

Research Article

Semisupervised Association Learning Based on Partial Differential Equations for Sparse Representation of Image Class Attributes

Wei Song,¹ Guang Hu ,² Liuqing OuYang,³ and Zhenjie Zhu³

¹School of Sports Engineering and Information Technology, Wuhan Sports University, Wuhan Hubei 430079, China

²School of Resource and Environmental Sciences, Wuhan University, Wuhan Hubei 430072, China

³School of Physical Education, Wuhan Sports University, Wuhan Hubei 430079, China

Correspondence should be addressed to Guang Hu; 00008620@whu.edu.cn

Received 27 August 2021; Accepted 8 September 2021; Published 19 October 2021

Academic Editor: Miaochao Chen

Copyright © 2021 Wei Song et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Semisupervised learning is an idea that addresses how to use a large number of unlabeled samples and a limited number of labeled samples to learn decision knowledge together. In this paper, we propose a multitask multiview semisupervised learning model based on partial differential equation random field and Hilbert independent standard probability image genus attribute model, i.e., shared semantics. In the framework of the image-like genus attribute model, data from different data sources are generated by their shared hidden space representation. Different from the traditional model, this paper uses the Hilbert independence criterion to inscribe the shared relationship of hidden expressions. Meanwhile, to exploit the correlations between labels in the label space as well, this paper uses the partial differential equation random field to inscribe the correlations between different kinds of labels in the label space and the correlations between hidden features and labels. Using the variational expectation-maximization algorithm, the whole generative process model can be inferred. To verify the effectiveness of the model, two artificial datasets and three real datasets are tested in this paper, and the experimental results verify the effectiveness of the algorithm in the paper. On the one hand, it not only improves the classification accuracy of the multiclassification problem and the multilabel problem; it also outputs the association structure between different kinds of labels and between hidden features and labels.

1. Introduction

Data contains a rich value, and nowadays, in the era of big data, the application of massive high-dimensional data lacks suitable means. The multilabel learning framework is to address these multisemantic phenomena. In the framework, each data object is described by an example (feature vector), which can belong to multiple categories. When machine learning and data mining techniques are applied to high-dimensional multilabel data, an important issue is a dimensional catastrophe. Therefore, multilabel feature selection techniques have emerged. In the past few years, multilabel feature selection has attracted the attention of many researchers and some excellent algorithms have emerged [1]. However, they still have some problems that are difficult

to solve: (1) to select features, existing feature selection algorithms usually adopt one of these two strategies: selecting a common subset of features for all tags that are discriminative to all tags (shared features) or selecting features for each tag separately that are discriminative to themselves (class features), these important features play an important role during the label recognition process, and they are important for the recognition ability of the selected features; (2) exploring and exploiting label relevance in feature selection are considered an important idea to improve the performance of the algorithm; although existing algorithms have achieved good results, it is necessary to explore new methods to improve the performance of the algorithm recently. In addition, existing multilabel feature selection algorithms tend to exploit label global correlation. However, label relevance is usually

local and shared by local regions of the dataset; (3) existing multilabel feature algorithms are usually modeled based on the raw label information of the data; however, this label information cannot fully express the rich semantics of the object; on the one hand, the relevant labels are of different importance to the example because the relevant labels of the example usually describe it with different degrees. On the other hand, the label importance cannot be provided directly by the data annotator.

In traditional supervised learning, the learner learns from a large number of labeled examples to build a model for predicting future example labels. The “label” here is the output corresponding to the example, which is the category of the example in classification problems, and the real-valued output corresponding to the example in regression problems. At the same time, with the rapid development of data collection and storage technologies, it has become quite easy to collect a large number of unlabeled examples, while it is relatively difficult to obtain a large number of labeled examples because obtaining these labels can be labor-intensive [2]. The large amount and low cost of unlabeled data can be used to assist supervised learning to improve the prediction efficiency and accuracy as well as reduce the prediction cost. If only a small number of labeled examples are used, it is often difficult to train a learning system with strong generalization ability; on the other hand, if only a small number of “expensive” labeled examples are used instead of a large number of “cheap” unlabeled examples, it is a great waste of data resources. On the other hand, it is a great waste of data resources if only a small number of “expensive” labeled examples are used instead of a large number of “cheap” unlabeled examples. Therefore, how to use a large number of unlabeled examples to improve the learning performance when there are few labeled examples has become one of the most concerning issues in current machine learning research [3]. In this paper, we study a semisupervised association learning method based on partial differential equations for sparse representation of image class attributes, focusing on the problems of solving semisupervised multilabel learning and semisupervised multiclassification learning.

2. Related Work

The main traditional machine learning is supervised learning and unsupervised learning. Among the classic scenarios of supervised learning are the two main categories of classification and regression. Semisupervised learning, which has received a lot of attention in the last decade or so, is dedicated to using a large number of unlabeled samples to complement a relatively small number of manually labeled samples, thus hopefully training a more accurate classifier than if only manually labeled samples were used.

Self-training methods, proposed in the literature [4], were the first methods to use samples without class labels for supervised learning. This class of methods mainly uses the idea of iteration, where supervised learning is repeated continuously, and the resulting optimally labeled results are applied to the next round and added to the sample set

along with the class labels to continue iteration and iterative supervision. The advantage of this method is its simplicity and ease of operation, but it is prone to erroneous errors thus leading to a vicious circle in the iteration. The literature [5] first proposed “semisupervised” and can use semisupervised for classification. In [6], it is shown that the use of unlabeled samples can mitigate the “Hughes” phenomenon in small samples, and this idea has led to widespread interest in unlabeled samples and semisupervised learning. The literature [7] proposed semisupervised learning for deep generative models. The first semisupervised distance metric learning method was proposed in the literature [8]. The literature [9] proposed a particle swarm optimization algorithm based on a semisupervised classifier for solving the classification problem of Chinese text. The literature [10] proposed a semisupervised hashing method for dealing with the retrieval problem of large-scale images. The literature first proposed the minimum partitioning operator, where the source node is a positively labeled sample instance and the target node is a negatively labeled sample instance, to find a set of edges that can split the source and target nodes after deletion, and this set of edges in the graph cut, and the graph is also split into two independent parts. After that, there are other algorithms that gradually emerged; the literature [11] studied the energy function minimization and confirmed the high efficiency of the graph cut algorithm. The literature [12] proposed the proportional cut method as well as the normalized cut method. The literature [13] summarized the streamlined semisupervised learning method and proposed the popular regularization method. After that, the literature [14] proposed the regularization of online stream shapes, which improves the applicability of stream shape regularization in large-scale data. The literature [15] used strong domain knowledge to construct graphs and then performed semisupervised learning for character recognition based on graphs. The edges in the graph are a combination of temporal, color, and face edges, such that the graph reflects strong domain knowledge, a deep understanding of the problem structure, and how to use unannotated data. The literature [16, 17] improves the problem of overadjustment of affiliation in AFCC algorithms and proposes an improved class of semisupervised fuzzy clustering algorithms. The literature [18] investigated the effect of pairwise constraint attributes on the effect of semisupervised clustering.

3. Semisupervised Association Learning Based on Partial Differential Equations for Sparse Representation of Genus Attributes of Image Classes

3.1. Algorithm for Sparse Representation of Attributes of Image Classes Based on Partial Differential Equations. The sparse representation of image class properties based on partial differential equations is the problem of solving the optimal solution of an energy generalized function about images, which is an ill-posed inverse problem, so regularization theory is applied to transform the ill-posed problem into an ill-

posed problem. The following convex combined variational regularization model is first proposed [19]:

$$\begin{aligned} \min K(x) &= \frac{pb\sqrt{b^2+7ac}}{3a} \int_{\Omega} (n-rx)^m \nabla y dy dx \\ &+ \rho \int_{\Omega} (k-ry)^m \nabla y dy dx, \\ R^m f(x) &= \int \int g(t) dt = \left(\frac{m+\gamma}{n} \right) \cdot \sum (x^m - t^{m-1}) f(t), \quad (1) \end{aligned}$$

where p is the definition domain of the image; r is the image after noise reduction; a is the noisy image; b is the gradient operator; c is the regularization parameter; and ρ is the variational order control parameter; the first term on the right side of the equation is the loyalty term, and the second term is the regularization term that has a soothing effect on the image. The model in this section has the following cases.

- (1) When $m = 1$, the model in this section can be rewritten as

$$\begin{aligned} \min K(x) &= \frac{pb\sqrt{b^2+7ac}}{3a} \int_{\Omega} (n+\rho k - rx - rpy) \nabla y dy dx, \\ Rf(x) &= \int \int g(t) dt = \left(\frac{1+\gamma}{n} \right) \cdot \sum (x-1) f(t) \quad (2) \end{aligned}$$

At this point, the model degenerates to a TV model with the regularization operator as a first-order variation. From the previous section, it is clear that the model has good edge-preserving performance but produces a "step effect."

- (2) When $m = 0$, the model in this section can be rewritten as

$$\begin{aligned} \min K(x) &= \frac{pb\sqrt{b^2+7ac}}{3a} \int_{\Omega} (1+\rho) \nabla y dy dx, \\ f(x) &= \int \int g(t) dt = \frac{\gamma}{n} \cdot \sum (1-t^{-1}) f(t) \quad (3) \end{aligned}$$

- (3) When $m \in (0, 1)$, the model in this section is similar to the TVBH model, taking into account both first-order and second-order variables and is a fusion of the TV and BH models

In summary, the choice of parameter m determines the filtering form and filtering performance of the new model. As the parameters are often tried to be scraped together by a large number of experiments or the best value is obtained empirically, it is a rough evaluation of the global content and ignores the local features of the image.

Next, the local features of the image are considered to improve the adaptivity of the model in this section by replacing the constant m with an edge diffusion function $m(z)$, and a variable-order variational model is proposed, which is optimized in the following new model form [20].

$$\begin{aligned} I^\alpha f(\eta) &= m(z)(i\omega)^\alpha f(\omega) = |\omega|^\alpha e^{i2\pi\alpha} f(\omega), \\ F(\rho) &= \int k^{-i2\pi\omega x} m(z) f(x) dx, \quad (4) \end{aligned}$$

of which

$$\theta(z) = \frac{z_i}{\sum_{i=1}^n X_i^2}. \quad (5)$$

To better detect detailed information such as edge texture contained in the image, the construction of the feature detection factor λ concerning the edge diffusion function $m(\lambda)$ incorporates both gradient and local entropy feature detection operators.

3.1.1. Image Gradient. The gradient table characterizes the magnitude and direction of change of the image gray value, so the gradient modulus is commonly used to distinguish the edge regions and nonedge regions of the image. The gradient of edge regions is larger, and the gradient of flat regions is smaller, but the gradient of some detail information is not much different from that of flat regions, and the gradient at noise points is even larger than that of edges, so that the gradient edge detection operator may misjudge the weak edge regions and strong noise points in the image with rich details, resulting in the loss of detail information or incomplete noise reduction of the processed image.

3.1.2. Image Local. Local entropy characterizes how drastically the grayscale values of pixels in local regions of an image change and thus can reflect the richness of the information contained in the image. The entropy value of a grayscale image f of size $k * l$ is defined as

$$\begin{aligned} KL(x, y) &= \frac{\delta y}{\delta x} \cdot \sum_{i=1}^n X_i Y_i, \\ P_{ab}^A(t) &= \frac{\partial \left([\alpha_a^2(t)]^k [\beta_b^2(t)]^j / \sum_{m \neq a, n \neq b} [\alpha_m^2(t)]^k [\beta_n^2(t)]^j \right)}{\partial t}, \quad (6) \end{aligned}$$

where $f(i, j)$ denotes the gray value of the pixel located at the point (i, j) of the image; $P(i, j)$ denotes the distribution probability of the gray value of the pixel at the point (i, j) in a local neighborhood of size $k * l$; H denotes the local entropy of the image. Through the local entropy, the local characteristics of the image can be effectively determined, and the local entropy value is larger in the edge detail region with complex grayscale distribution and smaller in the flat region with the uniform grayscale distribution. In addition, local entropy has strong noise immunity, and independent noise

points have little effect on it. Therefore, local entropy can be widely used in image processing.

The continuous first-, second-, and fourth-order differential operators as well as the scattering operator are first discretized. In addition, to further improve the speed of the split Bregman algorithm computation, periodic boundary conditions are used to make the FFT applicable to the split Bregman algorithm. Let γ be a two-dimensional grayscale image region of size $k * l$, and the coordinates of the image column and row directions are denoted by x and y , respectively. The first-order forward differencing at pixel (i, j) along the coordinate x and y directions is noted as

$$E_{tx}(x, d) = \begin{cases} E_e \times x + \varepsilon_{\text{free}} \times y^2 \times d^{1/2} + \gamma, & d \leq d_0, \\ E_e \times x + \varepsilon_{\text{decay}} \times y^2 \times d^2 + \gamma, & d > d_0, \end{cases}$$

$$E_{\text{all}} = \sum_{i=0}^R \left(E_{\text{contest}} \times x + E_e \times x + \varepsilon_{\text{free}} \times x^2 \times d_i^{1/2} \right) + \sum_{i=0}^K \left(4E_e \times x + \varepsilon_{\text{decay}} \times x^2 \times d_i^2 \right). \quad (7)$$

The first-order backward differential is noted as

$$E_{tx}(x, d) = \begin{cases} E_e \times x - \varepsilon_{\text{free}} \times y^2 \times d^{1/2} + \gamma, & d \leq d_0, \\ E_e \times x - \varepsilon_{\text{decay}} \times y^2 \times d^2 + \gamma, & d > d_0, \end{cases}$$

$$E_{\text{all}} = \sum_{i=0}^R \left(E_{\text{contest}} \times x - E_e \times x + \varepsilon_{\text{free}} \times x^2 \times d_i^{1/2} \right) - \sum_{i=0}^K \left(4E_e \times x + \varepsilon_{\text{decay}} \times x^2 \times d_i^2 \right). \quad (8)$$

3.2. Semisupervised Associative Learning in the Sparse Representation of Image Class Attributes Based on Partial Differential Equations. Data from different regions, generations, and individuals are characterized by a huge amount of data and also contain a great deal of information. In the new technological era, there is an urgent need to analyze data from different sources and to integrate them efficiently to obtain information about their intrinsic structure. In the face of certain complex challenges, it is possible to understand and analyze them in steps and solve them one by one in a smaller individual perspective so that the original challenges are solved. A comparison of unsupervised image classification and supervised image classification processes based on deep learning is shown in Figure 1. In recent years, deep learning algorithms have been continuously updated with the latest performance results for image classification tasks, showing a strong burst of power, but they also face some specific challenges and still have room for improvement. In addition to common problems such as time-consuming training process, high hardware requirement

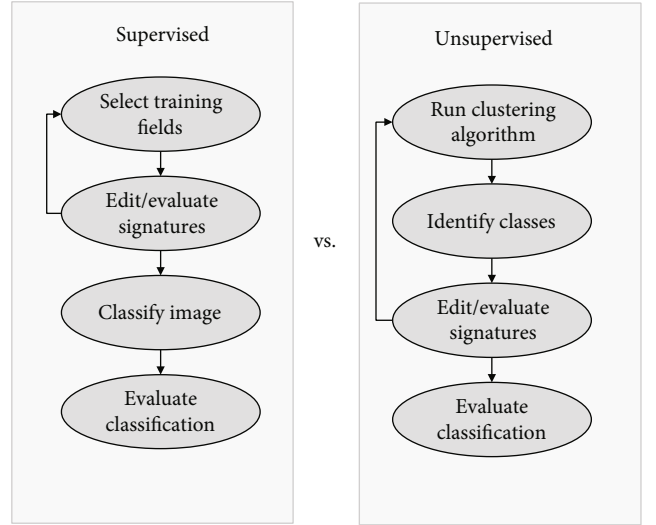


FIGURE 1: Comparison of unsupervised image classification and supervised image classification process.

standards, and difficulty in portability, there are also some domain-specific problems [21].

Certain teams in machine learning work on theories and algorithms related to semisupervised associative learning of sparse representations of image-like attributes based on partial differential equations, integrating multiple single solutions into a comprehensive answer that yields a convincing answer for all viewers. It possesses better accuracy and robustness and is more stable than using a particular model. This model has been successfully implemented in several directions. The tasks of semisupervised associative learning are mainly classified into classification tasks, clustering tasks, and semisupervised learning tasks, which specifically enable specific processes such as collaborative filtering, anomaly detection, distributed computing, and multisource data fusion, making it a powerful tool for data analysis. By the assistance of a single learning model, one is exploring the unknown dataset from a single perspective and can only get a one-sided learning result, but if one can brainstorm and explore the unknown dataset through several different perspectives with the help of a semisupervised associative learning model, then one can perform several learning processes simultaneously and eventually, one or more learning results can be obtained as well. The overall framework of semisupervised associative learning is shown in Figure 2.

Semisupervised clustering is mainly guided by supervised information to the traditional clustering algorithm, and the two types of supervised information are class labels and pairwise constraints. When introducing supervised information, the supervised information needs to be selected properly, and the amount of effective information selected is high, which has a positive impact on time and results in the subsequent clustering process; at the same time, it is also necessary to identify and consider whether the supervised information is reliable and avoid wrong supervised information as much as possible or too redundant amount of information, such as two samples labeled as must-link constraints, but the amount of information of two samples even if the unlabeled

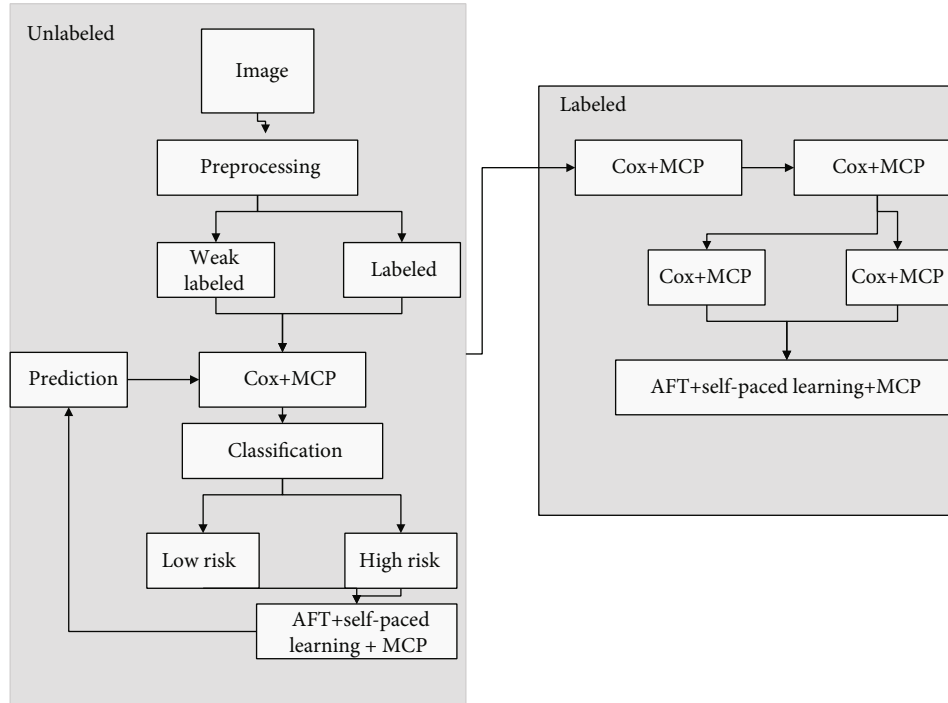


FIGURE 2: Overall framework of semisupervised associative learning.

information is also in a cluster, which will not have any effect on the results, and at this point, the labeling cost is considered wasteful. To optimize the supervised information, scholars try to select the supervised information through active learning to achieve more accurate screening results. Two typical algorithms that combine active learning with semisupervised clustering are the APCKmeans algorithm and the IASSCF framework. Each marker has its original characteristics, which have an important role and function for the inherent properties of the marker itself, providing stronger evidence for the existence of certain marker properties in the sample. Therefore, multitag learning can be performed more effectively through the study of class attributes. Feature selection is achieved through the study of class attributes; however, some of the features processed through the class attribute approach may still have redundancy, and the redundancy in the feature space can be effectively addressed through mutual information theory. Mutual information can be formulated as a mainstream statistical algorithm based on the extension and expansion of information theory and statistical theory, which can provide an accurate description of the correlation that exists between most samples and categories with each other. First, a sparse representation of class attributes is performed. However, for the features processed through the class genus attributes, there may still exist a certain redundancy; therefore, combining information entropy to all features in the new feature, space separately calculates their mutual information with the marker space based on the size of the mutual information the sequential arrangement of features obtains the relevant feature subset; and then further considers the redundancy of the feature space through the mutual information theory based on the study of the class genus attributes. This is more effective to improve the multitag classification accuracy.

The fine-grained features are the most characteristic and important information in the process of image sparse representation. To further learn finer discriminative features, local regions with discriminative features are further localized, finer fine-grained features are learned, and fine-grained features of different scales are fused for classification. Specifically, firstly, different regions in the image are scored using anchors of different sizes, and the discriminative local regions in the original image are initially filtered, and the local regions with less information are filtered out to reduce the interference of the regions with poor effective information on the classification results and effectively reduce the computational cost. Secondly, zooming in on the images of key regions after filtering and locating them with discriminative regions for the second time enables the designed model to capture finer features in the images and obtain higher-quality fine-grained feature information. Finally, the weights between images of different scales are learned, and the interscale weights are used to fuse the fine-grained image information of different scales to provide a rich decision basis for the final fine-grained image classification results. The fine-grained image information at different scales collaborates to jointly correct the final classification results. In fine-grained images, individual local regions contain different amounts of information, and thus, they contribute differently to the accurate recognition of various fine-grained images. The information-rich local regions contain more discriminative feature information which contributes to enhancing the correct recognition of the model for different fine-grained images. Therefore, for the final feature representation, the analysis focuses on the fine-grained features of the effective information-rich local regions, while weakening the information of those local regions with less information. To further improve the classification performance of the model and

effectively fuse the feature information of discriminative regions at different scales, it is necessary to constrain the weights of the learned fine-grained features at different scales using partial differential equations.

It is known that the proportion of pairwise constraint information in the total data sample is very small, so we add the concept of partial differential equation to describe the data samples with constraints and adjust the weight measure of pairwise constraint information, and consider the case that the boundary points of clusters have fuzzy divisions, and force the active addition of constraint information to the fuzzy boundary, and propose an improved partial differential equation-based active semisupervised fuzzy clustering algorithm is proposed, hoping to improve the traditional SFCEM algorithm and MEC algorithm. The ASFCM-CE algorithm is improved mainly from two perspectives: (1) the self-information and constraint information are described by partial differential equations, and weights θ are added to adjust the objective function; (2) there must be pairwise constraints to control the clustering boundary; i.e., pairwise constraints are added actively for fuzzy boundary points. In general, the amount of labeled data in the dataset is much smaller than the amount of unlabeled data, and at this time, it is not enough to guide the pairwise constraints only by θ . At this time, we adjust the weights of the labeled data so that the constraint information can better guide the subsequent iterations.

The proposed semisupervised association learning algorithm process based on partial differential equations for sparse representation of image class attributes is divided into three main stages. First, features are randomly selected from all candidate features to form a series of random feature subspaces. Second, weighted constraint selection and constraint projection are performed on the above subspaces to improve the clustering quality. Third, a scheme is designed for integrating the clustering solutions generated in each subspace to obtain a more robust uniform clustering solution.

4. Experimental Design and Conclusion

To verify the clustering effect of the proposed partial differential equation-based semisupervised association learning algorithm for sparse representation of image attributes, the clustering effect of the ARSCE method on several real data sets is evaluated based on normalized mutual information in this section. To ensure the validity of the experimental results and avoid the influence of chance, the method is run 20 times for each experiment, and the average of the 20 times is calculated as the final experimental result. The extraction rate of the used pairwise constraint set is set to 0.2; i.e., 20% of the real label set is extracted to construct must-link and cannot-link constraint sets.

The effect of sampling rate on clustering performance is first explored in terms of standard mutual information (NMI), where the sampling rate determines the number of features in each subspace. This experiment was conducted on 3 datasets, namely, Alizadeh-2000-v3, Armstrong-2002-v2, and lymphoma. Here, the sampling rate was varied between 4 and 8. Figure 3 demonstrates the effect of sam-

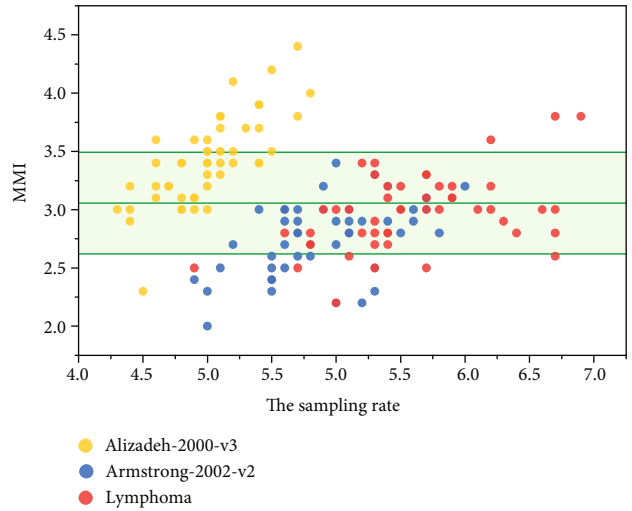


FIGURE 3: Effect of feature sampling rate on clustering performance.

pling rate on clustering performance. As can be seen from the figure, in general, the performance improves with an increased sampling rate. This means that more and more informative features are selected to facilitate clustering. However, when the sampling rate reaches a certain value, a clear downward trend in the clustering performance can be noticed. A possible reason for this is the selection of redundant features in this setup, which negatively affects the clustering. In most cases, the optimal sampling rate is between 3 and 4, while for the Armstrong-2002-v2 dataset, the optimal value of the sampling rate is between 2 and 3. By another way, different data have their own more desirable sampling rates. In this case, the optimal sampling rate needs to be chosen specifically. From the feature selection perspective, this paper argues that it is necessary to explore a more reasonable strategy in constructing a random feature subspace by selecting more efficient information features. Thus, multiple different clustering partitions with satisfactory performance can be generated.

Next, the effect of pairwise constraints on clustering performance is explored by increasing the percentage of pairwise constraints. In general, a larger percentage of pairwise constraints indicates that more supervised information is available to drive the clustering method for better clustering performance. Figure 4 shows the effect of pairwise constraints on performance for the six datasets. From the figure, it can be observed that the overall clustering performance shows an increasing trend at different levels as the number of pairwise constraints increases. This implies that these pairwise constraints provide effective supervisory information, which helps the clustering process when discovering the clustering-friendly space. When the number of paired constraints is all set to a tolerance of 10, with equal division in the range 0-100, the five compared algorithms show higher overall accuracy on large data samples than on small data samples, with the ASFCM-CE algorithm being the most prominent among the five. However, when observed separately on the large data sample dataset, the ASFCM-CE algorithm is less stable compared to the other algorithms, and there are even several instances where the SCE-SSC algorithm

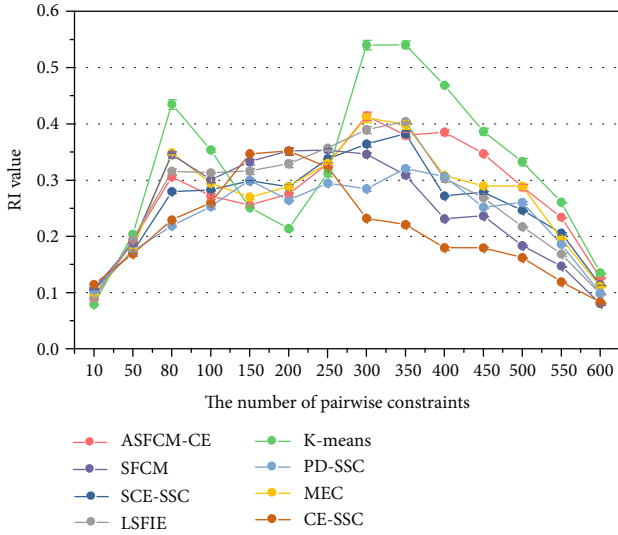


FIGURE 4: Effect of the number of pairwise constraints on the performance of the algorithm.

is more accurate than the ASFCM-CE algorithm. Although the ASFCM-CE algorithm is less stable in the 0-100 pairs, because of the large dataset, the overall accuracy still shows an increasing and stable trend if divided by 0-300 pairs.

To verify the stability of different multitag feature selection algorithms, the stability of the algorithms can be expressed by iterative validation. Since predictive classification has a large variation in results on different datasets using different evaluation metrics, the results are all normalized between 1 and 10 as a general criterion. Finally, the stability index is represented by the standardized values. Figure 5 shows the stability of the six algorithms on the six datasets. In the figure, the LSFIE algorithm provides a very stable solution on 5 datasets and the stability index is also between 8.2 and 9.8. On the Genbase dataset, the stability index is also between 7.3 and 8.4, which also yields fairly stable results. In summary, the results show that the LSFIE algorithm is better stable and its stability index values do not fluctuate much and are better. The results of the LSFIE algorithm are unstable on very few datasets but are more stable and slightly better on most datasets.

This paper also explores the effect of the number of integrated members on the clustering performance based on normalized mutual information (NMI), as shown in Figure 6. From the figure, it can be seen that the performance shows an increasing trend as the number of integration members increases. This means that more integration members can provide more informative and auxiliary information for better clustering. When the number of members reaches a certain threshold, it will reduce the improvement in performance, in line with the law of diminishing marginal benefit, and reaching the same performance improvement at this time means adding more cost, which needs to be balanced between computational cost and performance improvement.

The effect of the feature detection factor λ on the clustering performance is studied according to normalized mutual information (NMI), as shown in Figure 7. In this study,

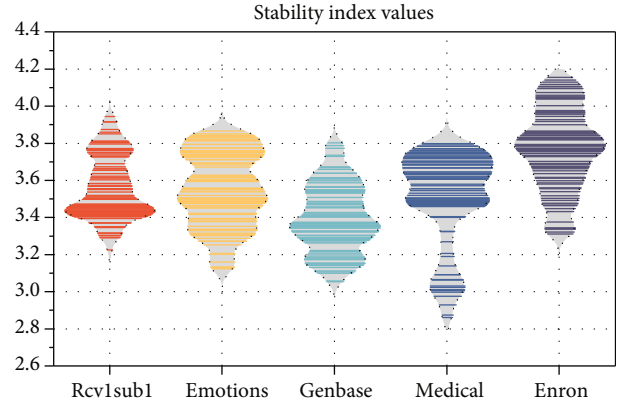


FIGURE 5: Stability index values.

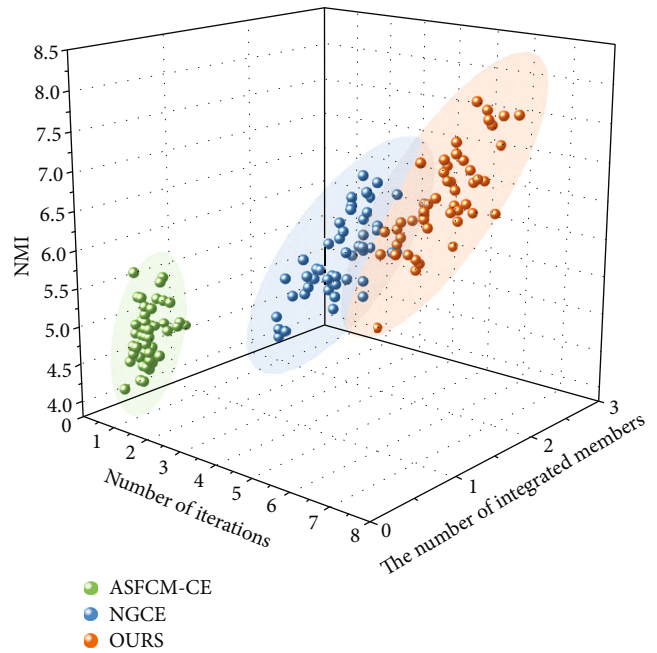


FIGURE 6: Effect of the number of integrated members on the clustering performance.

the value of λ was varied by changing it between 0.1 and 0.9. It can be noted from Figure 7 that when λ increases, the clustering performance shows a rapid upward trend before peaking, followed by a decreasing trend of varying degrees. In addition to this observation, it can be found that for different datasets, there are respective appropriate λ . This suggests that the method in this paper is sensitive to λ , which is used to control the distribution of the weight association graph in the newly learned space. The optimal value of the equilibrium parameter λ is between 0.4 and 0.6 in most cases, except on the dataset nci9 where the optimal value of λ is between 0.6 and 0.8. In this paper, we argue that the distribution of data samples on different clusters is somewhat different from the distribution of other data sets. Therefore, it is necessary to choose its more optimal value to adjust the weights of the learned association graph to obtain better performance.

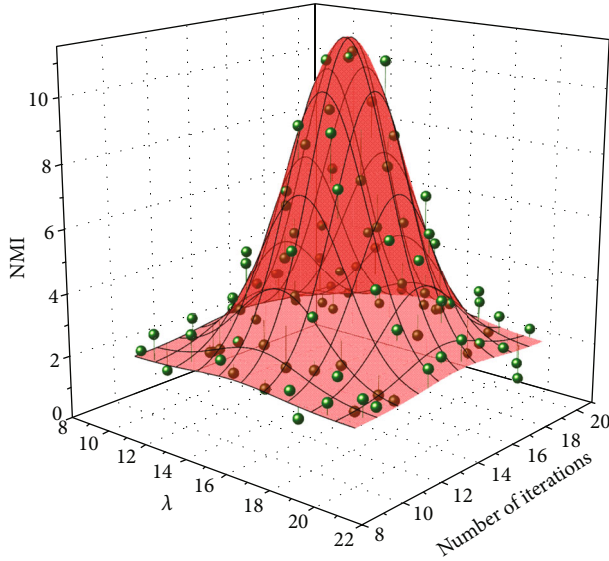


FIGURE 7: The effect of λ on the clustering effect.

In general, the more the number of pairwise constraints, the better the clustering effect should be; however, according to the above analysis, the clustering effect of both PD-SSC and CE-SSC algorithms tends to decrease with the increase of the number of pairwise constraints. The possible reason for this phenomenon is that to simplify the experimental process, a fixed penalty coefficient is chosen, and when the number of pairwise constraints increases, the weight of the penalty term in the objective function will increase, which affects the clustering effect. In this case, the penalty coefficient decreases as the number of pairwise constraints increases, and the weight of the penalty term is reduced to obtain a better clustering effect. Taking the iris-wine data set as an example, the penalty coefficients are reduced appropriately as the number of pairwise constraints increases, and the penalty coefficients are 0.7 for the CE-SSC algorithm and 0.9 for the PD-SSC algorithm, where N is the number of pairwise constraints. After the penalty coefficients were adjusted, the values of the three indicators of the PD-SSC algorithm showed a fluctuating upward trend, and the clustering effect was significantly improved. In Figure 8, the index value of the CE-SSC algorithm increases continuously when the number of pairwise constraints is greater than 35; the index value of the PD-SSC algorithm increases slightly when the number of pairwise constraints is greater than 15; the index value of the CE-SSC algorithm decreases continuously when the number of pairwise constraints is 20-80 and then increases slightly. The index value of the PD-SSC algorithm reaches the highest value at the number of pairwise constraints of 20, but the overall trend is slightly increasing. This is one of the reasons for introducing the semisupervised correlation learning algorithm based on partial differential equations for the sparse representation of attributes of image classes.

In addition, the comparative results of recent semisupervised clustering integration methods and the proposed method are analyzed. The homogeneous algorithms include neural gas-based clustering integration algorithm (NGCE), stochastic K -means-based clustering integration algorithm

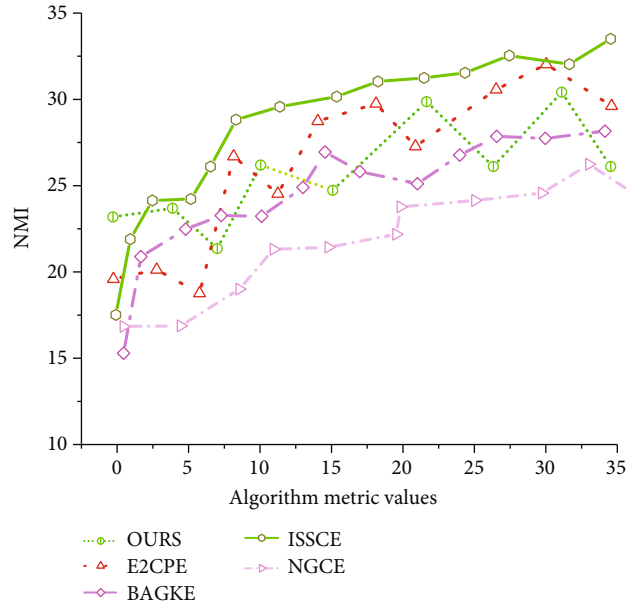


FIGURE 8: Different algorithm metric values after adjusting the penalty factor.

(RSKE), bagging-based K -means clustering integration algorithm (BAGKE), hierarchical clustering integration algorithm (HCCE), clustering integration algorithm using constraint propagation (E2CPE), incremental semisupervised clustering integration algorithm (ISSCE), and double weighted semisupervised integration clustering algorithm (DCCP). The following observations can be obtained: (1) E2CPE can achieve better performance compared to NGCE, RSKE, BAGKE, and HCCE methods because E2CPE methods use constraint propagation techniques to make better use of supervised information, which helps guide the clustering process. This illustrates the effectiveness of pairwise constraints to improve the quality of clustering. (2) Both constraint weighting and constraint projection weighting transform the feature subspace into a space that is friendly to clustering, yielding high-quality clustering solutions with sufficient diversity. This can be seen from the fact that both ISSCE and DCECP achieve better performance than E2CPE on most datasets. (3) The proposed method in this paper achieves the best or at least better performance on all datasets, which indicates the necessity of using adaptive clustering integration to assign appropriate weights to the underlying clustering solutions and combine them to form a better clustering partition. In other words, it validates the effectiveness of the diffusion fusion approach.

5. Conclusion

In this paper, semisupervised associative learning based on partial differential equations for image genus attribute coefficient representation is used as the research background, and the problems of partial differential equations for image genus attribute representation, solving semisupervised multilabel learning, and semisupervised multiclassification learning are primarily studied. As the data magnitude increases and the

structure becomes increasingly complex, a new semisupervised associative learning integration method based on partial differential equation image class attribute coefficient representation is proposed in this paper to better handle the data clustering problem. These problems are very common in the field of machine learning, and there is a large amount of related work. Unlike traditional approaches to these individual problems, this paper is a reinterpretation of these two problems from a new perspective by combining matrix complementation and generative models, respectively, and tests the effectiveness on several simulated and real datasets. The goal of feature selection is to obtain a subset of features that satisfy the conditions under some specific evaluation metric criteria, which is essentially a comprehensive optimization problem for the objective. The amount of information of each feature is generally calculated when performing feature selection, and then, all features are ranked according to their information size, and the desired number of features is selected. In this paper, when performing multitag feature selection, rough set theory can effectively evaluate imprecise and unstable data, to analyze and process the data more efficiently, uncover the potential connotation, and reveal the potential law. According to the principle of maximum correlation and minimum redundancy, the correlation between features and tokens is calculated based on the affiliation in the rough set, and then, the Kendall correlation coefficient is used to measure the redundancy between unselected features and selected features, and finally, the correlation and redundancy are made to calculate the difference, and the differences are ranked for magnitude, and the desired features are selected. Finally, the experimental results on multiple datasets illustrate the effectiveness of the algorithm. Traditionally in multitag learning, tags are predicted from the same set of attributes, ignoring certain features of the tags themselves. These unique attributes have strong discriminative power for the tokens, so strengthening the study of class attributes can be more effective for multitoken learning. The proposed algorithm in this paper, after a sparse representation of the class attributes, then computes the mutual information between the features and the token space, then ranks the features according to the magnitude of the mutual information, and selects the desired subset of features. The experiments also verify that the proposed algorithm is feasible.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of 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

The study was supported by the National Social Science Fund (Grant No. 19BTY046) and Educational Commission of Hubei Province (Grant No. B2019197).

References

- [1] S. S. Sawant and M. Prabukumar, "A review on graph-based semi-supervised learning methods for hyperspectral image classification," *The Egyptian Journal of Remote Sensing and Space Science*, vol. 23, no. 2, pp. 243–248, 2020.
- [2] Y. Chong, Y. Ding, Q. Yan, and S. Pan, "Graph-based semi-supervised learning: a review," *Neurocomputing*, vol. 408, pp. 216–230, 2020.
- [3] B. Xu, Z. Zeng, C. Lian, and Z. Ding, "Semi-supervised low-rank semantics grouping for zero-shot learning," *IEEE Transactions on Image Processing*, vol. 30, pp. 2207–2219, 2021.
- [4] C. Kunselman, V. Attari, L. McClenny, U. Braga-Neto, and R. Arroyave, "Semi-supervised learning approaches to class assignment in ambiguous microstructures," *Acta Materialia*, vol. 188, pp. 49–62, 2020.
- [5] D. Hong, N. Yokoya, N. Ge, J. Chanussot, and X. X. Zhu, "Learnable manifold alignment (LeMA): a semi-supervised cross-modality learning framework for land cover and land use classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 147, pp. 193–205, 2019.
- [6] Y. Li, Y. Wang, Q. Liu, C. Bi, X. Jiang, and S. Sun, "Incremental semi-supervised learning on streaming data," *Pattern Recognition*, vol. 88, pp. 383–396, 2019.
- [7] Y. Qin, S. Ding, L. Wang, and Y. Wang, "Research progress on semi-supervised clustering," *Cognitive Computation*, vol. 11, no. 5, pp. 599–612, 2019.
- [8] F. Dornaika, A. Baradaaji, and Y. El Traboulsi, "Semi-supervised classification via simultaneous label and discriminant embedding estimation," *Information Sciences*, vol. 546, pp. 146–165, 2021.
- [9] H. C. Li, G. Yang, W. Yang, Q. Du, and W. J. Emery, "Deep nonsmooth nonnegative matrix factorization network with semi-supervised learning for SAR image change detection," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 160, pp. 167–179, 2020.
- [10] J. Wang, L. Xia, X. Hu, and Y. Xiao, "Abnormal event detection with semi-supervised sparse topic model," *Neural Computing and Applications*, vol. 31, no. 5, pp. 1607–1617, 2019.
- [11] W. Wang, H. Wang, Z. Zhang, and Y. Gao, "Semi-supervised domain adaptation via Fredholm integral based kernel methods," *Pattern Recognition*, vol. 85, pp. 185–197, 2019.
- [12] A. S. Qureshi, A. Khan, N. Shamim, and M. H. Durad, "Intrusion detection using deep sparse auto-encoder and self-taught learning," *Neural Computing and Applications*, vol. 32, no. 8, pp. 3135–3147, 2020.
- [13] S. Wang, Z. Wang, and W. Guo, "Accelerated manifold embedding for multi-view semi-supervised classification," *Information Sciences*, vol. 562, pp. 438–451, 2021.
- [14] C. Chen, Y. Liu, M. Kumar, J. Qin, and Y. Ren, "Energy consumption modelling using deep learning embedded semi-supervised learning," *Computers & Industrial Engineering*, vol. 135, pp. 757–765, 2019.
- [15] J. Tanha, "A multiclass boosting algorithm to labeled and unlabeled data," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 12, pp. 3647–3665, 2019.
- [16] X. Xin, J. Wang, R. Xie, S. Zhou, W. Huang, and N. Zheng, "Semi-supervised person re-identification using multi-view clustering," *Pattern Recognition*, vol. 88, pp. 285–297, 2019.
- [17] M. Zhao, Y. Zhang, Z. Zhang, J. Liu, and W. Kong, "ALG: adaptive low-rank graph regularization for scalable semi-

- supervised and unsupervised learning,” *Neurocomputing*, vol. 370, pp. 16–27, 2019.
- [18] T. Zhao, H. Huang, X. Yao, and X. Fu, “Predicting individual socioeconomic status from mobile phone data: a semi-supervised hypergraph-based factor graph approach,” *International Journal of Data Science and Analytics*, vol. 9, no. 3, pp. 361–372, 2020.
- [19] Z. Lin, M. Luo, Z. Peng, J. Li, and Q. Zheng, “Nonlinear feature selection on attributed networks,” *Neurocomputing*, vol. 410, pp. 161–173, 2020.
- [20] R. Zhu, F. Dornaika, and Y. Ruichek, “Learning a discriminant graph-based embedding with feature selection for image categorization,” *Neural Networks*, vol. 111, pp. 35–46, 2019.
- [21] L. Chum, A. Subramanian, V. N. Balasubramanian, and C. V. Jawahar, “Beyond supervised learning: a computer vision perspective,” *Journal of the Indian Institute of Science*, vol. 99, no. 2, pp. 177–199, 2019.