RESEARCH PAPER

RS-SVM Machine Learning Approach Driven by Case Data for Selecting Urban Drainage Network Restoration Scheme

Li Jiang[1], Zheng Geng[1], Dongxiao Gu[1†], Shuai Guo[2], Rongmin Huang[3], Haoke Cheng[3], Kaixuan Zhu[4]

[1]School of Management, Hefei University of Technology, Hefei 230002, China

[2]College of Civil Engineering, Hefei University of Technology, Hefei 230002, China

[3]Yangtze Ecology and Environment Co.,Ltd., Wuhan 430062, China

[4]Luddy School of Intelligent System and Engineering, Indianan university, Bloomington, Indiana 47404, USA

Keywords: Drainage pipe network; Machine learning; Rough set; Multilevel SVM; Restoration scheme

Citation: Jiang, L., Geng, Z., Gu, D.X. et al.: RS-SVM machine learning approach driven by case data for selecting urban drainage network restoration scheme. Data Intelligence 5(2), 413-437 (2023). doi: 10.1162/dint_a_00208
Submitted: Janurary 15, 2023; Revised: February 14, 2023; Accepted: March 10, 2023

ABSTRACT

Urban drainage pipe network is the backbone of urban drainage, flood control and water pollution prevention, and is also an essential symbol to measure the level of urban modernization. A large number of underground drainage pipe networks in aged urban areas have been laid for a long time and have reached or practically reached the service age. The repair of drainage pipe networks has attracted extensive attention from all walks of life. Since the Ministry of ecological environment and the national development and Reform Commission jointly issued the action plan for the Yangtze River Protection and restoration in 2019, various provinces in the Yangtze River Basin, such as Anhui, Jiangxi and Hunan, have extensively carried out PPP projects for urban pipeline restoration, in order to improve the quality and efficiency of sewage treatment. Based on the management practice of urban pipe network restoration project in Wuhu City, Anhui Province, this paper analyzes the problems of lengthy construction period and repeated operation caused by the mismatch between the design schedule of the restoration scheme and the construction schedule of the pipe network restoration in the existing project management mode, and proposes a model of urban drainage pipe network restoration scheme selection based on the improved support vector machine. The validity and feasibility of the model are analyzed and verified by collecting the data in the project practice. The research results show that the model has a favorable effect on the selection of urban drainage pipeline restoration

† Corresponding author: DongXiao Gu (e-mail: dongxiaogu@yeah.net; ORCID: 0000-0003-3557-009X).

schemes, and its accuracy can reach 90%. The research results can provide method guidance and technical support for the rapid decision-making of urban drainage pipeline restoration projects.

1. INTRODUCTION

With the rapid urbanization, the short board of urban drainage pipe network construction is showing up daily. Various structural and functional defects of the drainage pipe network caused by aging easily cause problems such as urban waterlogging, sewage overflow, and ground subsidence [1]. Therefore, completing the urban drainage pipe network detection and repair is critical for realizing city sewage quality and efficiency, which helps promote the development of high-quality urban governance. However, the urban drainage pipe network has long miles, complicated structure, uncertainty factors, and extensive range, which significantly influences residents' living environment and urban traffic [2]. Thus, how to determine the drainage pipe network's status and performance, design reasonable repair scheme rapidly, shorten the whole pipeline repair project period, and reduce the project construction according to its testing results can greatly affect the social environment and sustainable development.

In recent years, the introduction of pipe network performance evaluation technology has made the best solution for pipe repair, which has become a research hot spot among domestic and foreign scholars and experts. The existing research mainly uses qualitative and quantitative research methods to establish the prediction model of pipe network performance indicators and formulate maintenance plans. However, a difficult problem remains, which is the low efficiency of the decision making. Using machine learning-related technologies in mining case history, the literature that studies the quick decisions for fixing urban drainage pipe networks is scant.

Thus, this research put forward RS-SVM machine learning approach driven by case data for selecting urban drainage network restoration scheme. The main contribution of this study is threefold. First, we combine the attribute reduction based on RS technology [3] and the SVM technology [4] to give full play to their technological advantages. The minimalist data set of the excellent classification characteristics is used as the input of the SVM. Second, we propose an RS-SVM model for selecting an urban drainage pipe network repair scheme. The basic idea is to collect history data set from urban pipeline repairing project management practice for a pipeline, use RS theory to reduce the sample's attributes, use the indirect method combining two classifiers to construct a multi-level SVM scheme selection model, and then use the built model for scheme selection of the test sample to solve the matching analysis. Finally, we select Wuhu's drainage pipeline repair engineering case data for big data analysis in Anhui Province. The effectiveness of the proposed model and method is verified. This study provides decision support for the quick selection of drainage pipeline repair schemes and has a certain application value.

2. RELATED WORKS

As for the technology on predicting the state of pipe network and developing repair strategies, Altarabsheh A et al. [5] used the Markov model to predict pipeline networks in the future and choose the most

appropriate operational plan by GA, according to the whole life cycle of a sewage pipe network, considering the construction cost, operation cost, and expected benefits. Hernández N et al. [6] used the differential evolution method as an optimization tool for hyperparameter combination and combined it with the SVM model for two different management objectives (network and pipe levels). This model was applied to Colombia's main cities of Bogotá and Medellin, resulting in a less than 6% deviation in the prediction of structural conditions in both cities at a network level. Wang Y J et al. [7] proposed an XGBoost-based MICC model with the benefits of hyperparameters autooptimization. Yu A L et al. [8] put forward a method to carry out the five directions of research for the drainage pipeline repair scheme. Intelligent decision provides the basis. To predict the future performance of trenchless rehabilitations, Ibrahim B et al. [9] presented condition prediction models for the chemical grouting rehabilitation of pipelines and manholes in the city of Laval, Quebec, Canada. Bakry I et al. [10] presented condition prediction models for CIPP rehabilitation of sewer mains, and the models can predict the structural and operational conditions of CIPP rehabilitation on the basis of basic input, such as pipe material, and rehabilitation type and date.

As for the technology on development of decision support tools for drainage network repair, Cai X T et al. [11] proposed a sensitivity-based adaptive procedure (SAP), which can be integrated with optimization algorithms. SAP was integrated with non-dominated sorting genetic algorithm II (NSGA-II) and multiple objective particle swarm optimization (MOPSO) methods. Ulrich A et al. [12] studied a novel solution combining both approaches (pipes and tanks) and proposed a decision support system based on the NSGA-II for the rehabilitation of urban drainage networks through the substitution of pipes and the installation of storage tanks. Debères P et al. [13] used multiple criteria to locate the repair section and made a repair plan according to the pipeline inspection report and the economic, social, and environmental indicators. Ramos-Salgado et al. [14] developed a decision support system (DSS) to help water utilities design intervention programs for hydraulic infrastructures. Chen S B et al. [15] summarized the characteristics, causes, and evolution mechanism of typical defect types of coastal urban drainage network, which can provide technical guidance for the evaluation and repair of drainage network in this area or similar cities. Based on the investigation of the current situation of the existing underground drainage network in a certain area of Chongqing, Liu W et al. [16] used the AHP and entropy weight method to study the urban drainage pipe network risk rating and provide decision support for pipeline repair plan design. In view of the urban drainage pipe network deterioration, Wang J L et al. [17] used AHP and fuzzy comprehensive evaluation methods to research the urban drainage pipe network status and operational efficiency to provide decision support for drainage pipe network maintenance and repair plan.

This study is inspired by Xie B et al. [18] but different from the previous works that focused on designating remediation solutions based on current drainage inspection results but does not sufficiently explore the value of historical cases to solve this current problem. The choice of an urban drainage pipeline repair plan belongs to the category of multiple attribute decision making [19]. For multiple attribute decision-making problems, policymakers tend to be objective in formulating the optimal alternatives [20]. Support vector machine (SVM) is established on VC dimension theory and structural risk minimization principle based on machine learning methods [21]. It has better properties, especially in quickly setting the optimal alternative

for current cases based on multi-attribute case history data [22]. As the SVM's input, the case history data sets frequently have redundant attributes [23]. Redundant attributes can increase the complexity of SVM training, extend the SVM training time, and reduce the SVM decision efficiency. A rough set (RS) deals with the uncertainty problem of mathematical tools [24]. RS attribute reduction algorithm can effectively handle attribute redundancy; it reduces redundant attributes that interfere with the SVM. Thus, on the basis of the combination of rough set and SVM technology, we propose an RS-SVM model for selecting an urban drainage network restoration scheme in this study.

3. METHODOLOGY

The detailed process of combining RS and SVM to select an urban drainage pipe network repair scheme consists of four components. The structure of which is shown in Figure 1, and the role of each component is as follows:

- Collecting historical data sets and then standardizing them.
- Using the RS attribute reduction algorithm to reduce the redundant attributes contained in the data set.
- Training the model of multi-level SVM classification.
- Using the trained model to match the new detection results.

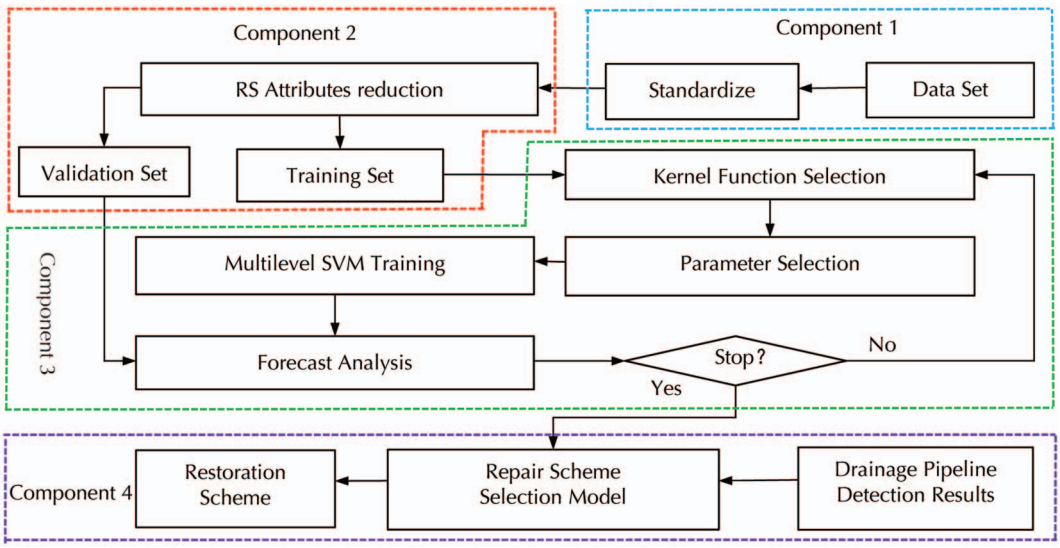

Figure 1. Detailed process of combining RS and SVM to select an urban drainage pipe network repair scheme.

The related parameters described in the model are as follows: $\{Z_1, Z_2, \ldots, Z_i, \ldots, Z_m\}$ are the historical data sets of drainage pipeline repair projects. The target case denoted by Z_0 is the current case reflecting the

need for a repair scheme. $\{C_1^p, C_2^p,..., C_j^p,..., C_n^p\}$ is the attribute set of drainage pipe network detection results. $(a_{01}, a_{02},..., a_{0j},..., a_{0n})$ is the attribute value vector of Z_0, where a_{0j} is the C_j^p value corresponding to Z_0. In this study, the drainage pipe network detection properties are divided into two types, numeric and symbols. For example, the length, diameter, various defects, and quantity of the pipes belong to numeric data. The material of the pipes belongs to symbol data. $\{S_1,S_2,...,S_k,...,S_g\}$ is the history solution set of drainage pipeline repair projects. According to the characteristics of the urban pipeline repair project, we presume that a solution may apply to multiple cases, with each case only a final implementation scheme.

3.1 Standardizing

To eliminate the influence of the dimension, we need to standardize the data set. Symbol variables are denoted by $\{F_i|i\in T\}$, where T equals {1, 2, 3, ..., t}. Thus, we set the symbol sequence within the specified symbol in advance. The sequence number of the Sign F_i is $seq(F_i)$, where $F_i\in F$, $seq(F_i)\in T$. Standardizing the symbolic variables uses Equation (1):

$$F_i^{'} = \frac{seq(F_i)}{t} \tag{1}$$

We use the normal method to standardize some numeric variables. This method is suitable for indicators with a nonzero-range. The standardized variable values are between 0 and 1. Standardizing the symbolic variables uses Equation (2):

$$x_i^{'} = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)} \tag{2}$$

3.2 RS Attributes Reduction

Selecting as many attributes that have a greater impact on the scheme as possible can avoid missing crucial ones. If there are a large number of attributes, the complexity of the model inevitably increases, and the prediction performance is reduced, supposing all attributes are input into the SVM model. Some attributes have little influence on the selection of repair schemes for urban drainage networks, and the repair schemes are determined to some extent by a few key attributes that best reflect the characteristics of the categories [25]. Therefore, attribute reduction is conducive to improving the prediction accuracy and efficiency of the SVM model. In light of the influence of the definition of attribute importance and reduction rules, the result of attribute reduction is often not unique, and finding a minimum reduction has been proven to be an NP-hard problem [26]. The general method to solve this problem is to find the optimal or suboptimal reduction by heuristic search method [27]. Thus, we propose the attribute reduction algorithm based on attribute similarity. The basic flow of the algorithm is as follows:

Step 1: All conditional attributes are reduced using discernible attribute matrix.

In rough set theory, tuples S=(U, A, V, f) are defined as information systems, where U is a discussion domain of A, A=C∪D is an attribute set, C represents conditional attributes, and D represents target attributes. In the information system S, the differentiation determined by RA is: $A=U/RA=\{K_i: i\leq l\}$, and f_l

(K_i) is used to represent the value of attribute cl with respect to the object in K_i, called $I(K_i, K_j)$ (Equation (3)). It is the discernable attribute matrix of K_i and K_j, and the main diagonal element of I is the empty set.

$$I(K_i, K_j) = \{c_l \in A : f_l(K_i) \neq f_l(K_j)\} \tag{3}$$

The identification equation is as follows:

$$M = \underset{i \neq j}{\wedge}(\vee D(c_i, c_j)) \tag{4}$$

All attributes of the information system can be reduced using Equation (4). For example, in the information system shown in Table 1, {c_1, c_2, c_3, c_4} is the conditional attribute and D is the decision attribute.

Table 1. Information system.

code	c_1	c_2	c_3	D	code	c_1	c_2	c_3	D
Z_1	2	1	3	0	Z_5	1	1	2	4
Z_2	3	2	1	1	Z_6	1	1	4	3
Z_3	2	1	3	0	Z_7	1	2	3	2
Z_4	1	1	4	3	Z_8	1	2	3	2

According to Table 1, the samples with the same attribute value are merged to obtain the simplified information system shown in Table 2.

Table 2. Simplified information system.

code	c_1	c_2	c_3	D
$K_1 = \{Z_1, Z_3\}$	2	1	3	0
$K_2 = \{Z_2\}$	3	2	1	1
$K_3 = \{Z_7, Z_8\}$	1	2	3	2
$K_4 = \{Z_4, Z_6\}$	1	1	4	3
$K_5 = \{Z_5\}$	1	1	2	4

From Table 2 and Formula (3), the discernable attribute matrix **I** shown in Table 3 is obtained.

Table 3. Discernible attribute matrix **I**.

I	K_1	K_2	K_3	K_4	K_5
K_1	∅	$\{c_1, c_2, c_3\}$	$\{c_1, c_2\}$	$\{c_1, c_3\}$	$\{c_1, c_3\}$
K_2	$\{c_1, c_2, c_3\}$	∅	$\{c_1, c_3\}$	$\{c_1, c_2, c_3\}$	$\{c_1, c_2, c_3\}$
K_3	$\{c_1, c_2\}$	$\{c_1, c_3\}$	∅	$\{c_2, c_3\}$	$\{c_2, c_3\}$
K_4	$\{c_1, c_3\}$	$\{c_1, c_2, c_3\}$	$\{c_2, c_3\}$	∅	$\{c_3\}$
K_5	$\{c_1, c_3\}$	$\{c_1, c_2, c_3\}$	$\{c_2, c_3\}$	$\{c_3\}$	∅

In Table 3, according to the discernable attribute matrix **I** and Equation (4), all the reductions of the information system are obtained as $G_1 = \{c_1, c_3\}$ and $G_2 = \{c_2, c_3\}$.

Step 2: The core of the information system should be found.

Equation (5) is used to find the core of the information system, where $\{G_i: i \leq l\}$ represents all reductions of the information system.

$$\text{core}() = \bigcap_{i=1}^{l} G_i \tag{5}$$

According to the result of step 1 and Equation (5), the core of the information system can be obtained as $\text{core}()=G_1 \cap G_2=\{c_1, c_3\} \cap \{c_2, c_3\}=\{c_3\}$.

Step 3: The similarity of relatively necessary attributes is calculated with respect to decision attribute D.

If the core is not empty, any reduction contains the core. So the properties in the core are absolutely necessary properties. Equation (6) is used to represent the set of relatively necessary attributes that appear in some reductions.

$$\text{rna}() = \bigcup_{i=1}^{l} G_i - \bigcap_{i=1}^{l} G_i \tag{6}$$

According to the result of Step 2 and Equation (6), the relatively necessary attributes of the information system can be obtained: $\text{rna}()=\{c_1, c_3\} \cup \{c_2, c_3\}-\{c_1, c_3\} \cap \{c_2, c_3\}=\{c_1, c_2\}$.

Step 4: The similarity between each relatively necessary attribute and decision attribute is gradually added to the core in order from high to low to form set R until R satisfies the indistinguishable relation: IND (R) = IND(C). Therefore, R is the relative minimum.

The similarity between conditional attribute {c} and decision attribute D is calculated by Equation (7).

$$S(\{c\},D) = \frac{\left|\text{IND}(D \cup \{c\})\right|}{\sqrt{\text{IND}(D)} \cdot \sqrt{\text{IND}(\{c\})}} \tag{7}$$

According to Table 1, $S(c_1,D)$ and $S(c_2,D)$ are calculated by the following equation:

$$S(c_1,D) = \frac{\left|\text{IND}(D \cup c_1)\right|}{\sqrt{\text{IND}(c_1)} \cdot \sqrt{\text{IND}(D)}} = 0.683$$

$$S(c_2,D) = \frac{\left|\text{IND}(D \cup c_2)\right|}{\sqrt{\text{IND}(c_2)} \cdot \sqrt{\text{IND}(D)}} = 0.642$$

By virtue of $S(c_1, D)>S(c_2, D)$, adding the relatively necessary attribute c_1 to core() can form the set $R_1=\{c_1, c_3\}$. Since $\text{IND}(R_1)=\text{IND}(C)=14$, the indistinguishable relation is satisfied. Thus, $\{c_1, c_3\}$ is the relative minimum for this example.

3.3 Selecting Kernel Function

The prerequisite for SVM classification is that the sample space is linearly separable. However, the complexity of data spatial distribution increases along with the increase of the sample space dimension. For example, in Figure 2, points are linearly indivisible in a two-dimensional plane. In Figure 3, a kernel function is adopted to transform the sample dimensionally. The requirement of linear divisibility is satisfied after the sample set is mapped to a higher-dimensional space. Different kernel functions can be used to construct and realize different types of nonlinear decision surface learning machines in the input space, thus generating different support vector algorithms [28]. We select several kernel functions commonly used in classification problems. Functions (8)-(11) are expressed as follows:

Linear kernel function: $K(\boldsymbol{x}_i \ \boldsymbol{x}_j) = \boldsymbol{x}_i^T \cdot \boldsymbol{x}_j$ (8)

Poly kernel function: $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i^T \cdot \boldsymbol{x}_j + 1)^d, d \geq 0$ (9)

RBF kernel function: $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = exp(-\gamma || \boldsymbol{x}_i - \boldsymbol{x}_j ||^2), \gamma > 0$ (10)

Sigmoid kernel function: $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = tanh(k\boldsymbol{x}_i^T \cdot \boldsymbol{x}_j + c), k > 0; c > 0$ (11)

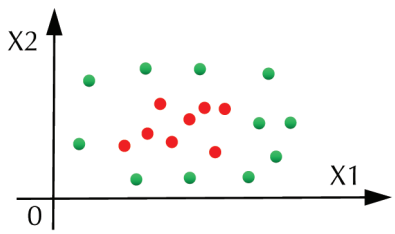

Figure 2. Linearly indivisible points.

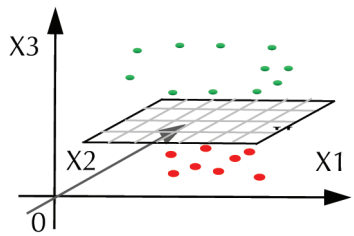

Figure 3. Linearly separable points.

3.4 Training Multi-level SVM Classification Model

3.4.1 Binary Classification Algorithm Based on Historical Cases

Before building a multi-level SVM classification model, a binary classification model should be built. Let the scheme set of the sample be {E, F}. In Figure 4, the binary classification problem is to build a binary classifier that can distinguish schemes E and F through historical cases. The theoretical derivation of the binary classifier is as follows:

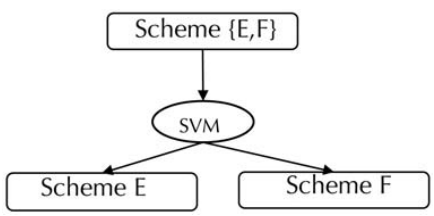

Figure 4. Work flow of binary classifier.

We set some sample points in the sample space $\{x_i|i=1, 2, 3, ..., n\}$, where $\mathbf{x}_i$ is the vector corresponding to each sample point. y_i is the class value corresponding to each sample point, where $y_i \in \{-1, 1\}$. The equations of positive hyperplane H_1, negative hyperplane H_2 and decision hyperplane H_0 are as follows:

$$\begin{cases} H_1 : \boldsymbol{w}\boldsymbol{x} + b = +1 \\ H_0 : \boldsymbol{w}\boldsymbol{x} + b = 0 \\ H_2 : \boldsymbol{w}\boldsymbol{x} + b = -1 \end{cases} \tag{12}$$

The constrained optimization problem for the maximum value L is written in the following form:

$$\max\ L = \frac{2}{\|\boldsymbol{w}\|} \quad s.t.\ (\boldsymbol{w} \cdot \boldsymbol{x}_i + b) y_i \geq 1 \tag{13}$$

To facilitate calculation, Equation (13) is rewritten as

$$Lar(\boldsymbol{w}, b, \lambda_i, p_i) = \frac{\boldsymbol{w} \cdot \boldsymbol{w}^T}{2} - \sum_{i=1}^{n} \lambda_i [(\boldsymbol{w} \cdot \boldsymbol{x}_i + b) y_i - 1] \quad \lambda_i \geq 0 \tag{14}$$

The dual problem of Equation (14) is as follows:

$$\max_{\ddot{e}}(\min_{\boldsymbol{w},b}(Lar(\boldsymbol{w}, b, \lambda_i, p_i)) \quad s.t.\ \lambda_i \geq 0 \tag{15}$$

Considering the influence of **w** and b on Lagrange function (15), the following function is constructed.

$$f(\boldsymbol{w}, b) = (\frac{\boldsymbol{w} \cdot \boldsymbol{w}^T}{2} - \sum_{i=1}^{n} \lambda_i ((\boldsymbol{w} \cdot \boldsymbol{x}_i + b) y_i - 1)) \ , \lambda_i \geq 0 \tag{16}$$

The argument of function f(**w**,b) is unconstrained and the function is a convex function about the argument $(w_1, w_2, w_3, ..., w_s, b)$, so it has a unique minimum point. Its gradient expression is as follows:

$$\nabla f(\boldsymbol{w}, \mathrm{b}) = \begin{bmatrix} \dfrac{\partial f(\boldsymbol{w}, \mathrm{b})}{\partial \boldsymbol{w}} \\ \dfrac{\partial f(\boldsymbol{w}, \mathrm{b})}{\partial b} \end{bmatrix} = \begin{bmatrix} \boldsymbol{w}^T - \sum_{i=0}^{n} \lambda_i y_i \boldsymbol{x}_i \\ \sum_{i=0}^{n} \lambda_i y_i \end{bmatrix}, \lambda_i \geq 0 \tag{17}$$

Let $\nabla f(\boldsymbol{w}, \mathrm{b}) = \boldsymbol{0}$; Equation (18) can be obtained according to Equation (17):

$$\begin{cases} \boldsymbol{w}^T = \sum_{i=1}^{n} \lambda_i y_i \boldsymbol{x}_i \\ \sum_{i=0}^{n} \lambda_i y_i = 0 \end{cases}, \lambda_i \geq 0 \tag{18}$$

After substituting Equation (18) into the function and introducing some kernel function, Equation (15) can be written as

$$\max_{\ddot{e}}(\sum_{k=1}^{n} \lambda_k - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j y_i y_j K(\boldsymbol{x}_i \boldsymbol{x}_j)) \quad s.t. \quad \lambda_i, \lambda_j \geq 0 \tag{19}$$

After solving for $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3, ..., \lambda_n)^T$, we can solve for **w*** in terms of $\boldsymbol{w}^T = \sum_{i=1}^{n} \lambda_i y_i \boldsymbol{x}_i$, We just use the support vector $\mathbf{x}_i$* in the computation of **w***. Positive hyperplane H_1: **w*****x**+b*=1. Negative hyperplane H_2: **w*****x**+b*=–1. Decision hyperplane H_0: **w*****x**+b*=0. After the three hyperplane equations are determined, the binary classifier is constructed.

3.4.2 Multi-Classification Method Based on Combinatorial Thinking

The repair schemes of urban drainage pipe networks are diverse, so a single binary classifier cannot complete the classification of all schemes; combining multiple binary classifiers is necessary. The common combination methods of binary classification include indirect and direct methods [29]. The indirect method is to construct a series of binary classifiers in a certain way and combine them to achieve multi-class classification. The other combines the parameter solutions of multiple classification surfaces into an optimization problem and realizes the classification of multiple classes by solving the optimization problem. Although the direct method looks simple, the variables in the optimized problem-solving process are significantly more than the indirect method, and the training speed and classification accuracy are not as good as the first method. This problem is more prominent when the training sample size is large. Therefore, in Figure 5, the indirect method and combinatorial thinking are adopted in this study to build a binary tree multi-level classifier. Let the scheme set of the sample be {scheme 0, scheme 1, scheme 2..., scheme g-1}. The number of schemes is g. In Figure 6, the multi-level SVM classification model aims to differentiate all schemes by combining multiple binary classifiers step by step.

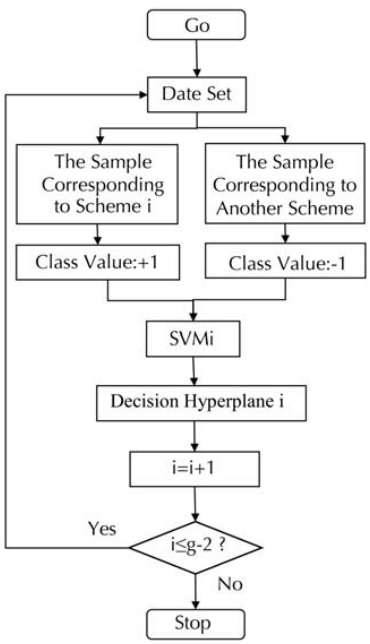

Figure 5. The algorithm flow of multilevel classifier.

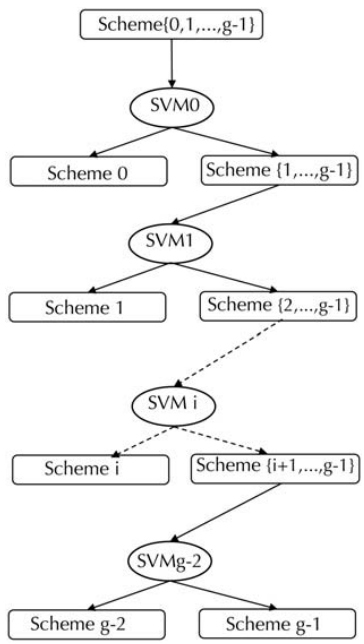

Figure 6. The work flow of multilevel classifier.

4. EXPERIMENTS

We take the management practice of urban pipe network restoration project in Wuhu City, Anhui Province as the background. Basing on sklearn.SVM and sklearn.metrics packages, we draw the conclusion by comparing the RS-SVM machine learning method proposed in this paper with other algorithms (SVM without attribute reduction, logistic algorithm without attribute reduction, logistic algorithm with attribute reduction). The data set selected in this study contains 1500 samples, of which 1000 samples are randomly selected as the training set and others as the test set. All experiments are conducted on a personal desktop computer with 2 GHz Intel(R) Xeon (R) E5-2620 CPU, 8 GB RAM and Python 3.8.

4.1 Data Set

Referring to "Technical Specification for Inspection and Evaluation of Urban Drainage Pipes (CJ181-2012)", 68 conditional attributes $\{c_1^p, c_2^p, ..., c_{68}^p\}$ are shown in Table 4.

Table 4. Sixty-eight conditional attributes.

Attribute	Code	Attribute	Code
Pipe length	c_1^p	Transformation I–IV	c_9^p–c_{12}^p
Buried depth	c_2^p	Corrode I–IV	c_{13}^p–c_{16}^p
Pipe diameter	c_3^p	Wrong mouth I–IV	c_{17}^p–c_{20}^p
Material	c_4^p	Ups and downs I–IV	c_{21}^p–c_{24}^p
Rupture I–IV	c_5^p–c_8^p	Disconnect I–IV	c_{25}^p–c_{28}^p
Interface material shedding I–IV	c_{29}^p–c_{32}^p	Scaling I–IV	c_{49}^p–c_{52}^p
Branch pipe dark connection I–IV	c_{33}^p–c_{36}^p	Obstacle I–IV	c_{53}^p–c_{56}^p
Foreign body penetration I–IV	c_{37}^p–c_{40}^p	Residual wall and dam root I–IV	c_{57}^p–c_{60}^p
Leakage I–IV	c_{41}^p–c_{44}^p	Root I–IV	c_{61}^p–c_{64}^p
Deposition I–IV	c_{45}^p–c_{48}^p	Scum I–IV	c_{65}^p–c_{68}^p

The sample has four fixes. To facilitate differentiation, schemes are numbered in this study, and the numbered values are taken as target attribute values, as shown in Table 5. According to Figure 6, three classifiers should be built for the four schemes: SVM_0, SVM_1, and SVM_2.

Table 5. Value of each scheme target attribute.

Scheme	Code	Value
Not repair	Scheme 0	0
Excavation and reconstruction	Scheme 1	1
Ultraviolet light polymerization	Scheme 2	2
Spot in situ curing	Scheme 3	3

4.2 Data Standardization

To compress the data distribution, eliminate the impact of dimension, and improve the efficiency of attribute reduction and classification of SVM, we standardize the attribute values of pipeline material (c_4^p)

according to Equation (1) and standardize those of other attributes according to Equation (2). After standardizing, the distribution of sample attribute values corresponding to the four repair schemes is shown in Figure 7. In a repair scheme, the larger the sample size corresponding to the attribute value under an attribute, the darker the color of the area. If the attribute value under an attribute corresponds to a smaller sample size, the color of the area is lighter.

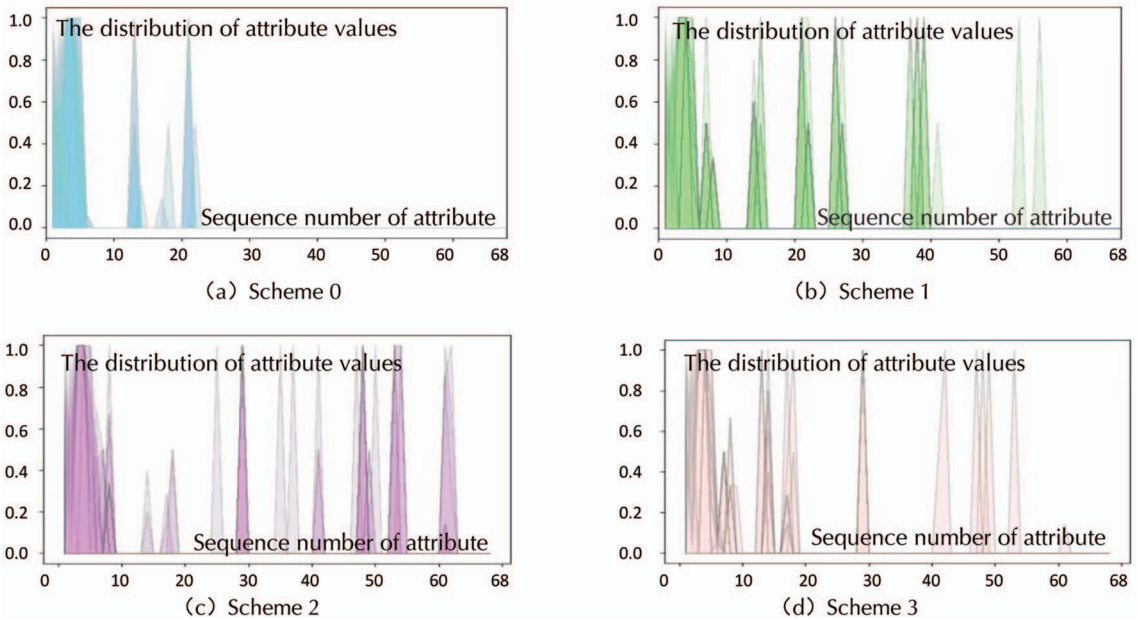

Figure 7. Distribution of attribute values of four restoration schemes.

4.3 Attributes Reduction

The pipeline defect type shown in Table 6 does not appear in all sample pipe segments of the pipeline network repair project in Wuhu City, Anhui Province. Therefore, under the existing samples, the attributes in Table 6 do not affect the selection of repair schemes and can be preliminarily reduced.

Table 6. Type of defect not present in samples.

Attribute	Code	Attribute	Code
Deformation III	c_{11}^{p}	Mismouth IV	c_{20}^{p}
Metamorphosis IV	c_{12}^{p}	Disjunction IV	c_{28}^{p}
Etch IV	$_{c16}^{p}$	Interface material off IV	c_{32}^{p}
Misopening III	c_{19}^{p}	The branch pipe is secretly connected to IV	c_{36}^{p}
Foreign body is penetrated into IV	c_{40}^{p}	Residual wall and dam root I-IV	c_{57}^{p}-c_{60}^{p}
Leakage IV	c_{44}^{p}	Root III	c_{63}^{p}
Scaling III	c_{51}^{p}	Root IV	c_{64}^{p}
Scaling IV	c_{52}^{p}	Scum I-IV	c_{65}^{p}-c_{68}^{p}

According to Equation (3)–(5), the core of the sample information system is obtained: core()={c_2^p, c_3^p, c_5^p, c_6^p, c_7^p,c_8^p, c_9^p,c_{13}^p,c_{14}^p,c_{15}^p,c_{17}^p,c_{18}^p,c_{21}^p,c_{22}^p,c_{24}^p,c_{25}^p,c_{26}^p,c_{27}^p,c_{29}^p,c_{30}^p,c_{33}^p,c_{34}^p,c_{35}^p,c_{37}^p,c_{38}^p,c_{39}^p,c_{41}^p,c_{42}^p,c_{43}^p,c_{45}^p,c_{46}^p,c_{47}^p,c_{48}^p,c_{49}^p,c_{50}^p,c_{53}^p,c_{54}^p,c_{55}^p,c_{56}^p,c_{61}^p,c_{62}^p}. According to Equation (6), the relatively necessary attributes are obtained: rna()={c_1^p,c_4^p, c_{10}^p,c_{23}^p,c_{31}^p}. All reductions of the sample information system are G_1={core(),c_1^p} , G_2={core(),c_4^p} , G_3={core(),c_{10}^p}, G_4={core(),c_{23}^p}, and G_5={core(),c_{31}^p}. According to Formula (7), $S(c_4^p,D)>S(c_{10}^p,D)>S(c_{31}^p,D)>S(c_{23}^p,D)>S(c_1^p,D)$; therefore, the relatively necessary attribute c_4^p is first added to core() to form the set R={c_4^p,core()}. As IND(R)=IND(C) satisfies the indiscriminability relation, R={c_4^p, core()} is the relative minimalism of this example, and {c_1^p,c_{10}^p,c_{23}^p,c_{31}^p} is further reduced as a redundant attribute. Afterward, the 42 attributes in R are sorted from smallest to largest according to the subscript. The distribution of sample attribute values corresponding to the four repair schemes is shown in Figure 8. The attribute numbers from 42 to 68 in Figure 8 is the redundant attribute set reduced, and there is no peak value. By comparing Figures 7 and 8, we see that attribute reduction can remove redundancy and compress and reduce the dimension of the distribution of data sets.

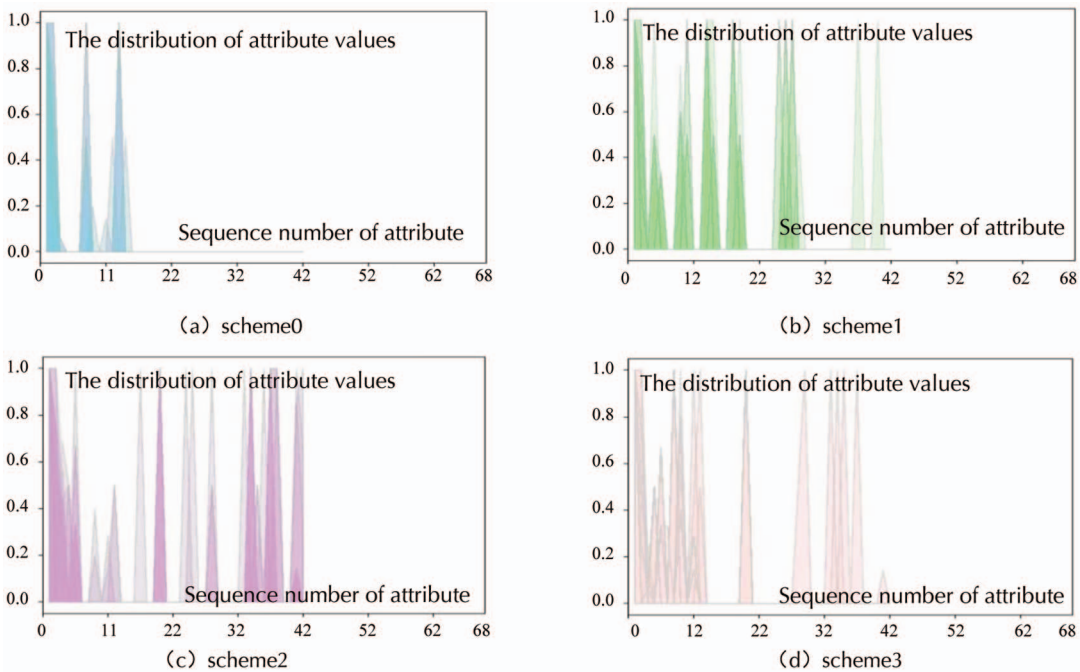

Figure 8. Sample attribute value distribution of the four restoration schemes after reduction.

4.4 Kernel Parameter Optimization

In recent years, many methods have been proposed by domestic and foreign experts and scholars to evaluate the prediction effect of SVM, most of which are the Macro-Average-ROC [30]. The curves can be used to qualitatively evaluate the prediction effect of different SVM classification models [31]. The area

under the curve (Macro-Average-AUC) can quantitatively evaluate the prediction accuracy of different SVM classification models. To facilitate the narration, we use MAR to represent Macro-Average-ROC and MAA to represent Macro-Average-AUC.

Suppose the number of test samples is n and the number of categories is g. After the training is completed, the probability of each test sample under each category is calculated and a matrix **P** with n rows and g columns is obtained. Each line of **P** represents the probability value of a test sample under each category. Accordingly, the labels of each test sample are converted to a binary like form. Each position is used to mark whether it belongs to the corresponding category. Thus, a label matrix **L** with n rows and g columns can be obtained.

The basic idea of macro averaging is as follows: Under each category, you can get the probability that n test samples are of that category (the columns in the matrix **P**). Therefore, according to each corresponding column in the probability matrix **P** and label matrix **L**, false positive rate (FPR) and true positive rate (TPR) under each threshold can be calculated, thus drawing a ROC curve. Thus, a total of g ROC curves can be plotted. FPR_all is obtained by merging, de-duplicating, and sorting all FPR of these ROC curves. The FPR and TPR of the current class determine the ROC curve of the current class. The linear interpolation method [32] is used to interpolate the horizontal coordinate that does not exist relative to FPR_all in FPR, and the TPR′ after interpolation is obtained. TPR_mean is obtained by arithmetic averaging TPR′ after class g interpolation. Finally, MAR curves are drawn according to FPR_all and TPR_mean. The equations of TPR_mean and MAR are as follows:

$$\mathrm{TPR\_mean}=\frac{1}{g}\sum_{i=1}^{g}\mathrm{interp}(\mathrm{FPR\_all}_i,\mathrm{FPR}^i,\mathrm{TPR}^i) \tag{20}$$

$$\mathrm{MAA}=\sum_{i=2}^{\mathrm{len(FPR\_all)}}(\mathrm{TPR\_mean}_i+\mathrm{TPR\_mean}_{i+1})(\mathrm{FPR\_all}_{i+1}-\mathrm{FPR\_all}_i)/2 \tag{21}$$

Among the four SVM kernel functions described in Section 3.3, we are uncertain which one is most suitable for this data set. The parameter values of each kernel function are also uncertain. Thus, the training set containing 1000 samples is used to optimize the selection of the optimal kernel function and parameters, as shown in Figure 1. Specifically, in the parameter definition domain, we draw an image of the MAA of the model after attribute reduction as the parameter values changed (Figure 9) so as to determine the optimal parameters of each kernel function after attribute reduction and the maximum MAA. To compare and analyze the influence of attribute reduction on the SVM classification effect, we draw an image of the model's AUC changing with parameter values in the case of no attribute reduction (Figure 10). We then determine the optimal parameter and maximum MAA of each kernel function without attribute reduction.

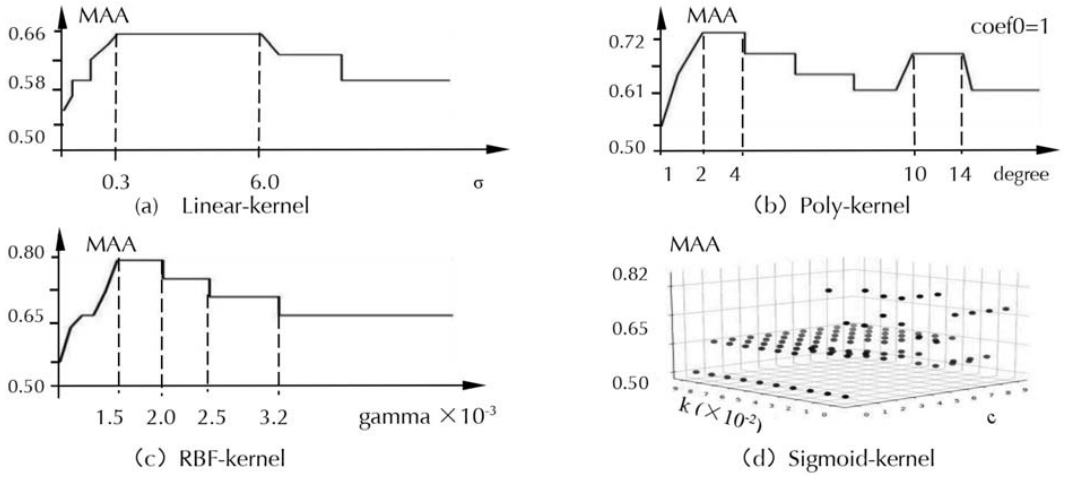

Figure 9. Parameter optimization results after attribute reduction.

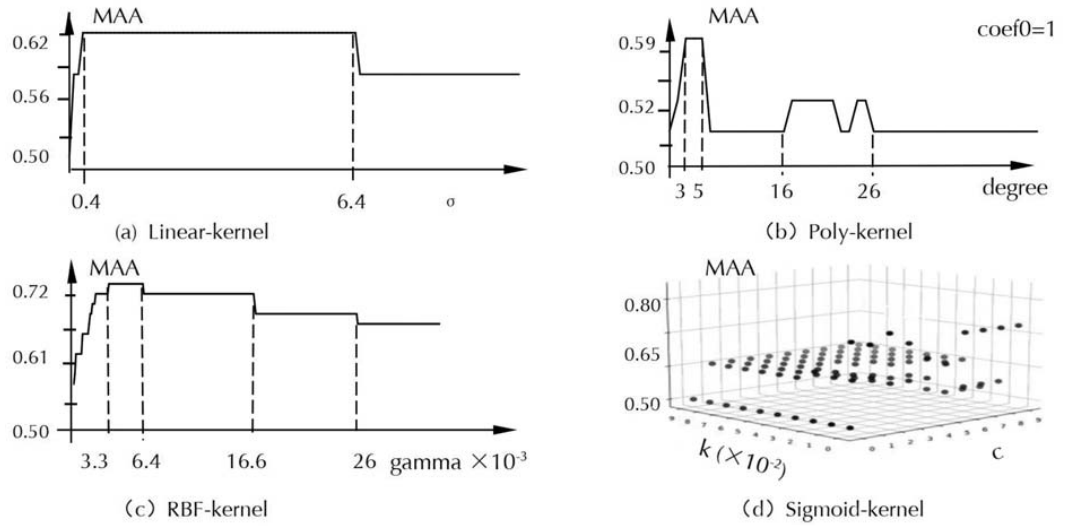

Figure 10. Parameter optimization results before attribute reduction.

(1) Linear kernel parameter optimization

The penalty coefficient σ is an important parameter of linear kernel function. The greater the σ is, the greater the penalty for misclassification. Hence, the accuracy of the training set is high, but the generalization ability is weak. The smaller the σ value; the lower the penalty for misclassification, allowing fault tolerance, and treating them as noise points; the stronger the generalization ability [33]. For linear kernel function, before attribute reduction (Figure 10 (a)), if $0<\sigma<0.4$, the value of MAA increases with the increase of σ. If

$0.4 \leq \sigma \leq 6.4$, MAA stabilizes around 0.62. If $\sigma > 6.4$, the data will be overfitted with the increase of σ, and the value of MAA first decreases and then stabilizes around 0.56. After attribute reduction (Figure 9 (a)), if $0 < \sigma < 0.3$, the value of MAA increases with the increase of σ. If $0.3 \leq \sigma \leq 6.0$, MAA stabilizes around 0.66. If $\sigma > 6.0$, the data are overfitted with the increase of σ, and the value of MAA first decreases and then stabilizes around 0.58. Thus, before attribute reduction, the optimal parameter interval of linear kernel is [0.4,6.4], and then the value of MAA can reach 0.612. After attribute reduction, the optimal parameter interval of linear kernel is [0.3,6.0], and then the value of MAA can reach 0.656.

(2) Poly kernel parameter optimization

High-order coefficients degree and coef0 are important parameters of the poly kernel function. The greater the degree, the higher the spatial dimension after mapping, the higher the complexity of calculating the polynomial [34]. In particular, if degree=1 and coef0=0, the poly kernel is equivalent to the linear kernel. In the case of coef0=1, we study the influence of degree on the classification accuracy of the poly kernel function. Before attribute reduction (Figure 10 (b)), if $1 \leq \text{degree} \leq 3$, the value of MAA increases with the increase of degree. If $3 \leq \text{degree} \leq 5$, the classification accuracy is stable at around 0.59. If $16 \leq \text{degree} \leq 26$, the MAA value fluctuates slightly between 0.51 and 0.53. With the increase of degree, the computational complexity increases, and the value of MAA is stable around 0.51. After attribute reduction (Figure 9 (b)), if $1 \leq \text{degree} \leq 2$, the value of MAA increases with the increase of degree. If $2 \leq \text{degree} \leq 4$, the value of MAA is stable around 0.72. If $10 \leq \text{degree} \leq 14$, MAA values show another peak, but this peak is slightly lower than the first peak of 0.72. With the increase of degree, the computational complexity increases, and the value of MAA is stable around 0.61. The optimal parameter interval of the poly kernel before attribute reduction is [3,5], and then the value of MAA can reach 0.583. After attribute reduction, the optimal parameter interval of the poly kernel is [2,4], and then the value of MAA can reach 0.716.

(3) RBF kernel parameter optimization

Gamma, an important parameter of the RBF kernel function, controls the penalty threshold range and mainly defines the influence of a single sample on the entire classification hyperplane. If gamma is small, a single sample has little influence on the whole classification hyperplane, so selecting it as a support vector is hard. Conversely, a single sample has a greater impact on the whole classification hyperplane and is more likely to be selected as a support vector, or the whole model will have more support vectors [33]. In this study, the optimal gamma value is searched within $[10^{-3},+\infty)$, and the search ends when the value of MAA becomes stable with the increase of gamma. Before attribute reduction (Figure 10 (c)), if $0 < \text{gamma} < 3.3\times10^{-3}$, the value of MAA increases with the increase of gamma. If $3.3\times10^{-3} \leq \text{gamma} \leq 6.4\times10^{-3}$, MAA is stable around 0.72. If $\text{gamma} \geq 2.6\times10^{-2}$, the value of MAA is stable around 0.67 with the increase of gamma. After attribute reduction (Figure 9 (c)), if $0 < \text{gamma} < 1.5\times10^{-3}$, the value of MAA increases with the increase of gamma. If $1.5\times10^{-3} \leq \text{gamma} \leq 2.0\times10^{-3}$, MAA is stable around 0.79. If $2.0\times10^{-3} < \text{gamma} < 3.2\times10^{-3}$, the value of MAA showed a decreasing trend. If $\text{gamma} \geq 3.2\times10^{-3}$, the value of MAA is stable around 0.66. It can be seen that the optimal value interval of the RBF kernel parameter before attribute reduction is $[3.3\times10^{-3},6.4\times10^{-3}]$, and then the value of MAA can reach 0.718. After attribute

reduction, the optimal parameter interval of RBF kernel is [1.5×10^{-3},2.0×10^{-3}], and then the value of MAA can reach 0.793.

(4) Sigmoid kernel parameter optimization

k (gamma) and c (coef0) are two important parameters of the sigmoid kernel function [34]. To comprehensively consider the influence of k and c on classification accuracy, this study takes (k,c) as the independent variable and the value of MAA as the dependent variable and calculates the value of MAA of sigmoid kernel function within the search range of $0<k\leq0.1$ and $0\leq c\leq10$. Moreover, we draw a spatial scatter plot containing 100 points (Figure 9 (d) and Figure 10 (d)). Before attribute reduction (Figure 9 (d)), if $0.04\leq k\leq0.07$ and $3\leq c\leq6$, the MAA value of the sigmoid kernel function reaches 0.735. If (k, c) takes other values, the MAA values of sigmoid kernel functions are all below 0.735. After attribute reduction (Figure 9 (d)), if $0.03\leq k\leq0.05$ and $1\leq c\leq5$, the MAA value of the sigmoid kernel function reaches 0.825. If (k, c) takes other values, the MAA values of the sigmoid kernel function are all below 0.825. Thus, before attribute reduction, the optimal value interval of the Sigmoid-kernel parameter is $k\in[0.04,0.07]$ and $c\in[3,6]$, and then the value of MAA can reach 0.732. After attribute reduction, the optimal value interval of the sigmoid kernel parameter is $k\in[0.03,0.05]$ and $c\in[1,5]$, at which time the value of MAA reaches 0.825.

4.5 Comparison Models

The validation set containing 500 samples is used to test the prediction effect of our proposed RS-SVM classification model. To verify the effectiveness and feasibility of the proposed RS-SVM machine learning method in the urban drainage network repair scheme selection model, we also compare this algorithm with the Laplacian logistic regression algorithm [35]. The results of parameter optimization in Section 4.4 and the prediction results of logistic regression are shown in Table 7. According to Table 7, we draw the MAR curve of each kernel function with optimal parameters and the MAR curve of logistic regression, as shown in Figure 11. Figure 12 shows the average time of each kernel function and the average time of logistic regression change with the sample size.

Table 7. Comparison models.

Kernel	SVM						Laplacian Logistic Regression			
	Optimal parameter		MAA		Average time consuming (second)		MAA		Average time consuming (second)	
	before	after	before	after	before	after	before	after	before	after
Linear	[0.4,6.4]	[0.3,6.0]	0.612	0.656	129.4	85.6				
Poly	[3,5]	[2,4]	0.583	0.716	159.3	103.9				
RBF	[0.0033, 0.0064]	[0.0015, 0.0020]	0.718	0.793	161.3	100.7	0.564	0.598	523.3	483.1
Sigmoid	$0.04\leq k\leq0.07$; $3\leq c\leq6$	$0.03\leq k\leq0.05$; $1\leq c\leq5$	0.732	0.825	150.5	90.6				

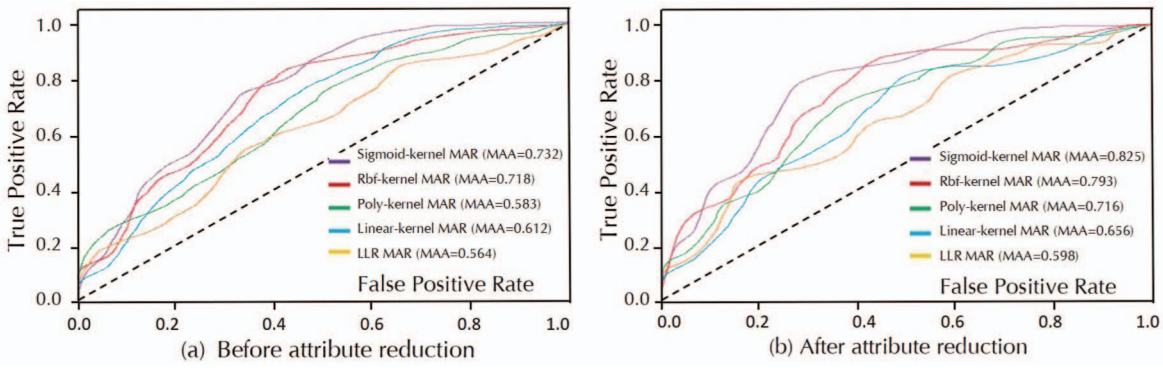

Figure 11. MAR curves of different classification models.

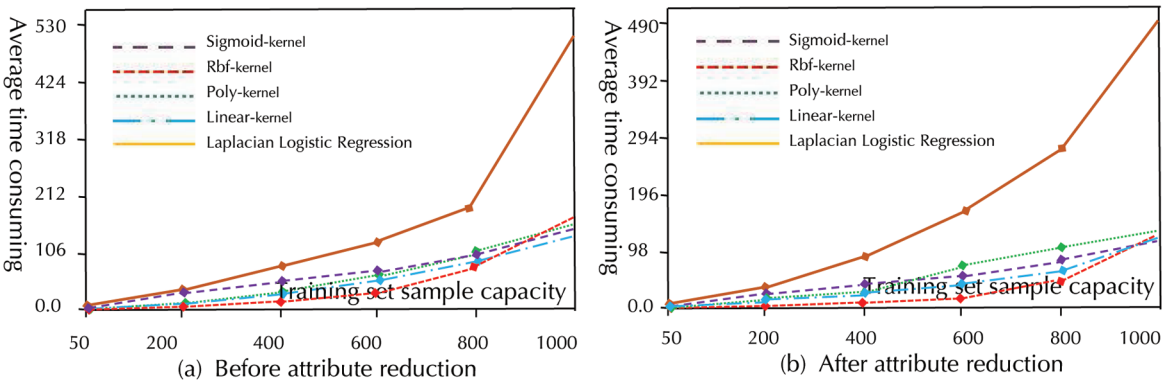

Figure 12. Average time of different classification models.

From the perspective of the MAR curve, before attribute reduction (Figure 11 (a)), according to the value of MAA, four kernel functions are sorted in descending order: Sigmoid, RBF, Linear, Poly. After attribute reduction (Figure 11 (b)), four kernel functions are sorted from largest to smallest according to the value of MAA: Sigmoid, RBF, Poly, Linear. In terms of average time spent before attribute reduction (Figure 12 (a)), classification time spent by SVM and Laplacian logistic regression shows an increasing trend with the increase in sample size, and the growth rate of Laplacian logistic regression is significantly faster than SVM. After attribute reduction (Figure 12 (b)), the classification time of SVM and Laplacian logistic regression also show an increasing trend with the increase in sample size, and the growth rate of Laplacian logistic regression is also significantly faster than that of SVM. From the perspective of the MAR curve and average time, although the Sigmoid classification time is not the lowest, it has little difference from the classification time of other kernel functions because Sigmoid can guarantee a high value of MAA. Therefore, we can consider that the Sigmoid-kernel algorithm with attribute reduction is more suitable to select an urban drainage network repair scheme under the current data set than other algorithms.

Through comparative analysis, the detailed contributions of the rough set and SVM model are as follows:

(1) The rough set attribute reduction algorithm can effectively reduce the sample attributes and improve the classification efficiency and accuracy of multilevel SVM algorithm. Additionally, comparative analysis reveals that the parameter optimization results of the four SVM kernel functions are different before and after attribute reduction. The classification time of the same classification model under different sample sizes is also different.

(2) Compared with Laplacian logistic regression algorithm, SVM adopts a kernel function mechanism. When calculating the decision surface, only samples representing the support vector in the SVM algorithm participate in the calculation, which can guarantee higher accuracy and shorter running time overall and effectively reduce overfitting. At the same time, SVM can build a linear learning machine in a high-dimensional feature space, which avoids the 'disaster of dimension' to some extent.

5. CONCLUSION AND FUTURE WORK

In view of the difficulty of urban drainage network repair detection and the mismatch between repair scheme design and construction schedule, this study organically combines the rough set attribute reduction algorithm based on attribute similarity with a multi-level support vector machine to construct an RS-SVM model for selecting an urban drainage network repair scheme. We select the case data of the Wuhu network repair project for big data analysis. In general, the scheme selection model proposed in this study can effectively predict the urban drainage network repair scheme and provide a basis for the rapid selection of urban drainage network repair schemes. At the same time, the method can be extended to the fields of disease diagnosis, project risk assessment, fire identification and bank loan risk prediction; moreover, it has high practical application value. However, in collecting case data, we did not consider the situation of multi-scheme combination repair nor the situation of using different kernel function combinations to search for optimization in the classification algorithm. Therefore, future studies can be carried out from the following aspects:

(1) According to the characteristics of the project, the urban pipeline repair scheme is further subdivided, and the pipeline sections repaired by combining multiple schemes are taken as samples into the training set to study a more accurate pipeline repair construction scheme.

(2) The prediction efficiency and accuracy of RS-SVM machine learning can be further improved by combining multiple kernel functions for SVM multi-level classifiers.

(3) The sample data of multiple pipeline repair projects are selected, and other prediction methods, such as random forest, are selected for comparative experiments, which not only enriches the experimental results but also improves the accuracy of pipeline repair schemes and provides auxiliary support for designers to make quick decisions.

ACKNOWLEDGMENTS

This work is supported by the Funds for the Anhui Provincial Science and Technology Innovation Strategy and Soft Science Research Project (Grant Number 202206f01050017), National Natural Science Foundation of China (Grant Number 72131006, 6201101347, 72071063), Fundamental Research Funds for the Central Universities (Grant Number JS2021ZSPY0037), Research Project of China Three Gorges Corporation (Grant Number 202103355), Yangtze Ecology and Environment Co.,Ltd. (Grant Number HB/AH2021039), and Power China Huadong Engineering Corporation Limited (KY2019-ZD-03).

AUTHOR CONTRIBUIONS

Li Jiang (E-mail: jiangli@hfut.edu.cn): has participated in the proposed model design and writing of the manuscript.

Zheng Geng (E-mail: gengzheng2023@163.com): has participated in the coding, the experiment and analysis, writing the manuscript.

DongXiao Gu (E-mail: dongxiaogu@yeah.net): has participated in the part of the experiment and analysis.

Shuai Guo (E-mail: guoshuai@hfut.edu.cn): has participated in the part of the experiment and analysis.

RongMin Huang (E-mail: huang_rongmin@ctg.com.cn): has participated in the writing and revision of the manuscript.

HaoKe Cheng (E-mail: cheng_haoke@ctg.com.cn): has participated in the writing and revision of the manuscript.

KaiXuan Zhu (E-mail: zkx9116@126.com): has participated in the writing and revision of the manuscript.

REFERENCE

[1] Zhang, Z. Y.: Can the Sponge City Project improve the stormwater drainage system in China?—Empirical evidence from a quasi-natural experiment. International Journal of Disaster Risk Reduction 5(102980), 1–9 (2022)

[2] Liu, Y. X., Ye, S. Z., Lv, B., et al.: Information solution for Intelligent detection of drainage pipe network defects. China Water & Wastewater 37(8), 32–36 (2021)

[3] Yan, C., Li, Z., Boota, M. W., et al.: River pattern discriminant method based on Rough Set theory. Journal of Hydrology: Regional Studies 5(101285), 1–14 (2023)

[4] Boran, S., Yoney, K. E., Kamil, D., et al.: Comparative evaluation and comprehensive analysis of machine learning models for regression problems. Data Intelligence 4(3), 620–652 (2022)

[5] Altarabsheh, A., Kandil, A., et al.: New multi-objective optimization approach to rehabilitate and maintain sewer networks based on whole life-cycle behavior. Journal of Computing in Civil Engineering 32(1), 1–20 (2018)

[6] Hernández, N., Caradot, N., Sonnenberg, H., et al.: Optimizing SVM models as predicting tools for sewer pipes conditions in the two main cities in Colombia for different sewer asset management purposes. Structure and Infrastructure Engineering 17(2), 156–169 (2021)
[7] Wang, Y. J, Su, F., Guo, Y., et al.: Predicting the microbiologically induced concrete corrosion in sewer based on XGBoost algorithm. Case Studies in Construction Materials 17(01649), 1–17 (2022)
[8] Yu, A. L.: Forecasting and decision optimization theory and methods based on artificial intelligence. Journal of Management Science 35(1), 60–66 (2022)
[9] Ibrahim, B., Hani, A., Khalid, K., et al.: Condition prediction for chemical grouting rehabilitation of sewer networks. Journal of Performance of Constructed Facilities 30(04016042), 1–11 (2016)
[10] Bakry, I., Alzraiee, H., Masry, M. E., et al.: Condition prediction for cured-in-place pipe rehabilitation of sewer mains. Journal of Performance of Constructed Facilities 30(04016016), 1–12 (2016)
[11] Cai, X. T., Shirkhani, H., et al.: Sensitivity-based adaptive procedure (SAP) for optimal rehabilitation of sewer systems. Urban Water Journal 19(9), 889–899 (2022)
[12] Ulrich, A., Ngamalieu-Nengoue, F., Javier, M., et al.: Multi-objective optimization for urban drainage or sewer networks rehabilitation through pipes substitution and storage tanks installation. Water 11(5), 935–949 (2019)
[13] Debères, P., Ahmadi, M., et al.: Deploying a sewer asset management strategy using the indigo decision support system. Optics Express 16(9), 5997–6007 (2013)
[14] Ramos, S., Cristobal, M., Jesus, A., et al.: A decision support system to design water supply and sewer pipes replacement intervention programs. Reliability Engineering & System Safety 216(107967), 1–16 (2021)
[15] Chen, S. B., Yang, Y. X., Wang, H., et al.: Typical defect types and cause mechanism of sewer pipelines in southern coastal city. Water & Wastewater Engineering 48(1), 464–470 (2022)
[16] Liu, W., et al.: Risk assessment on the drainage pipe network based on the AHP-entropy weight method. Journal of Safety and Environment 21(3), 949–956 (2021)
[17] Wang, J. L., Xiong, Y. H., Zhang, X. G., et al.: Evaluation of the state and operational effectiveness of urban drainage pipe network based on AHP-fuzzy comprehensive evaluation method: taking Huai'an District of Huai'an City as an example. Journal of Environmental Engineering Technology 12(4), 1162–1169 (2022)
[18] Xie, B., Xiang, T., Liao, X. F., et al.: Achieving privacy-preserving online diagnosis with outsourced SVM in internet of medical things environment. IEEE Transactions on Dependable and Secure Computing 19(6), 4113–4126 (2022)
[19] Wu, F., Li, Z. Q., et al.: Research on application of pipeline repair technology in urban water environment management. Water & Waste water Engineering 48(1), 471–475 (2022)
[20] Mpimis, T., Kapsis, T. T., Panagopoulos, A. D., et al.: Cooperative D-GNSS aided with multi attribute decision making module: a rigorous comparative analysis. Future Internet 14(7), 195–195 (2022)
[21] Vapnik, V. N.: The nature of statistical learning theory. Springer Verlag, New York (1995)
[22] Fan, H. W., Xue, C. Y., Ma, J. T., et al.: A novel intelligent diagnosis method of rolling bearing and rotor composite faults based on vibration signal-to-image mapping and CNN-SVM. Measurement Science and Technology 34(4), 44008–44022 (2023)
[23] Yu, R., Kong, X. H., et al.: Optimizing the diagnostic algorithm for pulmonary embolism in acute COPD exacerbation using fuzzy rough sets and support vector machine. COPD: Journal of Chronic Obstructive Pulmonary Disease 20(1), 1–8 (2023)
[24] Pawlak, Z.: Rough sets. International Journal of Computer and Information Sciences 11(2), 341–356 (1982)
[25] Li, Y. J., Quan, J. S., Tan, Y. Y., et al.: Attribute reduction for high-dimensional data based on bi-view of similarity and difference. Journal of Computer Applications 12(5), 1–16 (2022)

[26] Zhang, X. Y., et al.: A novel rough set method based on adjustable-perspective dominance relations in intuitionistic fuzzy ordered decision tables. International Journal of Approximate Reasoning 154(2), 218–241 (2023)
[27] Lu, J. R., Chen, L., Meng, K. M., et al.: Identifying user profile by incorporating self-attention mechanism based on CSDN data set. Data Intelligence 1(2), 160–175 (2019)
[28] Zhang, Y. X., Zhang, W. D., Wang, L. K., et al.: Study on stress state of loaded concrete based on PSO-SVM. Word Transportation Convention 2022 (WCT 2022), 334–339 (2022)
[29] Zhang, Y. Y., Chen, Y., Yu, S. K., et al.: Bi-GRU relation extraction model based on keywords attention. Data Intelligence 4(3), 552–572 (2022)
[30] Muschelli, J.: ROC and AUC with a binary predictor: a potentially misleading metric. Journal of Classification 37(3), 696–708 (2020)
[31] Li, B., Gatsonis, C., Dahabreh, I. J., et al.: Estimating the area under the ROC curve when transporting a prediction model to a target population. Biometrics 10(5), 1–12 (2022)
[32] Huang, S. X.: Reading the moody chart with a linear interpolation method. Scientific Reports 12(1), 6587–6599 (2022)
[33] Wu, S. Z., Wang, X. W., Wang, Z. N., et al.: Prediction model of bank lending risk using rough set and support vector machine. Journal of Chengdu University of Technology (Science & Technology Edition) 49(2), 249–256 (2022)
[34] Ding, X., Zhao, X. D., Wu, X. J., et al.: Landslide susceptibility assessment model based on muli-class SVM with RBF kernel. China Safety Science Journal 32(3), 194–200 (2022)
[35] Tian, X. C., et al.: Cost-Sensitive Laplacian logistic regression for ship detention prediction. Mathematics 11(1), 119–134 (2022)

AUTHOR BIOGRAPHY

Li Jiang is an associate professor in the School of Management at Hefei University of Technology. She obtained her Doctor's degree at University of Science and Technology of China. Her particular research interests are in Intelligent logistics, artificial intelligence and information management.

Zheng Geng is currently a master student in the School of Management at Hefei University of Technology. He received his Bachelor's Degree from Anhui University of Technology in 2021. His research interests include machine learning, intelligent decision making, and logistics engineering and management.

Dongxiao Gu is a professor in the School of Management at Hefei University of Technology. He obtained his Doctor's degree at Hefei University of Technology. His particular research interests are in artificial intelligence and machine learning, digital economy and service innovation, big data analysis and knowledge engineering management, smart medical and health management, smart social governance and public policy.

Shuai Guo is an associate Professor in the School of Engineering at Hefei University of Technology where he has been a faculty member since 2017. He is the Deputy Dean of the Department of Municipal Engineering. The author completed his Ph.D. at Zhejiang University and his undergraduate studies at Tongji University. His research interests lie in the area of water engineering, ranging from urban hydraulics (ground water infiltration/ inflow assessment, sinkhole problem) to Sponge city design.

Rongmin Huang is a senior engineer in Yangtze Ecology and Environment Co. Ltd. He obtained his master's degree at Wuhan University. His particular research interests are in Municipal engineering and engineering management.

Haoke Cheng is a senior engineer in Yangtze Ecology and Environment Co. Ltd. He obtained his Doctor degree at Hohai University. His particular research interests are in Municipal engineering and engineering management.

Kaixuan Zhu is a master student in the school of Intelligent System and Engineering at Indiana University. He obtained his another Master's degree at Northeastern University in 2020. His research interests include artificial intelligence and machine learning, signal processing and neural networks.