در فضای ورودی با ابعاد پایین به یک رگرسیون خطی در فضای مشخصه با ابعاد بالا تبدیل می‌شود. ضرایب $\{w_i\}_{i=1}^{D}$ را می‌توان با کمینه کردن تابع ذیل تعیین کرد.

$$R[w] = \frac{1}{N}\sum_{i=1}^{N} |f(x_i) - y_i|_\varepsilon + \lambda \|w\|^2 \qquad (۶)$$

که در آن λ ثابت هموارسازی و$|f(x_i) - y_i|_\varepsilon$ تابع خطای $\varepsilon -$ $sensitive$ بوده و به‌صورت زیر تعریف می‌شود.

$$|f(x_i) - y_i|_\varepsilon = \begin{cases} |f(x_i) - y_i|_\varepsilon - \varepsilon & for\ |f(x_i) - y_i|_\varepsilon \geq \epsilon \\ 0 & otherwise \end{cases} \qquad (۷)$$

می‌توان نشان داد تابع تخمین به‌صورت زیر خواهد بود.

$$f(x, \alpha, \alpha^*) = \sum_{i=1}^{N} (\alpha_i, \alpha_i^*) k(x_i, x) + b \qquad (۸)$$

که در آن $(\alpha_i, \alpha_i^* = 0, i = 1,2, \dots, N)$ بوده و $k(x_i, x)$ تابع کرنل است و بیان‌کننده ضرب داخلی در فضای مشخصه D بعدی است.

$$k(x, y) = \sum_{j=1}^{D} \phi_j(x) \phi_j(y) \qquad (۹)$$

۴-۳- بهینه سازی گروه ذرات PSO [17]

PSO از دسته الگوریتم های بهینه سازی است که بر مبنای تولید تصادفی جمعیت اولیه عمل می کنند. در این الگوریتم با الگوگیری و شبیه سازی رفتار پرواز دسته جمعی (گروهی) پرندگان یا حرکت دسته جمعی (گروهی) ماهی ها بنا نهاده شده است. هر عضو در این گروه توسط بردار سرعت و بردار موقعیت در فضای جستجو تعریف می گردد. در هر تکرار زمانی، موقعیت جدید ذرات با توجه به برار سرعت و بردار موقعیت در فضای جستجو تعریف می گردد. در هر تکرار زمانی، موقعیت جدید ذرات با توجه به بردار سرعت فعلی، بهترین موقعیت یافت شده توسط آن ذره و بهترین موقعیت یافت شده توسط بهترین ذره موجود در گروه، به روز رسانی می گردد.

با فرض کنید یک فضای جستجوی d بعدی داریم. i اُمین ذره در این فضای d بعدی با بردار موقعیت X_i،بردار سرعت با V_i ، بهترین موقعیتی که ذره i اُم پیدا کرده با $P_{i\ best}$ و بهترین موقعیتی که بهترین ذره در بین کل ذرات پیدا کرده است با $P_{g\ best}$ به شکل زیر توصیف می گردد.

$$X_i = (x_{i_1}, x_{i_2}, x_{i_3}, \dots, x_{i_d}) \qquad (۱۰)$$

[13] Supervised Learning
[14].Machine Learning
[15].Reinforcement Learning
[16].Support Vector Regression
[17] . Particle Swarm Optimization

$$V_i = (v_{i_1}, v_{i_2}, v_{i_3}, \dots, v_{i_d}) \quad (۱۱)$$

$$P_{i\ .best} = (p_{i_1}, p_{i_2}, p_{i_3}, \dots, p_{i_d}) \quad (۱۲)$$

$$P_{g\ .best} = (p_{g_1}, p_{g_2}, p_{g_3}, \dots, p_{g_d}) \quad (۱۳)$$

همچنین برای به روز رسانی محل هر کدام از ذرات از رابطه زیر استفاده می کنیم.

$$V_i(t) = w \times V_i(t-1) + c_1 \times rand_1 \times (P_{i,best} - X_i(t-1)) + c_2 \times rand_2 \times (P_{g,best} - X_i(t-1)) \quad (۱۴)$$

$$X_i = X_i\ (t-1) + V_i(t) \quad (۱۵)$$

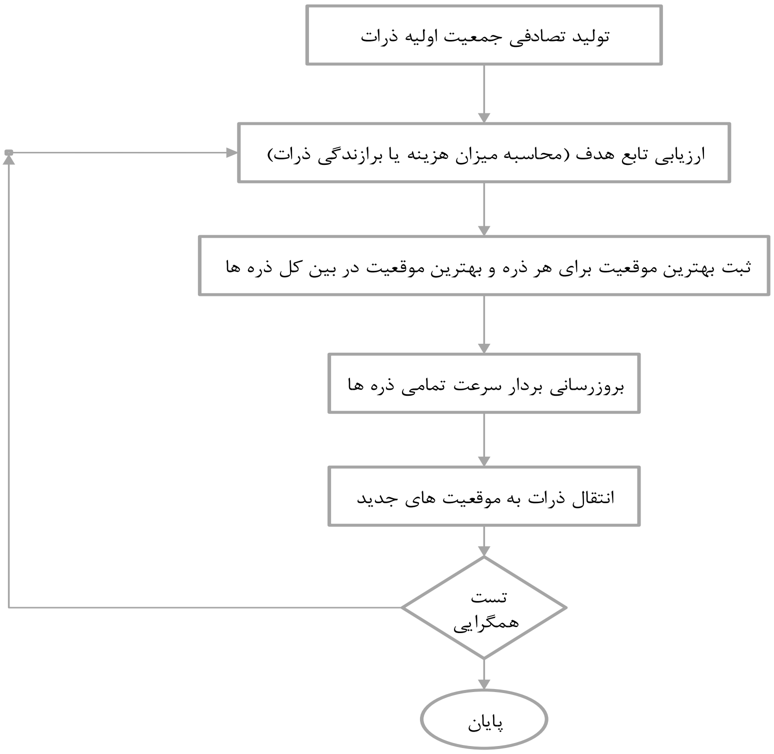
که در روابط فوق W ضریب وزنی اینرسی (حرکت در مسیر خودی) که نشان دهنده میزان تأثیر بردار سرعت تکرار قبل $(V_i(t))$ بر روی شکل۳ نشان داده شده است که شامل ۵ مرحله بوده و در ادامه تشریح می‌شود.

بردار سرعت در تکرار فعلی $(V_i(t+1))$ است، c_1 ضریب ثابت آموزش (حرکت در مسیر بهترین مقدار ذره مورد بررسی)، c_2 ضریب ثابت آموزش (حرکت در مسیر بهترین ذره یافت شده در بین کل جمعیت)، $rand_1, rand_2$ دو عدد تصادفی با توزیع یکنواخت در بازه ۰ تا ۱ ،$V_i\ (t-1)$ بردار سرعت در تکرار $(t-1)$ ام و$X_i\ (t-1)$ بردار موقعیت در تکرار $(t-1)$ ام می‌باشد.

برای جلوگیری از افزایش بیش از حد سرعت حرکت یک ذره در حرکت از یک محل به محل دیگر (واگرا شدن بردار سرعت)، تغییرات سرعت را به رنج V_{max} تا V_{min} محدود می کنیم؛ یعنی $V_{min} \leq V \leq V_{max}$. حد بالا و پایین سرعت با توجه به نوع مسئله تعیین می گردد.

۱-٤-۳- مراحل اجرای الگوریتم PSO

دیاگرام گردشی (فلوچارت) الگوریتم PSO در



شکل ۳: دیاگرام گردشی الگوریتم PSO.

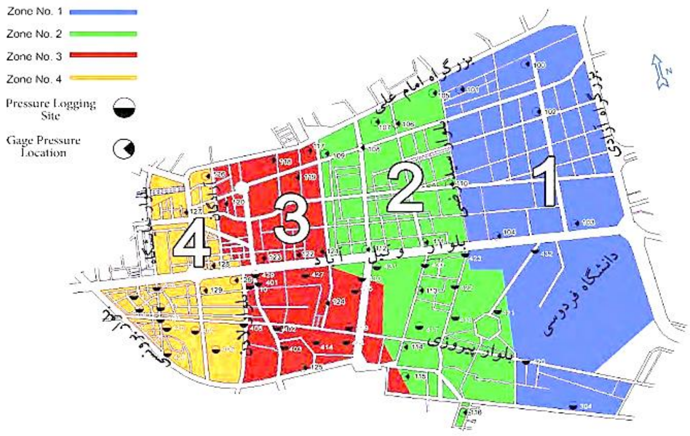
٤-۱- معرفی منطقه مطالعاتی

شهر مشهد به‌خاطر شرایط استراتژیکی و فرهنگی- مذهبی خود، نسبت به شهرهای دیگر ایران رشد بیشتری در زمینه‌ی ساخت‌وساز و افزایش جمعیت دارد و پیش‌بینی می‌شود در آینده بحران کمبود آب در این شهر معضلی بزرگ باشد. ازاین‌رو شهر مشهد به‌عنوان منطقه مورد مطالعه در این تحقیق در نظر گرفته شده است. مشهد با ۳۲۸ کیلومترمربع مساحت، دومین شهر پهناور ایران بعد از تهران است. بر اساس سرشماری عمومی نفوس و مسکن سال ۱۳۹۵، این شهر با ۳،۰۰۱،۱۸۴ نفر جمعیت، دومین شهر پرجمعیت ایران پس از تهران است. این شهر دارای یک‌میلیون و صد فقره انشعاب آب و ۴۱۳۳ کیلومتر طول کل شبکه و خطوط انتقال آب شرب است. ازآنجایی‌که داده‌های مربوط به

شبکه آب کل شهر، بسیار حجیم و محاسبات آن‌ها بسیار پیچیده و وقت‌گیر است، به سبب سهولت کار و انجام پذیر شدن تحقیق، یک ناحیه از شهر انتخاب شد؛ همان طور که

در این تحقیق ناحیه چهار آبفای مشهد، به‌عنوان منطقه مورد مطالعه در نظر گرفته شد. حداقل و حداکثر تراز ارتفاعی این ناحیه از سطح دریا به میلی‌متر است. نصب و اجرای لوله‌های شبکه در این ناحیه عموماً از سال ۱۳۶۰ شروع شده است. بر اساس آمار ثبت شده، این ناحیه دارای بیشترین نرخ شکست بوده که اکثریت آن هم مربوط به لوله‌های آزبست سیمانی بوده اما برای لوله‌های آهنی و چدنی اطلاعات کافی جهت ترتیب ۱۰۲۱ و ۱۲۴۱ متر بوده و دارای مساحتی بالغ بر ۲۴۱۸ هکتار و حدود ۹۳۷۱۹ فقره انشعاب آب می‌باشد. طول کل لوله‌های شبکه توزیع منطقه بالغ بر ۵۷۹۸۶ متر و شامل لوله‌های آهنی با قطرهای ۶۰۰، ۷۰۰ و ۸۰۰ میلی‌متر و لوله‌های آزبست و چدن با قطرهای ۸۰، ۱۰۰، ۱۵۰، ۲۰۰، ۲۵۰، ۳۰۰ و ۴۰۰

ساخت مدل پیش‌بینی نرخ شکست وجود ندارد. لذا صرفاً اطلاعات مربوط به لوله‌های آزبست سیمان ناحیه مورد استفاده قرار گرفته است. نیز نمونه‌ای از داده‌های به کار گرفته شده در این پژوهش بیان شده-است.



شکل ۴:محدوده مطالعاتی پژوهش.

جدول ۱: نمونه ای از داده‌های اطلاعات جمع آوری شده.

جنس لوله	طول (متر)	قطر (میلی متر)	سن (سال)	تعداد شکست	نرخ شکست تعداد شکست در سال در واحد طول (کیلومتر)
ACC	198/3	80	49	9	**4/54**
ACC	114/8	200	8	5	**4/36**
ACC	202/2	200	29	5	**2/47**
ACC	49/2	80	49	4	**8/13**
DCI	82/5	900	7	3	**3/63**
PE	97/8	110	8	3	**3/07**
ACC	123/9	80	39	3	**2/42**
ST	28/1	1000	8	3	**10/67**
PE	184/6	110	3	2	**1/08**

در این بخش ابتدا مدل‌های پیش‌بینی ماشین بردار پشتیبان (LS-SVM) بر روی کل داده‌ها اجرا شد و شاخص های ارزیابی استخراج گردید. سپس کل داده‌ها را با خوشه بندی مبتنی بر فازی (FCM) و خوشه‌بندی ترکیبی الگوریتم K-means با بهینه یاب PSO به چندین خوشه مختلف تقسیم کرده و در نهایت مدل‌های پیش بینی برای هر گروه اجرا شده و شاخص‌های ارزیابی هر خوشه به طور جداگانه محاسبه گردید.

پیاده سازی‌ها و آزمایشات مختلف با نرم‌افزار Matlab 2016b انجام گرفت. سیستم مورد استفاده برای اجرای آزمایشات، سیستمی با حافظه رم ۸ گیگا بایت و پردازنده corei7 بود. برای ارزیابی مدل های تخمین، از سه معیار بهره‌وری (E)، ضریب همبستگی (R) و خطای جذر میانگین مربعات (RMSE) استفاده شد که به صورت زیر تعریف می‌شوند.

$$E = \frac{E_1 - E_2}{E_1} \qquad (۱۶)$$

$$E_1 = \sum_{i=1}^{N} (FR_{observed} - \overline{FR_{observed}})^2 \qquad (۱۷)$$

$$E_2 = \sum_{i=1}^{N} (FR_{observed} - FR_{estimate})^2 \qquad (۱۸)$$

$$RSME = \sqrt{\frac{1}{N} FR_{observed} - FR_{estimate}} \qquad (۱۹)$$

$$R = \frac{1}{N-1} \sum_{i=1}^{N} \left(\frac{FR_{observed} - \mu_{FR_{observed}}}{\sigma_{FR_{observed}}} \right) \left(\frac{FR_{estimate} - \mu_{FR_{estimate}}}{\sigma_{FR_{estimate}}} \right) \qquad (۲۰)$$

در این رابطه‌ها، $FR_{estimate}$ نرخ شکست برآورد شده، $FR_{observed}$ نرخ شکست مشاهده شده و $\overline{FR_{observed}}$ میانگین نرخ شکست های مشاهده شده خواهند بود. $\mu_{FR_{observed}}$ و $\mu_{FR_{estimate}}$ به ترتیب میانگین داده های $FR_{observed}$ و $FR_{estimate}$ هستند. همچنین $\sigma_{FR_{observed}}$ و $\sigma_{FR_{estimate}}$ به ترتیب انحراف معیار داده های $FR_{observed}$ و $FR_{estimate}$ تعریف شدند. ضریب همبستگی معیار آماری متداولی است که ارتباط نرخ شکست های تخمین زده شده و پیش بینی شده را بر روی یک رابطه خطی نشان می دهد. بهره وری E، معیاری برای ارزیابی عملکرد مدل ایجاد شده است. مقادیر E و R هرچه به ۱ نزدیک تر باشند، نشان دهنده عملکرد خوب مدل ایجاد شده، خواهند بود.

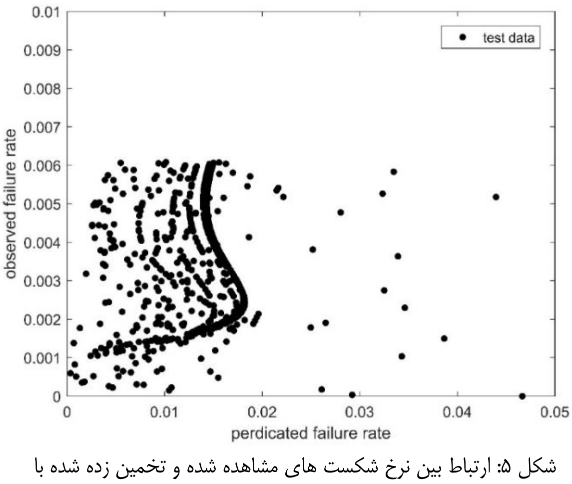
جدول ۲ ارائه شده است و شکل ۵ ارتباط بین نرخ شکست های پیش بینی شده و مشاهده را نشان می دهد

جدول ۲: نتایج مربوط به پارامترهای ارزیابی مربوط به روش LS-SVM

خطای جذر میانگین مربعات (RMSE)	ضریب همبستگی (R)	بهره وری (E)
۰,۰۱۴۷	۰,۴۱۲۶	۰,۴۴۵۱

۴-۲- تخمین نرخ شکست با استفاده از LS-SVM

برای تخمین نرخ شکست با استفاده از LS-SVM، دو آزمایش متفاوت در نظر گرفته شدند. در ابتدا مدل تخمین نرخ شکست با کل ۴۲۵۱ داده در نظر گرفته شدند (زمانی که خوشه بندی هنوز انجام نشده است). برای این آزمایش، از ۸۰ درصد داده ها به عنوان داده های آموزشی (۳۴۰۱ داده) برای آموزش مدل LS-SVM و از ۲۰ درصد باقی مانده داده های به عنوان داده های تست (۸۵۰ داده) برای تست مدل استفاده شد. نتایج مربوط به پارامترهای ارزیابی در



شکل ۵: ارتباط بین نرخ شکست های مشاهده شده و تخمین زده شده با استفاده از مدل *LS-SVM*

۴-۳- تخمین نرخ شکست با استفاده از LS-SVM و تلفیق با خوشه‌بندی ترکیبی K-means با PSO

داده ها به روش خوشه بندی مبتنی بر الگوریتم ازدحام ذرات (PSO) به زیر گروه هایی(که از لحاظ ویژگی بسیار شبیه به هم

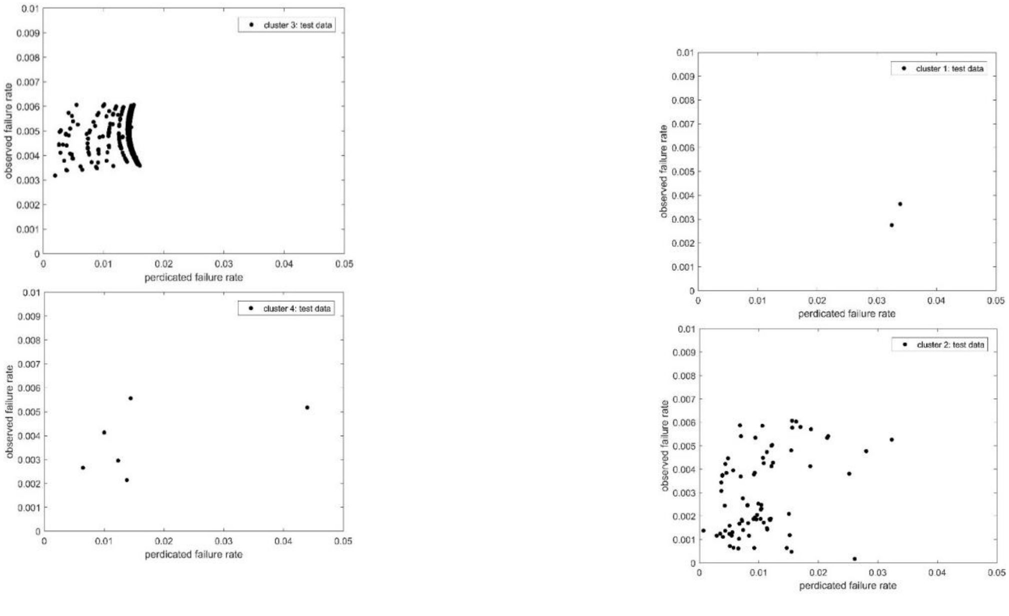
هستند) تقسیم شوند تا عملکرد مدل LS-SVM بهبود یافته و درنهایت نتایج حساس و قابل اطمینان به دست آید. مدل های تخمین نرخ شکست برای هر یک از ۸ زیرمجموعه (مجموعه C1-C8) که توسط روش خوشه بندی مبتنی بر الگوریتم ازدحام ذرات تعریف شده و مورد بررسی قرار گرفت. ۸۰ درصد داده ها برای آموزش مدل های تخمین نرخ شکست (در هر مجموعه داده) و ۲۰٪ باقیمانده برای آزمایش مدل ها مورد استفاده قرار گرفت. داده‌های مورد استفاده در آزمایش هر مدل به صورت تصادفی (از میان مجموعه داده‌ها) انتخاب شده و همچنین شایان ذکر است که این داده‌ها در آموزش مدل‌ها مورد استفاده قرار نگرفته‌اند. پس از انجام آزمایشات، مقدار نرخ شکست‌های پیش‌بینی بدست آمد که مقایسه این مقادیر با مقادیر نرخ شکست مشاهده شده در شکل ۶ نشان داده شده است. در

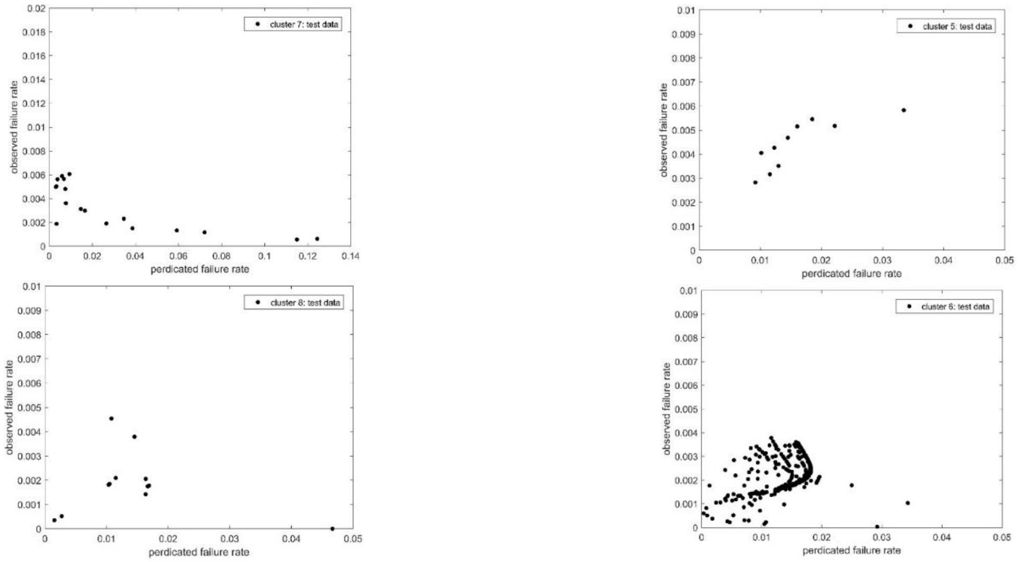
جدول ۳ (نتایج با اعمال خوشه بندی) مشاهده می‌شود؛ بهره وری ، ضریب همبستگی و خطای جذر میانگین مربعات نسبت به

جدول ۲ (نتایج بدون اعمال خوشه بندی) در تخمین نرخ شکست بهبود یافته است.

جدول ۳: نتایج مربوط به پارامترهای ارزیابی در مدل LS-SVMو خوشه بندی مبتنی بر PSO

شماره خوشه	بهره وری (E)	ضریب همبستگی (R)	خطای جذر میانگین مربعات (RMSE)
C1	۰,۴۵۱۶	۰,۴۸۲۶	۰,۰۱۶۹
C2	۰.۶۱۴۹	۰.۶۳۴۸	۰,۰۰۶۴
C3	۰,۵۹۸۳	۰,۵۵۹۸	۰,۰۰۹۸
C4	۰.۶۴۵۸	۰.۶۱۳۴	۰,۰۱۲۳
C5	۰.۶۳۸۷	۰.۶۴۸۱	۰,۰۱۷۸
C6	۰,۷۱۳۹	۰,۷۲۵۴	۰,۰۰۸۷
C7	۰,۷۵۹۸	۰,۷۳۶۲	۰,۰۰۵۸
C8	۰.۶۱۹۴	۰.۶۲۹۴	۰,۰۱۲۵





شکل ۶: ارتباط بین نرخ شکست‌های مشاهده شده و تخمین‌زده شده در هر خوشه با خوشه بند مبتنی بر pso

همچنین،

جدول ۴ نشان دهنده زمان اجرای مدل تخمین نرخ شکست برای دو حالت است؛ در حالت اول، زمان اجرای کل داده ها و در حالت دوم زمان اجرای هر خوشه از داده ها (خوشه بندی با استفاده از الگوریتم ازدحام ذرات) لحاظ شده است. نتایج جدول نشان می‌دهد اعمال خوشه بندی باعث کاهش زمان اجرای تخمین نرخ شکست نیز می گردد. در نهایت نتایج پیش بینی با اعمال خوشه بندی و بدون اعمال خوشه بندی در جدول ۵ نشان می دهد که پیش بینی با اعمال خوشه بندی نتایج بهتری حاصل می شود.

جدول ۴: نتایج مربوط به زمان اجرای روش LS-SVM با اعمال خوشه‌بندی PSO و بدون آن.

روش تخمین	زمان اجرا (ثانیه)	
Ls-svm بدون اعمال خوشه بندی	۳۹۱۲٫۶۵۹۵	
Ls-svm با اعمال خوشه بندی	خوشه بندی مبتنی بر pso	۱۰۶٫۳۶۶۸
	C1	۵٫۶۵۵۴
	C2	۱۵۱٫۲۶۵۴
	C3	۹۱۲٫۶۵۴۱
	C4	۵۴٫۹۴۵۲
	C5	۶۸٫۹۴۸۵
	C6	۱۰۱۶٫۸۴۱۶
	C7	۷۲٫۴۶۵۳
	C8	۵۲٫۱۹۴۵
	جمع	۲۴۴۱٫۳۳۶۸

جدول ۵ :مقایسه مدل ها

زمان اجرا (ثانیه)	بهره وری (E)	ضریب همبستگی (R)	خطای جذر میانگین مربعات (RMSE)	مدل	رتبه
۲۴۴۱	۰,۷۰۰۵	۰,۶۹۹۰	۰,۰۰۹۱	LS-SVM+ PSO	۱
۳۹۱۳	۰,۴۴۵۱	۰,۴۱۲۶	۰,۰۱۴۷	LS-SVM	۲

۵- نتیجه‌گیری

سیستم زیرساخت شهری از مهم‌ترین عناصر ساختاری شهر به‌حساب می‌آید. لذا بررسی وضعیت موجود و نظارت بر این زیرساخت‌ها اهمیت بالایی دارد. آگاهی از فراوانی نقاط شکست در شبکه (برای تعیین مکان‌هایی که احتمال شکست بیشتری دارد) و تعیین نرخ شکست، برای مدیریت بهتر شبکه و کاهش تلفات آب امری حیاتی است. امروزه کارشناسان با مدل‌های نوینی از حوادث و اتفاقاتی که در گذشته رخ داده است، مدلسازی کرده که باعث می‌شود با پیش‌بینی احتمال حوادث در آینده، مدیریت و نظارت بهبود یابد. سالانه شرکت آب و فاضلاب مبلغ قابل توجهی را صرف تعویض و نوسازی خطوط لوله می‌کند که مکان‌ها و لوله‌های انتخاب شده برای تعویض دقت مناسبی ندارند و درست کارشناسی نمی شوند. با استفاده از این مدل‌ها و تخمینی که درمورد نرخ شکست لوله‌ها می‌شود، می‌توان با دقت بیشتری لوله‌هایی که نیاز به تعویض بوده انتخاب شوند. همچنین برای مناطق تازه احداث نیز با درنظر گرفتن نرخ شکست و هزینه‌های بهره‌برداری، می‌توان جنس لوله مناسبی برای استفاده در شبکه پیشنهاد داد.

پیشنهاد می‌گردد در آینده با گسترش مدل‌ها، عوامل دیگری نظیر رطوبت خاک، دمای آب آشامیدنی، دمای آب‌وهوا، نوع پوشش خطوط، عمق لوله مدفون شده، فشار آب و مواردی از این دست نیز مورد بررسی قرار گیرند.

مراجع

[1] Aydogdu, M. & Firat, M., 2015. Estimation of Failure Rate in Water Distribution Network Using Fuzzy Clustering and LS-SVM Methods. *Water Resources Management*, 29(5), pp.1575–1590.

[2] Cooper, N.R. et al., 2000. The use of GIS to develop a probability-based trunk mains burst risk model. *Urban Water*, 2(2), pp.97–103.

[3] Nishiyama, M. & Filion, Y., 2013. Review of statistical water main break prediction models. *Canadian Journal of Civil Engineering*, 40(10), pp.972–979.

[4] Kleiner, Y. & Rajani, B., 2001. Comprehensive review of structural deterioration of water mains: statistical models. *Urban water*, 3(3), pp.131–150.

[5] Kleiner, Y., Nafi, A. & Rajani, B., 2010. Planning renewal of water mains while considering deterioration, economies of scale and adjacent infrastructure. *Water Science and Technology: Water Supply*, 10(6), pp.897–906.

[6] Shamir, U. & Howard, C.D.D., 1979. to Scheduling Pipe Replacement.

[7] Kleiner, Y., Rajani, B. & Sadiq, R., 2006. Failure risk management of buried infrastructure using fuzzy-based techniques. *Journal of Water Supply: Research and Technology-AQUA*, 55(2), pp.81–94.

[8] Oliveira, D., Garrett Jr, J.H. & Soibelman, L., 2009. Spatial clustering analysis of water main break events. In *Proc., ASCE Int. Workshop on Computing in Civil Engineering*.

[9] Toprak, S. et al., 2009. Pipeline damage assessment using cluster analysis. *Lifeline Earthquake Engineering in a Multihazard Environment*.

[10] Tran, T.N., Drab, K. & Daszykowski, M., 2013. Revised DBSCAN algorithm to cluster data with dense adjacent clusters. *Chemometrics and Intelligent Laboratory Systems*, 120, pp.92–96.

[11] Suykens, J.A.K. & Vandewalle, J., 1999. Least squares support vector machine classifiers. *Neural processing letters*, 9(3), pp.293–300.