

Research Paper

Identification of leakage area in large water distribution systems (Case study: Mahan city in Kerman province)

Yasaman-Sadat Hashemi-Poor¹, Gholam-Abbas Barani², Ehsan Fadaei-Kermani³

¹ M.Sc. Graduate of Water and Hydraulic Structures, Department of Civil Engineering, Faculty of Engineering, Shahid Bahonar University, Kerman, Iran, yas.hp@eng.uk.ac.ir.

² Professor, Department of Civil Engineering, Faculty of Engineering, Shahid Bahonar University, Kerman, Iran, gab@uk.ac.ir. (Corresponding author)

³ Assistant Professor, Department of Civil Engineering, Faculty of Engineering, Shahid Bahonar University, Kerman, Iran, e.fadaei@uk.ac.ir.



[10.22125/IWE.2023.173289](https://doi.org/10.22125/IWE.2023.173289)

Received:
December 8, 2021
Accepted:
February 26, 2022
Available online:
June 25, 2023

Keywords:
**Leakage area
identification, Water
distribution network,
Mahan city, Epanet,
Matlab.**

Abstract

Today, leak detection is an important issue in water distribution networks, because leakage increases costs and causes a waste of water resources. In this study, a new method was used to identify the leak area in the water distribution network at part of Mahan city. First, the network was divided into a number of areas by the K-means clustering algorithm. Then, training samples related to each area were made using random leakage simulation in the network hydraulic model. The number of each area was used as the classification label of the multi-class support vector machine and along with the training samples, the classification model was taught. Finally, the trained model was used as a leak area identification model and was applied to determine possible leak areas with the observed field samples. The results show that out of 10 structures used to build the model, only three structures include; "Three areas and radial basis kernel function", "Three areas and linear kernel function" and "Five areas and linear kernel function", respectively, make an acceptable rate of 99.27%, 99.18% and 98.99% for the "Accuracy" evaluation index. On the other hand, among these three structures, dividing the network into five leak areas and using the linear kernel function divides the network into more areas and makes leak detection easier in restricted areas. As a result, this structure was used to build the final model.

1. Introduction

Over the past decades, various methods have been proposed to identify the location of the leak. Because leaks are usually considered in nodes of a network, the goal of leak location approaches is also to find leak nodes in the network. Leak detection in large and complex water distribution systems, especially in large systems without district meter area (DMA), is associated with many challenges. The idea of locating a leak node in a predetermined area allows operators to limit the search for a leak node in an area and to find a leak node in the area with other leak detection methods. Hence, Zhang et al. (2016) proposed a new method for limiting the leakage area in large water distribution systems. First, a large water network in China was divided into a number of areas by K-means clustering. Then, using pressure monitoring points data and classification of multi-class support vector machines, they identified the area of possible leakage. This method limits the potential leak area before other methods can detect a leak. As a result, in this study, similar to the method presented by Zhang et al. (2016), the water distribution network of a part of Mahan city by K-means clustering is divided into clusters with

similar nodes in response to leakage, and then to minimize the search space, the leak area is identified by the classification algorithm.

2. Materials and Methods

Pressure differences of monitoring points due to artificial leakage created in each of the network nodes were entered as input data into the K-means clustering algorithm and the studied water distribution network was divided into three, four, five, six and seven leak areas, respectively. For each type of leakage area classification, 3725 samples of artificial leakage were generated based on simulation in hydraulic model and after pre-processing the samples, a number of samples that were not a good representative of the actual leak in the network were removed. Then 70% of the remaining samples were assigned to training samples and the rest to validation samples. Training samples were used to teach the model of multi-class support vector machines and validation samples were used to determine the type of kernel function, among linear and radial basis functions and to determine the number of leakage areas, according to the classification results of multi-class support vector machines. Then, the final classification model was taught by the selected structure and all the pre-processed samples for that structure (training samples and validation samples). Finally, the final classification model was tested by pressure differences at monitoring points during actual leakage.

3. Results

In this study, in order to determine the optimal structure for the leak area identification model, first 10 models with 10 different structures have been created and the optimal parameters (c, γ) have been identified for each structure. Table 1 shows the value of the optimized parameters in different structures. The evaluation index of multi-class classification models is "accuracy", which is derived from the "confusion matrix" and is determined for all 10 models. Figure 1 shows the "accuracy" graph for trained models.

Table 1. Value of optimized parameters in different structures

The value of the optimized parameters	Number of leak areas	kernel function
C=0/4	Three	Linear
C=65	Four	
C=21/8	Five	
C=0/1	Six	
C=1/1	Seven	
$\gamma=0/5, c=3$	Three	Radial basis
$\gamma=0/25, c=18$	Four	
$\gamma=1/02, c=4$	Five	
$\gamma=1/02, c=4/5$	Six	
$\gamma=1/38, c=1/4$	Seven	

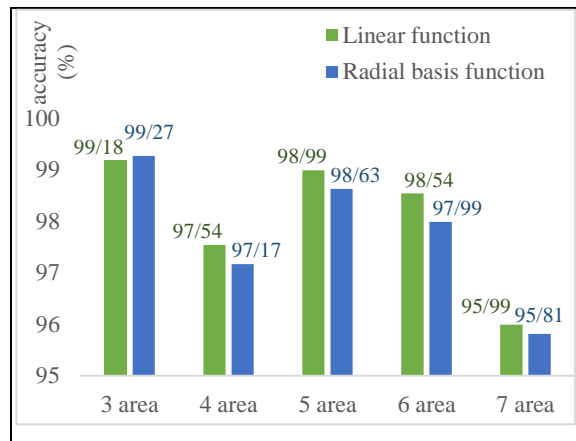


Figure 1. "Accuracy" graph for trained models

The structure chosen to build the final leak area identification model should have more leak areas in addition to providing sufficient "accuracy" for the model, because leak detection is easier in smaller areas. According to Figure 1, although the two models built with three leak areas had the highest degree of "accuracy" compared to the other models, they did not sufficiently limit the network to smaller areas. On the other hand, the model built with five leak areas and the linear kernel function, in addition to dividing the network into more areas, has an acceptable "accuracy" of 98.99% (approximately 99%). Therefore, in this study, the final model of leak area identification with a structure consisting of five leak areas and the linear kernel function is taught and according to Table 1, the value of the optimal parameter c for this model is 21.8.

4. Discussion and Conclusion

In this study, a leak area identification method based on multi-class support vector machines is presented. First, the K-means clustering algorithm was used to divide the network into a number of small leak areas and the number of each area was considered as a multi-class classification label. The effectiveness of the method was also evaluated in terms of the number of leakage areas and the type of kernel function. The results showed that dividing the network into five leak areas and using the linear kernel function, while creating an acceptable rate of 98.99% for the "accuracy" evaluation index, also divides the network into a number of suitable areas. Once the potential leak area has been determined, it is no longer necessary to search for all nodes in the water distribution network, and leaks should only be searched for nodes in the leak area. This reduces the search space and speeds up the search. The proposed method can be combined with common leak detection methods in a DMA and judge the leak node.

5. Six important references

- 1) Chamasemani, F. F., & Singh, Y. P. (2011). Multi-class support vector machine (SVM) classifiers-an application in hypothyroid detection and classification. In 2011 Sixth International Conference on Bio-Inspired Computing: Theories and Applications (pp. 351-356). IEEE.
- 2) Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. learn*, 20(3), 273-297.
- 3) Mac Kay, D. J. (2003). *Information theory, inference and learning algorithms*. Cambridge university press.
- 4) Quinones-Grueiro, M., Bernal-de Lázaro, J. M., Verde, C., Prieto-Moreno, A., & Llanes-Santiago, O. (2018). Comparison of classifiers for leak location in water distribution networks. *IFAC-PapersOnLine*, 51(24), 407-413.
- 5) Weston, J., & Watkins, C. (1998). Multi-class support vector machines (pp. 98-04). Technical Report CSD-TR-98-04, Department of Computer Science, Royal Holloway, University of London, May.

- 6) Zhang, Q., Wu, Z. Y., Zhao, M., Qi, J., Huang, Y., & Zhao, H. (2016). Leakage zone identification in large-scale water distribution systems using multiclass support vector machines. *Journal of Water Resources Planning and Management*, 142(11), 04016042.

Conflict of Interest

Authors declared no conflict of interest.

شناسایی منطقه نشت در سیستم‌های توزیع آب بزرگ (مطالعه موردی: شهر ماهان در استان کرمان)

یاسمن سادات هاشمی پور^۱، غلامعباس بارانی^۲، احسان فدائی کرمانی^۳

تاریخ ارسال: ۱۴۰۰/۰۹/۱۷

تاریخ پذیرش: ۱۴۰۰/۱۲/۰۷

مقاله پژوهشی

چکیده

امروزه شناسایی نشت مسئله مهمی در شبکه‌های توزیع آب می‌باشد، زیرا نشت هزینه‌ها را افزایش داده و موجب اتلاف منابع آب می‌شود. در این مطالعه، از یک روش جدید برای شناسایی منطقه نشت در شبکه توزیع آب قسمتی از شهر ماهان استفاده شد. ابتدا توسط الگوریتم خوشه‌بندی K-means، شبکه به تعدادی از مناطق تقسیم شد. سپس نمونه‌های آموزش مربوط به هر منطقه، با استفاده از شبیه‌سازی تصادفی نشت در مدل هیدرولیکی شبکه ساخته شد. شماره هر منطقه به عنوان برجسب طبقه‌بندی ماشین بردار پشتیبان چندکلاسی مورد استفاده قرار گرفت و به همراه نمونه‌های آموزش، مدل طبقه‌بندی آموزش داده شد. در نهایت، مدل آموزش دیده به عنوان مدل شناسایی منطقه نشت استفاده شد و برای تعیین مناطق احتمالی نشت با نمونه‌های میدانی مشاهده شده، اعمال شد. نتایج نشان می‌دهد که از میان ۱۰ ساختار استفاده شده برای ساخت مدل، تنها سه ساختار شامل: «سه منطقه و تابع هسته پایه شعاعی»، «سه منطقه و تابع هسته خطی» و «پنج منطقه و تابع هسته خطی»، به ترتیب میزان قابل قبول ۹۹/۲۷٪، ۹۹/۱۸٪ و ۹۸/۹۹٪ را برای شاخص ارزیابی «دقت» ایجاد می‌کنند. از طرف دیگر، از میان این سه ساختار، تقسیم شبکه به پنج منطقه نشت و استفاده از تابع هسته خطی، شبکه را به تعداد مناطق بیشتری تقسیم می‌کند و نشت‌یابی را در مناطق محدود شده آسان‌تر می‌کند. در نتیجه برای ساخت مدل نهایی از این ساختار استفاده شد.

واژه‌های کلیدی: شناسایی منطقه نشت، شبکه توزیع آب، شهر ماهان، ایپانت، متلب.

^۱ کارشناسی ارشد آب و سازه‌های هیدرولیکی، گروه مهندسی عمران، دانشکده فنی مهندسی، دانشگاه شهید باهنر، کرمان، ایران. ۰۳۴۳۳۲۲۰۰۵۴، yas.hp@eng.uk.ac.ir

^۲ استاد، گروه مهندسی عمران، دانشکده فنی مهندسی، دانشگاه شهید باهنر، کرمان، ایران. ۰۳۴۳۳۲۲۰۰۵۴، gab@uk.ac.ir (نویسنده مسئول)

^۳ استادیار، گروه مهندسی عمران، دانشکده فنی مهندسی، دانشگاه شهید باهنر، کرمان، ایران. ۰۳۴۳۳۲۲۰۰۵۴، e.fadaei@uk.ac.ir



مقدمه

شبکه‌های توزیع آب سیستم‌های پویای غیرخطی پیچیده‌ای هستند که به طور مداوم آب آشامیدنی را به مصرف‌کنندگان مختلف می‌رسانند (کوینونس گروپرو و همکاران، ۲۰۱۹). لوله‌های آسیب‌دیده تأثیر زیادی بر کیفیت آب و هیدرولیک سیستم توزیع آب دارند. با این حال، به دلیل گسترش سریع محدوده شهرها، اغلب شبکه‌های آبرسانی برنامه‌ریزی مشخصی ندارند و لوله‌های شبکه مستعد نشت می‌باشند (هو و همکاران، ۲۰۲۱). نشت‌ها اغلب کشف نشده باقی می‌مانند و باعث تلفات آب و در نتیجه افزایش هزینه برای اپراتورهای شبکه توزیع آب می‌شوند (سوفوکلئوس و همکاران، ۲۰۱۹). بنابراین، یافتن سریع محل نشت می‌تواند از تلفات آب و تبدیل یک نشت کوچک به ترکیدهای بزرگ جلوگیری کند.

در سال‌های اخیر، تکنیک‌های بسیاری مبتنی بر سخت افزار و نرم افزار، جهت شناسایی نشت در هنگام وقوع ترکیدهای لوله در شبکه‌های توزیع آب، ارائه شده است. تکنیک‌های مبتنی بر سخت‌افزار، از جمله روش‌های صوتی، الکترومغناطیسی و اپتیک، به طور گسترده‌ای برای شناسایی دقیق نشت‌ها به کار گرفته شده‌اند (بوهورکز و همکاران، ۲۰۲۰؛ کاتالدو و همکاران، ۲۰۱۷؛ گونگ و همکاران، ۲۰۱۸). با این حال، این روش‌ها تنها می‌توانند برای تشخیص نشت محلی استفاده شوند و استفاده از آن‌ها در شبکه‌های توزیع آب مقیاس بزرگ به علت زمان‌بر بودن نصب و تشخیص و هزینه‌های نگهداری، مقرون به صرفه نمی‌باشد.

تکنیک‌های مبتنی بر نرم‌افزار، از جمله روش‌های مبتنی بر کالیبراسیون و روش‌های داده محور، می‌توانند برای جبران کمبودهای تکنیک‌های مبتنی بر سخت‌افزار، در شبکه‌های توزیع آب مقیاس بزرگ به کار برده شوند. روش مبتنی بر کالیبراسیون به صورت یک مسئله معکوس برای شناسایی پارامترهای نشت ناشناخته، تعریف می‌شود که این مسئله با استفاده از داده‌های میدانی در یک مدل هیدرولیکی، حل می‌شود. شناسایی پارامترها به عنوان یک

مسئله بهینه‌سازی غیرخطی فرموله شده است که با استفاده از الگوریتم‌های بهینه‌سازی هوشمند حل می‌شود (وو، ۲۰۰۹؛ سانز و همکاران، ۲۰۱۵؛ سوفوکلئوس و همکاران، ۲۰۱۹). اشکال این روش این است که الگوریتم‌های بهینه‌سازی به ارزیابی مکرر و بعد وسیعی از متغیرهای تصمیم‌گیری نیاز دارند و همچنین احتمال همگرایی زودرس یک الگوریتم در فضای جستجوی بزرگ، وجود دارد (چن و همکاران، ۲۰۲۰). از آن‌جا که روش‌های مبتنی بر کالیبراسیون برای شبکه‌های مقیاس بزرگ مناسب نیستند، روش‌های داده محور مورد بررسی قرار می‌گیرند.

روش‌های داده محور معمولاً برای ساختن یک مدل داده محور نیاز به داده‌های زیادی دارند. این روش‌ها شامل؛ روش‌های مبتنی بر تئوری یادگیری تجربی، مانند شبکه‌های عصبی مصنوعی (ANN^۱) و روش شناسایی بیز^۲، و همچنین روش‌های یادگیری ماشینی مبتنی بر تئوری یادگیری آماری، مانند ماشین‌های بردار پشتیبان (SVM^۳)، هستند.

روش بیز توزیع احتمال داده‌های آموزش را فرض می‌کند و پارامترهای تابع چگالی احتمال را براساس نمونه‌های آموزش تخمین می‌زند. نمونه‌های آزمایش با توجه به حداکثر احتمال مشخص می‌شوند.

پولاکیس و همکاران (۲۰۰۳) چهارچوب احتمالی بیز را برای پیش بینی محتمل‌ترین محل نشت و شدت آن در شبکه توزیع آب پیشنهاد کردند و تجزیه و تحلیل قابلیت اطمینان را از نظر خطای اندازه‌گیری، خطای مدل‌سازی، شدت نشت و ترتیب حسگر انجام دادند، تا مبنای نظری برای این روش فراهم شود.

ژو و همکاران (۲۰۱۱) یک رویکرد استدلال بیز را براساس الگوریتم بازگشتی^۴ برای تشخیص نشت ارائه دادند، که این الگوریتم بازگشتی می‌تواند رابطه بین جریان‌ها یا فشارها در نقاط نظارت و اندازه نشت را یاد بگیرد.

کستانزو و همکاران (۲۰۱۴) یک روش جدید کالیبراسیون بیز را برای تشخیص نشت و کالیبراسیون ایجاد کردند. ابتدا شبکه با توجه به زبری لوله‌ها به تعدادی از

^۳ Support vector machines
^۴ Recursive algorithm

^۱ Artificial neural networks
^۲ Bayes identification method

کردند. تابع نشت از تابع فاصله و تابع اطمینان ساخته شده است. ارزش تابع فاصله، وابسته به فاصله بین نقاط نظارت و نقاط نشت است. هرچه نقطه نظارت به یک نشت نزدیکتر باشد، مقدار تابع فاصله بیشتر است. ارزش تابع اطمینان با افزایش زمان اطلاع، افزایش می‌یابد. به عبارت دیگر، با گذشت زمان نشت‌یابی آسان‌تر می‌شود.

اگرچه شبکه عصبی مصنوعی در بسیاری از مسائل آب کاربرد داشته و به نتایج خوبی رسیده است، اما بیش از حد به نمونه‌های آموزش اعتماد می‌کند. اگر اندازه نمونه‌های آموزش کوچک باشد، دقت پیش‌بینی شبکه عصبی کم است. برای دستیابی به دقت کافی، نمونه‌های آموزش کافی لازم است. در ضمن، زمان آموزش شبکه عصبی با افزایش اندازه نمونه‌های آموزش افزایش می‌یابد. همچنین هنگامی که شرایط فیزیکی یک سیستم آب تغییر می‌کند، شبکه عصبی احتیاج به آموزش مجدد دارد.

روش‌های مرسوم اغلب مبتنی بر تجزیه و تحلیل آماری هستند و فقط در مواردی که اندازه نمونه‌های آموزش به مقدار کافی بزرگ باشد، می‌توانند نتایج خوبی کسب کنند و هنگامی که اندازه نمونه‌های آموزش اندک باشد، عملکرد این روش‌ها تحت تاثیر قرار می‌گیرد.

برای حل مشکل نمونه‌های کوچک، وپنیک و کوتز (۱۹۸۲) تئوری یادگیری آماری را ایجاد کردند. کورتس و وپنیک (۱۹۹۵) ماشین‌های بردار پشتیبان را به عنوان یک روش جدید یادگیری ماشین ایجاد کردند.

مونس و همکاران (۲۰۱۱) روش جدیدی را بر اساس ماشین‌های بردار پشتیبان برای تشخیص ناهنجاری ناشی از نشت، در سطح یک DMA پیشنهاد دادند. در این روش، از پردازش داده‌های سری زمانی جریان و فشار در ورودی DMA استفاده می‌شود.

ماشین‌های بردار پشتیبان فقط می‌توانند مسئله‌های طبقه‌بندی دودویی را حل کنند و برای حل مسئله‌های طبقه‌بندی چندکلاسی وستون و واتکینز (۱۹۹۸) و چماسمانی و سینگ (۲۰۱۱) الگوریتم‌هایی را پیشنهاد

مناطق تقسیم می‌شود و سپس زبری‌ها با استفاده از روش بیز کالیبره می‌شوند. منطقه نشت با مقایسه زبری در وضعیت عادی و زبری در وضعیت غیرطبیعی مشخص می‌شود.

رومانو و همکاران (۲۰۱۰) تجزیه و تحلیل موجک^۱، شبکه‌های عصبی مصنوعی، کنترل فرآیند آماری (SPC^۲) و سیستم‌های استنتاج بیز را به کار بردند و آن‌ها را در یک چهارچوب یکپارچه برای یافتن نشت ادغام کردند. این روش در سطح یک DMA^۳ مورد آزمایش قرار گرفت و نتایج نشان داد که این روش می‌تواند به سرعت نشت را پیدا کند و با موفقیت در سطح یک DMA استفاده شود.

در نتیجه، روش بیز از توزیع احتمال برای نشان دادن همه اشکال عدم اطمینان استفاده می‌کند و فرآیند یادگیری یا استدلال از طریق احتمال اجرا می‌شود. این می‌تواند از عدم قطعیت‌های اجتناب ناپذیر در فرآیند اندازه‌گیری و مدل‌سازی جلوگیری کند. با این حال، این روش دارای معایبی است. اول اینکه، باید توزیع احتمال داده‌های آموزش را فرض کرد. دقت تشخیص معمولاً زمانی که توزیع احتمال فرض شده با توزیع احتمال واقعی متفاوت باشد، بسیار کم است. دوم اینکه، تخمین دقیق ماتریس واریانس الگوریتم بیز دشوار است. بنابراین زمانی که اندازه نمونه کوچک باشد، تخمین چگالی احتمال از طریق نمونه‌های آموزش، یک مسئله دشوار است. در نتیجه برای دستیابی به دقت، روش بیز به حجم زیادی از نمونه‌های آموزش احتیاج دارد.

در سال‌های اخیر، شبکه‌های عصبی مصنوعی با موفقیت بسیار زیادی در صنعت آبرسانی به کار گرفته شده‌اند.

فنگ و ژانگ (۲۰۰۶) برای شناسایی نشت خط لوله، رویکردی مبتنی بر شبکه‌های عصبی فازی^۴ پیشنهاد دادند. بر همین اساس، مونس و همکاران (۲۰۱۰) یک سیستم هوش مصنوعی برای تشخیص برخط نشت ایجاد کردند. این سیستم از شبکه‌های عصبی مصنوعی و سیستم‌های استنتاج فازی^۵ تشکیل شده است.

آکسلا و همکاران (۲۰۰۹) یک تابع نشت را با یک نگاه خود سازمانده (SOM^۶) برای تشخیص نشت ترکیب

^۴ Fuzzy neural networks
^۵ Fuzzy inference systems
^۶ Self organizing map

^۱ Wavelet analysis
^۲ Statistical process control
^۳ District meter area



همچنین ارزیابی روش ارائه شده را از سه جنبه انجام دادند: تعداد مناطق نشت، نوع تابع هسته و تعداد و محل نقاط نظارت فشار.

نتایج نشان داد که تعداد مناسب برای مناطق نشت با توجه به دقت طبقه‌بندی تعیین می‌شود و تابع هسته خطی و پایه شعاعی عملکرد بهتری را نسبت به تابع هسته سیگموئید و چند جمله‌ای نشان می‌دهند. تعداد و محل نقاط نظارت فشار بهینه‌شده، عملکرد بهتری نسبت به حالت بهینه نشده دارد. این روش قبل از تشخیص کانون نشت توسط سایر روش‌ها، منطقه احتمالی نشت را محدود می‌کند.

به دلیل پیچیدگی شبکه‌های بزرگ توزیع آب و تعداد محدود فشارسنج‌ها، پاسخ نقاط نظارت فشار به نشت‌هایی که در گره‌های مجاور وارد می‌شود، بسیار مشابه خواهد بود. در نتیجه، در نظر گرفتن همه گره‌های شبکه برای شناسایی نشت در شبکه‌های بزرگ توزیع آب، باعث افزایش پیچیدگی مدل و کاهش دقت مدل طبقه‌بندی می‌شود. اما با خوشه‌بندی گره‌های دارای پاسخ مشابه به نشت، می‌توان دقت مدل طبقه‌بندی را بهبود بخشید. در نتیجه، در این مطالعه نیز مشابه روش ارائه شده توسط ژانگ و همکاران (۲۰۱۶)، شبکه توزیع آب قسمتی از شهر ماهان توسط خوشه‌بندی K-means، به خوشه‌ها با گره‌هایی مشابه در پاسخ به نشت، تقسیم می‌شود و سپس برای کوچک کردن فضای جستجو، توسط الگوریتم طبقه‌بندی ماشین‌های بردار پشتیبان چند کلاسی، منطقه نشت شناسایی می‌گردد.

مواد و روش‌ها

منطقه مورد مطالعه

سیستم توزیع آب مورد مطالعه، واقع در قسمت جنوبی شهر ماهان می‌باشد. این سیستم تقریباً در یک شبانه‌روز ۳۲۰۰ متر مکعب آب را بین مصرف‌کنندگان توزیع می‌نماید و شامل یک مخزن، ۸۱۲ لوله با ۷۴۵ گره، ۵۸ دریچه کنترل آب (TCV^۴) و هفت فشارسنج می‌باشد. محل و تعداد فشارسنج‌ها قبل از جایگذاری در شبکه بهینه نشده است و

دادند. این الگوریتم‌ها به صورت کلی به عنوان ماشین‌های بردار پشتیبان چندکلاسی (M-SVM^۱) شناخته می‌شوند. مامو و همکاران (۲۰۱۴) یک تکنیک برای تشخیص نشت و طبقه‌بندی آن مبتنی بر ماشین‌های بردار پشتیبان چند کلاسی، در سطح یک DMA ایجاد کردند. وضعیت عملکرد ابتدا به شش کلاس تقسیم شد: طبیعی، غیر-طبیعی، شکستن یا ترکیدن، نشت زیاد، نشت متوسط و نشت کم. سپس از ماشین‌های بردار پشتیبان چند کلاسی برای شناسایی وضعیت عملکرد DMA با توجه به جریان‌ها و فشارهای مشاهده شده میدانی از نقاط نظارت، استفاده شد.

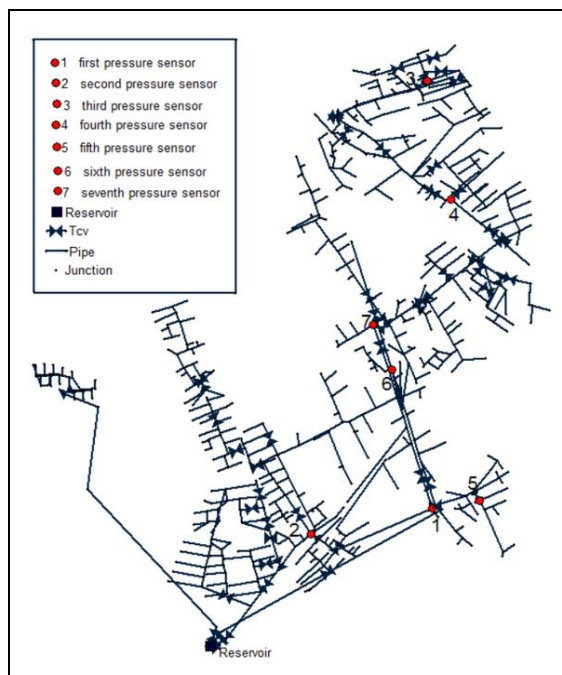
کویونس گروپرو و همکاران (۲۰۱۸) چهار ابزار طبقه‌بندی نظارت شده، که به‌طور گسترده در چهارچوب تشخیص الگو استفاده می‌شوند را برای نشت‌یابی در شبکه‌های توزیع آب مقایسه کردند. این طبقه‌بندی‌کننده‌ها شامل؛ کی-نزدیک‌ترین همسایه (KNN^۲)، طبقه‌بند بیز، شبکه‌های عصبی مصنوعی و ماشین‌های بردار پشتیبان بودند. برتری طبقه‌بندی‌کننده‌ها، با توجه به عملکرد نشت-یابی تحت عدم قطعیت پارامترهای مدل هیدرولیکی، عدم قطعیت تقاضای مصرف‌کننده، عدم قطعیت مقدار نشت و خطای حسگر مقایسه شد. نتایج نشان داد که ماشین‌های بردار پشتیبان در صورت وجود تمام عدم قطعیت‌ها، عملکرد مشابه یا بهتر از دیگر طبقه‌بندی‌کننده‌ها دارند.

روش‌های تشخیص نشت را می‌توان با موفقیت در سیستم‌های توزیع آب کوچک به کار برد، اما چالش‌های بزرگی بخصوص برای سیستم‌های بزرگ بدون DMA وجود دارد.

از این رو، ژانگ و همکاران (۲۰۱۶) روش جدیدی برای محدود کردن منطقه نشت در سیستم‌های توزیع آب بزرگ ارائه دادند. ابتدا یک شبکه بزرگ آب در یکی از شهرهای کشور چین را توسط خوشه‌بندی K-means^۳، به تعدادی از مناطق تقسیم کردند. سپس با استفاده از داده‌های نقاط نظارت فشار و طبقه‌بندی ماشین‌های بردار پشتیبان چند کلاسی، منطقه نشت احتمالی را تشخیص دادند. آن‌ها

^۳ K-means clustering
^۴ Throttle control valve

^۱ Multiclass support vector machines
^۲ K-nearest neighbor



شکل (۱): مدل هیدرولیکی شبیه‌سازی شده

الگوریتم خوشه‌بندی k-means

الگوریتم خوشه‌بندی K-means (مک‌کی، ۲۰۰۳)، یک مسئله یادگیری غیر نظارت شده می‌باشد. این الگوریتم n داده را در k خوشه ($k \leq n$) تقسیم می‌کند. تعداد خوشه‌ها (k) را کاربر مشخص می‌کند. دلیل نام‌گذاری K-means این است که در این الگوریتم مراکز خوشه‌ها به‌گونه‌ای انتخاب می‌شوند که میانگین اعضا خوشه باشند:

$$c = \{c_1, c_2, \dots, c_k\}$$

هر خوشه با یک بردار m_k مشخص می‌شود و هر داده x_i یک بردار d بعدی است:

$$x_i = (x_{i1}, x_{i2}, \dots, x_{id})$$

$$X^{(n)} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1d} \\ x_{21} & x_{22} & \dots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nd} \end{bmatrix} \quad (1)$$

یک تابع فاصله نیز به صورت زیر تعریف می‌شود:

$$d(x, y) = \frac{1}{2} \sum_1 (x_i - y_i)^2 \quad (2)$$

x و y می‌توانند هر یک از نقاط x_1, x_2, \dots, x_n باشند. در واقع این تابع فاصله بین x و y را مشخص می‌کند.

براساس تجربه برای نظارت بر سیستم قرار داده شده‌اند. آن‌ها اغلب در محدوده گره‌ها با ارتفاع بالا یا در نزدیکی تقاضاهای زیاد هستند.

در این مطالعه برای شبیه‌سازی هیدرولیکی شبکه توزیع آب از نرم‌افزار ایپانت استفاده شده است. در شروع شبیه‌سازی، ابتدا در تنظیمات مدل هیدرولیکی، واحد دبی جریان (لیتر بر ثانیه) و فرمول افت هد (هیزن ویلیامز) تعیین شد و همچنین الگوی تقاضا در هر یک از ۲۴ ساعت شبانه روز برای مدل هیدرولیکی تعریف شد. سپس هر یک از مؤلفه‌های شبکه جایگذاری شدند و مشخصات مربوط به هر یک شامل؛ «هد کل» برای مخزن، «طول»، «قطر» و «ضریب زبری» برای لوله‌ها، «ارتفاع»، «تقاضای پایه» و «الگوی تقاضا» برای گره‌ها و «قطر» و «ضریب افت» برای دریچه‌های کنترل آب، وارد شد. با ساخت مدل هیدرولیکی از شبکه و اجرای آن، می‌توان فشار در هر گره و سایر خروجی‌های مورد نیاز را از هر مؤلفه در شبکه، در هر یک از ۲۴ ساعت شبانه روز، دریافت کرد. در نهایت، توسط فشارهای برداشت شده از فشارسنج‌ها در مدت ۲۴ ساعت (یک شبانه روز) که در آن هیچ نشستی رخ نداده است، مدل هیدرولیکی ساخته شده از شبکه کالیبره شد. در واقع، ضریب افت برای دریچه‌های کنترل آب به‌گونه‌ای تنظیم شد که فشار شبیه‌سازی شده در گره‌هایی که فشارسنج بر روی آن‌ها نصب است، با فشار نمایش داده شده توسط همان فشارسنج، یکسان شود.

در شکل ۱ مدل هیدرولیکی شبیه‌سازی شده، نمایش داده شده است.



$$y = f(x) = (w^T x) + b \quad (۶)$$

که در آن w ، وزن؛ w^T ، ترانهاده وزن؛ x ، مقدار ورودی نمونه؛ و b عرض از مبدأ می‌باشد.

تابع تصمیم غیرخطی باید تابع هدف زیر را به حداقل برساند:

$$\min \frac{1}{2} (w^T \cdot w) + c \sum_i \xi_i \quad (۷)$$

که در آن c ، ضریب جریمه خطا؛ و $\sum_i \xi_i$ ، خطا آموزش می‌باشد.

در نهایت تابع هدف با جایگزینی $w = \sum_{i=1}^l a_i y_i x_i$ ، به صورت زیر در می‌آید:

$$\min \frac{1}{2} \sum_{i,j=1}^l y_i y_j a_i a_j k(x_i, x_j) - \sum_{i=1}^l a_i \quad (۸)$$

همراه با محدودیت‌های

$$\sum_{i=1}^l y_i a_i = 0 \quad (۹)$$

$$0 \leq a_i \leq c \quad (۱۰)$$

و تابع تصمیم غیرخطی مطابق زیر است:

$$f(x) = \text{sign} \left\{ \left[\sum_{i=1}^l a_i y_i k(x, x_i) \right] + b \right\} \quad (۱۱)$$

که در آن y ، برچسب کلاس؛ a ، ضریب لاگرانژ؛ b ، عرض از مبدأ؛ و $k(x, y)$ تابع هسته می‌باشد.

چهار نوع تابع هسته به شرح زیر می‌باشد:

- خطی:

$$k(x, x_i) = x^T x_i \quad (۱۲)$$

- چند جمله‌ای:

$$k(x, x_i) = (\gamma x^T x_i + r)^p, \quad \gamma > 0 \quad (۱۳)$$

- تابع پایه شعاعی:

$$k(x, x_i) = \exp(-\gamma \|x - x_i\|^2), \quad \gamma > 0 \quad (۱۴)$$

- سیگموئید:

$$k(x, x_i) = \tanh(\gamma x^T x_i + r) \quad (۱۵)$$

الگوریتم ماشین‌های بردار پشتیبان چند کلاسی را می‌توان به دو دسته کلی تقسیم کرد:

- رویکرد مستقیم

- رویکرد غیرمستقیم

در نتیجه، خوشه‌بندی K-means در چهار مرحله انجام می‌شود:

- مرحله اولیه: k نقطه از بین $x^{(n)}$ به عنوان مرکز خوشه‌ها مشخص می‌شود.

- مرحله واگذاری: هر یک از n داده به نزدیک‌ترین مرکز اختصاص داده می‌شود.

$$c_i = [x_p : d(x_p, m_i) \leq d(x_p, m_j) \forall j, 1 \leq j \leq k] \quad (۳)$$

که c_i مرکز i امین خوشه است ($1 \leq i \leq k$) و هر x_p فقط به یک خوشه اختصاص داده می‌شود.

- مرحله بروزرسانی: میانگین هر خوشه به عنوان مرکز جدید خوشه تعیین می‌شود و تابع بروزرسانی مطابق زیر است:

$$m_i = \frac{1}{|c_i|} \sum_{x_p \in c_i} x_p \quad (۴)$$

که $|c_i|$ ، تعداد اعضا خوشه c_i می‌باشد.

- در نهایت، مرحله واگذاری و مرحله بروزرسانی تا جایی تکرار می‌شوند که دوبار متوالی واگذاری‌ها تغییر نکنند.

الگوریتم ماشین‌های بردار پشتیبان چند کلاسی

ماشین‌های بردار پشتیبان فقط می‌توانند مسئله‌های طبقه‌بندی دودویی را حل کنند و برای جدا کردن نمونه‌ها با ابعاد زیاد فضای ویژگی به دو کلاس می‌باشند. اما برای حل مسئله‌های طبقه‌بندی چند کلاسی، از ماشین‌های بردار پشتیبان چند کلاسی استفاده می‌شود.

ابتدا ماتریس S با استفاده از نمونه‌های آموزش تشکیل می‌شود:

$$S = \{(x_1, y_1), \dots, (x_l, y_l)\} \\ = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1d} & y_1 \\ x_{21} & x_{22} & \dots & x_{2d} & y_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{l1} & x_{l2} & \dots & x_{ld} & y_l \end{bmatrix} \quad (۵)$$

که x_i ، نمونه i ام ($i=1,2,\dots,l$) و برداری d بعدی است:

$$x_i = (x_{i1}, x_{i2}, \dots, x_{id})$$

همچنین y_i کلاس نمونه است:

$$y_i \in \{1, 2, \dots, k\}$$

از یک تابع تصمیم غیرخطی با استفاده از نمونه‌های آموزش، برای طبقه‌بندی نمونه‌های آزمون استفاده می‌شود:

موجود طبقه‌بندی می‌شود و نتیجه هر طبقه‌بندی به صورت زیر ثبت می‌شود:

Flag-data =
[num-class1,num-class2,...,num-classK]
هر کدام از num-class ها که مقدار بیشتری داشته باشند، نمونه آزمون به همان کلاس تعلق می‌گیرد. اما اگر همه num-class ها مقدار یک را داشته باشند، نمونه در مرز classK,...,class2,class1 قرار می‌گیرد، در این حالت نمونه متعلق به اولین کلاس، یعنی class1 است.

انتخاب پارامترهای مدل

براساس تحقیقات چاماسمانی و سینگ (۲۰۱۱) و مشفورد و همکاران (۲۰۱۲) می‌توان دریافت که کارایی ماشین‌های بردار پشتیبان چند کلاسی به طور کلی به نوع تابع هسته، انتخاب پارامترهای هسته و ضریب جریمه c بستگی دارد. ژانگ و همکاران (۲۰۱۶) در مطالعه‌ای که براساس این روش داشتند به این نتیجه رسیدند که توابع هسته خطی و پایه شعاعی دقت طبقه‌بندی را نسبت به توابع هسته سیگموئید و چند جمله‌ای، افزایش می‌دهند. بنابراین، در این مطالعه برای طبقه‌بندی چند کلاسی با استفاده از ماشین‌های بردار پشتیبان چند کلاسی توابع هسته خطی و پایه شعاعی مقایسه شده است. علاوه بر این، تقسیم‌بندی مناطق نشت می‌تواند بر دقت طبقه‌بندی ماشین‌های بردار پشتیبان چند کلاسی تاثیر بگذارد. هرچه تعداد مناطق بیشتر شود، دقت طبقه‌بندی کاهش می‌یابد. پس باید تعداد مناطق نشت را تا حدی کم کرد که دقت طبقه‌بندی استاندارد را برآورده کند. به منظور انتخاب تعداد مناطق نشت بهینه، تقسیم بندی شبکه با تعداد مناطق نشت مختلف انجام شد و مقادیر ۳ و ۴ و ۵ و ۶ و ۷ برای k استفاده شد.

انتخاب پارامترهای مدل معمولاً به دو روش زیر انجام می‌شود:

- روش اعتبارسنجی متقابل^۱ (رفاعیل‌زاده و همکاران، ۲۰۰۹)

- الگوریتم گرادیان کاهشی^۲ (چاپل و همکاران، ۲۰۰۲)

روش مستقیم، مستقیماً تابع تصمیم ماشین‌های بردار پشتیبان را اصلاح می‌کند و پارامترهای گوناگون طبقه‌بندی را با یک مسئله بهینه‌سازی ترکیب می‌کند و سپس با حل مسئله بهینه‌سازی، طبقه‌بندی چند کلاسی را بدست می‌آورد. تعداد متغیرها در روش مستقیم از روش غیرمستقیم بیشتر است و سرعت آموزش روش مستقیم کمتر از سرعت آموزش روش غیرمستقیم است و همچنین دقت طبقه‌بندی روش مستقیم از دقت طبقه‌بندی روش غیر مستقیم کمتر است.

روش غیرمستقیم با استفاده ترکیبی از طبقه‌بندی کننده‌های دودویی گوناگون، به صورت غیرمستقیم طبقه‌بندی کننده‌های چند کلاسی را می‌سازد. این روش به دو روش زیر تقسیم می‌شود:

- روش یکی در مقابل همه

- روش یکی در مقابل یکی

میلگرام و همکاران (۲۰۰۶) این دو روش را با هم مقایسه کردند و نتایج نشان داد که روش یکی در مقابل یکی، سرعت آموزش بیشتری دارد و در مسائل با تعداد زیادی کلاس، این روش ارجحیت دارد. با توجه به اینکه یک شبکه بزرگ توزیع آب به مناطق مختلفی تقسیم می‌شود، در این مطالعه از روش یکی در مقابل یکی برای طبقه‌بندی ماشین‌های بردار پشتیبان چند کلاسی استفاده شده است. در این روش نمونه‌های آموزش در k کلاس مختلف هستند و دارای برچسب‌های زیر می‌باشند:

$$y \in \{class1, class2, \dots, classk\}$$

نمونه‌ها با برچسب یکسان در یک گروه قرار می‌گیرند و در نهایت k گروه را تشکیل می‌دهند:

$$train1, train2, \dots, trainK$$

این روش از n کلاسبند دو کلاسه استفاده می‌کند و یک مسئله چند کلاسه را حل می‌کند:

$$n = \frac{k(k-1)}{2} \quad (16)$$

هر بار دو گروه از نمونه‌های آموزش انتخاب می‌شود و یک کلاسبند دو کلاسه با استفاده از آن‌ها ساخته می‌شود. در نتیجه، n مدل طبقه‌بندی دودویی ساخته می‌شود و پس از آن، هر بار یک نمونه آزمون با استفاده از n کلاسبند



۱- در شرایطی که در شبکه نشت وجود ندارد، فشار برای تمام نقاط نظارت برداشت می‌شود:

$$p^0 = \{p_{11}^0, p_{12}^0, \dots, p_{in}^0\} \quad (17)$$

که در آن p_{ij}^0 ، فشار در نقطه نظارت j ام در شرایط بدون در نظر گرفتن تقاضا برای نشت است.

۲- در یک گره، تقاضای اضافی (نشت) کمتر از $6/53$ لیتر بر ثانیه ایجاد می‌شود. سپس مدل هیدرولیکی اجرا می‌شود و فشار برای هر نقطه نظارت برداشت می‌شود:

$$p^1 = \{p_{11}^1, p_{12}^1, \dots, p_{in}^1\} \quad (18)$$

که در آن p_{ij}^1 ، فشار در نقطه نظارت j ام در اثر نشت ایجاد شده در گره i ام است.

۳- تغییرات فشار محاسبه می‌شود:

$$\Delta p_i = \{p_{11}^1 - p_{11}^0, p_{12}^1 - p_{12}^0, \dots, p_{in}^1 - p_{in}^0\} \quad (19)$$

۴- تقاضای اضافی (نشت) ایجاد شده برای این گره برداشته می‌شود و مدل هیدرولیکی اجرا می‌شود.

تمام مراحل ۲ تا ۴ باید برای همه گره‌های سیستم طی شوند و نتایج در ماتریس زیر ریخته شود:

$$s = \begin{bmatrix} p_{11}^1 - p_{11}^0 & p_{12}^1 - p_{12}^0 & \dots & p_{1n}^1 - p_{1n}^0 & 1 \\ p_{21}^1 - p_{21}^0 & p_{22}^1 - p_{22}^0 & \dots & p_{2n}^1 - p_{2n}^0 & 2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ p_{m1}^1 - p_{m1}^0 & p_{m2}^1 - p_{m2}^0 & \dots & p_{mn}^1 - p_{mn}^0 & m \end{bmatrix} \quad (20)$$

۵- با حذف ستون آخر ماتریس s ، ماتریس x با ابعاد $m \times n$ تشکیل می‌شود، از ماتریس x به عنوان ماتریس ورودی الگوریتم K-means استفاده می‌شود و سرانجام این الگوریتم گره‌ها را به k خوشه تقسیم می‌کند و شماره هر خوشه به عنوان برچسب طبقه‌بندی ماشین‌های بردار پشتیبان چند کلاسی استفاده می‌شود.

تولید نمونه‌های آموزش

نمونه‌های آموزش با در نظر گرفتن سه عامل انتخاب می‌شوند:

- اعتبار؛ نمونه‌های آموزش انتخاب شده، باید به طور صحیح قانون طبیعی سیستم‌های توزیع آب را منعکس کنند. به عبارت دیگر، احتمال ایجاد نمونه‌های آموزش در سیستم‌های توزیع آب واقعی وجود داشته باشد.

نتایج بهینه‌سازی روش الگوریتم گرادیان کاهش تا حد زیادی به میزان خطا نمونه‌های مشاهده شده بستگی دارد. بنابراین، در این مطالعه از روش اعتبارسنجی متقابل برای انتخاب ساختار بهینه استفاده شده است.

تحلیل شبکه

براساس میانگین چند روز بدون نشت، حداقل تقاضای کل شبکه ۲۶ لیتر بر ثانیه و در ساعات ۱۲ نیمه شب تا ۴ صبح می‌باشد و حداکثر تقاضای کل شبکه ۴۹ لیتر بر ثانیه و متوسط آن در روز ۳۷ لیتر بر ثانیه می‌باشد.

در این شبکه هیچ معیاری برای تعیین دامنه نشت وجود ندارد و حداقل تخلیه نشت، وابسته به تغییرپذیری طبیعی نقاط نظارت فشار تنظیم می‌شود. باید اطمینان حاصل کرد که تغییر فشار ناشی از نشت، با نوسانات طبیعی فشار قابل تشخیص است. در نتیجه، براساس تجربیاتی از شبکه در دوره‌ای که در آن نشت رخ نداده، یک مقدار حداکثر (آستانه) برای تغییرات طبیعی فشار در نقاط نظارت مشخص شده است. بنابراین، مقدار آستانه $0/15$ مترآب در نظر گرفته شده است. در نتیجه، فقط هنگامی که نوسان فشار نقاط نظارت از $0/15$ مترآب فراتر رود، می‌توان آن را نشت در نظر گرفت. از طرف دیگر، دقت فشارسنج‌های موجود در شبکه $0/01$ مترآب می‌باشد، پس اگر تغییرات فشار کمتر از $0/01$ مترآب باشد، فشارسنج تغییرات فشار را صفر در نظر می‌گیرد. حداکثر تخلیه نشت در شبکه قابل پیش‌بینی نمی‌باشد و وقایع نشت توسط مصرف‌کنندگان گزارش می‌شود. براساس نتایج بانک جهانی (کینگدام و همکاران، ۲۰۰۶) حدود 15% تا 35% از آب تصفیه شده، در سیستم‌های توزیع آب از بین می‌رود. در این مطالعه حداکثر تخلیه نشت، 15% متوسط جریان ورودی به سیستم یعنی تقریباً به میزان $6/53$ لیتر بر ثانیه در نظر گرفته شده است.

تقسیم‌بندی شبکه به مناطق نشت

در این مطالعه، مدل هیدرولیکی دارای m گره ($m=745$) و n نقطه نظارت فشار ($n=7$) می‌باشد. همچنین تعداد مناطق نشت (k) توسط کاربر مشخص می‌شود. تقسیم‌بندی شبکه به مناطق نشت در پنج مرحله انجام می‌شود:

$$T = \begin{bmatrix} p_{11}^1 - p_1^0 & p_{12}^1 - p_2^0 & \dots & p_{1n}^1 - p_n^0 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ p_{L_1 1}^1 - p_1^0 & p_{L_1 2}^1 - p_2^0 & \dots & p_{L_1 n}^1 - p_n^0 & 1 \\ p_{11}^2 - p_1^0 & p_{12}^2 - p_2^0 & \dots & p_{1n}^2 - p_n^0 & 2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ p_{L_2 1}^2 - p_1^0 & p_{L_2 2}^2 - p_2^0 & \dots & p_{L_2 n}^2 - p_n^0 & 2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ p_{11}^k - p_1^0 & p_{12}^k - p_2^0 & \dots & p_{1n}^k - p_n^0 & k \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ p_{L_k 1}^k - p_1^0 & p_{L_k 2}^k - p_2^0 & \dots & p_{L_k n}^k - p_n^0 & k \end{bmatrix} \quad (22)$$

المان $P_{hj}^k - P_j^0$ در سطر h ام و ستون j ام ($j \leq n$) از ماتریس T ، تغییرات فشار در نقطه نظارت j ام، در اثر نشت ایجاد شده در منطقه k ام است و آخرین ستون T برچسب نمونه‌ها می‌باشد. در نهایت، نمونه‌های نشت موجود در ماتریس T به صورت زیر پیش پردازش می‌شوند:

- در همه درایه‌های ستون اول تا n ام ماتریس T ، اگر قدر مطلق تغییرات فشار بدست آمده کمتر از دقت فشارسنج ($0/01$ مترآب) باشد، آن درایه به صفر تبدیل می‌شود.

- در هر سطر از ماتریس T ، حداقل قدر مطلق یکی از درایه‌ها باید از مقدار آستانه ($0/15$ مترآب) فراتر باشد، که نشانگر وجود نشت در شبکه باشد. در غیر این صورت، نمونه آموزش که در آن سطر قرار دارد از ماتریس T حذف می‌شود.

ماتریس T پیش پردازش شده، به عنوان ماتریس نمونه‌های آموزش، برای انتخاب پارامترهای مدل با استفاده از روش اعتبارسنجی متقابل و آموزش ماشین‌های بردار پشتیبان چند کلاسی با استفاده از بهترین پارامترها استفاده می‌شود.

مدل شناسایی منطقه نشت

تغییرات فشار نقاط نظارت در اثر نشت تصادفی ایجاد شده در هر یک از گره‌های شبکه، به عنوان داده‌های ورودی به الگوریتم خوشه‌بندی K -means وارد شد و شبکه توزیع آب مورد مطالعه به ترتیب به سه، چهار، پنج، شش و هفت منطقه نشت تقسیم شد و برای هر نوع تقسیم‌بندی منطقه نشت، ۳۷۲۵ نمونه براساس شبیه‌سازی در مدل هیدرولیکی تولید شد و پس از پیش‌پردازش نمونه‌ها، تعدادی از نمونه‌ها حذف شدند. سپس ۷۰٪ نمونه‌های باقی مانده به نمونه‌های

- تناسب؛ نمونه‌های آموزش انتخاب شده، باید تا حد امکان شامل احتمال موقعیت‌های مختلف در سیستم‌های توزیع آب باشند.

- اندازه؛ اندازه نمونه‌های آموزش، دقت طبقه‌بندی ماشین‌های بردار پشتیبان چند کلاسی را تعیین می‌کند.

همان‌طور که قبلاً نیز گفته شد، ماشین‌های بردار پشتیبان چند کلاسی در مقایسه با شبکه‌های عصبی و مصنوعی یا الگوریتم بیز، اندازه نمونه‌های آموزش مورد نیاز را به میزان قابل توجهی کاهش داده‌اند. با این وجود، تعداد کمی نشت واقعی در شبکه وجود دارد. علاوه بر این، گاهی بازه زمانی حوادث نشت بسیار طولانی است. اگرچه داده‌های نشت واقعی دارای اعتبار هستند، اما آن‌ها الزامات تناسب و اندازه را برآورده نمی‌کنند. در نتیجه، تعدادی داده نشت مصنوعی تصادفی، به روش مونت کارلو از طریق شبیه‌سازی در مدل هیدرولیکی تولید شدند. فرآیند تولید نمونه‌های آموزش به صورت زیر است:

۱- در شرایطی که در شبکه نشت وجود ندارد، فشار برای تمام نقاط نظارت برداشت می‌شود:

$$p^0 = \{p_1^0, p_2^0, \dots, p_n^0\} \quad (21)$$

که در آن P_j^0 ، فشار در نقطه نظارت j ام در شرایط بدون در نظر گرفتن تقاضا برای نشت است.

۲- منطقه نشت i ام انتخاب می‌شود ($i=1,2,\dots,k$).

۳- توسط روش مونت کارلو در آن منطقه نشت، یک گره به طور تصادفی انتخاب می‌شود و به آن گره تقاضای اضافی (نشت) کمتر از $6/53$ لیتر بر ثانیه وارد می‌شود.

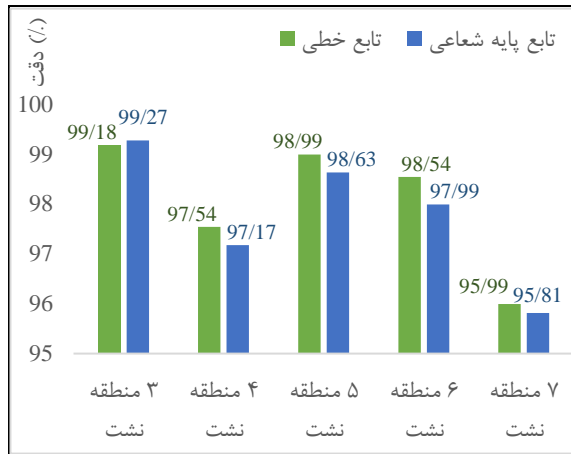
۴- مدل هیدرولیکی اجرا می‌شود و فشارهای جدید در نقاط نظارت مشاهده می‌شود و تغییرات فشارهای جدید با حالت بدون نشت محاسبه می‌شود.

۵- تقاضای اضافی (نشت) ایجاد شده در این گره برداشته می‌شود و مدل هیدرولیکی اجرا می‌شود.

مراحل ۳ تا ۵، L_i مرتبه در منطقه نشت i ام تکرار می‌شود. در این مطالعه L_i پنج برابر تعداد گره‌های منطقه i ام در نظر گرفته شده است. پس از آن که مراحل ۲ تا ۵ برای تمام مناطق نشت اجرا شوند، کلیه نتایج محاسبه شده در مرحله ۴ در ماتریس T ریخته می‌شود.



است و برای هر ۱۰ مدل تعیین شده است. در شکل ۲ نمودار میزان «دقت» برای مدل‌های آموزش‌دیده، نمایش داده شده است.



شکل (۲). نمودار میزان «دقت» برای مدل‌های آموزش‌دیده

ساختاری که برای ساخت مدل نهایی شناسایی منطقه نشت انتخاب می‌شود، باید علاوه بر ایجاد «دقت» کافی برای مدل، تعداد مناطق نشت بیشتری داشته باشد، زیرا نشت-یابی در مناطق کوچک‌تر، آسان‌تر می‌باشد. با توجه به شکل ۲، با این‌که دو مدل ساخته شده با سه منطقه نشت، بیشترین میزان دقت را نسبت به سایر مدل‌ها داشته‌اند، اما به اندازه کافی شبکه را به مناطق کوچک‌تر محدود نمی‌کنند. از طرفی مدل ساخته شده با پنج منطقه نشت و تابع هسته خطی، علاوه بر این‌که شبکه را به تعداد مناطق بیشتری تقسیم می‌کند، دقت قابل قبول ۹۸/۹۹٪ (تقریباً ۹۹٪) را نیز دارد. بنابراین در این مطالعه، مدل نهایی شناسایی منطقه نشت با ساختاری شامل پنج منطقه نشت و تابع هسته خطی آموزش داده شده است و با توجه به جدول ۱، مقدار پارامتر c بهینه برای این مدل ۲۱/۸ می‌باشد.

در شکل ۳ شبکه با پنج منطقه نشت تفکیک شده نمایش داده شده است.

آموزش و مابقی به نمونه‌های صحت‌سنجی تعلق گرفت. از نمونه‌های آموزش، برای آموزش مدل ماشین‌های بردار پشتیبان چند کلاسی استفاده شد و از نمونه‌های صحت-سنجی برای انتخاب ساختار بهینه (تعیین نوع تابع هسته، از میان توابع خطی و پایه شعاعی و تعیین تعداد مناطق نشت)، با توجه به نتایج طبقه‌بندی ماشین‌های بردار پشتیبان چند کلاسی به کار گرفته شد. سپس توسط ساختار بهینه و کل نمونه‌های پیش‌پردازش شده برای آن ساختار (نمونه‌های آموزش و نمونه‌های صحت‌سنجی)، مدل طبقه-بندی نهایی آموزش داده شد. در نهایت، توسط تغییرات فشار نقاط نظارت در زمان نشت واقعی، مدل طبقه‌بندی نهایی ساخته شده، آزمایش شد.

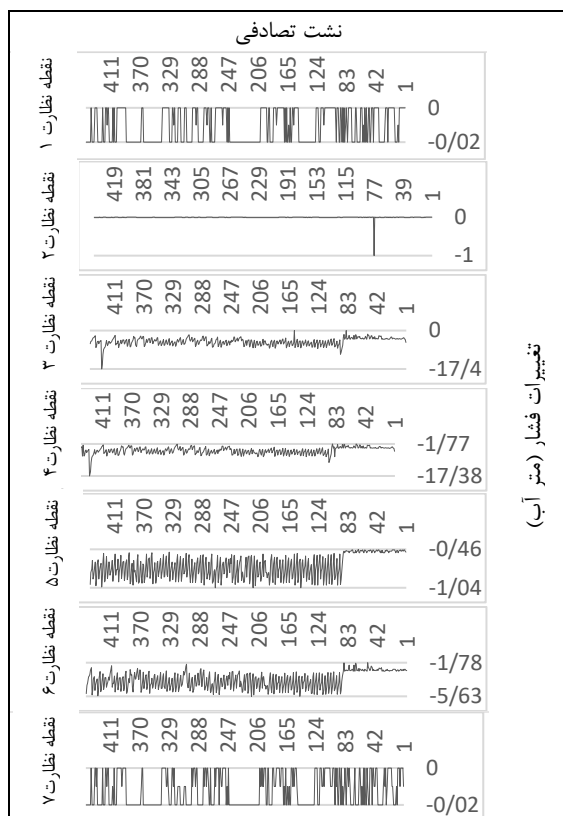
نتایج و بحث

در این مطالعه به منظور تعیین ساختار بهینه برای مدل شناسایی منطقه نشت، ابتدا با ۱۰ مدل با ۱۰ ساختار مختلف ایجاد شده است و برای هر ساختار پارامترهای بهینه (c, γ) مشخص شده‌اند. در جدول ۱ مقدار پارامترهای بهینه شده در ساختارهای مختلف، نشان داده شده است.

جدول (۱): مقدار پارامترهای بهینه‌شده در ساختارهای مختلف

مقدار پارامترهای بهینه‌شده	تعداد مناطق نشت	تابع هسته
$c = 0/4$	سه	خطی
$c = 65$	چهار	
$c = 21/8$	پنج	
$c = 0/1$	شش	
$c = 1/1$	هفت	
$\gamma = 0/5, c = 3$	سه	پایه شعاعی
$\gamma = 0/25, c = 18$	چهار	
$\gamma = 1/02, c = 4$	پنج	
$\gamma = 1/02, c = 4/5$	شش	
$\gamma = 1/38, c = 1/4$	هفت	

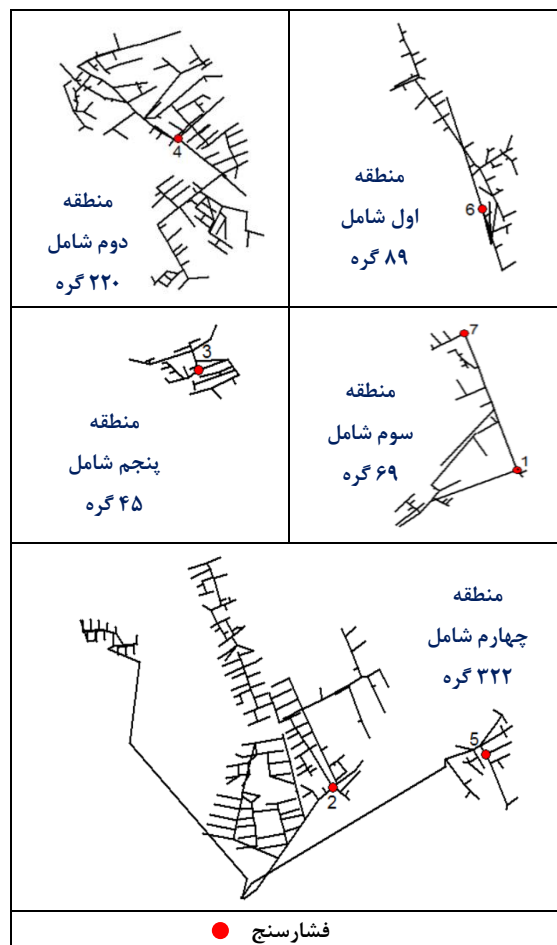
شاخص ارزیابی مدل‌های طبقه‌بندی چندکلاسی «دقت^۱» می‌باشد که مستخرج از «ماتریس درهم‌ریختگی^۲»



شکل (۴): نمودار نمونه‌های آموزش تولیدشده در منطقه اول

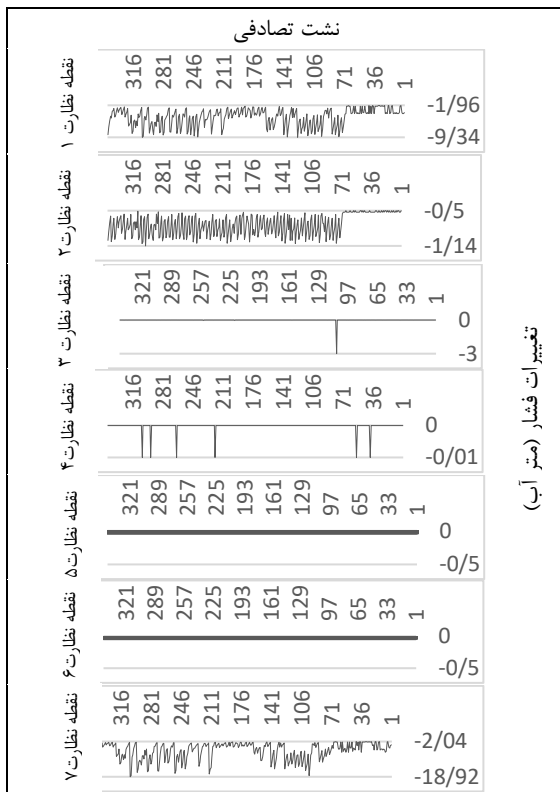
همان‌طور که در نمونه‌های آموزش موجود در شکل‌های ۴ تا ۸ مشاهده می‌شود، برای هر نمونه آموزش حداقل در یکی از نقاط نظارت، تغییرات فشار از آستانه (۰/۱۵ متر آب) فراتر رفته است و بیانگر وجود نشت در منطقه مورد نظر می‌باشد.

مدل طبقه‌بندی نهایی آموزش داده شده، توسط نشستی که در ساعت ۰۹:۲۷:۰۰ صبح ۱۳۹۸/۸/۲ در انتهای خیابان امام خمینی گزارش شده بود، آزمایش شد. زمان دقیق رخداد این نشت و همچنین مشخصات لوله دارای نشت و گره‌های متصل به آن در گزارشات ثبت نشده است. اما گره‌هایی از شبکه که در انتهای خیابان امام خمینی واقع شده‌اند، براساس خوشه‌بندی شبکه با پنج منطقه نشت، در منطقه چهارم قرار دارند. در شکل ۹ محدوده گره‌های واقع شده در انتهای خیابان امام خمینی در منطقه چهارم، توسط کادر قرمز نشان داده شده است.

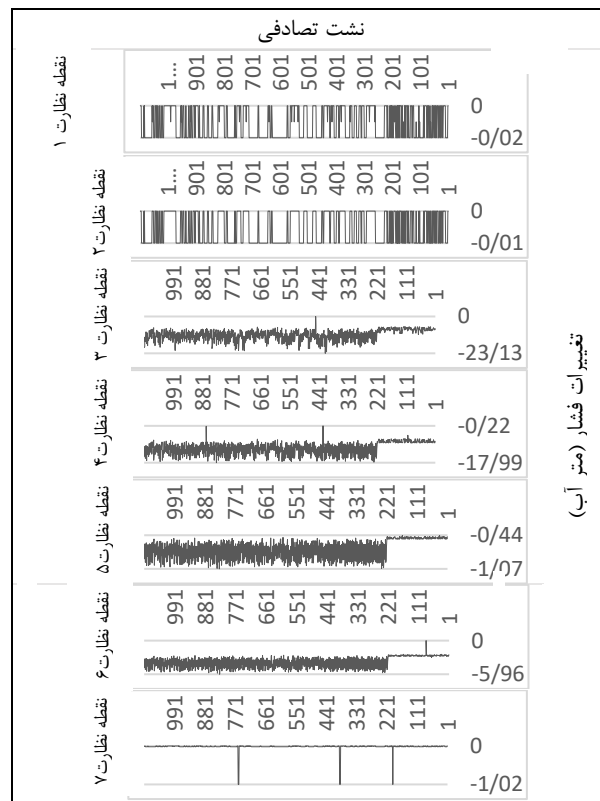


شکل (۳): شبکه با پنج منطقه نشت تفکیک شده

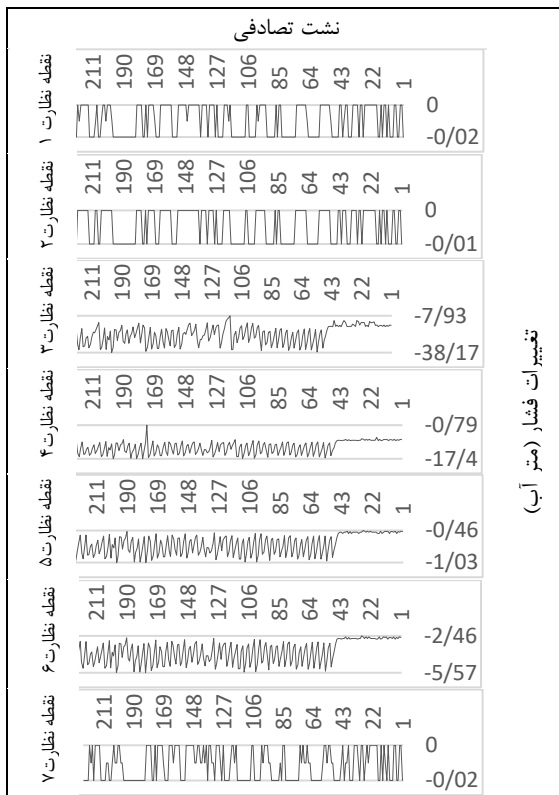
همان‌طور که گفته شد، مدل نهایی شناسایی منطقه نشت با ساختاری شامل پنج منطقه آموزش داده شد. در این ساختار مجموعاً ۳۷۲۵ نمونه در پنج منطقه تولید شد و پس از پیش‌پردازش نمونه‌ها، ۳۶۶۵ نمونه باقی ماند که به عنوان نمونه‌های آموزش این ساختار استفاده شد. در شکل‌های ۴ تا ۸ نمودار نمونه‌های آموزش تولید شده در هر یک از پنج منطقه، نمایش داده شده است.



شکل (۶): نمودار نمونه‌های آموزش تولیدشده در منطقه سوم

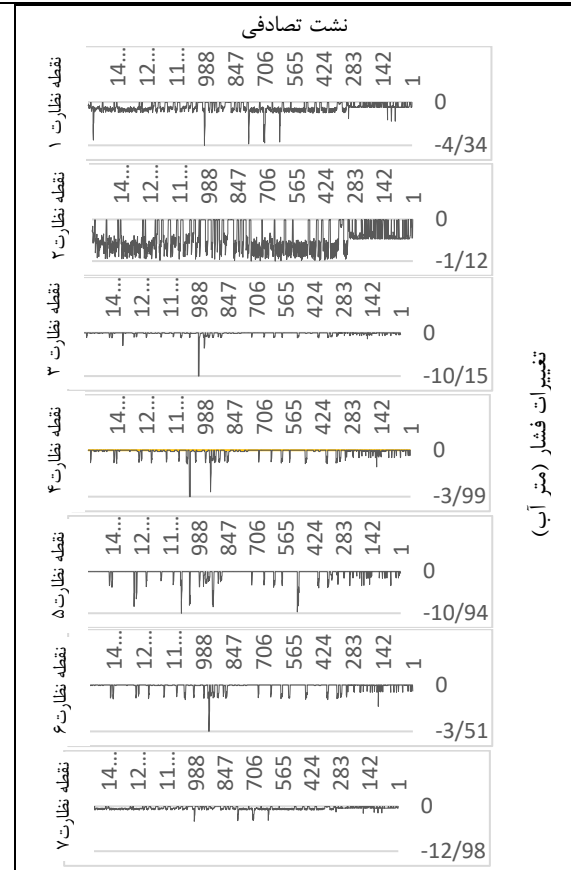


شکل (۵): نمودار نمونه‌های آموزش تولیدشده در منطقه دوم



شکل (۸): نمودار نمونه‌های آموزش تولیدشده در منطقه

پنجم



شکل (۷): نمودار نمونه‌های آموزش تولیدشده در منطقه

چهارم

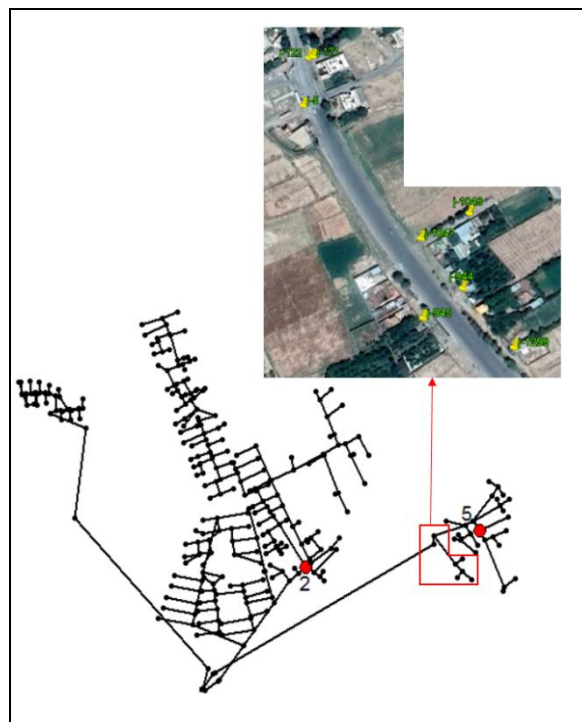


جدول (۲): تغییرات فشار نقاط نظارت در ساعات ۱۲ نیمه -

شب تا ۴ صبح دوم آبان ۱۳۹۸

نقاط نظارت	تغییرات فشار (متر آب)			
	۱۲ نیمه شب	صبح ۱	صبح ۲	صبح ۳
اول	۰	۰	-۰/۰۲	-۰/۰۲
دوم	۰	۰	-۰/۰۱	-۰/۰۱
سوم	-۰/۰۵	-۱/۰۲	-۲/۳۱	-۳/۳۱
چهارم	-۰/۰۵	-۱/۰۲	-۲/۲۸	-۳/۲۹
پنجم	-۰/۰۹	-۱/۶۷	-۳/۸۷	-۵/۸۳
ششم	-۰/۰۶	-۱/۰۲	-۲/۲۱	-۳/۱۴
هفتم	-۰/۰۱	۰	-۰/۰۲	-۰/۰۲

همان طور که در جدول ۲ مشاهده می‌شود، در ساعات ۱۲ نیمه شب تغییرات فشار کلیه نقاط نظارت از آستانه (۰/۱۵ متر آب) کمتر بوده است، در نتیجه تا این ساعت نشستی در شبکه رخ نداده است، اما در ساعات ۱ تا ۴ صبح، تغییرات فشار نقاط نظارت سوم، چهارم، پنجم و ششم از آستانه (۰/۱۵ متر آب) فراتر رفته است، که نشان‌گر وجود نشت در شبکه می‌باشد. همان‌طور که گفته شد، زمان دقیق رخداد این نشت مشخص نمی‌باشد، اما با توجه به تغییرات فشار نقاط نظارت، نشت در بین ساعات ۱۲ نیمه شب تا ۱ صبح اتفاق افتاده است. در نتیجه، پس از وارد کردن تغییرات فشار نقاط نظارت در ساعات ۱۲ نیمه شب تا ۴ صبح، به عنوان نمونه‌های آزمایش به مدل طبقه‌بندی نهایی، این مدل در ساعت ۱۲ نیمه شب هیچ منطقه نشستی را در شبکه اعلام نکرده است، اما برای ساعت ۱ تا ۴ صبح منطقه چهارم را به درستی به عنوان منطقه نشت تشخیص داده است. از طرف دیگر، با مقایسه بازه تغییرات فشار نقاط نظارت در نمونه‌های آموزش منطقه چهارم، که در شکل ۷ نشان داده شده است، با تغییرات فشار نقاط نظارت در هر یک از نمونه‌های آزمایش نشان داده شده در جدول ۲، می‌توان دریافت که نمونه‌های آزمایش ساعت ۱ تا ۴ صبح به ازای هر نقطه نظارت، در بازه تغییرات همان نقطه نظارت در نمونه‌های آموزش ایجاد شده در منطقه چهارم قرار می‌گیرند، که این نشان‌دهنده صحت عملکرد مدل طبقه‌بندی نهایی می‌باشد.



شکل (۹): محدوده گره‌های واقع شده در انتهای خیابان امام خمینی در منطقه چهارم

همان‌طور که گفته شد، زمان دقیق رخداد این نشت مشخص نمی‌باشد، اما از آن‌جا که در ساعاتی از شب تقاضای مصرف‌کنندگان بسیار کم می‌باشد و فشار در شبکه زیاد می‌شود، اغلب نشت‌ها در این ساعات اتفاق می‌افتند. براساس مدل هیدرولیکی کالیبره شده، از ساعت ۱۲ نیمه شب تا ۴ صبح، شبکه دارای حداقل جریان و حداکثر فشار می‌باشد و احتمال رخداد نشت در این ساعات بیشتر می‌باشد. بنابراین در جدول ۲ تغییرات فشار نقاط نظارت در ساعات ۱۲ نیمه شب تا ۴ صبح دوم آبان ۱۳۹۸، نشان داده شده است.

نتیجه گیری

در این مطالعه، یک روش شناسایی منطقه نشت مبتنی بر ماشین‌های بردار پشتیبان چند کلاسی ارائه شده است. ابتدا از الگوریتم خوشه‌بندی K-means برای تقسیم شبکه به تعدادی از مناطق نشت کوچک استفاده شد و شماره هر منطقه به عنوان برجسب طبقه‌بندی چند کلاسی در نظر گرفته شد. همچنین اثربخشی روش از دو جنبه‌ی تعداد مناطق نشت و نوع تابع هسته، مورد ارزیابی قرار گرفت. نتایج نشان داد که، تقسیم شبکه به پنج منطقه نشت و

استفاده از تابع هسته خطی، علاوه بر این که میزان قابل قبول ۹۸/۹۹٪ را برای شاخص ارزیابی «دقت» ایجاد می‌کند، شبکه را نیز به تعداد مناطق مناسبی تقسیم می‌کند. پس از تعیین منطقه احتمالی نشت، دیگر نیاز به جستجو در همه گره‌های شبکه توزیع آب نمی‌باشد و نشت را فقط باید در گره‌های منطقه نشت جستجو کرد. با این کار فضای جستجو کوچک می‌شود و به جستجو سرعت بخشیده می‌شود. روش ارائه‌شده را می‌توان با روش‌های رایج تشخیص نشت در یک DMA ترکیب کرد و در مورد گره نشت قضاوت کرد.

منابع

- Aksela, K., Aksela, M., & Vahala, R. (2009). Leakage detection in a real distribution network using a SOM. *Urban Water Journal*, 6(4), 279-289.
- Bohorquez, J., Alexander, B., Simpson, A. R., & Lambert, M. F. (2020). Leak detection and topology identification in pipelines using fluid transients and artificial neural networks. *Journal of Water Resources Planning and Management*, 146(6), 04020040.
- Cataldo, A., De Benedetto, E., Cannazza, G., Leucci, G., De Giorgi, L., & Demitri, C. (2017). Enhancement of leak detection in pipelines through time-domain reflectometry/ground penetrating radar measurements. *IET Science, Measurement & Technology*, 11(6), 696-702.
- Chamasemani, F. F., & Singh, Y. P. (2011). Multi-class support vector machine (SVM) classifiers-an application in hypothyroid detection and classification. In *2011 Sixth International Conference on Bio-Inspired Computing: Theories and Applications* (pp. 351-356). IEEE.
- Chapelle, O., Vapnik, V., Bousquet, O., & Mukherjee, S. (2002). Choosing multiple parameters for support vector machines. *Machine learning*, 46(1), 131-159.
- Chen, J., Feng, X., & Xiao, S. (2020). An iterative method for leakage zone identification in water distribution networks based on machine learning. *Structural Health Monitoring*, 1475921720950470.
- Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. learn.*, 20(3), 273-297.
- Costanzo, F., Morosini, A. F., Veltri, P., & Savić, D. (2014). Model calibration as a tool for leakage identification in WDS: A real case study. *Procedia Engineering*, 89, 672-678.
- Feng, J., & Zhang, H. (2006). Algorithm of pipeline leak detection based on discrete incremental clustering method. In *International Conference on Intelligent Computing* (pp. 602-607). Springer, Berlin, Heidelberg.
- Gong, J., Png, G. M., Arkwright, J. W., Papageorgiou, A. W., Cook, P. R., Lambert, M. F., ... & Zecchin, A. C. (2018). In-pipe fibre optic pressure sensor array for hydraulic transient measurement with application to leak detection. *Measurement*, 126, 309-317.
- Hu, X., Han, Y., Yu, B., Geng, Z., & Fan, J. (2021). Novel leakage detection and water loss management of urban water supply network using multiscale neural networks. *Journal of Cleaner Production*, 278, 123611.
- Kingdom, W. D., Limberger, R., and Marin, P. (2006). The Challenge of Reducing NRW in Developing Countries. In *WSS Sector Board Discussion, Paper No. 8*, World Bank, Washington, DC.
- MacKay, D. J., & Mac Kay, D. J. (2003). *Information theory, inference and learning algorithms*. Cambridge university press.
- Mamo, T. G., Juran, I., & Shahrou, I. (2014). Virtual DMA municipal water supply pipeline leak detection and classification using advance pattern recognizer multi-class SVM. *Journal of Pattern Recognition Research*, 1, 25-42.

- Mashford, J., De Silva, D., Burn, S., & Marney, D. (2012). Leak detection in simulated water pipe networks using SVM. *Applied Artificial Intelligence*, 26(5), 429-444.
- Milgram, J., Cheriet, M., & Sabourin, R. (2006). "One against one" or "one against all": Which one is better for handwriting recognition with SVMs?. *Training*, 195, 143-160.
- Mounce, S. R., Boxall, J. B., & Machell, J. (2010). Development and verification of an online artificial intelligence system for detection of bursts and other abnormal flows. *Journal of Water Resources Planning and Management*, 136(3), 309-318.
- Mounce, S. R., Mounce, R. B., & Boxall, J. B. (2011). Novelty detection for time series data analysis in water distribution systems using support vector machines. *Journal of hydroinformatics*, 13(4), 672-686.
- Poulakis, Z., Valougeorgis, D., & Papadimitriou, C. (2003). Leakage detection in water pipe networks using a Bayesian probabilistic framework. *Probabilistic Engineering Mechanics*, 18(4), 315-327.
- Quinones-Grueiro, M., Bernal-de Lázaro, J. M., Verde, C., Prieto-Moreno, A., & Llanes-Santiago, O. (2018). Comparison of classifiers for leak location in water distribution networks. *IFAC-PapersOnLine*, 51(24), 407-413.
- Quiñones-Grueiro, M., Verde, C., & Llanes-Santiago, O. (2019). Novel Leak Location Approach in Water Distribution Networks with Zone Clustering and Classification. In *Mexican Conference on Pattern Recognition* (pp. 37-46). Springer, Cham.
- Refaeilzadeh, p., Tang, L., and Liu, H. (2009). *Cross-validation*, Springer, Berlin, 532-538.
- Romano, M., Kapelan, Z., & Savić, D. A. (2010). Real-time leak detection in water distribution systems. In *Water Distribution Systems Analysis 2010* (pp. 1074-1082).
- Sanz, G., Pérez, R., Kapelan, Z., & Savic, D. (2016). Leak detection and localization through demand components calibration. *Journal of Water Resources Planning and Management*, 142(2), 04015057.
- Sophocleous, S., Savić, D., & Kapelan, Z. (2019). Leak localization in a real water distribution network based on search-space reduction. *Journal of Water Resources Planning and Management*, 145(7), 04019024.
- Vapnik, V. N., and Kotz, S. (1982). *Estimation of dependences based on empirical data: springer series in statistics*, Springer, Berlin.
- Weston, J., & Watkins, C. (1998). Multi-class support vector machines (pp. 98-04). Technical Report CSD-TR-98-04, Department of Computer Science, Royal Holloway, University of London, May.
- Wu, Z. Y. (2009). Unified parameter optimisation approach for leakage detection and extended-period simulation model calibration. *Urban Water Journal*, 6(1), 53-67.
- Zhang, Q., Wu, Z. Y., Zhao, M., Qi, J., Huang, Y., & Zhao, H. (2016). Leakage zone identification in large-scale water distribution systems using multiclass support vector machines. *Journal of Water Resources Planning and Management*, 142(11), 04016042.
- Zhou, Z. J., Hu, C. H., Xu, D. L., Yang, J. B., & Zhou, D. H. (2011). Bayesian reasoning approach based recursive algorithm for online updating belief rule based expert system of pipeline leak detection. *Expert Systems with Applications*, 38(4), 3937-3943.



Identification of leakage area in large water distribution systems (Case study: Mahan city in Kerman province)

Yasaman-Sadat Hashemi-Poor¹, Gholam-Abbas Barani², Ehsan Fadaei-Kermani³

Abstract

Today, leak detection is an important issue in water distribution networks, because leakage increases costs and causes a waste of water resources. In this study, a new method was used to identify the leak area in the water distribution network at part of Mahan city. First, the network was divided into a number of areas by the K-means clustering algorithm. Then, training samples related to each area were made using random leakage simulation in the network hydraulic model. The number of each area was used as the classification label of the multi-class support vector machine and along with the training samples, the classification model was taught. Finally, the trained model was used as a leak area identification model and was applied to determine possible leak areas with the observed field samples. The results show that out of 10 structures used to build the model, only three structures include; "Three areas and radial basis kernel function", "Three areas and linear kernel function" and "Five areas and linear kernel function", respectively, make an acceptable rate of 99.27%, 99.18% and 98.99% for the "Accuracy" evaluation index. On the other hand, among these three structures, dividing the network into five leak areas and using the linear kernel function divides the network into more areas and makes leak detection easier in restricted areas. As a result, this structure was used to build the final model.

Keywords: Leakage area identification, Water distribution network, Mahan city, Epanet, Matlab.

¹ M.Sc. Graduate of Water and Hydraulic Structures, Department of Civil Engineering, Faculty of Engineering, Shahid Bahonar University, Kerman, Iran, 03433220054, yas.hp@eng.uk.ac.ir.

² Professor, Department of Civil Engineering, Faculty of Engineering, Shahid Bahonar University, Kerman, Iran, 03433220054, gab@uk.ac.ir. (Corresponding author)

³ Assistant Professor, Department of Civil Engineering, Faculty of Engineering, Shahid Bahonar University, Kerman, Iran, 03433220054, e.fadaei@uk.ac.ir.