



Zero Flow Rate Detection of Ultrasonic Water Meter Using Machine Learning Techniques

Alkım Gökçen^{*}, Bahadır Yeşil²

^{1*} BAYLAN Water Meters, Department of Research and Development, Izmir, TURKEY, (ORCID: 0000-0002-8131-388X), a.gokcen@baylanwatermeters.com

² BAYLAN Water Meters, Department of Research and Development, Izmir, TURKEY, (ORCID: 0000-0002-9622-2593), b.yesil@baylanwatermeters.com

(3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications June 11-13, 2021)

(DOI: 10.31590/ejosat.961090)

ATIF/REFERENCE: Gökçen, A. & Yeşil, B. (2021). Zero Flow Rate Detection of Ultrasonic Water Meter Using Machine Learning Techniques. European Journal of Science and Technology, (26), 477-481.

Abstract

This paper presents a classification-based method to detect false-positive ultrasonic sensor measurements when the flow rate is zero while the pipeline is full of water caused by the physical disturbances in flow metering. Due to the outlined disturbances, these false-positive readings of the meters may cause the wrong billing. To overcome this problem, ultrasonic sensor measurements are collected as timeseries data at variously different water flow rates. Numerical and statistical measures of the timeseries are computed to construct an input-output relation. Hence, the feature extraction process is performed. For the modeling phase, both zero flow rate and non-zero flow rate parts of the dataset, and its corresponding features are labeled. Logistic Regression (LR), Support Vector Machine (SVM), and Linear Discriminant Analysis (LDA) algorithms are employed to classify the flow status in the MATLAB environment. Model performances are compared in terms of accuracy, sensitivity, specificity, and precision. For the investigation of the availability of the implementation of an embedded system, model complexities are discussed. Selected model parameters are embedded in a water meter, and consumption values are compared to a water meter without the detection algorithm in the same test bench underfilled pipeline with zero flow rate condition. To simulate the physical disturbance conditions, and observe the effect of the false-positive detection algorithm on flow metering, the flow pipeline is vibrated periodically. Consumption loggings of the water meters are tabled, model performance results are discussed, and test results are shared. According to the results, the radial basis kernel SVM algorithm performs better in terms of all metrics. LR algorithm is employed for the real plant experiment when its model complexity is considered.

Keywords: Flow Metering, False positive detection, Ultrasonic sensor measurement, Classification, Machine Learning.

Makine Öğrenmesi Tekniklerini Kullanarak Ultrasonik Su Sayacının Duragan Su Tespiti

Öz

Bu çalışma, akış hattının su ile dolu olduğu ve akışın sıfır olduğu durumlarda fiziksel bozucu etkilerden kaynaklı yanlış-pozitif ultrasonik sensör okumalarının sınıflandırma tabanlı tespit metodu sunulmaktadır. Özetlenen bozucu etkiler nedeniyle bu sayaç yanlış-pozitif okumaları yanlış faturalandırmaya sebep olabilir. Bu sorunu aşmak için, ultrasonik sensör ölçümleri farklı su akış hızlarında zamanseri verileri olarak toplandı. Verilerin nümerik ve istatistiksel ölçümleri, bir giriş/çıkış ilişkisi kurmak adına hesaplanmıştır. Bu nedenle, nitelik çıkarma işlemleri yapılmıştır. Modelleme fazında akışın olup/olmadığı her iki durum için veriler ve denk geldikleri nitelikler etiketlenmiştir. Lojistik regresyon (LR), Destek Vektör Makineleri (DVM), ve lineer diskriminant analizi (LDA) algoritmaları akış durumlarını sınıflandırmak amacıyla MATLAB ortamında kullanılmıştır. Model başarımları doğruluk,

* Alkım GÖKÇEN: a.gokcen@baylanwatermeters.com

hassasiyet, özgüllük ve kesinlik performans metrikleri ile karşılaştırılmıştır. Ayrıca algoritmaların gerçek sistem çalışmalarında uygulanabilirliği model kompleksiteleri incelenerek tartışılmıştır. Seçilen model bir su sayacına uygulanmış, ve tüketim ölçümleri algoritmanın uygulanmadığı bir başka sayaç ile ayı test masasında karşılaştırılmıştır. Fiziksel bozucu etkilerin benzetimi için akış hattı düzenli olarak titreştirilmiştir. Sayaçların tüketim kayıtları tablolanmış, model performansları tartışılmış ve sonuçlar paylaşılmıştır. Sonuçlara göre, radyal tabanlı DVM algoritması bütün metrikler anlamında en iyi sonucu vermiştir. Model karmaşıklığı açısından oldukça basit olan LR algoritması gerçek sistem çalışması için seçilmiştir.

Anahtar Kelimeler: Akış ölçümü, Yanlış pozitif tespiti, Ultrasonik sensör ölçümü, Sınıflandırma, Makine öğrenmesi.

1. Introduction

Occurrence of anomaly and/or fault is a widely encountered situation in the real-world sensor measurement applications including the fields of medicine (Chauhan and Vig, 2015), biomedical (Jha et al., 2013), automotive (Capriglione et al., 2003), chemical process (Russel et al., 2012) and physical process (Amaral et al., 2015). Among these areas, measurement of the water flow using the piezoelectric materials is one of the commonly used methods (Kroemer, et al., 2019). Due to the speed of sound in the water medium is increased dramatically, water flow measurement using piezoelectric materials becomes challenging (Tawackolian et al., 2013). Furthermore, the detection of anomalies and/or faults is important for the sake of measurements, due to the possible disturbance factors. These can be arranged in to three as electromagnetic noise effects, ultrasonic aftershock wave reflection effects on the internal pipe line walls, and physical effects such as water hammer or externally occurred ones that cause the vibration and instant displacements.

Related works in the literature are investigated (Loureiro et al., 2016; Wu et al., 2016; Wang et al., 2020; Mounce et al., 2010; Palau et al., 2012; Mounce and Machell, 2006; Romano et al., 2010). A flow monitoring system is designed, and an outlier region definition-based algorithm is employed to detect anomalous of flow by Loureiro et. al. (2016). Wu et. al. (2016) developed a clustering-based detection algorithm which considers historical data of district flow area, and investigated the clustering performance on anomaly detection. Wang et. al. (2020) proposed to use a deep learning model which considers the extracted features of flow to predict the flow rate, and employed a thresholding method to detect bursts in the pipelines. An online flow rate modeling artificial neural network (ANN) based approach is proposed by Mounce et al. (2010) to detect bursts on a server where the flow rate data is received. Palau et. al. (2012) proposed to use principal component analysis to detect the bursts in water networks by investigating separation of the variation of the hourly collected flow data on the server. An Auto-regressive-integral-moving-average (ARIMA) type model is trained by considering the delayed-flow data and seasonality variables to model flow behavior, and an ANN-based classification algorithm is employed to detect bursts in the distribution system by Mounce and Machell (2006). Romano et. al. (2010) performed a Wavelet-based data pre-processing to train an ANN model which predicts the flow in water distribution system, and empolyses a thresholding method to detect possible leakage.

In this study, Time-of-Flight Difference (TOF-Diff) measurements are collected as a timeseries dataset using Time-to-Digital Converter (TDC). TOF-Diff dataset behavior is investigated and features are extracted by using the statistical methods on the MATLAB environment. Necessary pre-processing methods are carried out, and parameter

determinations are performed. Logistic Regression, Support Vector Machine (SVM) using both linear and Radial Basis kernel (RBK) and Discriminant Analysis methods are implemented to extract features to classify that the water is flowing or not. Model performances are compared and analyzed by investigating the performance metrics accuracy, specificity, sensitivity and precision. To test the algorithm on the real-plant application, an appropriate model for the embedded system on the water meter is selected considering not only the performance but also the model complexity. Two water meters, one is with the detection algorithm, flow readings are logged and the results are discussed.

Remaining parts of the paper are organized as follows: In Section 2, obtain dataset is explained. Firstly, statistical measures, which are computed for feature extraction process, are explained. Afterward, pre-processing procedure of the dataset is covered. Machine Learning models (ML) and traning procedures are explained briefly. Section 3 presents the train-test-validation phases, and model performance evaluations. Considered performance evaluation metrics are explained, and train/test results are shared. Real system experiment procedure is explained in Section 3 also. Inferences of the study are evaluated and possible future directions of the work are given in Section 4.

2. Material and Method

In this part of the study, necessary informations about dataset, preprocessing, and ML models are given.

2.1. Dataset

The dataset is collected over a TDC kit GP22-Development Kit produced by Acam-Messelectronic gmbH (Zuo et al, 2020). TOF data is processed in the GP22-EVA-KIT software, and logged to the computer. TOF-Diff value is logged with a sampling time of 0.1 seconds. By changing the water flow rate manually in an interval of 0-600 L/h, TOF-Diff values are collected as a timeseries data.

2.2. Feature Extraction

In this part of the paper, numerical and statistical methods are employed to obtain useful data used as features of the model (Tefai et al., 2020). These considerations of the feature extraction are Absolute Mean, Absolute Sample Difference, Standard Deviation, Skewness, Kurtosis and Squared-RMS Level. For a timeseries vector of 1-by-n, related methods can be computed as:

$$Absolute\ Mean = \frac{1}{N} \sum_{i=1}^N |x_i| \quad (1)$$

$$\text{Sample Difference} = \frac{1}{N} \sum_{i=2}^N |x_i - x_{i-1}| \quad (2)$$

$$\text{Standard Deviation} = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N - 1}} \quad (3)$$

$$\text{Skewness} = \frac{N}{(N - 1)(N - 2)} \sum_{j=1}^N \left(\frac{x_j - \bar{x}}{S}\right)^3 \quad (4)$$

$$\text{Kurtosis} = N \frac{\sum_{i=1}^N (x_i - \bar{x})^4}{\sum_{i=1}^N (x_i - \bar{x})^2} \quad (5)$$

$$\text{Squared RMS Level} = \frac{1}{N} \sum_{i=1}^N |x_i|^2 \quad (6)$$

2.3. Feature Extraction

Normalization of the dataset is performed to increase the convergence performance and the training speed (Jo, 2019). For this purpose, the unity normalization method is used. The normalization method that ranges the inputs-output pairs in to 0-1 interval is as shown:

$$\hat{x}_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (7)$$

where \hat{x}_i is the normalized timeseries.

2.4. Machine Learning Models

Supervised learning is a method that constructs a numerical expression between input and output pairs (Gökçen and Şahin, 2019; Taşören et al., 2020). LR is a widely used supervised ML algorithm that performs binary classification (Chen et al., 2019). LR is a statistical model that learns probabilistic behavior of a system, and performs binary regression using Logistic function. LR model can be described as the following equation:

$$y = h(W^T X + b) \quad (8)$$

$$h(x) = \frac{1}{1 + e^{-x}} \quad (9)$$

$$y = h(W^T X + b) \quad (8)$$

$$h(x) = \frac{1}{1 + e^{-x}} \quad (9)$$

where $h(x)$ is the hyperbolic tangent (Logistic) function, W is the $n - by - 1$ matrix, X is the $n - by - m$ feature matrix, b is the bias term, n is the number of regressors and m is the sample size. Optimal model parameters of W and b are computed using gradient descent optimization algorithm with momentum term (Liu et al., 2019).

To be able to identify a hyper-plane on the sample space, SVM algorithm considers a quadratic constrained optimization problem (Elmaz et al., 2020). The characteristic of this hyper-plane is defined by a function called kernel. SVM model and kernel functions can be described as:

$$f(x) = a_i K(x_i, x) + b \quad (10)$$

$$K(x_i, x_j) = x_i^T x_j \quad (11)$$

$$K(x_i, x_j) = e^{-\|x_i - x_j\|^2} \quad (12)$$

where $a_i (i = 1, 2, 3 \dots n)$ is the constant weights and, b is the bias term and $K(x_i, x_j)$ is the kernel function. In this work, both linear and RBK function is used to classify the ultrasonic measurements.

The goal of the Linear Discriminant Analysis (LDA) is to find a linear relation of the features that describes the classes of a given data-set using statistical measures (Li et al., 2020). It estimates the probability of input (feature) belonging to a class by maximizing the ratio of between class variance and within class variance. Increasing this ratio provides a separable sample space for each class. LDA considers the covariances of the classes of the dataset where the discriminant function can be explained as:

$$\delta_k(x) = x^T Q^{-1} \mu_k - 1 - \frac{1}{2} \mu_k^T Q^{-1} \mu_k + \log \pi_k \quad (13)$$

where subindex k represents each class, μ represents the means of the classes, π is the prior probability of membership in a class, x is the vector includes samples, and Q represents the covariance matrix.

K-fold cross validation is employed for the train-test procedures of the dataset. This method is used to assess the predictive performance of the models and calculate how they perform outside of the training data (Elmaz et al., 2019). The dataset is divided into K equal parts, and a model is trained for $K - 1$ part, and the remaining part of the dataset is used to test the trained models. In this work, K-fold cross validation method is used for $K = 10$.

3. Results and Discussion

In this part of the paper, related model achievements are investigated considering the model performance metrics. Also, a developed method is employed for test procedure on a test bench.

3.1. Modeling Results

Necessary model training procedures are performed in the MATLAB environment. Table 1 indicates that the performance metrics evaluations for each constructed model.

Table 1. Model performance metrics.

	Accuracy		Precision		Sensitivity		Specificity	
	Train	Test	Train	Test	Train	Test	Train	Test
Logistic Regression	99.3%	98.7%	0.9815	0.9839	0.9876	0.9839	0.9941	0.9880
Linear Kernel SVM	99.6%	99.4%	0.9817	0.9818	1.0000	0.9939	0.9941	0.9941
RBK SVM	98.4%	99.6%	0.9934	0.9878	0.9379	0.9939	0.9980	0.9961
LDA	91.6%	99.4%	0.9817	0.9818	1.0000	0.9939	0.9941	0.9941

Table 2. Test bench readings.

Time (min)	0	15	30	45	60	75	90	105	120	135	150	175
SW1 (L)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SW2 (L)	0.000	0.007	0.011	0.012	0.015	0.017	0.019	0.019	0.021	0.022	0.023	0.036

According to the results given in Table 1, employed model have close performances interns of all metrics. RBK-based SVM model has better accuracy performance on test scores. Evaluation of precision and specificity metrics indicates that RBK-based SVM model has better performance which means that this model is more succesfull on true-positive and true-negative cases. Evaluation of sensitivity metric depicts that the linear hyperplane-based models are expressively successful in detecting true positive data points. Although all methods are close performances, LR-based classification method will be employed for the test procedure due to the its less model complexity, ease of implement and less memory requirement.

3.1. Real Plant Test

Two water meters are connected on the test bench, and meters are labeled as “WM1” and “WM2”. Zero water flow rate detection algorithm is embedded in WM1, and WM2 is used with the default measurement mode. During the test bench study, a case, where the flow rate is zero while the pipeline is full of water, is realized. WM1 and WM2 readings are logged periodically as 15 minutes, and are as shown in Table 2. It may be said that the false-reading of the ultrasonic sensor caused by the physical disturbances can be easily detected by the proposed method. Although the reading differences are in tolerable region, it might be a significant difference for long term usages. In either case, larger amplitude disturbance effects might be removed.

4. Conclusions and Recommendations

In this work, a developed method for the detection of zero flow rate case false-positive ultrasonic sensor measurements of a water meter system is presented. A TDC module is used, and TOF-Diff samples are collected as timeseries to generate a dataset. For a predetermined window length, related statistical and numerical measures are computed and features are determined. To increase the model performances and training speed, a normalization method is applied to the dataset. Different ML algorithms are implemented, and performance metrics

accuracy, sensitivity, specificity and precision are computed to compare the model performances. Despite the SVM classifier performing better, other model performances are sufficient to detect water flow status. Thus, Logistic regression-based detection algorithm is selected for the real plant implementation work due to the model simplicity. Readings of a water meter with the embedded detection algorithm are compared to another simultaneously, and results are tabled. It is observed that the detection algorithm may detect the false-positive sensor measurements under zero flow rate conditions. Also, measurement due to the disturbances, such as external physical effects, can be detected by these methods according to the results. By this way, it might prevent not only the over-read bills of water consumptions, but also losses in utilities including the physical escape of water from the pressurized piping system as leakage. As a future direction, a method that detects true-negative measurements may be developed. Also, reinforcement learning based detection algorithms might be a new way of a research area for the measuring systems.

References

Chauhan, S., & Vig, L. (2015, October). Anomaly detection in ECG time signals via deep long short-term memory networks. In *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)* (pp. 1-7). IEEE.

Zhang, M., Raghunathan, A., & Jha, N. K. (2013). MedMon: Securing medical devices through wireless monitoring and anomaly detection. *IEEE Transactions on Biomedical circuits and Systems*, 7(6), 871-881.

Capriglione, D., Liguori, C., Pianese, C., & Pietrosanto, A. (2003). On-line sensor fault detection, isolation, and accommodation in automotive engines. *IEEE Transactions on Instrumentation and Measurement*, 52(4), 1182-1189.

Russell, E. L., Chiang, L. H., & Braatz, R. D. (2012). *Data-driven methods for fault detection and diagnosis in chemical processes*. Springer Science & Business Media.

Amaral, G. C., Garcia, J. D., Herrera, L. E., Temporao, G. P., Urban, P. J., & von der Weid, J. P. (2015). Automatic fault detection in WDM-PON with tunable photon counting OTDR. *Journal of Lightwave Technology*, 33(24), 5025-5031.

Kroemer, H., Oefelein, W., & Huenenberger, P. (2019). *U.S. Patent No. 10,458,824*. Washington, DC: U.S. Patent and Trademark Office.

Tawackolian, K., Büker, O., Hogendoorn, J., & Lederer, T. (2013). Calibration of an ultrasonic flow meter for hot water. *Flow Measurement and Instrumentation*, 30, 166-173.

Loureiro, D., Amado, C., Martins, A., Vitorino, D., Mamade, A., & Coelho, S. T. (2016). Water distribution systems flow monitoring and anomalous event detection: A practical approach. *Urban Water Journal*, 13(3), 242-252.

Wu, Y., Liu, S., Wu, X., Liu, Y., & Guan, Y. (2016). Burst detection in district metering areas using a data driven clustering algorithm. *Water research*, 100, 28-37.

Wang, X., Guo, G., Liu, S., Wu, Y., Xu, X., & Smith, K. (2020). Burst detection in district metering areas using deep learning method. *Journal of Water Resources Planning and Management*, 146(6), 04020031.

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.

Palau, C. V., Arregui, F. J., & Carlos, M. (2012). Burst detection in water networks using principal component analysis. *Journal of Water Resources Planning and Management*, 138(1), 47-54.

Mounce, S. R., & Machell, J. (2006). Burst detection using hydraulic data from water distribution systems with artificial neural networks. *Urban Water Journal*, 3(1), 21-31.

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).

Zuo, J., Luan, C., & Zhang, Y. (2020, June). TDC-GP22 high-precision time measurement based on FPGA and linear regression mapping. In *2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)* (Vol. 1, pp. 1284-1289). IEEE.

Tefai, H. T., Saleh, H., Tekeste, T., Alqutayri, M., & Mohammad, B. (2020, October). ASIC Implementation of a Pre-Trained Neural Network for ECG Feature Extraction. In *2020 IEEE International Symposium on Circuits and Systems (ISCAS)* (pp. 1-5). IEEE.

Jo, J. M. (2019). Effectiveness of normalization pre-processing of big data to the machine learning performance. *The Journal of the Korea institute of electronic communication sciences*, 14(3), 547-552.

Gökçen, A., & Şahin, S. (2019, October). Design of chaotic system based pacemaker on field programmable analog array board. In *2019 Medical Technologies Congress (TIPTEKNO)* (pp. 1-4). IEEE.

Taşören, A. E., Gökçen, A., Soydemir, M. U., & Şahin, S. Artificial Neural Network-Based Adaptive PID Controller
e-ISSN: 2148-2683

Design for Vertical Takeoff and Landing Model. *Avrupa Bilim ve Teknoloji Dergisi*, 87-93.

Chen, W., Sun, Z., & Han, J. (2019). Landslide susceptibility modeling using integrated ensemble weights of evidence with logistic regression and random forest models. *Applied sciences*, 9(1), 171.

Liu, Y., Gao, Y., & Yin, W. (2020). An improved analysis of stochastic gradient descent with momentum. *arXiv preprint arXiv:2007.07989*.

Elmaz, F., Büyükçakır, B., Yücel, Ö., & Mutlu, A. Y. (2020). Classification of solid fuels with machine learning. *Fuel*, 266, 117066.

Li, H., Zhang, L., Huang, B., & Zhou, X. (2020). Cost-sensitive dual-bidirectional linear discriminant analysis. *Information Sciences*, 510, 283-303.

Elmaz, F., Yücel, Ö., & Mutlu, A. Y. (2019). Evaluating the Effect of Blending Ratio on the Co-Gasification of High Ash Coal and Biomass in a Fluidized Bed Gasifier Using Machine Learning. *Mugla Journal of Science and Technology*, 5(1), 1-12.