ONE-CLASS SVM – LEAK DETECTION IN WATER DISTRIBUTION SYSTEMS

Roya Cody[1], Sriram Narasimhan[2], Bryan Tolson[2]
[1] PhD candidate, [2]Associate Professor, Civil and Environmental Engineering, University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1
[1]rcody@uwaterloo.ca

ABSTRACT

Acoustic leak detection in water distributions systems has been reviewed and validated for decades in various laboratory and field settings. However, the existing systems rely heavily on detailed knowledge of the pipe system, an assumption of ideal conditions, as well as direct access to infrastructure pipelines. This paper presents an experimental investigation that addresses the need of minimally invasive water distribution monitoring in cold climates. Monitoring in cold climates is achieved with a permanent dry barrel hydrant mounted passive sensor system. The sensor system sits within the water column while still being accessible via the hydrant. Lab tests utilize a retrofitted hydrant and pipe system. Experiments show the effectiveness of using fire hydrant mounted sensors in leak detection. Acoustic signals due to simulated leaks are measured, and a one-class support vector machine (OCSVM) classification methodology is applied. Results showed that a simulated leak can be detected with a 97% classification accuracy.

Keywords: One Class SVM, Water Distribution System, Dry Barrel Fire Hydrant Mounted Sensor

1 INTRODUCTION

Growing populations, climate change, and deteriorating water supply infrastructure are exerting unprecedented demands on water resources worldwide. A core component of water supply infrastructure includes the extensive network of aging underground pipes that deliver treated drinking water to consumers across cities. Facing this reality government/ regulatory bodies and water utilities are becoming increasingly aware of the importance of effectively assessing and controlling water losses. Water losses, referred to as unaccounted-for-water or non-revenue water, are categorized into physical losses and non-physical losses from their water distribution networks (WDNs). Physical losses include leakage in transmission and distribution lines, leakage and overflows at storage tanks, and leakage on service connections up to consumer meters, while non-physical losses may include unauthorized consumption or metering inaccuracies.

While the primary cause of leaks in water distribution pipelines are largely speculative, it is widely assumed that the main contributing factors include temperature (seasonal freeze-thaw), water demand stress, the occurrence of hydraulic transients, and pipeline deterioration and corrosion. Most leak, large or small, detection methods in use today are inspection based and not meant for long-term monitoring. These inspection techniques are fairly accurate and well established, however, they are time consuming and therefore are quite costly. They involve an inspector on site with their inspection tool of choice to inspect regions of pipe one section at a time. Large lengths of pipe are routinely excavated in order to find and repair small defective sections, often as a result of complaints once a leak (or burst) has surfaced and is evident visually. While effective, many current inspection methods are intrusive and often require a part of the WDN to be shut down temporarily during investigations. Long-term monitoring for leaks requires a fundamental re-thinking of both the technology as well as the application procedure. Ideally, results of such long-term monitoring should then be followed by inspection methods to pin-point the exact location(s) and repair. Any alterations to existing water distribution infrastructure must be done mindfully so as to not disrupt the pipe system. Most importantly, such a monitoring system should be be capable of operating year-round. While the cost may be higher for the initial installation, the long-term costs would be reduced when the detrimental effects of leaks and the cost of spot inspections across the system are taken into account.

While acoustic leak detection in water distribution systems is a well-researched topic [1, 2, 3], a long-term monitoring system capable of year-round operation in cold climates is unavailable. From an analysis point of view, most well-established research methodologies fall primarily into the model or predictive-based methods, briefly discussed below. Literature pertaining to data-driven learning methods with water system applications is comparatively limited. A number of machine learning methods can be applied to leak detection in pipes, the most common and promising of which include Support Vector Machines (SVM), Neural Networks (NN), and Bayesian Learning. The application of a classification method involves knowing the number of classes (e.g., a leak versus no leak class), gathering system response data across all classes and then an extensive supervised training period. In the lab, conducting these experiments to gather the necessary training data across all classes can be a challenge. In the field, it is usually not practical or possible to collect enough system response data in all classes making it quite difficult to implement these methodologies in the field. In this paper, we present the identification of leaks as an event detection task, carried out through anomaly detection using a one-class support vector machine (OCSVM).

In what follows, we describe the related state of the art relevant research, and present the methodology and results for application of a OCSVM in a laboratory setting.

1.1 Review of Related Work

A number of studies, e.g., [4], have applied artificial neural networks (ANNs) to detecting pipe bursts. This process yields effective classification, however, significant historical data is generally needed for the training process. This is not always available, limiting its application, or requiring an extended training period. [5, 6, 7] also reviewed the application of ANN method and yielded promising results for pipe burst detection, however the same limitations were found. ANN is a good method for obtaining reasonable predictions, however when applied to WDNs, extensive data history and extended training periods, making this training process generally very computationally expensive. In the realm of model-based data-driven methods, [8] found that the adaptive Kalman filtering improved the performance of ANNs while reducing the training period time.

SVMs as binary classification method has been reviewed by many researchers, including [9, 10, 11]. Others, such as [12],who reviewed traditional two-class SVM in the time domain, in which leaks and non leak cases are trained and tested, yielded high accuracies of 97\%. This was done in a field case study situation and used an interesting Principal Component Analysis (PCA) based feature selection method. [13] yielded similar results of 78% - 94% accuracy when applying k-nearest neighbor (KNN), SVM and Gaussian mixture (GM) models for the binary classification of leak and non-leak cases. Another non-numerical modeling method, namely Bayesian inference, involves the probabilistic classification of a current state belonging to one of the previously known cases, the case in which the current state shares the highest probability. This has been reviewed extensively by [14].

A major limitation in all the aforementioned classification techniques is the necessity of known classes. While studies have effectively classified cases with high accuracy, they all require knowledge of all possible cases during the training period. This is not always readily available. Furthermore, it is not robust or easily adaptable.

This paper presents the application of an anomaly detection algorithm in order to detect the presence of a leak. This method involves modeling the normal state of the system, enabling deviations from the known norm to be detected, i.e. the detection of an event which strays from the system's normal state. This is an advantage compared to traditional classifications algorithms as it can detect previously unknown events. The one-class SVM methodology developed by [15] has been applied to a number of structural health monitoring applications, however it has not yet been applied to water, WDNs, or WDS event detection.

2 OVERVIEW OF LEAK DETECTION METHODOLOGY

The event detection methodology presented in this paper was utilized to study the effectiveness of a fire hydrant mounted hydrophone sensor, as seen in Figure 1, for leak detection with a OCSVM used for data processing.

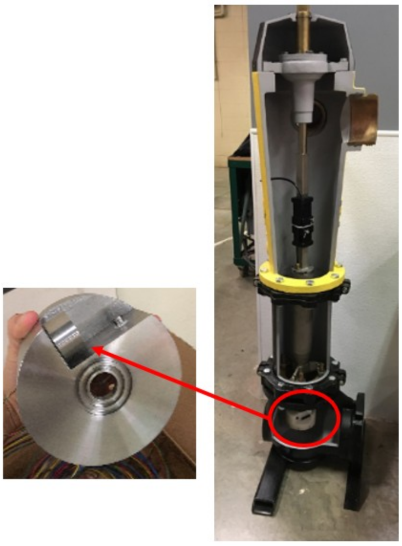

Figure 1. Fire hydrant mounted hydrophone sensor

The event detection approach requires the following four steps, the first three of which pertain to training and the fourth encompasses testing. These steps are as follows,

1) acquisition of raw hydro-acoustic data from the water distribution system using the hydrant mounted sensor;
2) acoustic data feature extraction;
3) statistical pattern recognition via a OCSVM classification scheme using the above features, training solely with ambient data; and,
4) testing on ambient and non-ambient cases.

This classifier is trained using feature sets extracted from time windows of raw hydro-acoustic data gathered from tests on the baseline condition of the system.

2.1 Raw Data Acquisition

The hydrant mounted sensor system overcomes many challenges and limitations associated with long term passive monitoring of WDN systems. The primary benefit in the use of hydrophones is that leaks typically generate frequencies which can be carried much further inside the pressurized water pipes. The associated hardware for processing includes a raspberry Pi in which the data is acquired and stored onto, wirelessly transmitted and automatically processed on another computer.

A laboratory experimental set up was designed to evaluate this sensor system in a section of a water distribution system, as seen in Figure 2. In this system the simulated leaks are represented by ½ inch valves. This system relatively represents field conditions, as it links to the city's water system, this all pressure fluctuations and background noise are relatively representative of conditions that will be experienced in the field.

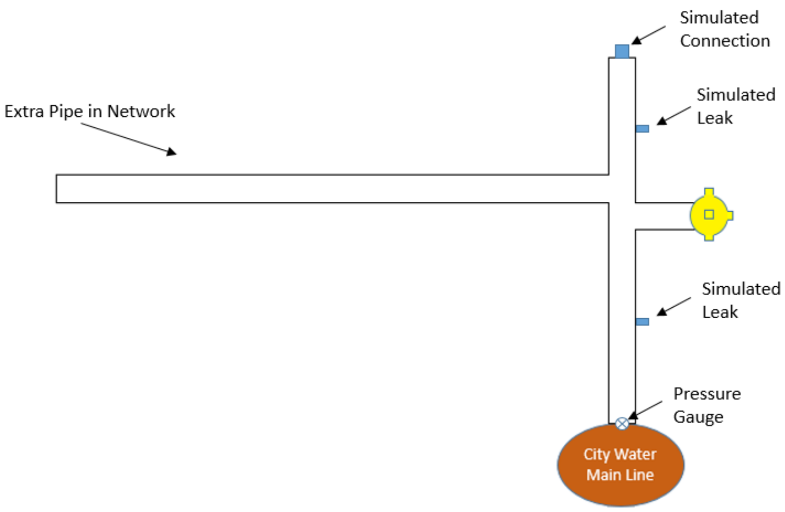

Figure 2. Laboratory Test Setup Schematic, approximately 50 ft in length

Baseline data is collected while the system is pressurized and all valves are closed. This is required to train the classifier prior to the live monitoring phase. This baseline is not meant to represent the perfectly intact, original state of the system. In fact, it captures the current state of the system in order to determine when it degrades from its current state. As such, it is not meant to be able to determine the presence of existing leaks, only the development of new leaks should be detected in this system.

A parametric study was carried out in order to determine the effect of time resolution on classification accuracy. Data sets of 30 seconds were collected, thus time length were varied up to 30 seconds. Experimentally it was determined that 1.25 second data set lengths yielded the highest accuracies, with anywhere from 1 to 3 seconds yielding excellent accuracies.

The raw data consists of data sets 1.25 seconds in length, sampled at 1350 Hz, and is pre-processed by shifting the mean to zero and removing outliers at 3 standard deviations from the data; this can be seen in Figure 3. The fire hydrant mounted custom hydrophone also includes 20 dB gain amplification built in.

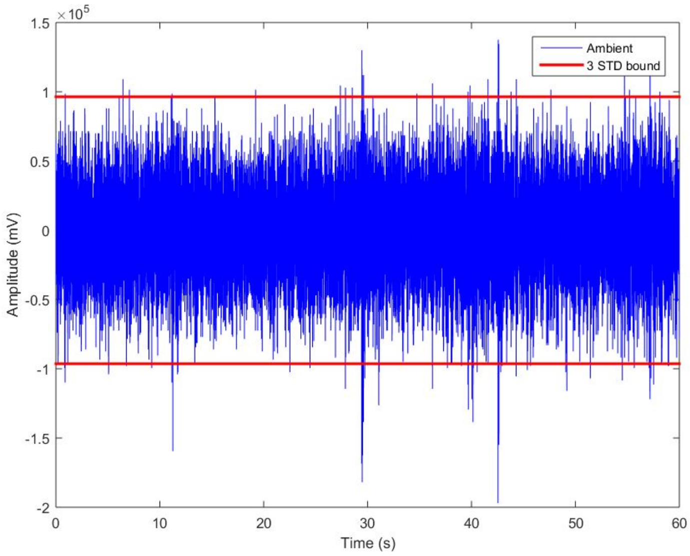

Figure 3. Sample time domain of ambient data, with 3 STD threshold

2.2 Feature Extraction

The statistical pattern recognition methodology applied in this paper is one-class kernel based, utilizing information pertaining only to the sensor at one specific location. That is, the selected features must not require comparison with previously computed features and must only require information from the one location and not multiple sensor data sets. The features used for the classification analyses are listed in Table 1. Among all features considered, using the mutual information algorithm [16] was applied in order to determine which features were the most statistically independent of one another, this yielded the list of 6 features found in Table 1. These features are then converted into a vector that represent each 1.25 second data set. These feature vectors provide more comprehensive information than any one feature individually [17].

Table 1. Time Domain Features

Feature	Expression
Maximum	$x_{mx} = \max(\lvert x_i \rvert)$
Minimum	$x_{mn} = \min(\lvert x_i \rvert)$
Standard Deviation	$x_{sd} = (\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \frac{\sum x_i}{n})^2)^{1/2}$
Root-Mean Square	$x_{rm} = (\frac{1}{n}\sum_{i=1}^{n} x_i^2)^{1/2}$

Amplitude Square	$x_{as} = \sum_{i=1}^{n} x_i^2$
Root Amplitude Square	$x_{ra} = (\frac{1}{n}\sum_{i=1}^{n} \lvert x_i \rvert^{\frac{1}{2}})^2$

2.3 Anomaly Detection

Anomaly detection is a small subclass of machine learning within supervised and unsupervised learning methods. It encompasses the field of outlier detection. It is the identification of events which do not conform to an expected pattern in a data set. This is extremely useful for applications in which the damaged data is not available a priori. It is typically unrealistic to train for all possible scenarios, since in many complex systems all scenarios simply are not known, similarly their signatures may change based on proximity or system material. This application of an anomaly detection algorithm as opposed to a multi-class classification algorithm is more realistic for complex, real-world systems.

In this experimental test case, in order to maximize the robustness of implementation of the leak detection system, an unsupervised anomaly detection system was chosen for this application. This includes over half a dozen possible general algorithms. The preference of different methods depends primarily on the data set and parameters. The different methods have little advantage over one another when compared across many data sets and parameters. Since the long term field implementation of this system will comprise of a relatively small data set, the OCSVM learning methods was selected as a well-established anomaly detection algorithm which is ideal and applicable given the data set.

The feature sets are used as input in the OCSVM anomaly detection methodology. This method requires only training data from the baseline state of the system in order to determine if a new instance feature vector is abnormal. This is done by modeling the feature vectors in a non-linear feature space and finding the location and size of the circular hyperplane that encompasses the baseline case. It is problem related to density estimation, enclosing an area of high density. The boundary is chosen based on the probability of a point landing within the region.

2.4 OCSVM Background

Building on the foundation of the two class support vector machine method [18], the OCSVM methodology maximizes the distance from the spherical hyperplane to the origin. Essentially it created a two class SVM in which the training point are represented as one class, and everywhere else represents the second class. The minimization function changes slightly creating a hypersphere characterized by a center, *a,* and a radius, R, in which R^2 will be minimized [19].

The original two class SVM has an objective function minimization formula as follows,

$$\min_{\omega,b,\xi_i} \frac{\|\omega\|^2}{2} + C\sum_{i=1}^{n} \xi_i$$

Subject to: $y_i(\omega^T\phi(x_i)+b)\geq 1-\xi_i$, $\xi_i\geq 0$

Where *y* is +1 or -1 depending on classification, ω is the normal vector to the hyperplane, $\phi(x)$ is the kernel transform in hyperspace, *b* is related to the offset of the hyperplane from the origin, ξ is the minimum number of nonnegative numbers, and *C* is the parameter which is used to determine the hyperplane's smoothness.

The OCSVM's quadratic program then becomes [19],

$$\min_{R,a} R^2 + C\sum_{i=1}^{n}\xi_i$$

Subject to: $\|x_i - a\|^2 \leq R^2 - \xi_i$, $\xi_i \geq 0$

3 EXPERIMENTAL RESULTS

The efficacy of leak detection using the OCSVM methodology will now be demonstrated using data acquired from the experimental laboratory test set up depicted in Figure 2, developed specifically for the purpose of testing and validating the leak detection methodology with hydrant mounted sensors.

Experiments were carried out in which acoustic data was recorded during a simulated leak event from a ½ inch value. The data set consists of 75 ambient cases and 30 leak cases, 60% of the ambient cases were used for training and 40% for testing, making the test set consist of 30 ambient and 30 leak cases. The system is also set up to determine the state to determine the state of a new instance as it is collected.

The leak detection via OCSVM anomaly detection yielded a 97% classification accuracy. This accuracy is detailed in the truth table shown as Table 2. A truth table is a better representation of accuracy, breaking it down into accurate and inaccurate classifications of the two cases; in which basic accuracy comes from the number of correctly classified points, or "True" values, while the misclassified points, or "False" cases, represent the remaining percentage of the points.

Table 2. OCSVM Classification Truth Table

Case	% of case correctly classified (True)	% of case incorrectly classified (False)
Ambient	97%	3%
Leak	97%	3%

These results are depicted in Figure 4, using a dimensionality reduction algorithm in order to represent the clusters in two dimensional feature space. These results show a very low false positive rate, and false negative rate. There is a trade-off between these two cases in which it is a waste of resources to deploy inspectors to check in a false negative scenario, however in a false positive scenario a leak is occurring and the system simply isn't detecting it. Thus ideally the false positive rate should be minimized above all else, provided the false negative is acceptable infrequent.

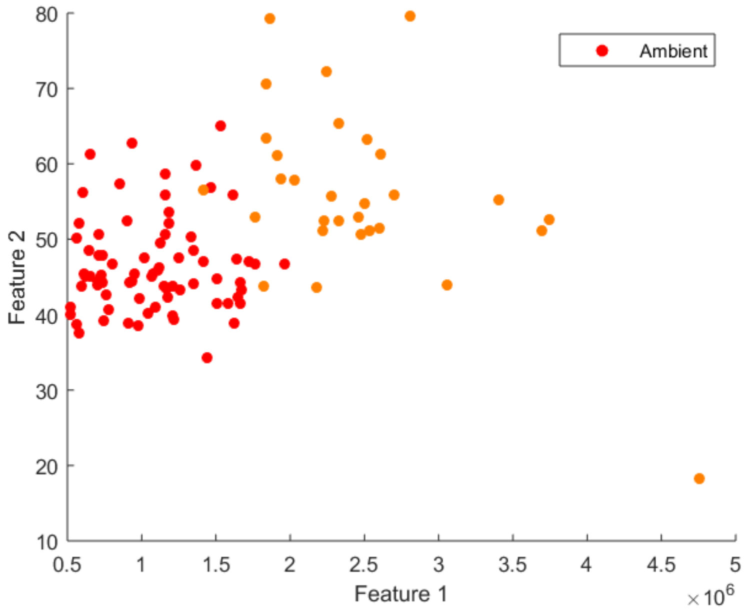

Figure 4. OCSVM Ambient versus Leak Classification Results, including both training and testing data sets

4 CONCLUSIONS

A novel application of OCSVM anomaly detection has been reviewed in this paper. Using this algorithm, a fully automated anomaly detection system has been implemented. The efficacy of the algorithm for this application is demonstrated using data taken from a field representative, laboratory test bed with a ½ inch valve used to represent a relatively small leak scenario. A binary decision boundary was created using only baseline data in order to classify if the new event is typical for baseline data or abnormal. In conclusion, this methodology is ideal for leak detection in water distribution networks since it is robust enough to be deployed at any location as its baselines as trained based on the location in which it is placed, and anomalies detected based solely off these baselines.

References

[1] Y. A. Khulief, A. Khalifa, R. Ben Mansour, and M. A. Habib, "Acoustic detection of leaks in water pipelines using measurements inside pipe," *Journal of Pipeline Systems Engineering and Practice,* vol. 3, no. 2, pp. 47-54, 2011

[2] O. Hunaidi, W. Chu, A. Wang, and W. Guan, "Detecting Leaks in Plastic Water Distribution Pipes," *Journal AWWA*, vol 92, no. 2, pp. 82-94, 2000

[3] J. Yang, Y. Wen, P. Li, and X. Wang, "Study on an improved acoustic leak detection method for water distribution systems," *Urban Water Journal,* vol. 10, no. 2, pp. 71-84, 2013
http://i.imgur.com/CNy9J.jpg, date accessed Nov 2016.

[4] S. R. Mounce, J. B. Boxall and J. Machelle, "Development and Verification of an Online Artificial Intelligence System for Detection of Bursts and Other Abnormal Flows," *Journal of Water Resources Planning and Management,* vol 136, no. 3, pp. 309-318, 2010

[5] K. Aksela, M. Aksela and R. Valhala, “Leak detection in a real distribution network using SOM,” *Urban Water Journal,* vol 64, (January), pp. 279-289, 209

[6] M. Romano, Z. Kapelan and D. Savic, “Real-Time Leak Detection in Water Distribution Systems,” *Journal of Water Distribution Systems Analysis,* pp. 1074-1082, 2011

[7] J. Yang, Y. Wen and P. Li, “Approximate entropy-based leak detection using artificaial neural network in water distribution pipelines,” *International Conference on Control Automation Robotics Vision (ICARCV),* pp. 1029-1034, 2010

[8] G. Ye, and R. A. Fenner, “Kalman Filtering of Hydraulic Measurements for Burst Detection in Water Distribution Systems,” *Journal of Pipeline Systems Engineering and Practice,* vol 2, no. 1, pp. 14-22, 2011

[9] A. E. U. Salam, M. Tola, M. Selintung, and F. Maricar, “A leak detection system on the Water Pipe Network through Support Vector Machine method,” *Makassar International Conference on Electrical Engineering and Informatics (MICEEI),* pp. 161-165, 2014

[10] J. Mashford, D. De Silva, D. Burns, and D. Marney, “Leak Detection in Simulated Water Pipe Networks Using SVM,” *Journal of Applied Artificial Intelligence,* vol 26, no. 5, pp. 429-444, 2012

[11] Q. Zhang, Z. Y. Wu, M. Zhao, and J. Qi, “Leak Zone Identification in Large-Scale Water Distribution Systems Using Multiclass Support Vector Machines,” *Journal of Water Resources Planning and Management,* vol 142, no. 11, pp. 1-15, 2010

[12] Y. Terao, and A. Mitra, “Robust water leakage detection approach using the sound signals and pattern recognition,” *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems,* vol 6932, pp. 69322D-69322D-9, 2008

[13] S. Rashid, U. Akram, and S. A. Khan, “WML: Wireless sensor network based machine learning for leak detection and size estimation,” *Procedia Computer Science,* vol 63, pp. 171-176, 2015

[14] S. Leu, and Q. Bui, “Leak Prediction Model for Water Distribution Networks Created Using a Bayesian Network Learning Approach,” *Journal of Water Resources Management,* vol 30, no. 8, pp. 2719-2733, 2016

[15] B. Scholkopf, J. Platt, J. Shawe-Taylor, A. Smola, and R. Williamson, “Estimating the support of a high-dimensional distribution,” *Journal of Neural Computation,* vol 13, no. 7, pp. 1443-1471, 2001

[16] G. Chandrashekar, and F. Sahin, “A survey on feature selection methods,” *Elsevier – Computers & Electrical Engineering,* vol 40, no. 1, pp. 16-28, 2014

[17] J. Long, and O. Buyukozturk, “Deccentralized one-class kernel classification-based damage detection and localization,” *Journal of Structural Control and Health Monitoring,* 2016

[18] V. Vapnik, “The Nature of Statistical Learning Theory,” *Springer science & business media*, 2000

[19] D. Tax, and R. Duin, “Support vector method for novelty detection,” *Journal of Machine Learning,* vol 54, no.1, pp. 45-66, 2004