ISSN 2616-5775 Vol. 7, Issue 4: 10-16, DOI: 10.25236/AJCIS.2024.070402

Weighted Multi-view Feature Selection with Genetic Algorithm

Menghan Cui[1,*]

[1]*College of Information Engineering, Nanjing University of Finance and Economics, Nanjing, China*
[*]*Corresponding author: boredcui@foxmail.com*

Abstract: *The feature selection problem of multi-view data has received widespread attention from researchers in recent years. However, existing multi-view approaches suffer from two main issues in weight optimization: (1) Weight coupling problem, where the weights of different views may be coupled, meaning that changing the weight of one view may affect the weights of other views. In such cases, the weight optimization process may be constrained, leading to suboptimal weight allocation. (2) Lack of handling instability, where some algorithms may not fully consider the instability in the weight optimization process, such as noise and changes in data distribution. This can result in unstable weight selections that cannot cope with uncertain data environments. To address these issues, we propose Weighted Multi-view Feature Selection with Genetic Algorithm (WMFS-GA). Specifically, our algorithm combines feature selection results from multiple views and encodes the selected features as initial features. This enables a more comprehensive utilization of information from multiple views, improving the accuracy and robustness of feature selection. We then employ an improved genetic algorithm for weight optimization, allowing for reasonable weight allocation for features from different views during the feature selection process, enhancing the integration and accuracy of multi-view data. Experimental comparisons with several state-of-the-art multi-view feature selection algorithms demonstrate significant advantages in classification performance for our proposed algorithm. Code for this paper available on: https://github.com/boredcui/WMFS-GA.*

Keywords: *Multi-view, Feature selection, Genetic Algorithm*

1. Introduction

With the development of data collection technologies, researchers can use various measuring instruments to collect data on the same entity. Data collected from multiple perspectives or sources is referred to as multi-view data, which typically encompasses various data types and feature representations, such as images, text, audio, etc., providing multiple descriptions of the same object or scene from different angles. Each perspective or source provides different views or observation angles about the same object or scene. These views may include different data types, feature representations, time points, or spatial locations [1]. The characteristics of multi-view data include: (1) Consistency, different views in multi-view data may exhibit consistency, meaning their provided information is not entirely independent but rather interconnected. This correlation can be leveraged to uncover the latent structure and relationships in the data by analyzing the interrelationships between different views. (2) Complementarity, different views may exhibit complementary relationships, where information from one view can complement or compensate for deficiencies in another view, thereby providing a more comprehensive and accurate description of the data [2]. Therefore, multi-view data finds extensive applications in various fields, including computer vision, natural language processing, bioinformatics, etc. Utilizing multi-view data for modeling and learning can offer more comprehensive and accurate information, aiding in addressing complex real-world problems [3]. However, multi-view data faces the challenge of high dimensionality, as each view may contain a large number of features or data dimensions. This may lead to challenges in data processing and analysis, necessitating appropriate dimensionality reduction or feature selection techniques to handle the data.

In the early stages, researchers primarily focused on feature selection methods for single-view data. These methods were typically based on information theory, statistics, or heuristic algorithms such as information gain, analysis of variance, forward selection, etc. However, these methods could only handle single-view data and could not effectively utilize information from multi-view data. As multi-view data became widely used across various fields, researchers began exploring methods to integrate multiple

ISSN 2616-5775 Vol. 7, Issue 4: 10-16, DOI: 10.25236/AJCIS.2024.070402

views for feature selection. At this stage, the main approach involved simple integration of feature selection results from different views, such as arranging features from different views or using a voting mechanism to combine feature selection results from different views. With a deeper understanding of the correlations in multi-view data, researchers started exploring feature selection methods based on inter-view correlations. These methods utilize the correlations between different views to guide the feature selection process, for example, using correlation matrices, joint probability distributions, etc., to evaluate the degree of correlation between different views and perform feature selection accordingly. Graph-based methods emerged as one of the mainstream approaches in multi-view feature selection algorithms. These methods first construct a graph structure for each view and then integrate graph information from different views for feature selection. These methods leverage graph structures to better capture relationships between views and similarities between samples, thus achieving more accurate and robust feature selection [4]. Learning-based methods have emerged as one of the emerging approaches in multi-view feature selection in recent years. Learning-based methods use machine learning models to learn feature selection patterns from multi-view data, such as ensemble learning, deep learning, etc. These methods can automatically learn complex patterns and rules from the data and perform feature selection accordingly [5].

In summary, with a deeper understanding of the characteristics of multi-view data and the continuous development of technology, multi-view feature selection algorithms have been continuously evolving and improving. From simple integration methods to complex graph-based algorithms and learning-based approaches, they have provided more effective tools and techniques for processing and analyzing multi-view data. However, existing methods still suffer from issues such as weight coupling and lack of handling instability. This paper proposes Weighted Multi-view Feature Selection with Genetic Algorithm (WMFS-GA) to address these challenges.

2. Method

As shown in Figure 1, WMFS-GA consists of two main steps. Firstly, feature selection is performed on each view of the multi-view data using the mRMR [6] algorithm, followed by the calculation of weights for each feature. Next, the selected features are encoded as initial features. Secondly, an improved genetic algorithm [7] is employed to further optimize the initial feature population and select the optimal feature subset. My approach comprehensively considers the characteristics of multi-view data and enhances feature selection by effectively leveraging multi-view information and weight optimization.

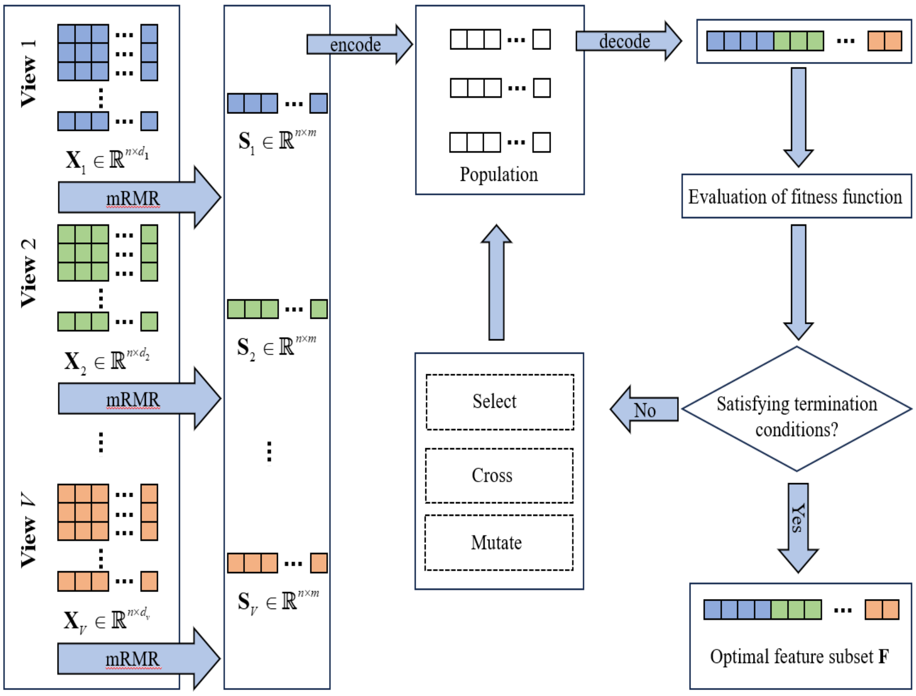


Figure 1: Weighted Multi-view Feature Selection with Genetic Algorithm (WMFS-GA).

ISSN 2616-5775 Vol. 7, Issue 4: 10-16, DOI: 10.25236/AJCIS.2024.070402

Algorithm 1 provides a detailed overview of the WMFS-GA algorithm process. In summary, the algorithm can be divided into the following 4 steps:

(1) Multi-view Feature Selection: Firstly, for each view of the multi-view data, a set of informative features is selected using the minimum Redundancy Maximum Relevance (mRMR) algorithm. This step ensures that each view retains the most relevant and discriminative features related to the target variable.

(2) Weight Calculation: After feature selection, the algorithm calculates the weight of each feature based on its relevance and importance within each view. These weights are then used to guide the optimization process in the subsequent stage.

(3) Genetic Algorithm Optimization: Once features from all views are selected, an initial feature population is formed. Then, a Genetic Algorithm (GA) is employed to further optimize this feature set. The GA iteratively evolves populations of feature subsets, where each individual in the population represents a set of features. The evaluation of individuals is based on their fitness, combining classification accuracy with a weighted penalty term based on feature weights. The algorithm uses tournament selection, crossover, and mutation operations to generate new individuals, preserving the best-performing individuals through elitism. The optimization process continues until a termination condition is met, such as reaching the maximum number of iterations or stagnation in performance improvement.

(4) Result Evaluation: Finally, the algorithm returns the selected features along with relevant evaluation metrics such as classification accuracy, standard deviation, and the elapsed time of the optimization process.

Algorithm 1: Weighted Multi-view Feature Selection with Genetic Algorithm

Input: Multi-view data sets $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_V]$ with the corresponding label **Y**. The number of features to be selected m.

Output: Data matrix of selected features **F**.

```
For each view:
    Perform feature selection using mRMR algorithm
    Calculate accuracy of selected features using SVM
    Store selected features and their accuracies
End
Calculate view weights based on accuracies
Calculate feature weights based on view weights
Initialize feature set and index set with selected features
Initialize a genetic algorithm:
    Set parameters: population size, number of generations, tournament size, mutation rate, crossover rate
    Create initial population of binary feature vectors
    For each generation:
        Evaluate fitness of each individual in the population
        Select parents for reproduction using tournament selection
        Generate offspring through crossover and mutation operations
        Implement elitism by preserving the best individual
        Replace population with offspring
        Check for convergence or maximum generations reached
End
Identify the best individual from the final population
Extract selected features from the best individual
Return selected features F and associated metrics (accuracy, standard deviation)
```

Genetic Algorithm Optimization is the process of evolving and optimizing the initial feature subset. Here is a detailed description:

ISSN 2616-5775 Vol. 7, Issue 4: 10-16, DOI: 10.25236/AJCIS.2024.070402

(1) Initialization of Population: Firstly, the algorithm creates an initial population containing multiple individuals. Each individual represents a set of feature subsets, where the selection of features is represented by binary encoding. The population size is set to 50.

(2) Fitness Evaluation: For each individual, the algorithm computes its fitness. Fitness is typically measured by evaluating the performance of individual feature subsets on the training data. Here, fitness is a weighted sum of classification accuracy and a penalty term based on feature weights, which promotes diversity and generalization of feature subsets.

(3) Selection Operation: Employing the tournament selection strategy, individuals with higher fitness are selected as parents from the population to generate offspring. In the tournament selection process, a group of individuals is randomly selected as competitors, and the one with the highest fitness is chosen as the winner. This process is repeated until a sufficient number of parents are selected.

(4) Crossover Operation: After selecting parents, offspring are generated using crossover operation. An improvement is made to the traditional crossover strategy, where genes that appear at the same position in the parents, referred to as dominant genes, are preserved, while different genes, referred to as non-dominant genes, are available for crossover. This operation helps retain the favorable characteristics of parent individuals and introduce new diversity.

(5) Mutation Operation: To introduce diversity into the population, the algorithm introduces mutation operation in the offspring generated after crossover. With a certain probability, the mutation operation randomly changes certain features of an individual, resulting in new feature subsets. This helps prevent getting trapped in local optima and may discover better feature combinations. The mutation rate is set to 0.01.

(6) Elitism Preservation Strategy: In the generated offspring, the best individuals are retained to ensure that the algorithm maintains or improves the performance of the best individual. This ensures that the population does not lose excellent feature combinations during the evolutionary process.

(7) Iterative Update: These steps constitute one iteration of the genetic algorithm. These operations are repeated multiple times until a termination condition is met, such as reaching the maximum number of iterations or no significant improvement in fitness. Once the termination condition is met, the genetic algorithm optimization part stops and returns the final optimal feature subset along with relevant performance evaluation metrics. The maximum number of iterations is set to 100.

3. Experiments

In this chapter, to validate the effectiveness of the proposed multi-view feature selection algorithm, we conducted a series of experiments on multiple publicly available datasets. Firstly, we will introduce the information of the datasets used in the experiments and the methods used to evaluate the algorithm's performance. Then, we will introduce several state-of-the-art feature selection methods and experimental parameter settings used for comparison. Finally, we will analyze the experimental results.

3.1. Datasets

To evaluate the proposed algorithm, we used four publicly available multi-view datasets. Table 1 summarizes detailed information about the datasets, including categories, sample sizes, and views, as well as the types and dimensions of features in each view. Brief descriptions of each dataset are as follows:

(1) MSRCv1 [8]: This is an object recognition dataset with seven categories. Each category consists of 30 images. Visual features extracted from each image include 256 LBP, 100 HOG, 512 GIST, 48 color moments, 1302 CENTRIST, and 210 SIFT features.

(2) ORL [9]: This dataset contains 40 different subjects, each with 10 different images. The images were taken at different times, under different lighting conditions, facial expressions, and facial details. All images were captured against a dark uniform background with the subjects in an upright frontal position.

(3) Yale [9]: The dataset consists of 165 grayscale images in GIF format. There are a total of 15 subjects, with each subject having 11 images representing different facial expressions.

(4) Sources [10]: The dataset comprises 169 news articles from six categories, sourced from the British Broadcasting Corporation (BBC), Reuters, and The Guardian.

ISSN 2616-5775 Vol. 7, Issue 4: 10-16, DOI: 10.25236/AJCIS.2024.070402

Table 1: Descriptions of datasets

View	MSRCv1	ORL	Yale	Sources
View1	CMT(48)	Intensity(4096)	Intensity(4096)	BBC(3068)
View2	HOG(100)	LBP(3304)	LBP(3304)	Reuters(3631)
View3	LBP(256)	Gabor(6750)	Gabor(6750)	Guardian(3560)
View4	SIFT(210)			
View5	GIST(512)			
View6	CENT(1302)			
Samples	210	400	165	169
Classes	7	40	15	6

3.2. Experimental Setup

We compared our proposed method with 5 state-of-the-art algorithms, with their detailed information and parameter settings as follows:

(1) SCMvFS[11]: This algorithm adopt structure learning and feature selection into one framework to ensure joint optimization of each view. The parameter β is selected from {0.002, 0.006, 0.010, 0.014, 0.018}, while α and γ are selected from $\{10^{-2}, 10^{-1}, 1, 10^{1}, 10^{2}\}$.

(2) CCSFS[12]: This algorithm utilizes consensus clustering structure to guide multi-view learning. The parameters α, β, and λ are selected from the range $\{10^{-2}, 10^{-1}, 1, 10^{1}, 10^{2}\}$.

(3) JMVFG[13]: This algorithm constrains the feature selection matrix through orthogonal decomposition. The parameters η, β, and γ are selected from the range $\{10^{-3}, 10^{-2}, 10^{-1}, 1, 10^{1}, 10^{2}, 10^{3}\}$.

(4) MVSV[14]: This algorithm transfers knowledge from all available views to improve the feature selection process for specific views. In our experiments, to achieve the optimal level, we attempted all possible view training combinations on each multi-view dataset.

(5) MSFS[15]: This algorithm explores the relationships between different views and the multiple local geometric structures of different views to discriminate features. The parameters α, β, γ, and λ are selected from the range $\{10^{-3}, 10^{-2}, 10^{-1}, 1, 10^{1}, 10^{2}, 10^{3}\}$.

For each feature selection algorithm, we selected 40 features and then used the multi-class SVM classifier from the libsvm toolbox, maintaining default settings for classification. To ensure unbiased experimental results, we randomly selected 50% of the instances as training data and used the rest as testing data, repeating this process 10 times.

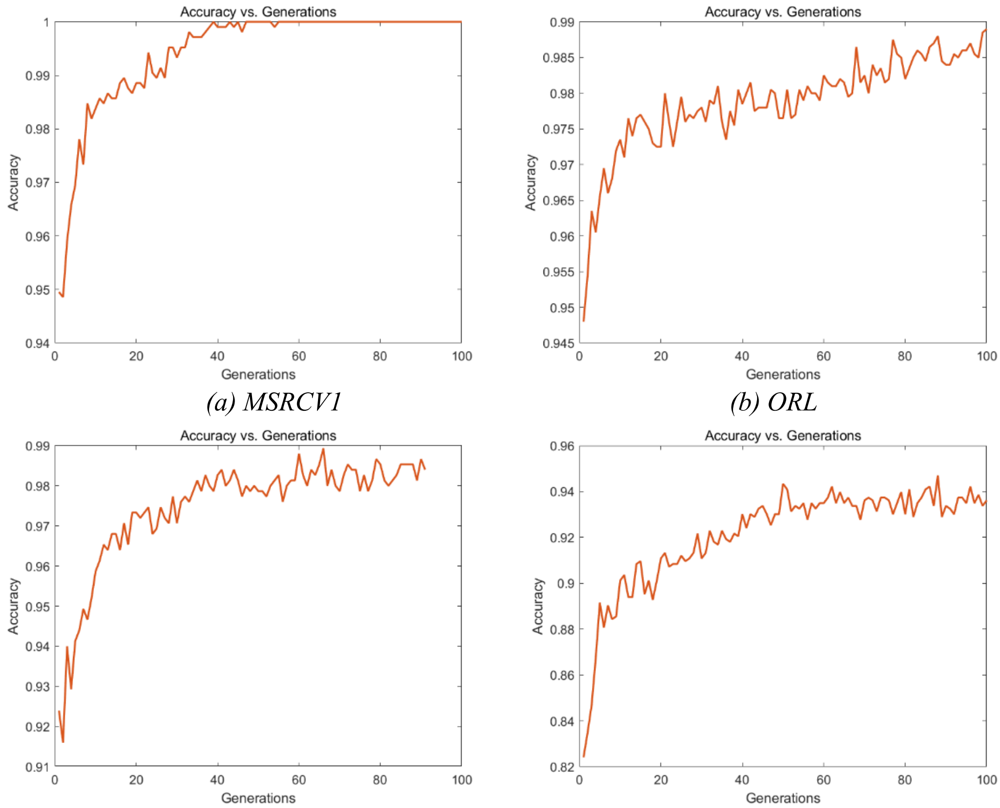
3.3. Experimental Results

As shown in Table 2, our method demonstrated significant advantages over 5 state-of-the-art multi-view feature selection algorithms in terms of multi-class SVM classification performance on the four datasets. Figure 2 illustrates the variations in accuracy during training iterations for our algorithm across the four datasets.

Table 2: Acc%±Std% using SVM classification of different methods on different datasets.

Methods	MSRCv1	ORL	Yale	Sources
SCMvFS	84.15±2.32	79.45±1.59	75.62±2.17	62.65±1.15
CCSFS	93.30±1.09	79.87±1.53	69.77±1.91	62.24±2.18
JMVFG	96.29±1.89	85.21±1.36	78.27±1.55	55.65±2.44
MVSV	92.61±3.58	88.91±2.23	91.87±1.78	88.22±1.17
MSFS	87.58±2.12	63.25±2.01	69.77±2.53	46.26±2.41
Ours	**100.00**±0.00	**98.90**±0.81	**98.93**±1.23	**94.70**±2.29

ISSN 2616-5775 Vol. 7, Issue 4: 10-16, DOI: 10.25236/AJCIS.2024.070402



(a) MSRCV1 (b) ORL

(c) Yale (d) Sources

Figure 2: Convergence curves on different datasets.

4. Conclusions

Multi-view data exhibits characteristics of consistency and complementarity, providing a more comprehensive and accurate data description. Early research primarily focused on feature selection methods for single-view data. However, with the widespread application of multi-view data, researchers began exploring how to integrate multiple views for feature selection. Graph-based and learning-based methods have emerged as main approaches for multi-view feature selection. These methods leverage relationships between views and similarities between samples to provide more accurate and robust feature selection. However, existing methods still suffer from issues such as weight coupling and instability. To address these challenges, this paper proposes Weighted Multi-view Feature Selection with Genetic Algorithm (WMFS-GA), which combines multi-view feature selection with genetic algorithm optimization to identify a compact and discriminative feature subset, thereby improving the classification performance of multi-view data.

References

[1] Yu Z, Li L, Xie J, et al. Pedestrian 3D Shape Understanding for Person Re-Identification via Multi-View Learning [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024.

[2] Hu Y, Wang Y, Li H, et al. Robust multi-view learning via M-estimator joint sparse representation[J]. Pattern Recognition, 2024: 110355.

[3] Sun S, Wang B, Tian Y. Decoupled representation for multi-view learning[J]. Pattern Recognition, 2024: 110377.

[4] Yu Y, Du Z, Meng L, et al. Adaptive online continual multi-view learning[J]. Information Fusion, 2024, 103: 102020.

[5] Khorasani M, Kahani M, Yazdi S A A, et al. Towards finding the lost generation of autistic adults: A deep and multi-view learning approach on social media[J]. Knowledge-Based Systems, 2023, 276: 110724.

[6] Ramírez-Gallego S, Lastra I, Martínez-Rego D, et al. Fast-mRMR: Fast minimum redundancy maximum relevance algorithm for high-dimensional big data[J]. International Journal of Intelligent

ISSN 2616-5775 Vol. 7, Issue 4: 10-16, DOI: 10.25236/AJCIS.2024.070402

Systems, 2017, 32(2): 134-152.
[7] Lambora A, Gupta K, Chopra K. Genetic algorithm-A literature review[C]//2019 international conference on machine learning, big data, cloud and parallel computing (COMITCon). IEEE, 2019: 380-384.
[8] Winn J, Jojic N. Locus: Learning object classes with unsupervised segmentation[C]//Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1. IEEE, 2005, 1: 756-763.
[9] Cao X, Zhang C, Fu H, et al. Diversity-induced multi-view subspace clustering[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 586-594.
[10] Wei H, Chen L, Chen C L P, et al. Fuzzy clustering for multiview data by combining latent information [J]. Applied Soft Computing, 2022, 126: 109140.
[11] Cao Z, Xie X. Structure learning with consensus label information for multi-view unsupervised feature selection [J]. Expert Systems with Applications, 2024, 238: 121893.
[12] Cao Z, Xie X, Sun F, et al. Consensus cluster structure guided multi-view unsupervised feature selection [J]. Knowledge-Based Systems, 2023, 271: 110578.
[13] Fang S G, Huang D, Wang C D, et al. Joint multi-view unsupervised feature selection and graph learning [J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2023.
[14] Komeili M, Armanfard N, Hatzinakos D. Multiview feature selection for single-view classification [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 43(10): 3573-3586.
[15] Zhang Y, Wu J, Cai Z, et al. Multi-view multi-label learning with sparse feature selection for image annotation [J]. IEEE Transactions on Multimedia, 2020, 22(11): 2844-2857.