Control Engineering Practice 107 (2021) 104691

Contents lists available at ScienceDirect

Control Engineering Practice

journal homepage: www.elsevier.com/locate/conengprac

A novel multi-leak sensor deployment strategy in water distribution networks based on the LSDR-JMI method

Juan Li [a], Cong Wang [a], Weidong Zhang [a], Changgang Lu [b,*]

[a] College of Communication Engineering, Jilin University, Changchun, China
[b] College of Automotive Engineering, Jilin University, Changchun, China

ARTICLE INFO

Keywords:
Optimal sensor deployment
Water distribution networks
Multi-point leak localization
LSDR-JMI
Multi-label feature selection

ABSTRACT

Water leakage in water distribution networks (WDNs) can cause serious economic and social impacts. Therefore, it is urgent to discover the efficient sensor deployment methods to monitor the leakage of WDNs. While related research studies mainly consider the single-point leakage, this paper firstly presents a pressure sensor deployment strategy to deal with the problem of the multi-point leakage occurred in WDNs. The novel proposed label space dimension reduction joint mutual information method (LSDR-JMI) is based on a filter multi-label feature selection algorithm. The label space dimension reduction (LSDR) method is based on the dependence of the label matrix on the leak cases to obtain the optimal label subspace. Then the optimal subset of feature matrix is calculated according to the relevant information between the optimal label subspace and the feature matrix. This method is applied to two hydraulic networks simulated by EPANET to demonstrate its effectiveness and advantage.

1. Introduction

Water distribution networks (WDNs) are critical to guarantee the transportation, distribution and provision of potable water in urban areas (Berkel, Caba, Bleich, & Liu, 2018; Sankar, Kumar, Narasimhan, Narasimhan, & Bhallamudi, 2015; Wang, Puig, & Cembrano, 2017). In many WDNs, water losses due to leaks are estimated to account for 30% of the total amount of extracted water (Kang, Park, Lee, Wang, & Eom, 2017; Puust, Kapelan, Savic, & Koppel, 2010). Hence, it is urgent and vital to carry out effective monitoring and real-time control management to stop the leakage from getting worse.

In Galuppini et al. (2019), through the research and analysis of hydraulic and automatic control, it was showed that real-time pressure control is a method often used in solving water supply pipe network problems, so as to inhibit the leakage of water supply pipe network. In Galuppini, Magni, and Creaco (2020), the authors provided further insight about the control algorithms to ensure safe operations. Effective monitoring sensor placement is the crucial problem of real-time control.

Recently, several studies have addressed the optimal sensor placement problem in WDNs to some extent. Most of the works focus on the contamination monitoring (Aral, Guan, & Maslia, 2010; Winter, Palleti, Worm, & Kooij, 2019). With regard to the issue of leak localization, fewer approaches have been proposed to address the problem of monitoring sensor deployment. For instance, in Farley, Mounce, and Boxall (2010), the identification method based on complete enumeration studies using hydraulic model simulations was proposed to determine the optimal locations of pressure sensor instruments for the detection of leak/burst events. In Sarrate, Nejjari, and Rosich (2012), a strategy based on the diagnosability maximization taking into account the fixed sensor configuration cardinality constraint was presented to optimally locate sensors. The sensor deployment problem was reformulated as a multi-objective optimization and the clustering technique was combined to reduce the size and the complexity of the optimization problem in Blesa, Nejjari, and Sarrate (2016). In Cugueró-Escofet, Puig, and Quevedo (2015), a sensor placement methodology for leak diagnosis in WDNs was presented, which considered real aspects of leak mislabeling, the obtainment of large number of possible leak locations and the practical application. Also, a hybrid feature selection method for identifying the optimal sensor placement in WDNs was introduced in Soldevila, Blesa, Tornil-Sin, Fernandez-Canti, and Puig (2018). More recently, in our previous work (Li, Wang, Qian, & Lu, 2019), the authors have proposed a semi-supervised strategy to solve the optimal sensor placement problem on account of partial leak locations may become lost or not be recorded.

To the best of our knowledge, in previous studies of optimal sensor placement techniques used in leak localization, it is mainly assumed that there is only single leak in WDNs. In other words, these works have overlooked the conditions that the leaks occurring on multiple positions. Besides, only one paper deals with the multi-leak detection and localization issue based on water transportation pipelines. The experimental water pipeline model in Kayaalp, Zengin, Kara, and Zavrak (2017) only contains 3 leakage nodes simulated by 3 taps. Other

* Corresponding author.
E-mail addresses: ljuan@jlu.edu.cn (J. Li), 2567933078@qq.com (C. Wang), 1339676841@qq.com (W. Zhang), lucg@jlu.edu.cn (C. Lu).

https://doi.org/10.1016/j.conengprac.2020.104691
Received 9 May 2020; Received in revised form 3 November 2020; Accepted 14 November 2020
Available online 27 November 2020
0967-0661/

multi-leak detection methods on water pipelines only consider the long straight pipeline (Zhou, Liu, Wang, She, & Lau, 2018) or the looping one (Torres, Jiménez-Cabas, González, Molina, & López-Estrada, 2020). However, the topology of WDN is complicated, and the deployment of pressure sensors needs to be further considered.

According to the previous discussion, the research in the aspect of multi-leak conditions in the WDNs is almost a blank. Although pressure sensors can be deployed everywhere in the WDN between every node, we can save a lot of money and effort by using only a few sensors to monitor the whole WDN. Besides, there are not only single point leakage, but also multi-point leakage, which cannot be handled by single point leakage monitoring. Therefore, in this paper, to solve this problem, the optimal sensor placement issue is firstly formulated as a multi-label feature selection problem. All columns of the feature matrix correspond to the pressure sensitivity of the nodes in the WDN. And the features selected are the nodes where the pressure sensors are ought to be deployed. The goal of our paper is to place the designated number of sensors to monitor multi-leak in WDNs.

Furthermore, there are three main categories of feature selection techniques surveyed in the literature (Guyon & Elisseeff, 2003; Saeys, Inza, & Larranaga, 2007): filter based methods (Ding & Peng, 2005; Hoque, Bhattacharyya, & Kalita, 2014; Peng, Long, & Ding, 2005), wrapper methods (Blanco, Larranaga, Inza, & Sierra, 2004; Li, Weinberg, Darden, & Pedersen, 2001; Ooi & Tan, 2003) and, embedded methods (Lu, 2019; Peralta & Soto, 2016; You, Liu, Li, & Chen, 2012). The filter methods are independent from the classifier design and tend to simultaneously select highly predictive but uncorrelated features. The advantages of these methods are the low computational cost and it is less likely to lead to over-fitting, while the main drawback is that the selection does not take into account the posterior use of the data by the model. Wrapper methods are the second type of methods which incorporate a classifier hypothesis into a close-loop search for an optimal feature subset. Due to the fact that a new model has to be built for each subset, the shortcoming of these methods are computationally expensive, but they usually provide the best results for the particular type of model used. The third type of methods, embedded methods, provide a trade-off solution between filter methods and wrapper methods by embedding feature selection into the model learning.

Based on the previous discussion and the complexity of WDNs, this paper has developed an optimal sensor placement strategy for multi-leak localization in WDNs which uses a filter multi-label feature selection algorithm. The destination is to reduce the computation cost of high-dimensional pressure data while maintaining the good performance of the selected monitoring nodes. Meanwhile, the proposed multi-label feature selection method is diverse from the previous studies. Fairly good conventional feature selection methods based on the criterion of mutual information capture the relevancy of each feature and each target variable, the average redundancy between each feature and the already selected features, and the average complementarity between the feature and the already selected features. The method in Sechidis, Spyromitros-Xioufis, and Vlahavas (2019) not only considered these 3 terms, but also captured various high-order target interactions. And these properties are both helpful to select the optimal features to solve the multi-label feature selection issue. However, in practice, the dimensions of feature matrix and label matrix are pretty large in monitoring sensor deployment problems of the WDNs. Therefore, it may lead to parameter adjustment time-consuming and be full of uncertainty. The proposed method in this paper has adopted a label space dimension reduction (LSDR) method to deal with the above problems. The pseudo-label is generated to reduce the dimension of original label space. And the low-dimensional label space can save the sufficient and effective relevant information of targets. Therefore, the algorithm has low time complexity and very strong practicability in the leak detection of the WDNs.

The rest of the paper is organized as follows. Section 2 introduces the concept of previous feature selection methods based on mutual information theory; Section 3 presents the proposed pressure sensor deployment strategy; Section 4 presents the results and their corresponding discussion, and Section 5 concludes this thesis.

2. Basic theory

In this paper, the sensor deployment problem is converted to the issue of multi-label feature selection. The nodes corresponding to the selected features are the sensor deployment locations. To formally describe the multi-label feature selection methods, $\mathbf{X} = [x_1, x_2, \ldots, x_d]$ denotes the d-dimensional domain of examples, and $\mathbf{Y} = [y_1, y_2, \ldots, y_q]$ denotes the q-dimensional domain of output. As a matter of convenience, the output space is binary, that is, the alphabet $\mathbf{Y}$ is $\{-1, 1\}^q$ where, $y_i = 1$ if and only if the example $\mathbf{x}$ has assigned the target y_i. The ultimate goal of the multi-label feature selection is to select a subset of $\mathbf{K}$ features $\mathbf{X}_\beta \subset \mathbf{X}$, where $\mathbf{X}_\beta$ contains useful information for the multi-label classification.

Mutual information (MI) is a measure of interdependence between variables. The mutual information of two discrete random variables X and Y can be defined as:

$$I(X;Y) = \sum_{y\in Y}\sum_{x\in X} p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right), \tag{1}$$

where $p(x, y)$ is the joint probability distribution function of X and Y, $p(x)$ and $p(y)$ are the marginal probability density functions of X and Y, respectively.

2.1. Single-label feature selection criteria

Single-label feature selection problem can be regarded as an original and special multi-label feature selection issue whose output $\mathbf{Y}$ is a one-dimensional vector. A stable strategy is the joint mutual information (JMI) criterion (Brown, Pocock, Zhao, & Lujan, 2012), where the score is conditional on the set already chosen:

$$J_{\mathrm{JMI}}(\mathbf{X}_k) = \sum_{\mathbf{X}_j\in\mathbf{X}_\beta} I(\mathbf{X}_j\mathbf{X}_k;\mathbf{Y}), \tag{2}$$

where I is the symbol of mutual information, $\mathbf{X}_k$ is the candidate feature, $\mathbf{X}_j$ indicates already selected feature, $\mathbf{X}_\beta$ is the collection of $\mathbf{X}_j$ and $\mathbf{Y}$ is the label matrix. Define $|\mathbf{X}_\beta|$ is the number of already selected features, the ranking equivalent expression for the criterion can be rewritten as

$$J_{\mathrm{JMI}}(\mathbf{X}_k) = I(\mathbf{X}_k;\mathbf{Y}) - \frac{1}{\left|\mathbf{X}_\beta\right|}\sum_{\mathbf{X}_j\in\mathbf{X}_\beta}\left(I(\mathbf{X}_k;\mathbf{X}_j) - I(\mathbf{X}_k;\mathbf{X}_j|\mathbf{Y})\right). \tag{3}$$

The process of proof about Eq. (3) can be found in Brown et al. (2012). The Eq. (3) contains 3 terms. The first term is the relevancy of each feature and each target variable, the second is the average redundancy between each feature the already selected features, the last is the average complementarity between the feature and the already selected features. From the intuitive expression of Eq. (3), the relevancy and the complementarity are helpful to select the competitive features. On the contrary, the redundancy indicates the information of the features to be determined is already included in the selected features, so it can be removed.

2.2. Multi-label feature selection criteria

For the single label classification problem, label $\mathbf{Y}$ is a column vector, and each row corresponds to the corresponding event of the feature matrix $\mathbf{X}$. The size of $\mathbf{X}$ is $\mathrm{M} \times p$. The size of $\mathbf{Y}$ is $\mathrm{M} \times q$. As for multi-label feature selection problems, label $\mathbf{Y}$ is a matrix. Each row of $\mathbf{Y}$ corresponds to the event of the feature matrix $\mathbf{X}$, and each column represents the label component. Sechidis, Nikolaou, and Brown (2014) puts forward two types of the JMI criterion. These two algorithms have adopted two completely different versions which rank the features according to the following scores:

$$J_{\mathrm{JMI}}^{\mathrm{Joint}}(\mathbf{X}_k) = \sum_{\mathbf{X}_j\in\mathbf{X}_\beta} I(\mathbf{X}_j\mathbf{X}_k;\mathbf{Y}), \tag{4}$$

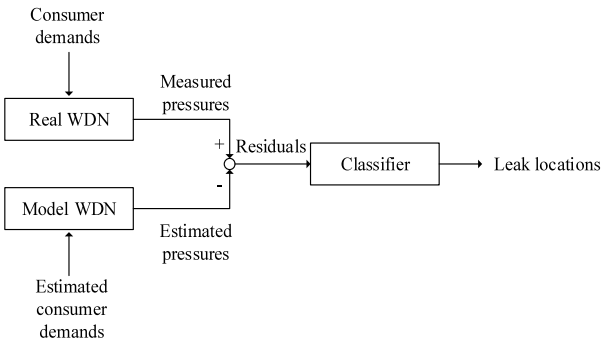

Fig. 1. Leak localization scheme.

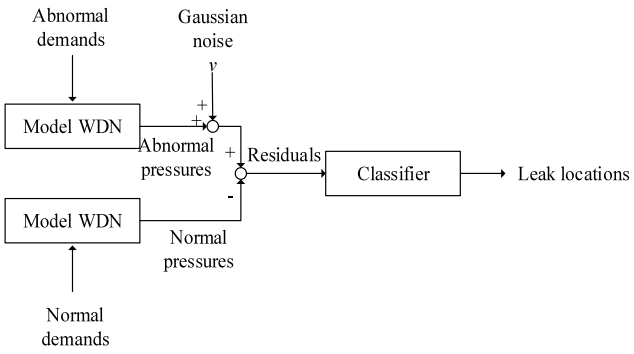

Fig. 2. The leak localization method.

$$J_{\mathrm{JMI}}^{\mathrm{Single}}(\mathbf{X}_k) = \sum_{\mathbf{X}_j \in \mathbf{X}_\beta} \sum_{\mathbf{Y}_i \in \mathbf{Y}} I(\mathbf{X}_j \mathbf{X}_k; \mathbf{Y}_i), \tag{5}$$

where $\mathbf{Y}_i$ is the ith column of $\mathbf{Y}$. These two versions of the JMI criterion can be seen as two extreme cases. The Eq. (4) considers the relevance of $\mathbf{Y}$ without taking into account the independence and original label $\mathbf{Y}$ is directly processed in the calculation. However, the Eq. (5) assumes all the components of $\mathbf{Y}$ are dependent and each column of label $\mathbf{Y}$ is processed in each calculation iteration. It is obviously that these two expressions both possess a few disadvantages to some extent. The main shortcoming of the Joint-JMI algorithm is that as the dimensionality of target $\mathbf{Y}$ grows, it is less reliable to estimate the MI expressions. Similarly, the Single-JMI method also considers one aspect of the target $\mathbf{Y}$ and thoroughly ignores the possible useful interaction among all the components of target $\mathbf{Y}$. In Zhou et al. (2018), the authors have shown that the Single-JMI algorithm outperforms the Joint-JMI algorithm and explained that the low-dimensional MI expression of the Single-JMI algorithm is more reliable to estimate than the Joint-JMI algorithm. On these foundations, Zhou et al. (2018) have changed the expression of Eq. (5) to derive a novel multi-label feature selection criteria, which is named Group-JMI method:

$$J_{\mathrm{JMI}}^{\mathrm{Group}}(\mathbf{X}_k) = \sum_{\mathbf{X}_j \in \mathbf{X}_\beta} \sum_{\tilde{\mathbf{Y}}_i \in \tilde{\mathbf{Y}}} I(\mathbf{X}_j \mathbf{X}_k; \tilde{\mathbf{Y}}_i). \tag{6}$$

The difference between Eqs. (5) and (6) is the specific form of the target. In Eq. (5), the target $\mathbf{Y}$ is the original label which consists of "−1" and "1". Different from this formulation, the target $\tilde{\mathbf{Y}}$ in the Eq. (6) considers the dependence of the target. The computing of the target $\tilde{\mathbf{Y}}$ is a two-step process:

Step 1: Creat q groups of variables $\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_q$ and the size of $\mathbf{Z}_i \subset \mathbf{Y}$ is controlled by the Proportion of Targets (PoT). Besides, the columns of $\mathbf{Z}_i$ are randomly selected from $\mathbf{Y}$. The size of $\mathbf{Z}_i$ is $\mathrm{M} \times k$ ($k = q \times \mathrm{PoT}$).

Step 2: The Group-JMI method clusters the rows of $\mathbf{Z}_i$ with "similar" output vectors:

$$\tilde{\mathbf{y}}_i = k\text{-}means(\mathbf{Z}_i, \mathrm{N}), \forall i = 1, 2, \ldots, q,\ n = 1, 2, \ldots, \mathrm{N}, \tag{7}$$

where the parameter N is the Number of Cluster which is a priori and $\tilde{\mathbf{y}}_i$ is the clustering result of $\mathbf{Z}_i$.

After these 2 steps, the clustering results of each group are spliced into a matrix, the original target $\mathbf{Y}$ is transformed into the new pattern of target $\tilde{\mathbf{Y}}$. Due to random sampling of the **Step 1** and the similar examples are clustered, relevance is also taken into account apart from the inherent independence. Consequently, Eq. (6) has better experimental results than Eqs. (4) and (5). The MI estimator in the Eq. (6) is the plug-in estimator:

$$\hat{I}(\mathbf{X}_i \mathbf{X}_k; \tilde{\mathbf{Y}}_i) = \sum_{x_i \in \mathbf{X}_i} \sum_{x_k \in \mathbf{X}_k} \sum_{\tilde{y}_i \in \tilde{\mathbf{Y}}_i} \hat{p}(x_i, x_k, \tilde{y}_i) \ln \frac{\hat{p}(x_i, x_k, \tilde{y}_i)}{\hat{p}(x_i, x_k)\hat{p}(\tilde{y}_i)}, \tag{8}$$

where $\mathbf{X}_i$ takes the value x_i, $\mathbf{X}_k$ takes the value x_k and $\hat{p}(x_i, x_k, \tilde{y}_i)$ is the maximum likelihood estimate of the joint probability.

The variation of PoT, NoC can lead to the different results. Besides, once the size of feature matrix or the label space is big, it is not convenient to adjust the values of PoT, NoC.

3. Proposed solution: a filter multi-label feature selection algorithm

The pressure data is the direct reflection of WDN operation state. Leakage will lead to the change of pressure value of each node. WDN monitoring can be realized by processing the pressure data. The multi-leak pressure sensor deployment problem in the WDNs is considered to be a matter of the multi-label feature selection issue performed in a sequential way. First, the pressure sensitivity matrix $\mathbf{S}$ is constructed. The pressure sensitivity matrix $\mathbf{S}$ is the feature matrix $\mathbf{X}$ in the multi-label feature selection methods and the corresponding leak positions form the label matrix $\mathbf{Y}$. Second, a subset of features is remained according to the expression of mutual information (MI). The corresponding nodes of the preserved features are the exact monitoring locations to deploy the pressure sensors.

In Ferrandez-Gamot et al. (2015), a leak localization method is proposed which is depicted in Fig. 1. Consumer demands are the actual water demands of users and the estimated consumer demands are obtained from human experience. Residuals are calculated as the differences between the measurements of pressure sensors installed inside the WDN on the conditions of leak cases and no leak cases. However, the data available from the real WDN is limited. In practice, it is difficult to obtain enough leak cases to generate useful outputs of the classifier. The way to obtain the feature matrix is by using a hydraulic simulator depicted in Fig. 2, which is similar to the one presented in Fig. 1. Normal pressure refers to the pressure of WDN without leakage, and abnormal pressure is obtained under the condition of leak cases. In Rossman (2000), EPANET is a software application used throughout the world to model water distribution systems. First, the node and pipe structure of the WDN need to be constructed in EPANET. Then, the parameters of WDN need to be determined. Finally, the leakage can be obtained by changing the node water demand to simulate the leakage. With the increase of leakage, the pressure value of the node decreases. Besides, in this paper, the leakage is simulated by adding extra nodal demands in several nodes. More node water demand means more leakage, which leads to greater pressure change of some nodes. The pressure residuals can be obtained by calculating the pressure differences under the conditions of leak cases and no leak cases.

3.1. Constructing the pressure sensitivity matrix

In this paper, the sensor deployment problem of multi-leak cases is converted to the multi-label feature selection problem. The construction of the pressure sensitivity matrix $\mathbf{S}$ is useful in terms of the leak detection and localization because the matrix $\mathbf{S}$ is the feature matrix of the feature selection problem. It is calculated according to the change of pressure in different leak cases in WDNs.

In the related studies, two effective solutions are adopted to simulate the leakage. One way is to set a positive emit coefficient at nodes,

the other way is to add additional nodal demands. In this paper, the latter method is used owing to the convenience of accurately adjusting the magnitude of leakage. In the experiment, the normal nodal demands $(d_1, d_2, \ldots, d_s)$ and the abnormal nodal demands $(d_1', d_2', \ldots, d_s')$ are input into the EPANET simulator. Besides, other necessary conditions (for instance the reservoir pressure and the nodes' elevations) are also input into the EPANET. Therefore, the normal pressure and the abnormal pressure can be acquired, respectively. While the leakage is too small in the WDNs, it is no need to detect the size and the location. On the contrary, the huge leakage can be noticed easily in practice. Consequently, the leakage is set within a given reasonable range in this paper. The number of junctions in the WDNs is set to s. Supposing that the matrix $\mathbf{P}$ is the normal pressure matrix, the matrix $\mathbf{P}'$ is the abnormal pressure matrix. The rows of $\mathbf{P}$ and $\mathbf{P}'$ are the leak cases and the columns of them are the corresponding nodes. Then the pressure residual matrix $\mathbf{\Delta P}$ can be calculated by

$$\mathbf{\Delta P} = \mathbf{P} - \mathbf{P}'. \tag{9}$$

Subsequently, there are certain measurement errors while using the pressure sensors. Gaussian random noise v is added to simulate the measurement errors of pressure sensors. The mean value of Gaussian random noise is 0, and the variance is 5% of the mean value of $\mathbf{\Delta P}$ and then squared. Finally, the pressure sensitivity matrix $\mathbf{S}$ is obtained on the condition that the size of the leakage D is known and D_i is the leak size of leak case i.

$$S_{ij} = \left(\Delta P_{ij} + v\right) / D_i, \tag{10}$$

where S_{ij} is the value of pressure sensitivity matrix $\mathbf{S}$ in row i, column j, ΔP_{ij} is the value of the pressure residual matrix $\mathbf{\Delta P}$ in row i, column j, v is Gaussian random noise and D_i is the leak size of the leak event on row i.

3.2. The LSDR-JMI algorithm

In this paper, the reduction of the label space dimension is vital to reveal the relevancy between the local label spaces. The LSDR method is based on the dependence of the label matrix on the leak cases. Besides, each column of the label matrix corresponds to the pressure data of each node in the WDN. The labels of some nodes in the WDN can reflect the leakage of the whole network to the greatest extent, and the local label space, namely optimal label subspace, composed of these nodes can eliminate the interference of some node labels on the leakage event, so that the subsequent feature selection results are more accurate. The process of the label space dimension reduction joint mutual information method (LSDR-JMI) method is depicted in Fig. 3.

(1) The generating of the pseudo-label

The solution of the pseudo-label is the preprocessing procedure of the LSDR. For the pressure sensor deployment issue, all the rows of the pressure sensitivity matrix $\mathbf{S}$ are the nodal pressure sensitivity sets of the different leak cases. And all the rows of the label space $\mathbf{Y}$ are the indication sets of the corresponding different leak cases. The k-means method is used because of its robustness to outliers and low time complexity.

$$\mathbf{L} = k\text{-}means(\mathbf{Y}, \mathrm{N}), \tag{11}$$

where N is the number of the clustering. The rows of the label space $\mathbf{Y}$ are clustered into several combinations. The rows of $\mathbf{Y}$ in the same combination are the same leak cases. And the original label space $\mathbf{Y}$ can be represented by a column vector $\mathbf{L}$ (pseudo-label) to calculate the optimal label subspace. The size of $\mathbf{L}$ is $\mathrm{M} \times 1$.

(2) The acquisition of the optimal label subspace

In the detection of the leakage in the WDNs, the label space can reflect the leakage situation that "1" is the leak occurs at the node and "−1" is no leak at all. The rows of the label space $\mathbf{Y}$ are the different leak cases. However, the whole label space may "prevent" the result of multi-label feature selection method. In this paper, leak cases are generated by random simulation. Due to the random simulation, the leakage times of each node are different. Some columns of the pressure sensitivity matrix $\mathbf{S}$, which are more related to the leakage cases, are retained, that is, the pressure of the nodes corresponding to these columns of $\mathbf{S}$ is the most effective information to monitor node leakage. The pseudo-label $\mathbf{L}$ is the clustering result of all cases. Then the optimal label subspace can be calculated according the contribution of each column of the pressure sensitivity matrix $\mathbf{S}$ to the pseudo-label $\mathbf{L}$ in the process of feature selection. As for the traditional methods like PCA, the solution of covariance matrix is inevitable. However, in this paper, the pressure values of the same node are close under different leakage conditions. After calculating covariance matrix, the eigenvalues may reach infinity. Hence the JMI method is used to reduce the dimension of the label space because feature selection methods based on mutual information theory only deal with each column of feature matrix and label matrix respectively.

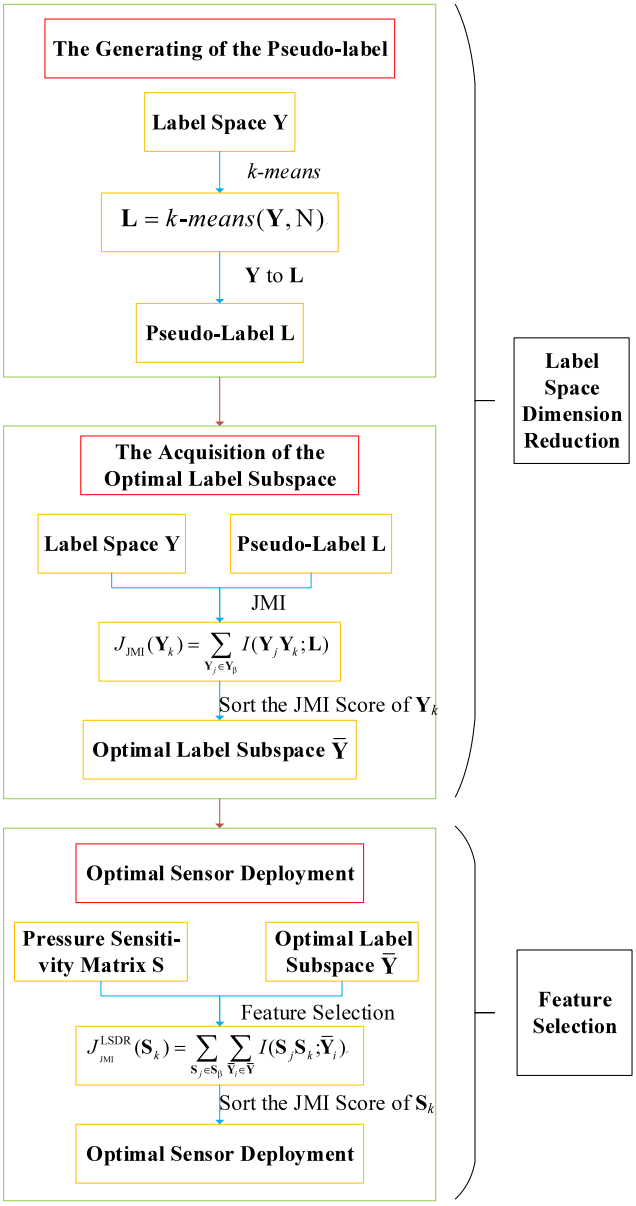

Fig. 3. The process of LSDR-JMI method.

According to Eq. (2), the LSDR method can be calculated by

$$J_{\mathrm{JMI}}(\mathbf{Y}_k) = \sum_{\mathbf{Y}_j \in \mathbf{Y}_\beta} I(\mathbf{Y}_j \mathbf{Y}_k; \mathbf{L}), \tag{12}$$

where $\mathbf{Y}_\beta$ indicates already selected the columns of the label space $\mathbf{Y}$. For convenience, Preservation Ratio (PR) is defined by using q*PR to control the ratio of saved columns of label space $\mathbf{Y}$. The benefit of Eq. (12) is that less relevant columns of $\mathbf{Y}$ with the pseudo-label $\mathbf{L}$ are removed. Apart from this, the time complexity can be reduced to a certain extent. All the selected columns of $\mathbf{Y}$ based on Eq. (12) formed the optimal label subspace $\overline{\mathbf{Y}}$.

(3) The optimal sensor deployment

The Group-JMI method in Eq. (6) captures the joint information of some groups of target variables to make up for the shortcoming of Single-JMI method. However, the solution of the novel categorical space $\tilde{\mathbf{Y}}$ is high time complexity and high computational complexity. The Group-JMI algorithm has sampled the set of target variables without replacement, and then the examples with "similar" output vectors are clustered. After that, the novel label space $\tilde{\mathbf{Y}}$ is generated with the same size of the original label space $\mathbf{Y}$. The pretreatment of $\tilde{\mathbf{Y}}$ has introduced the extra computing.

Based on the above discussion, this paper has adopted the LSDR-JMI method to select the pressure sensor deployment to reduce the time complexity of the algorithm and improve the accuracy of the algorithm. Just like the Group-JMI method, the equal-width discretization is used to discretize the values of the feature matrix. Next the pseudo-label $\mathbf{L}$ of $\mathbf{Y}$ is computed by Eq. (11). Then the optimal label subspace is derived according to Eq. (12). Finally, the feature selection of LSDR-JMI method can be rewritten as

$$J_{\mathrm{JMI}}^{\mathrm{LSDR}}(\mathbf{S}_k) = \sum_{\mathbf{S}_j \in \mathbf{S}_\beta} \sum_{\overline{\mathbf{Y}}_i \in \overline{\mathbf{Y}}} I(\mathbf{S}_j \mathbf{S}_k; \overline{\mathbf{Y}}_i). \tag{13}$$

According to the chain rule of MI, $I(\mathrm{AB;C}) = I(\mathrm{A;C}) + I(\mathrm{B;C} \mid \mathrm{A})$, the Eq. (13) can be represented as

$$J_{\mathrm{JMI}}^{\mathrm{LSDR}}(\mathbf{S}_k) = \sum_{\mathbf{S}_j \in \mathbf{S}_\beta} \sum_{\overline{\mathbf{Y}}_i \in \overline{\mathbf{Y}}} \left(I(\mathbf{S}_j; \overline{\mathbf{Y}}_i) + I(\mathbf{S}_k; \overline{\mathbf{Y}} | \mathbf{S}_j) \right). \tag{14}$$

The first term in Eq. (14) is constant with regard to $\mathbf{S}_k$ whose ranking is the final solution of the method in this paper, so we just consider the second term:

$$J_{\mathrm{JMI}}^{\mathrm{LSDR}}(\mathbf{S}_k) \sim \sum_{\mathbf{S}_j \in \mathbf{S}_\beta} \sum_{\overline{\mathbf{Y}}_i \in \overline{\mathbf{Y}}} I\left(\mathbf{S}_k; \overline{\mathbf{Y}} | \mathbf{S}_j\right), \tag{15}$$

where the symbol "∼" indicates a ranking equivalent expression for the criterion.

Based on the information theoretic identity, I (A;B | C) = I (A;B) − I (A;C) + I (A;C | B), the Eq. (15) can be rewritten as

$$J_{\mathrm{JMI}}^{\mathrm{LSDR}}(\mathbf{S}_k) \sim \sum_{\mathbf{S}_j \in \mathbf{S}_\beta} \sum_{\overline{\mathbf{Y}}_i \in \overline{\mathbf{Y}}} \left(I\left(\mathbf{S}_k; \overline{\mathbf{Y}}\right) - I\left(\mathbf{S}_k; \mathbf{S}_j\right) + I\left(\mathbf{S}_k; \mathbf{S}_j | \overline{\mathbf{Y}}\right) \right). \tag{16}$$

The first term in Eq. (16) does not contain $\mathbf{S}_j$, the Eq. (16) is calculated as

$$J_{\mathrm{JMI}}^{\mathrm{LSDR}}(\mathbf{S}_k) \sim |\mathbf{S}_\beta| \sum_{\overline{\mathbf{Y}}_i \in \overline{\mathbf{Y}}} I\left(\mathbf{S}_k; \overline{\mathbf{Y}}\right) - \sum_{\mathbf{S}_j \in \mathbf{S}_\beta} \sum_{\overline{\mathbf{Y}}_i \in \overline{\mathbf{Y}}} \left(I\left(\mathbf{S}_k; \mathbf{S}_j\right) - I\left(\mathbf{S}_k; \mathbf{S}_j | \overline{\mathbf{Y}}\right) \right), \tag{17}$$

where $|\mathbf{S}_\beta|$ indicates the number of already selected monitoring nodes (features). The first term of the proposed LSDR-JMI method in Eq. (17) reveals the relevancy between the feature $\mathbf{S}_k$ and all the columns of the target matrix. The second term represents the relevancy between the feature $\mathbf{S}_k$ and the feature $\mathbf{S}_j$, it is usually called the redundancy. And the last term is the complementarity between the feature $\mathbf{S}_k$ and all the chosen features on the condition that all the substitution matrix is given. Apart from these 3 terms, the optimal label subspace $\overline{\mathbf{Y}}$ can reflect the most useful information of the leak cases (pseudo-label $\mathbf{L}$) to some degree. The node positions of the feature $\mathbf{S}_k$ are the positions of the optimal sensor deployment.

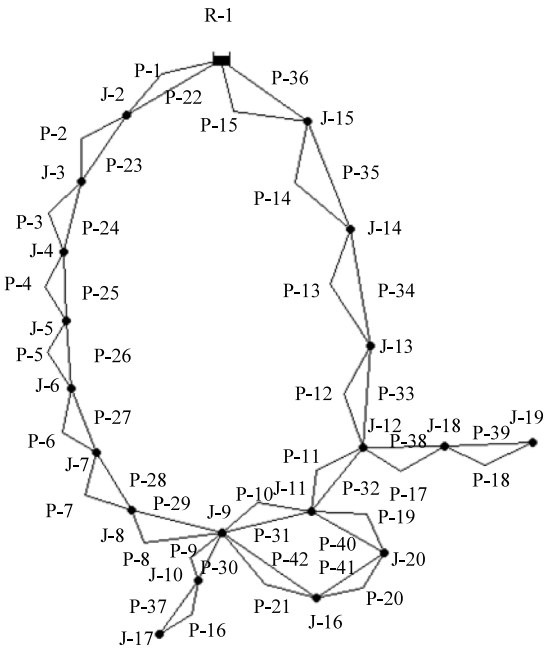

Fig. 4. The WDN of New York City.

Table 1
Nodal properties of the WDN in New York City.

Node ID	Elevation (ft)	Required demand (CFS)
J-2	255	92.4
J-3	255	92.4
J-4	255	88.2
J-5	255	88.2
J-6	255	128.2
J-7	255	88.2
J-8	255	88.2
J-9	255	170
J-10	255	1
J-11	255	170
J-12	255	117.1
J-13	255	117.1
J-14	255	132.4
J-15	255	92.4
J-16	260	210
J-17	272.8	57.5
J-18	255	117.1
J-19	255	117.1
J-20	255	170
R-1	300	N/A

4. Case studies

At present, the sensor deployment methods of WDN leakage monitoring mainly focus on the research of single point leakage, without considering the problem of multi-point leakage. In this section, an innovative method based on LSDR-JMI is applied to solve the problem of multi-point leakage monitoring. On the one hand, LSDR-JMI method makes use of the effective information contained in the label subspace to improve the effect of optimal sensor deployment strategy. On the other hand, it reduces the time complexity of the algorithm to ensure the efficiency of the WDN monitoring. The proposed multi-label feature selection method LSDR-JMI is applied to two WDNs. These two WDNs can be downloaded from "https://github.com/OpenWaterAnalytics/EPANET-Matlab-Toolkit/tree/master/networks/asce-tf-wdst". These two WDNs are simulated in the EPANET software. The EPANET simulation step is set to 1 h, and the simulation data is collected every hour. The pressure data is obtained through EPANET simulation and leakage is random. The data set contains 20,000 data samples, of which

Table 2
Pipe characteristics of the WDN in New York City.

Pipe ID	L (ft)	D (in)	C (HW)
P-1	11 600	204	100
P-2	19 800	204	100
P-3	7300	204	100
P-4	8300	204	100
P-5	8600	204	100
P-6	19 100	204	100
P-7	9600	204	100
P-8	12 500	204	100
P-9	9600	204	100
P-10	11 200	204	100
P-11	14 500	204	100
P-12	12 200	204	100
P-13	24 100	204	100
P-14	21 100	204	100
P-15	15 500	204	100
P-16	26 400	204	100
P-17	31 200	204	100
P-18	24 000	204	100
P-19	14 400	204	100
P-20	38 400	204	100
P-21	26 400	204	100
P-22	11 600	180	100
P-23	19 800	180	100
P-24	7300	180	100
P-25	8300	180	100
P-26	8600	180	100
P-27	19 100	180	100
P-28	9600	132	100
P-29	12 500	132	100
P-30	9600	180	100
P-31	11 200	204	100
P-32	14 500	204	100
P-33	12 200	204	100
P-34	24 100	204	100
P-35	21 100	204	100
P-36	15 500	204	100
P-37	26 400	72	100
P-38	31 200	72	100
P-39	24 000	60	100
P-40	14 400	60	100
P-41	38 400	60	100
P-42	26 400	72	100

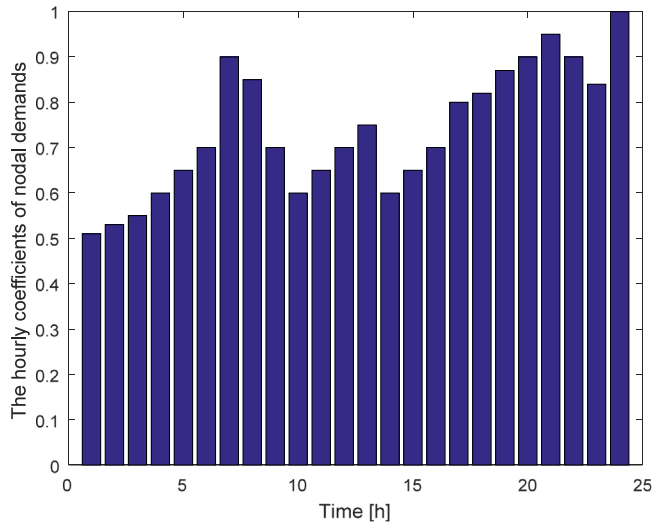

Fig. 5. The hourly coefficients of nodal demands in the WDN of New York City.

80% of the data samples are randomly selected as the training data set, and 20% of the data samples are used as the test data set. There is no difference between these two data sets. All the results have been obtained in the test phase. Besides, all the calculations have been performed by applying a desktop with an INTEL CORE i5-8400U CPU @ 2.80 [GHz], 8 [GB] of RAM memory as well as a Windows 10 Home 64 bits OS. Meanwhile, the MATLAB 2016a software is applied.

Table 3
The influence of the number of sensors on the performance of the algorithm.

Conventional indicator functions	Number of sensors		
	1	2	3
HammingLoss	0.0144	0.0121	0.0121
RankingLoss	0.0159	0.0121	0.0123
Coverage	0.9475	0.8726	0.8753
Average_Precision	0.9213	0.9349	0.9339
Macrof1	0.7992	0.8134	0.8136
Microf1	0.9152	0.9292	0.9294
Running time	6.235 s	8.915 s	11.461 s

Table 4
Results of three multi-label feature selection strategies in the WDN of New York City.

	Single-JMI	Group-JMI	LSDR-JMI
HammingLoss	0.0112	0.0112	**0.0077**
RankingLoss	0.0122	0.0122	**0.0064**
Coverage	0.8733	0.8733	**0.7656**
Average_Precision	0.9346	0.9346	**0.957**
Macrof1	0.818	0.818	**0.8721**
Microf1	0.9348	0.9348	**0.9562**
Node combinations	3,15	3,15	3,18
Running time	25.550 s	62.185 s	26.851 s

Table 5
The running time of different PoT values of Group-JMI in the WDN of New York City.

PoT values	Running time
0.5	62.185 s
0.6	63.216 s
0.7	67.960 s
0.8	70.671 s
0.9	72.266 s
1	76.090 s

Similar to the methods of single-label feature selection, the results of multi-label feature selection are usually evaluated by the results of the classifier. For multi-label data sets, assume that sample set x_i corresponds to label set Y_i, $|Y_i| = n$ represents the number of labels (i.e., the number of columns in the label matrix). Define R_i as the set of related labels of sample x_i, U_i the set of unrelated labels of sample x_i, and $|R_i|$ and $|U_i|$ respectively represent the number of related labels and the number of unrelated labels of sample x_i. Suppose the multi-label classifier is $h(.)$, then the prediction label of the sample is $h(\mathrm{x}_i)$; $f(\mathrm{x}_i, Y)$ is defined to represent the output of the prediction label corresponding to the label set Y. 6 conventional indicator functions have been adopted to evaluate the results of the multi-label feature selection.

(1) Hamming loss

Hamming loss is used to represent the proportion of false labels in prediction labels. N is the number of samples and Δ represents the symmetry difference between the two sets. The smaller the Hamming loss is, the better the classification effect is.

$$\text{Hamming Loss} = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|h(\mathrm{x}_i)\Delta Y_i\right|}{n}. \tag{18}$$

(2) Ranking loss (RL)

Ranking loss represents the average proportion of the irrelevant labels ranking before the related labels, and the smaller the ranking loss, the better the classification effect.

$$\mathrm{RL} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{|R_i|\,|U_i|}\left|\left\{\left(Y',Y''\right)|f\left(\mathrm{x}_i,Y'\right)\le f\left(\mathrm{x}_i,Y''\right),\left(Y',Y''\right)\in R_i\times U_i\right\}\right|. \tag{19}$$

Table 6
The running time and node combinations of different PR Values of LSDR-JMI in the WDN of New York City.

PR values	Running time	Node combinations
0.5	15.036 s	3,14
0.6	17.393 s	3,14
0.7	19.557 s	3,14
0.8	25.358 s	3,18
0.9	26.854 s	3,18
1	29.636 s	3,15

(3) Coverage

The coverage represents the average of the minimum number of tags required when all the real labels of the sample are covered. The smaller the coverage, the better the classification effect.

$$\text{Coverage} = \frac{1}{N}\sum_{i=1}^{N}\max_{Y'\in R_i} rank_f(\mathrm{x}_i, Y') - 1. \tag{20}$$

(4) Average precision (AP)

The average precision means that the ranking of sample related labels is larger than the average value of the given label sorting, and the higher the average accuracy, the better the classification effect.

$$\text{AP} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{|R_i|}\sum_{K\in R_i}\frac{\left|\left\{L'|rank_f(\mathrm{x}_i, L') \le rank_f(\mathrm{x}_i, K), L'\in R_i\right\}\right|}{\left|\left\{L'|rank_f(\mathrm{x}_i, L') \ge rank_f(\mathrm{x}_i, K), L'\in Y\right\}\right|}. \tag{21}$$

(5) Macro-F1

Macro-F1 aims to calculate the local F1 score of each label, which represents the predicted value of the first component of the label set corresponding to the sample, and represents the true value of the first component of the label set corresponding to the sample. The larger the Macro-F1 is, the better the classification effect is.

$$\text{Macro}-\text{F1} = \frac{1}{n}\sum_{j=1}^{n}\frac{2\sum_{i=1}^{N} Y_{ij}{}'\cdot Y_{ij}}{\sum_{i=1}^{N} Y_{ij}{}' + \sum_{i=1}^{N} Y_{ij}}. \tag{22}$$

(6) Micro-F1

Micro-F1 aims to calculate the global F1 score of all labels, and the larger the micro-F1, the better the classification effect.

$$\text{Micro}-\text{F1} = \frac{2\sum_{i=1}^{N}\sum_{j=1}^{n} Y_{ij}{}'\cdot Y_{ij}}{\sum_{i=1}^{N}\sum_{j=1}^{n} Y_{ij}{}' + \sum_{i=1}^{N}\sum_{j=1}^{n} Y_{ij}}. \tag{23}$$

4.1. Case study1

The WDN of New York City is shown in Fig. 4. The network consists of 1 reservoir, 19 nodes, and 42 pipes. The nodal properties of the WDN of New York City are given in Table 1 (R-1 in Table 1 is a reservoir); the pipe characteristics are listed in Table 2; the daily average demand is approximately 2017.5 CFS (1L/s = 2.12CFS).

The pressure residual matrix $\mathbf{\Delta P}$ is generated with the following uncertainties with the aim of generate realistic scenarios:
(1) The ratio of single leak cases: two-point leak cases: three-point leak cases to approximately 10:7:3.
(2) Leak size in the range between 60CFS and 120CFS, which is about 2.97% to 5.95% of the average demand in the WDN of New York City.
(3) Use 5% of the average of all pressure residuals as the standard deviation of Gaussian random noise to simulate the measurement error of the pressure sensor.
(4) Nodal demands are changed each hour according to the user's water consumption. Each node has a basic water demand, and the change of water demand in a day will be changed according to certain proportions, which are the hourly coefficients of nodal demands. The multiplication results of the hourly coefficients and basic water demand reflect the actual water demands of the node. The hourly coefficients of nodal demands for the 19 nodes in the WDN of New York City are depicted in Fig. 5.

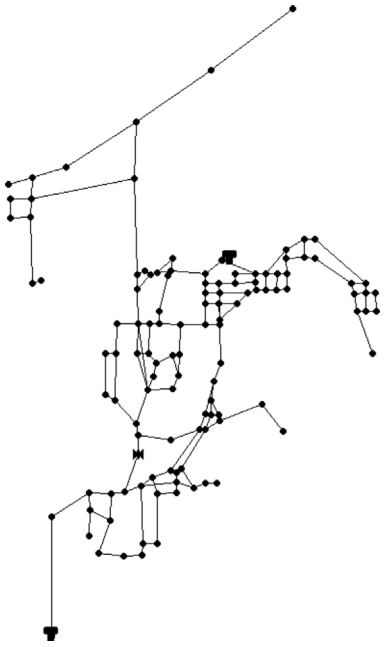

Fig. 6. The WDN of the Dakin Yew Zone, Bellingham City.

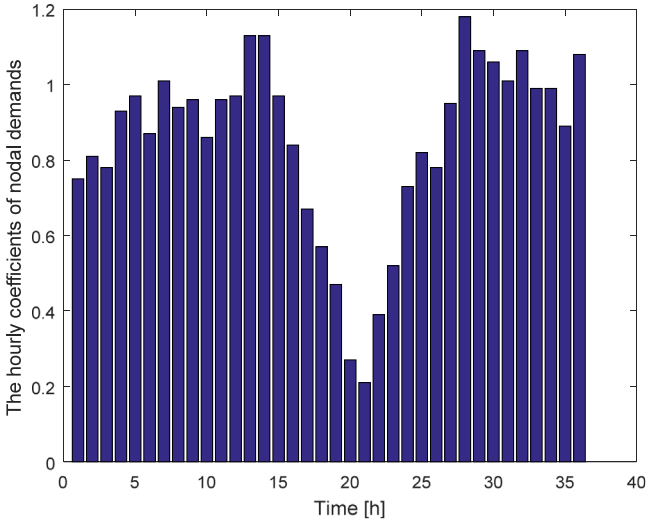

Fig. 7. The hourly coefficients of nodal demands in the WDN of Bellingham City.

Subsequently, the pressure sensitivity matrix $\mathbf{S}$ is calculated according to Eq. (10). The size of $\mathbf{S}$ is 16 000 × 19. The rows of $\mathbf{S}$ are the leak cases and the columns are the corresponding nodes.

Due to the WDN of New York City only contains 19 nodes, the experiment was done to consider the influence of the number of sensors on the performance of the algorithm, the results are shown in Table 3. According to the experimental data, we can see that the performance indicators of 2 sensors are significantly better than that of 1 sensor; and the running time of 2 sensors is significantly better than that of 3 sensors. Therefore, we think that 2 sensors are sufficient. All nodes have *Nc* combinations as follows:

$$Nc = \frac{n_f!}{n_s!(n_f - n_s)!}, \tag{24}$$

where n_f is the number of nodes in the WDNs, and n_s is the number of selected nodes. For the WDN of New York City, $n_f = 19$, $n_s = 2$, then $N_c = 171$. Once the network is big enough, the number of combinations is very huge. However, only a small number of these options are useful

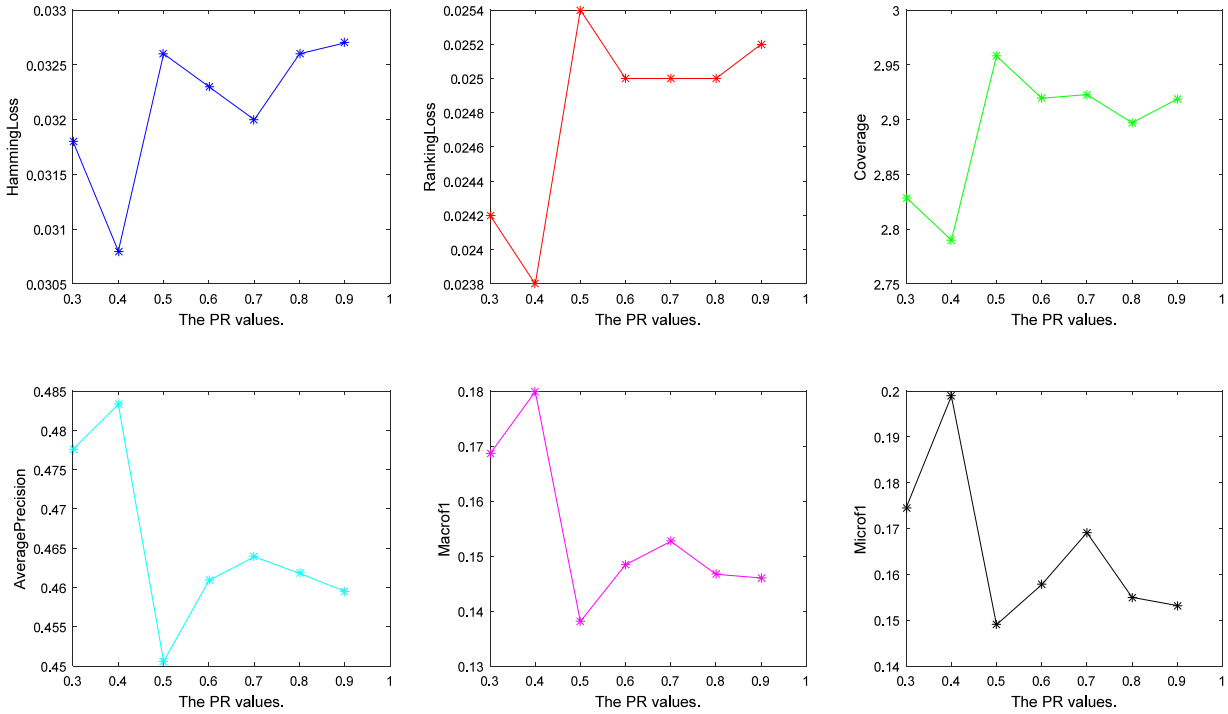

Fig. 8. The 6 indicator functions of the LSDR-JMI method in Dakin Yew Zone, Bellingham City.

for monitoring. So the pressure sensor deployment is necessary to select the useful monitoring nodes.

In this paper, we have considered one-point leak and multi-point leak at the same time. The results of three multi-label feature selection strategies are summarized in Table 4.

Single-JMI method deals with the original label space $\mathbf{Y}$ without any treatment, so this method has not considered the useful information of labels. As shown in Table 4, for the WDN of New York City, the running time of Single-JMI is shortest.

Group-JMI method handles the novel categorical space $\tilde{\mathbf{Y}}$ with taking into account the information of the labels. The node combinations are the set {3,15} which is same as the Single-JMI method. However, the running time is 62.185 s which is 2.434 times that of Single-JMI method. One thing to point out is that when the PoT is 0.5 to 1 and NoC is 100 (large enough), the node combinations are always the set {3,15}. Along with the growing of the PoT, the running time is longer as well. The running time of different PoT values are listed in Table 5. The values of the PoT control the saved scale of label space $\mathbf{Y}$ in each group. Once the PoT is bigger, the more columns of $\mathbf{Y}$ are reserved, this leads to the waste of time.

As for the LSDR-JMI, the chosen node combinations are the set {3,18}. The node 3 is the same as the other two methods, but the node 18 has improved the performances of all 6 conventional indicator functions in Table 4. This indicates that the effectiveness of the proposed pressure sensor deployment from 6 aspects. The running time and the different node combinations of different PR values are listed in Table 6. The different PR values mean the different preservation ratios of the label space $\mathbf{Y}$. And different PR values lead to different selected node combinations. While the PR values are less than 0.8, the node combinations are the set {3,14}. And the node combinations are the set {3,15} when the PR = 1. Only when the PR = 0.8 or PR = 0.9, the node combinations are the set {3,18} which are the optimal sensor deployment. From these results, it can be concluded that the low PR values may lead to the lack of effective information of $\mathbf{Y}$ and high PR values may result in the redundancy of the irrelevant labels. When the PR = 0.8 or PR = 0.9, optimal label subspace is saved, so the results are better than the other 2 methods. While the PR values are too small or too big, the node combinations are less effective than the other two method. Once the best PR value are determined, the proposed method can provide the optimal pressure sensor deployment.

4.2. Case study2

The WDN of the Dakin Yew Zone, Bellingham City depicted in Fig. 6, has two reservoirs, one valve, 121 consumer nodes and 168 pipes. The daily average demand is approximately 3073.35 GPM. For convenience, we have selected 84 nodes as the leak nodes. The pressure residual matrix $\mathbf{\Delta P}$ is generated with the following uncertainties with the aim of generate realistic scenarios:

(1) The ratio of single leak cases: two-point leak cases: three-point leak cases to approximately 42:20:15.

(2) Leak size in the range between 90 GPM and 180 GPM, which is about 2.93% to 5.85% of the average demand in the WDN of the Dakin Yew Zone, Bellingham City.

(3) Use 5% of the average of all pressure residuals as the standard deviation of Gaussian random noise to simulate the measurement error of the pressure sensor.

(4) Nodal demands are changed each hour. The hourly coefficients of nodal demands for the 84 nodes in the WDN of the Dakin Yew Zone, Bellingham City are depicted in Fig. 7.

Subsequently, the pressure sensitivity matrix S is calculated according to Eq. (11). The size of $\mathbf{S}$ is 11 088 × 84. The rows of $\mathbf{S}$ are the leak cases and the columns are the corresponding nodes. The results of three multi-label feature selection strategies are summarized in Table 7. From Table 7, it can be concluded that the optimal sensor combinations of the Single-JMI method and the Group-JMI method (PoT = 0.7) are similar. The set {44, 109, 132} is the same, and the other two nodes are close to each other. Therefore the results of two multi-label feature selection methods are similar. As discussed in Zhou et al. (2018), the Group-JMI method ought to be better than the Single-JMI method. However, in this paper, the 6 indicator functions of the Single-JMI method are a little bit better than that of the Group-JMI method. It can be explained that general multi-label feature selection methods usually save higher scale features than that of our paper. Only 2 monitoring nodes are saved for the WDN of New York City and 5 nodes for the Bellingham City, respectively. Besides, different multi-label data sets

Table 7
Results of three multi-label feature selection strategies in the WDN of the Dakin Yew Zone, Bellingham City.

	Single-JMI	Group-JMI	LSDR-JMI
HammingLoss	0.0326	0.0329	**0.0308**
RankingLoss	0.025	0.0256	**0.0238**
Coverage	2.9004	2.9524	**2.7899**
Average_Precision	0.4634	0.4542	**0.4833**
Macrof1	0.1535	0.1419	**0.1799**
Microf1	0.156	0.1491	**0.1989**
Node combinations	21,44,102,109,132	19,44,103,109,132	13,57,101,105,106
Running time (s)	678.643	1061.662	257.541

Table 8
The running time and node combinations of different PR values of LSDR-JMI in the WDN of New York City.

PR values	Running time (s)	Node combinations
0.3	192.212	21,44,102,106,132
0.4	268.425	13,57,101,105,106
0.5	328.361	31,44,57,103,132
0.6	427.691	44,57,105,111,132
0.7	471.976	13,46,101,103,109
0.8	547.988	18,46,101,109,132
0.9	629.801	18,44,101,109,132

may lead to the different results. In addition to the above points, the running of the Single-JMI method is 678.643 s and the running time of the Group-JMI is 1061.662 s. The running time of the Group-JMI method is too much longer but get the similar results as the Single-JMI method.

For the running time of the proposed LSDR method, it only costs 257.541 s for the optimal sensor deployment of the Bellingham City. The running time of the LSDR-JMI method is the 37.95% of the Single-JMI and the 24.26% of the Group-JMI. This is due to the 60% of the irrelevant are left out (PR = 0.4). The label subspace improves the accuracy of the node selection and saves a lot of time. Compared with the WDN of New York City, it is obviously that the proposed strategy can save more time when dealing with more complicated WDNs. We also change the PR values to evaluate the different results depicted in Fig. 8.

As shown in Fig. 8, different PR values may lead to different values of 6 indicator functions. The indicator functions may change a lot. However, once the optimal PR value is determined (0.4 in this paper), all the results of the 6 indicator functions are better than the Single-JMI method and the Group-JMI method. As for the time complexity and the optimal sensor deployment, the running time and the selected nodes of different PoT values are listed in Table 8.

In Table 8, different PR values mean different number of columns of the label space $\mathbf{Y}$ are saved. The bigger PR values lead to the more time complexity. And for the WDN of Dakin Yew Zone, Bellingham City, it is no need to save most of the columns of $\mathbf{Y}$. For a more complicated WDN, the lower proportion of $\mathbf{Y}$ can already meet the requirements of multi-label feature selection. This also shows that the applicability of the LSDR-JMI method in more complex WDNs is more obvious.

5. Conclusion

In this paper, a novel filter multi-label feature selection strategy is proposed to monitor the leakage of the WDNs. LSDR-JMI method is inspired by Group-JMI to explore the information of the label space. LSDR method is based on the dependence of label matrix on leakage events. In addition, the columns of the label matrix reflect the leakage of each node in WDN. Some node labels in WDN can not only reflect the leakage of the whole network, but also eliminate the interference of some node labels to the leakage event, so that the subsequent feature selection results are more accurate. In LSDR-JMI method, the dimensionalities of the label space are reduced by calculating the contribution of the label space to the pseudo-label. Besides, it is vital to reduce the time complexity and the computational complexity of the algorithm. Furthermore, the optimal PR values of two cases are different. Higher PR value in the WDN of New York City indicates that it is crucial for small-scale WDNs to save enough scale of label information. As for the WDN of the Dakin Yew Zone, Bellingham City, lower scale label matrix is adequate to represent the optimal subset of the label space. In addition, it has made outstanding contributions to reducing time complexity.

In an extensive empirical multi-label evaluation, we compared the LSDR-JMI method against a variety of other 2 methods. Experimental results demonstrate that the time complexity is low. Besides, the 6 conventional multi-label metrics are better than the other 2 multi-label feature selection methods. Consequently, it has been proven that the LSDR-JMI outperforms 2 existing multi-label feature selection methods in both the multi-label metrics and the time complexity.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This paper is supported by the key Science Foundation of the Department of Science and Technology of Jilin Province, China (Grant No. 20180201081SF, 20190303082SF) and Science and technology project of The Education Department of Jilin Province, China (Grant No. JJKH20200983KJ). Thanks for the permission to publish this paper.

References

Aral, M. M., Guan, J., & Maslia, M. L. (2010). Optimal design of sensor placement in water distribution networks. *Journal of Water Resources Planning and Management*, *136*(1), 5–18.

Berkel, F., Caba, S., Bleich, J., & Liu, S. (2018). A modeling and distributed MPC approach for water distribution networks. *Control Engineering Practice*, *81*, 199–206.

Blanco, R., Larranaga, P., Inza, I., & Sierra, B. (2004). Gene selection for cancer classification using wrapper approaches. *International Journal of Pattern Recognition and Artificial Intelligence*, *18*(8), 1373–1390.

Blesa, J., Nejjari, F., & Sarrate, R. (2016). Robust sensor placement for leak location: Analysis and design. *Journal of Hydroinformatics*, *18*(1), 136–148.

Brown, G., Pocock, A., Zhao, M. J., & Lujan, M. (2012). Conditional likelihood maximisation: A unifying framework for information theoretic Feature selection. *Journal of Machine Learning Research (JMLR)*, *13*, 27–66.

Cugueró-Escofet, M. À., Puig, V., & Quevedo, J. (2015). Optimal pressure sensor placement and assessment for leak location using a relaxed isolation index: application to the Barcelona water network. *Control Engineering Practice*, *63*, 1–12.

Ding, C., & Peng, H. (2005). Minimum redundancy feature selection from microarray gene expression data. *Journal of Bioinformatics and Computational Biology*, *3*(2), 185–205.

Farley, B., Mounce, S. R., & Boxall, J. B. (2010). Field testing of an optimal sensor placement methodology for event detection in an urban water distribution network. *Urban Water Journal*, *7*(6), 345–356.

Ferrandez-Gamot, L., Busson, P., Blesa, J., Tornil-Sin, S., Puig, V., Duviella, E., et al. (2015). Leak localization in water distribution networks using pressure residuals and classifiers. *IFAC-PapersOnLine*, *48*(21), 220–225.

Galuppini, Giacomo, Magni, Lalo, & Creaco, Enrico (2020). Stability and robustness of Real Time Pressure Control in Water Distribution Systems. *Journal of Hydraulic Engineering*, *146*(4), Article 04020023.

Galuppini, Giacomo, et al. (2019). Service pressure regulation in water distribution networks. *Control Engineering Practice*, *86*, 70–84.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, *3*, 1157–1182.

Hoque, N., Bhattacharyya, D. K., & Kalita, J. K. (2014). MIFS-ND: A mutual information-based feature selection method. *Expert Systems with Applications*, *41*(14), 6371–6385.

Kang, J., Park, Y. J., Lee, J., Wang, S. H., & Eom, D. S. (2017). Novel leakage detection by ensemble CNN-SVM and graph-based localization in water distribution systems. *IEEE Transactions on Industrial Electronics*, *65*, 4279–4289.

Kayaalp, F., Zengin, A., Kara, R., & Zavrak, S. (2017). Leakage detection and localization on water transportation pipelines: a multi-label classification approach. *Neural Computing and Applications*, *28*, 2905–2914.

Li, J., Wang, C., Qian, Z. H., & Lu, C. G. (2019). Optimal sensor placement for leak localization in water distribution networks based on a novel semi-supervised strategy. *Journal of Process Control*, *82*, 13–21.

Li, L. P., Weinberg, C. R., Darden, T. A., & Pedersen, L. G. (2001). Gene selection for sample classification based on gene expression data: study of sensitivity to choice of parameters of the GA/KNN method. *Bioinformatics*, *17*, 1131–1142.

Lu, M. (2019). Embedded feature Selection Accounting for unknown Data Heterogeneity. *Expert Systems with Applications*, *119*, 350–361.

Ooi, C., & Tan, P. (2003). Genetic algorithms applied to multi-class prediction for the analysis of gene expression data. *Bioinformatics*, *19*, 37–44.

Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information: Criteria of max-dependency, max-relevance and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*(8), 1226–1238.

Peralta, B., & Soto, A. (2016). Embedded local feature selection within mixture of experts. *Information Sciences*, *269*, 176–187.

Puust, R., Kapelan, Z., Savic, D. A., & Koppel, T. (2010). A review of methods for leakage management in pipe networks. *Urban Water Journal*, *7*(1), 25–45.

Rossman, L. A. (2000). *EPANET 2 User's Manual: EPA/600/R-00/ 057*, Cincinnati: U.S.EPA.

Saeys, Y., Inza, I., & Larranaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics*, *23*(19), 2507–2517.

Sankar, G. S., Kumar, S. M., Narasimhan, S., Narasimhan, S., & Bhallamudi, M. (2015). Optimal control of water distribution networks with storage facilities. *Journal of Process Control*, *32*, 127–137.

Sarrate, R., Nejjari, F., & Rosich, A. (2012). Sensor placement for fault diagnosis performance maximization in distribution networks. In *2012 20th mediterranean conference on control & automation (MED)* (pp. 110–115).

Sechidis, K., Nikolaou, N., & Brown, G. (2014). Information theoretic feature selection in multi-label data through composite likelihood. In *S+SSPR*. Berlin/Heidelberg, Germany: Springer.

Sechidis, K., Spyromitros-Xioufis, E., & Vlahavas, I. (2019). Information Theoretic multi-Target feature selection via output space quantization. *Entropy*, *21*(9).

Soldevila, A., Blesa, J., Tornil-Sin, S., Fernandez-Canti, R. M., & Puig, V. (2018). Sensor placement for classifier-based leak localization in water distribution networks using hybrid feature selection. *Computers & Chemical Engineering*, *108*, 152–162.

Torres, L., Jiménez-Cabas, J., González, O., Molina, L., & López-Estrada, F.-R. (2020). Kalman filters for Leak Diagnosis in pipelines: Brief history and Future Research. *Journal of Marine Science and Engineering*, *8*(3), 173.

Wang, Y., Puig, V., & Cembrano, G. (2017). Non-linear economic model predictive control of water distribution networks. *Journal of Process Control*, *56*, 23–34.

Winter, C. D., Palleti, V. R., Worm, D., & Kooij, R. (2019). Optimal placement of imperfect water quality sensors in water distribution networks. *Computers & Chemical Engineering*, *121*, 200–211.

You, M., Liu, J., Li, G. Z., & Chen, Y. (2012). Embedded feature selection for Multi-label Classification of music Emotions. *International Journal of Computational Intelligence Systems*, *5*(4), 668–678.

Zhou, B., Liu, A., Wang, X., She, Y., & Lau, V. (2018). Compressive sensing-based multiple-leak identification for smart water supply systems. *IEEE Internet of Things Journal*, *5*(2), 1228–1241.