Research Article

Multiview Sample Classification Algorithm Based on L1-Graph Domain Adaptation Learning

Huibin Lu, Zhengping Hu, and Hongxiao Gao

School of Information Science and Engineering, Yanshan University, Qinhuangdao 066004, China

Correspondence should be addressed to Zhengping Hu; hzp@ysu.edu.cn

Received 14 November 2014; Revised 4 February 2015; Accepted 7 February 2015

Academic Editor: Gerhard-Wilhelm Weber

Copyright © 2015 Huibin Lu 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.

In the case of multiview sample classification with different distribution, training and testing samples are from different domains. In order to improve the classification performance, a multiview sample classification algorithm based on L1-Graph domain adaptation learning is presented. First of all, a framework of nonnegative matrix trifactorization based on domain adaptation learning is formed, in which the unchanged information is regarded as the bridge of knowledge transformation from the source domain to the target domain; the second step is to construct L1-Graph on the basis of sparse representation, so as to search for the nearest neighbor data with self-adaptation and preserve the samples and the geometric structure; lastly, we integrate two complementary objective functions into the unified optimization issue and use the iterative algorithm to cope with it, and then the estimation of the testing sample classification is completed. Comparative experiments are conducted in USPS-Binary digital database, Three-Domain Object Benchmark database, and ALOI database; the experimental results verify the effectiveness of the proposed algorithm, which improves the recognition accuracy and ensures the robustness of algorithm.

1. Introduction

Traditional machine learning algorithms are usually applicable to the data, during which the training and testing samples are from the same characteristic space with the same distribution. With the change of characteristic and distribution, most statistical models need to be reconstructed by the new training sample collection. But, in practice, the training and testing samples are often collected in different periods and different environment; therefore, the distribution may be different. In order to solve this problem, the transfer learning theory is introduced, which aims to transform knowledge from the source domain marked to the target domain unmarked, for example, natural language learning processing [1], sentiment analysis [2], and image classification [3]. Using transfer learning to improve the recognition accuracy and robustness of different distribution is widely concerned by the domestic and foreign scholars.

In recent years, in view of the fact that the marked training samples and unmarked testing samples are from different domains, more and more researchers begin to attach importance to transfer learning [4, 5]. The basic idea of transfer

learning is that although the different data distribution in the source domain and the target domain exists, some associated domains can share some of the same knowledge structure, which can be used as a bridge of knowledge transformation from the source domain to the target domain. The existing methods usually look for these common structures through the optimization of predefined objective functions, including the maximum of empirical likelihood and the preservation of geometric structure. For example, from the perspective of empirical likelihood, Dai et al. proposed the coclustering model based on classification [6], which, through the source domain data, adds constraints to set of words to provide classification structure and part of classification information, and coclustering was regarded as a bridge, through which the classification structure and information are transformed from the source domain to the target domain. The shortcoming of the method is considering the same concept in the document only, so Zhuang et al. studied the relationship between set of words and document classification; some invariant factors were regarded as the bridge of information transformation from the source domain to the target domain by trifactorization [7], but the method only considers

the similar concepts in the text. With the similar idea, Wang et al. carried it out in the network data classification of different domains [8]. In consideration of the shortcoming of the above methods, a joint model of similar concepts and the same concepts was built [9], with synchronized learning of boundary condition distribution. Then, taking the unique concept in each domain document into consideration, Zhuang et al. studied with synchronization the sharing concept and the unique concept in all domains by matrix trifactorization [10], which was more flexible in the process of data-fitting so as to get a better recognition rate. From the geometric point of view, if the two sample data in the domain are similar in terms of the essence of data distribution and the geometric structure, their markers should also be similar [11]. To retain the essential structure, Ling et al. explored the consistency between the supervision of source domain and the essential structure of target domain through the spectrum learning [12]. And Pan et al. put forward the conversion component analysis [13], which aims to find a set of common conversion components in the two domains. The samples are projected into the subspace, which makes the reduction of different degrees of data distribution in different domains. With the same idea, Wang and Mahadevan presented the projection of different domains to the new potential space, synchronously matching the corresponding samples and preserving the geometric structure of each domain [14]. Accounting for the two views, a graph canonical transformation learning (GTL) [15] was put forward, which retained some geometric structures based on the maximum of empirical likelihood.

Based on study of joint canonical transformation learning and considering the fact that the L1-Graph [16, 17] has better adaptive ability and better stability compared with k neighbor graph, we put forward a multiview classification algorithm based on L1-Graph domain adaptation learning, which is based on Long et al. [15]. The idea is to construct a framework of nonnegative matrix trifactorization based on transfer learning, in which the unchanged information is regarded as the bridge of knowledge transformation from the source domain to the target domain, and the next step is to construct L1-Graph on the basis of sparse representation, so as to search for the nearest neighbor data with self-adaptation and preserve the samples and the geometric structure; finally, the paper uses the iterative algorithm to cope with the optimization issue of unified objective functions, and then the classification of the testing samples is completed.

2. The Multidomain and Multiview Sample Classification Based on L1-Graph Domain Adaptation Learning

2.1. Description of L1-Graph Domain Adaptation Learning. L1-Graph domain adaptation learning can be applied to all different domains, but, for purposes of explanation, two domains are presented here: the source domain D_s and the target domain D_t , and the domain index is expressed as $\ell = \{s, t\}$. Each domain D_π , $\pi \in \ell$, has a characteristic matrix X_π . In order to find the common structure, X_π will be broken down into three nonnegative matrices; that is, $X_\pi = U_\pi H V_\pi^T$,

among which $U_\pi = [u_{*1}^\pi, \dots, u_{*k}^\pi]$, $V_\pi = [v_{*1}^\pi, \dots, v_{*c}^\pi]$, and a_{*i} is the i column of the matrix A . In the domain D_π , the characteristic samples of matrix trifactorization are classified through the maximum of empirical likelihood; H represents the relationship between the characteristic set U_π and the sample classification V_π , and it can be regarded as a bridge of knowledge transformation because of the cross domain stability. In addition, the use of sparse representation is carried out to construct graphs and diagrams G_π^u and G_π^v to represent the geometry information of the characteristic space and the sample space in domain D_π , respectively. The basic idea of L1-Graph domain adaptation learning algorithm is presented in [15]. In general, similar characteristics represent the same meaning; likewise, similar samples have the same marker. These two graphs are regarded as joint regularization function so that the trifactorization model of learning successfully forecasts the sample marker in the case of retaining the essence of geometric structure, and the geometric information will be effectively integrated into the clustering process, thus ensuring that the common structure information H can effectively promote transfer learning.

2.2. Model of Matrix Trifactorization. Assuming that the source domain is D_s and the target domain is D_t , the domain index is expressed as $\ell = \{s, t\}$. D_s and D_t share the same characteristic space and marker space, including m characteristics and c categories in each characteristic space. $X_\pi = [x_{*1}^\pi, \dots, x_{*n_\pi}^\pi] \in R^{m \times n_\pi}$, $\pi \in \ell$, represents the characteristic sample matrix in domain D_π , in which x_{*i}^π indicates the i column of the domain D_π . $Y_\pi \in R^{n_\pi \times c}$ represents the sample marker of the source domain, as x_{*i}^π belongs to category j , $y_{ij}^\pi = 1$; otherwise, $y_{ij}^\pi = 0$.

A total of structural information exists in every associated domain D_π ; for the same structure information, the nonnegative matrix trifactorization is conducted as for the characteristic sample matrix X_π , $\pi \in \ell$, and the optimization is planned as follows:

$$\min_{U_\pi, H_\pi, V_\pi \geq 0} L_\pi = \|X_\pi - U_\pi H_\pi V_\pi^T\|^2 \quad (1)$$

in which $\|A\|$ is the Frobenius norm of matrix A ; $U_\pi = [u_{*1}^\pi, \dots, u_{*k}^\pi] \in R^{m \times k}$, and each u_{*i}^π represents semantic concepts, namely, characteristic clustering; $V_\pi = [v_{*1}^\pi, \dots, v_{*c}^\pi] \in R^{n_\pi \times c}$, and each v_{*i}^π represents a sample type. U_π and V_π are the clustering results as to characteristic and sample separately. $H_\pi \in R^{k \times c}$ represents the relationship between the characteristic set U_π and the sample classification V_π , which can keep better stability when compared with U_π and V_π in different domains. Therefore, assuming that $H_\pi = H$ adapts to each domain, then the collective matrix trifactorization is planned as follows:

$$\min_{U_\pi, H, V_\pi \geq 0} L = \sum_{\pi \in \ell} \|X_\pi - U_\pi H V_\pi^T\|^2. \quad (2)$$

The common structural information H , as the stable bridge of knowledge transformation, can be executed as the monitor information in the source domain, namely, executing $V_s \equiv Y_s$. Through the bridge, the knowledge marker

Y_s in source domain can be converted to the sample V_t in target domain, and the process is in correspondence to the maximum of multidomain empirical likelihood.

2.3. Sparse Representation of L1-Graph Structure. From the geometric point of view, it can be thought that the data points are sampled in the distribution, which is formed by a low dimensional manifold, and then embedded into a high dimensional data space. Therefore, in order to avoid the change of the essence of the data distribution and to hope that the geometric structure is preserved in the conversion process, the assumption is presented here; if the inherent geometric structure of data distribution of two samples x_{*i}^π and x_{*j}^π in domain D_π is close to each other, the markers of v_{*i}^π and v_{*j}^π should also be close to each other, so a model can be formed on the geometric structure of sample space. The traditional graph construction methods are often heavily dependent on the choice of parameters, and it is difficult to effectively reflect the complexity of the data distribution. According to the theory of sparse representation, any sample can be linearly reconstructed by other samples (allowing for some reconstruction error), and the sparse reconstruction coefficient of the sample can be obtained by coping with a L1 norm optimization issue. The reconstructed coefficients, as weights between two samples, can adjust the relationship between samples with self-adaption, so that the sparse graph, which represents the local relationship between samples, contains more useful structural information. From the perspective of duality between characteristic and sample, characteristic is also sampled from the distribution, which is supported by a low dimensional manifold, and then embedded into a high dimensional space. If the geometric structure of data distribution of two characteristics x_{i*}^π and x_{j*}^π in domain D_π is close to each other, their characteristic sets u_{i*}^π and u_{j*}^π should also be close to each other. Therefore, as for the characteristic space, a sparse graph based on the principle of sparse representation can be constructed to retain the characteristic geometric structure in each domain, just the same as retaining the geometric structure of sample space. The sparse graph construction of sample here is expressed as G_π^v , the corresponding sparse graph construction of characteristic is expressed as G_π^u , and the sparse graph construction of sample space is presented in the following steps.

(1) *Input.* Sample $X_\pi = [x_{*1}^\pi, \dots, x_{*n_\pi}^\pi] \in R^{m \times n_\pi}$, $\pi \in \ell$; each sample $x_{*i}^\pi \in R^m$ is normalized, so that $\|x_{*i}^\pi\|_2 = 1$.

(2) *To Solve the Reconstruction Coefficient.* The sparse reconstruction coefficient for each sample x_{*i}^π in each domain can be obtained by coping with the following optimization issue of minimization of L1 paradigm:

$$\begin{aligned} \min_{\alpha_i^\pi} \quad & \|\alpha_{*i}^\pi\|_1, \\ \text{s.t.} \quad & x_{*i}^\pi = B_i^\pi \alpha_{*i}^\pi \end{aligned} \quad (3)$$

in which $B_i^\pi = [x_{*1}^\pi, \dots, x_{*i-1}^\pi, x_{*i+1}^\pi, \dots, x_{*n_\pi}^\pi, I] \in R^{m \times (n_\pi - 1 + m)}$ is an overcomplete dictionary and $\alpha_{*i}^\pi \in R^{n_\pi - 1 + m}$

is a column vector which constitutes the reconstruction coefficient, indicating a relationship between sample x_{*i}^π and other samples.

(3) *To Set the Edge Weights of L1-Graph.* L1-Graph is expressed as $G_\pi^v = \{X_\pi, W_\pi^v\}$, in which X_π represents the set of all nodes in domain π and W_π^v represents the weight matrix of L1-Graph in the domain, that is, the similarity matrix. When $i > j$, $(W_\pi^v)_{ij} = |\alpha_{ij}^\pi|$; when $i < j$, $(W_\pi^v)_{ij} = |\alpha_{i(j-1)}^\pi|$; when $i = j$, $(W_\pi^v)_{ij} = 0$. The number of nearest neighbors of each sample is identified by optimizing the L1 paradigm issue instead of parameters, which are manually set up.

(4) *The Similarity of Symmetrization Matrix.* Consider $W_\pi^v = (W_\pi^v + (W_\pi^v)^T)/2$.

Similarly, the weights matrix of characteristic space W_π^u can also be obtained.

To preserve the L1-Graph regularization function of the geometric structure minimization sample in domain D_π ,

$$\begin{aligned} R_\pi^v &= \frac{1}{2} \sum_{ij} \|v_{i*}^\pi - v_{j*}^\pi\|^2 (W_\pi^v)_{ij} \\ &= \sum_i v_{i*}^\pi (v_{i*}^\pi)^T (W_\pi^v)_{ii} - \sum_{ij} v_{i*}^\pi (v_{j*}^\pi)^T (W_\pi^v)_{ij} \\ &= \text{tr}(V_\pi^T (D_\pi^v - W_\pi^v) V_\pi). \end{aligned} \quad (4)$$

To further preserve the L1-Graph regularization function of the geometric structure synchronization minimization characteristic in domain D_π ,

$$R_\pi^u = \frac{1}{2} \sum_{ij} \|u_{i*}^\pi - u_{j*}^\pi\|^2 (W_\pi^u)_{ij} = \text{tr}(U_\pi^T (D_\pi^u - W_\pi^u) U_\pi). \quad (5)$$

2.4. Joint Optimization. Evidently, (4) and (5) define that the geometric structure of sample and characteristic can be retained through the L1-Graph regularization function. Therefore, the two equations, as a combined L1-Graph regularization function, can be integrated into (2), which defines the optimization issue of L1-Graph coregularized collective matrix trifactorization (L1-GCMF) as follows:

$$\begin{aligned} \min_{U_\pi, H_\pi, V_\pi \geq 0} \quad & O = L + \sum_{\pi \in \ell} (\lambda R_\pi^u + \gamma R_\pi^v) \\ \text{s.t.} \quad & U_\pi^T 1_m = 1_k, \\ & V_\pi^T 1_{n_\pi} = 1_c, \\ & \forall \pi \in \ell \end{aligned} \quad (6)$$

in which, λ and γ are regularization parameters; the optimization issue can be better defined by constraining the norm ℓ_1 of each column of U_π and V_π . According to the optimization

results, the marking of a random sample x_{*i}^π in target domain D_π can be easily inferred by the following formula:

$$f(x_{*i}^\pi) = \arg \max_j (V_\pi)_{ij}. \quad (7)$$

In the process of optimization of formula (6), the common structure information H , which is found through the synchronous maximum of empirical likelihood and preserving the geometric structure, becomes more smooth in the process of conversion learning. LI-GCMF can be extended to handle multidomain issues and to research the commonality in the collection structure. The approaches to the optimization issue in (6) can be derived based on the constrained optimization theory. Specifically, the updating rules are deduced through fixing the rest of variables and optimizing one variable, and the process is repeated till the convergence. In consideration of the nonnegative and the constraint of norm ℓ_1 , a Lagrange function is constructed as follows:

$$\begin{aligned} \Phi = O + \sum_{\pi \in \ell} \text{tr} \left(\Lambda_\pi (U_\pi^T \mathbf{1}_m - \mathbf{1}_k) (U_\pi^T \mathbf{1}_m - \mathbf{1}_k)^T \right) \\ + \sum_{\pi \in \ell} \text{tr} \left(\Gamma_\pi (V_\pi^T \mathbf{1}_{n_\pi} - \mathbf{1}_c) (V_\pi^T \mathbf{1}_{n_\pi} - \mathbf{1}_c)^T \right) \end{aligned} \quad (8)$$

in which $\Lambda_\pi \in R^{k \times k}$ and $\Gamma_\pi \in R^{c \times c}$ are constrained Lagrange multipliers and $\mathbf{1}_m$ is 1 vector. By using the complementary conditions of Karush-Kuhn-Tucker (KKT), the constraint conditions of U_π are deduced as follows:

$$\begin{aligned} \frac{1}{2} \nabla_{U_\pi} \Phi \circ U_\pi = (V_\pi H^T U_\pi^T U_\pi H - X_\pi V_\pi H^T) \circ U_\pi \\ + (\lambda D_\pi^u U_\pi - \lambda W_\pi^u U_\pi + \mathbf{1}_m \mathbf{1}_m^T D_\pi^v \Lambda_\pi \\ - \mathbf{1}_m \mathbf{1}_k^T \Lambda_\pi) \circ U_\pi = 0. \end{aligned} \quad (9)$$

According to the KKT conditions, the updating plan is obtained as follows:

$$U_\pi \leftarrow U_\pi \circ \sqrt{\frac{[X_\pi V_\pi H^T + \lambda W_\pi^u U_\pi + \mathbf{1}_m \mathbf{1}_k^T \Lambda_\pi]}{[U_\pi H V_\pi^T V_\pi H^T + \lambda D_\pi^u U_\pi + \mathbf{1}_m \mathbf{1}_m^T D_\pi^v \Lambda_\pi]}}. \quad (10)$$

The computation of Λ_π can be avoided through the use of an iterative standardization technology. In the process of each iteration, each column of U_π can be normalized, so that $U_\pi^T \mathbf{1}_m = \mathbf{1}_k$. Then, two equal items, depending on Λ_π , can be obtained; namely, $\mathbf{1}_m \mathbf{1}_m^T D_\pi^v \Lambda_\pi = \mathbf{1}_m \mathbf{1}_k^T \Lambda_\pi$, which can be omitted in the updating plan with no influence on the convergence; therefore, the updating rules are presented as follows:

$$U_\pi \leftarrow U_\pi \circ \sqrt{\frac{[X_\pi V_\pi H^T + \lambda W_\pi^u U_\pi]}{[U_\pi H V_\pi^T V_\pi H^T + \lambda D_\pi^u U_\pi]}}. \quad (11)$$

Similarly, the updating rules can also be obtained as follows:

$$\begin{aligned} V_\pi \leftarrow V_\pi \circ \sqrt{\frac{[X_\pi^T U_\pi H + \gamma W_\pi^v V_\pi]}{[V_\pi H^T U_\pi^T U_\pi H + \gamma D_\pi^v V_\pi]}}, \\ H \leftarrow H \circ \sqrt{\frac{[\sum_{\pi \in \ell} U_\pi^T X_\pi V_\pi]}{[\sum_{\pi \in \ell} U_\pi^T U_\pi H V_\pi^T V_\pi]}} \end{aligned} \quad (12)$$

in which $\circ$ represents the element product, $[\cdot]/[\cdot]$ the element division, and $\sqrt{\cdot}$ the square root of elements.

2.5. Description of a Multiview Sample Classification Algorithm Based on L1-Graph Domain Adaptation Learning. See Algorithm 1.

3. Experiment Results and Analysis

3.1. Comparative Experiment Results Based on USPS-Binary Digital Database. The experimental samples are selected from the two digital databases: USPS and Binary. The USPS database contains 10 groups of handwritten digits: 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9, as shown in Figure 1(a), and each group contains 1100 gray scale samples. Some gray scale samples of one number are shown in Figure 1(b). The Binary database contains a total of 36 kinds of data, including the numbers 0–9 and the letters A–Z, and each kind of data includes 39 samples; only 10 groups of handwritten digits 0–9 are selected in the experiment. Figure 2(a) presents 10 categories of digits, and Figure 2(b) presents a part of two binary samples of one category.

All images are cut into the same dimension, SIFT characteristic extraction is made for each image, and then all the pictures are transferred into the 300-dimension characteristic histogram. The 10 categories from two databases were considered as the known ones; USPS database was treated as the source domain and Binary database the target domain; the data with the number $m = \{50, 100, 200\}$ were randomly selected from each category of data in USPS database to constitute the training samples, and 39 samples from each category of data in Binary database were treated as the observation samples. In order to obtain the optimal parameter values, we conducted the experiment 10 times to calculate the average recognition rate. The GCMF method obtains the optimal value when $\lambda = 0.1$, $\gamma = 1$, and the LI-GCMF method obtains the optimal value when $\lambda = 0.3$ and $\gamma = 1$. In all experiments, the number of k neighbor unified values is 10, and the number of iterations is 100. In view of the different values of m , 10 training samples were randomly selected from each group of data in USPS database to carry out the experiment, and Figure 3 showed the average recognition rate in correspondence with different training samples and different methods.

The experimental results from Figure 3 show that, for different number of training samples, the recognition rate of LI-GCMF algorithm is higher than that of GCMF [15] and ITML [18] algorithms, and, with the changing number

Input: data set $\{X_\pi\}_{\pi \in \ell}$, Y_s , parameters λ , γ , max $Iter$.

Output: the target domain classification results V_t

Start:

to construct the graph G_π^v , G_π^u through the principle of sparse representation; to normalize the data sets through $X_\pi \leftarrow X_\pi / \|X_\pi\|$, $\forall \pi \in \ell$.
to initialize $\{U_\pi\}_{\pi \in \ell}$, to generate H randomly, to obtain V_s through Y_s , to obtain V_t through the training of logistic regression in $\{X_s, Y_s\}$.
for $iter \leftarrow 1$ to max $Iter$ do
 foreach $\pi \in \ell$ do
 to update U_π , V_π and H through (11)~(12)
 to fix $V_s \equiv Y_s$
 to make the norm ℓ_1 for each column of U_π and V_π .
 to calculate the objective function O^{iter} through (6)

End

ALGORITHM 1: L1-GCMF multiview sample classification algorithm.

(a) Ten categories of USPS database

(b) Gray scale samples within one category

FIGURE 1: Samples of USPS database.

(a) Ten categories of Binary database

(b) Binary samples within one category

FIGURE 2: Samples of Binary database.

of training samples, L1-GCMF algorithm is more stable compared with the other two algorithms. Due to the use of L1-Graph, the method can automatically search for the nearest neighbor parameters, which is more conducive to the connection to the same type samples, so as to improve the classification accuracy of multiview samples and make it have stronger stability.

3.2. Comparative Experiment Results Based on Three-Domain Object Benchmark Database. The Three-Domain Object Benchmark database [19] contains three different domains of amazon, dslr, and webcam, and each domain contains a total of 4652 images from 31 different object categories. There are about 90 images for each category in the domain of amazon, some of which from one category are displayed in Figure 4(a); there are 30 images for each category in domains of dslr and webcam, some of which are separately shown in Figures 4(b) and 4(c).

With amazon as the source domain and webcam as the target domain, 20 categories from the two databases were selected for experiments, all of the selected category of images will be converted to SIFT characteristics. Then 10 samples were randomly selected as training samples from the source domain, and 5 samples as testing samples from the target domain. The GCMF method obtains the optimal value when $\lambda = 0.2$, $\gamma = 1$, and the L1-GCMF method obtains the

FIGURE 3: Average correctness recognition rate in USPS-VS-Binary.

optimal value when $\lambda = 0.1$ and $\gamma = 2$. A random selection of 10 training samples and testing samples was used for experiments, and the average recognition rate and standard deviation are shown in Table 1.

The experimental results from Table 1 show that the average recognition rate of L1-GCMF algorithm is relatively

FIGURE 4: (a) Samples of amazon domain; (b) samples of dslr domain; (c) samples of webcam domain.

FIGURE 5: (a) Samples of different angles; (b) samples of different light.

TABLE 1: Average correctness recognition rate in Three-Domain Object Benchmark database.

Algorithm	ITML [19]	GCMF [15]	LI-GCMF
Recognition rate (%)	10.2	24.9	50.55
Standard deviation	4.37	5.97	6.76

high, compared with information-theoretic metric learning (ITML) and GCMF algorithms. Different from the fact that the selection of the neighboring number of ITML and GCMF classification algorithms is dependent on the parameters, the algorithm presented in this paper can get higher recognition rate by using sparse principle of composition, which does not need to manually set up the parameters.

3.3. Comparative Experiment Results Based on ALOI Database. This set of experimental data is from the ALOI database. The database contains 1000 objects, and each

object contains images with different light and different angles. 50 object images were selected for experiments, and all of the selected object images were transferred into the 800-dimension SIFT characteristics. Images under different angles were treated as the source domain, as shown in Figure 5(a), and images under different light as the target domain, as shown in Figure 5(b).

30 data were randomly selected from each group in the source domain to be treated as the training samples, and 10 images were randomly selected from each group in the target domain as the observation samples; the data of each group were used to carry out 10 random experiments, so data in the results are the mean of 500 random tests of the 50 categories; the parameter values are the same as in experiment of Section 3.1, and the average recognition rate and the standard deviation are shown in Table 2.

Experimental results from Table 2 show that the average recognition rate of this method is higher than that of ITML and GCMF algorithms, and the robustness of the method is

TABLE 2: Average correctness recognition rate in ALOI database.

Algorithm	ITML [18]	GCMF [15]	LI-GCMF
Recognition rate (%)	14.95	37.3	55.88
Standard deviation	8.66	5.66	9.41

presented because of the smaller standard deviation than the other two methods. Compared with the other algorithms, the algorithm has strong adaptability, avoids the artificial setting of the parameters, and achieves better multiview sample classification, which explains the rationality of the algorithm using sparse composition.

4. Conclusion

Based on the research of marked and unmarked data from different distribution, the paper proposed the multiview sample classification algorithm based on LI-Graph domain adaptation learning. The method first builds a nonnegative matrix trifactorization framework, in which the common structural information was regarded as the bridge of knowledge transformation from the source domain to the target domain, and then constructs LI-Graph by using the principle of sparse representation, so that the neighboring numbers of each sample are determined by the optimization of LI paradigm issue, so as to search for the nearest neighbor data with self-adaptation. The next is to jointly optimize the objective function by using the iterative algorithm, and then the estimation of classification of the testing samples is completed. Comparative experimental results based on USPS-Binary digital database, Three-Domain Object Benchmark database, and ALOI database show the rationality of the method. However, compared with the traditional k nearest neighbor composition, the composition method based on sparse representation has relatively high complexity and relatively large computational cost, due to the need to calculate the reconstruction coefficients of all samples.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- [1] H. Daumé III, "Frustratingly easy domain adaptation," in *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL '07)*, pp. 256–263, June 2007.
- [2] T. Li, V. Sindhwani, C. Ding, and Y. Zhang, "Knowledge transformation for cross-domain sentiment classification," in *Proceedings of the 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '09)*, pp. 716–717, July 2009.
- [3] Y. Luo, D. Tao, B. Geng, C. Xu, and S. J. Maybank, "Manifold regularized multitask learning for semi-supervised multilabel image classification," *IEEE Transactions on Image Processing*, vol. 22, no. 2, pp. 523–536, 2013.
- [4] Y. Luo, T. Liu, D. Tao, and C. Xu, "Decomposition-based transfer distance metric learning for image classification," *IEEE Transactions on Image Processing*, vol. 23, no. 9, pp. 3789–3801, 2014.
- [5] S. Si, D. Tao, and B. Geng, "Bregman divergence-based regularization for transfer subspace learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 7, pp. 929–942, 2010.
- [6] W. Dai, G.-R. Xue, Q. Yang, and Y. Yu, "Co-clustering based classification for out-of-domain documents," in *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 210–219, 2007.
- [7] F. Zhuang, P. Luo, H. Xiong, Q. He, Y. Xiong, and Z. Shi, "Exploiting associations between word clusters and document classes for cross-domain text categorization," *Statistical Analysis and Data Mining*, vol. 4, no. 1, pp. 100–114, 2011.
- [8] H. Wang, H. Huang, F. Nie, and C. Ding, "Cross-language web page classification via dual knowledge transfer using nonnegative matrix tri-factorization," in *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '11)*, pp. 933–942, July 2011.
- [9] M. Long, J. Wang, G. Ding, W. Cheng, X. Zhang, and W. Wang, "Dual transfer learning," in *Proceedings of the 12th SIAM International Conference on Data Mining*, pp. 540–551, 2012.
- [10] F. Zhuang, P. Luo, C. Du, Q. He, and Z. Shi, "Triplex transfer learning: exploiting both shared and distinct concepts for text classification," in *Proceedings of the 6th ACM International Conference on Web Search and Data Mining (WSDM '13)*, pp. 425–434, it, February 2013.
- [11] D. Cai, X. He, X. Wang, H. Bao, and J. Han, "Locality preserving nonnegative matrix factorization," in *Proceedings of the 21st International Joint Conference on Artificial Intelligence*, pp. 1010–1015, July 2009.
- [12] X. Ling, W. Dai, G.-R. Xue, Q. Yang, and Y. Yu, "Spectral domain-transfer learning," in *Proceedings of the 14th ACMKDD International Conference on Knowledge Discovery and Data Mining*, pp. 488–496, 2008.
- [13] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp. 199–210, 2011.
- [14] C. Wang and S. Mahadevan, "Heterogeneous domain adaptation using manifold alignment," in *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*, pp. 1541–1546, July 2011.
- [15] M. Long, J. Wang, G. Ding, D. Shen, and Q. Yang, "Transfer learning with graph regularization," in *Proceedings of the 26th AAAI Conference on Artificial Intelligence and the 24th Innovative Applications of Artificial Intelligence Conference*, pp. 1033–1039, July 2012.
- [16] B. Cheng, J. Yang, B. Yan, and Y. Fu, "Learning with l^1 -graph for image analysis," *IEEE Transactions on Image Processing*, vol. 19, no. 4, pp. 858–866, 2010.
- [17] Y. Fang, R. Wang, B. Dai, and X. Wu, "Graph-based learning via auto-grouped sparse regularization and kernelized extension," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 1, pp. 142–154, 2015.
- [18] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon, "Information-theoretic metric learning," in *Proceedings of the 24th International Conference on Machine Learning (ICML '07)*, pp. 209–216, June 2007.

- [19] S. Kate, K. Brian, F. Mario, and D. Trevor, "Adapting visual category models to new domains," in *Proceedings of the 11th European Conference on Computer Vision, Part IV, Heraklion, Greece, September 2010*, vol. 6314, pp. 213–226, Springer, 2010.