International Journal of Wavelets, Multiresolution and Information Processing

1-D Embedding Multi-Category Classification Methods

Luoqing Li, and Chuanwu Yang
Hubei Key Laboratory of Applied Mathematics
Faculty of Mathematics and Statistics, Hubei University
Wuhan 430062, China
lilq@hubu.edu.cn; cw_yang0108@163.com

Qiwei Xie
Institute of Automation, Chinese Academy of Sciences
Beijing, 100190, China
Department of Electronics and Information, Toyota Technological Institute
Nagoya, Aichi, 468-8511, Japan
qiwei.xie@ia.ac.cn

Received 8 June 2015
Revised 12 September 2015
Accepted 12 October 2015

Abstract In this paper, we propose a novel semi-supervised multi-category classification method based on one-dimensional (1-D) multi-embedding. Based on the multiple 1-D embedding based interpolation technique, we embed the high-dimensional data into several different 1-D manifolds and perform binary classification firstly. Then we construct the multi-category classifiers by means of one-versus-rest and one-versus-one strategies separately. A weight strategy is employed in our algorithm for improving the classification performance. The proposed method shows promising results in the classification of handwritten digits and facial images.

Keywords: multi-category classification; 1-D embedding; semi-supervised learning.

AMS Subject Classification: 22E46, 53C35, 57S20

1. Introduction

Multi-category classification problems are widespread in our life. Various methods have been proposed to deal with the multi-category classification problems. Beyesian methods[4, 13, 15] based on the statistical theory use prior information to construct statistical models to estimate the unknown samples. These methods are valid when the number of training samples is sufficiently large and the samples obey some distributions. To deal with more general cases, k-nearest neighbor method,[19] neural networks,[12] support vector machine (SVM) methods[8] are proposed. Among these methods, SVM is widely used and shows good performance in handling multi-class classification tasks. Based on several binary classifiers, SVM uses one-versus-rest or one-versus-one strategy to construct the multi-category classifier. It disassembles the multi-category classification problem into several binary classification problems.[8]

When the dimensionality of data is huge and the number of available training samples is limited, the multi-category classification becomes a challenging task. With limited training samples, the performance of classifiers usually deteriorates as the dimensionality increases, which is the so called Hughes phenomenon.[9] To deal with the high dimensional data, some dimension reduction methods have been proposed to embed the high dimensional data into an underlying lower dimensional manifold.[14,16] To handle the small-sample-sized problems, some semi-supervised learning methods have been established to exploit both a large amount of unlabeled samples and limited labeled samples.[5,20] By utilizing the geometrical structure inferred from both labeled and unlabeled data, recently, a new one-dimensional (1-D) manifold embedding method has been proposed in Ref. 17 to solve semi-supervised classification problems. By embedding the high dimensional data into a 1-D space, the decision function on original high dimensional data can be substituted by a univariate function on a 1-D space. In this paper, we propose a 1-D multi-embedding multi-category classification method to deal with the high dimensional data. The proposed method constructs multi-category classifiers based on binary classifiers using one-versus-rest or one-versus-one strategy. It first maps all the samples into a line using multiple 1-D embedding, then generates binary classifiers by interpolation functions,[17,18] and finally constructs the multi-category classification decision function by one-versus-rest and one-versus-one methods. Based on 1-D multi-embedding multi-category classification method, the classification of data with high dimensional complex structure is simplified to an easier 1-D classification problem, which can be effectively solved by a linear interpolation technique. In addition, a novel one-versus-rest technique based on majority voting of autodidactic interpolation function values is constructed to handle the 1-D multi-category classification problem.

The rest of this paper is organized as follows. Section 2 gives a detailed description of multi-category classification method based on the 1-D multi-embedding. The experimental results and analysis are provided in Section 3. Section 4 gives a summary of our work.

2. Multi-category classification based on 1-D multi-embedding

Given a sample set $S = \{(x_1, y_1), \cdots, (x_n, y_n)\}$ with $x_i \in R^m$ and $y_i \in Y = \{1, \cdots, k\}$, the task of k-class classification is to construct a decision function labeling each test sample. In this section we will give a brief introduction on the 1-D embedding method[17] and propose an algorithm for multi-category classification.

2.1. *1-D multi-embedding*

Let $X = \{x_i\}_{i=1}^n \subset R^m$ be a given data set, and $d(\cdot,\cdot)$ be a distance function measuring the distance between points in the data set. The 1-D multi-embedding method for a given data set is proposed by J. Wang in Ref. 17. Here we give an outline for this method. Please refer to Ref. 17 for more details.

Assume that the initial order of samples in X is arranged as $\mathbf{x} = [x_1, x_2, \cdots, x_n]$. First one finds a permutation operator P (derived from the index permutation π), such that $\hat{\mathbf{x}} = PX = [\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_n]$ has a nearly shortest path, by solving the following minimization problem:

$$\hat{\mathbf{x}} = \arg\min_P \sum_{j=1}^{n-1} d(\tilde{x}_j, \tilde{x}_{j+1}), \qquad \tilde{x}_j = x_{\pi(j)}. \tag{2.1}$$

It leads to the ordered stack $\hat{\mathbf{x}} = [\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_n]$. Second, a mapping $h : X \to R$,

$$h(\hat{x}_j) = t_j,\ t_1 = 0,\ t_{j+1} - t_j = d(\hat{x}_{j+1}, \hat{x}_j), \quad j = 1, 2, \cdots, n-1 \tag{2.2}$$

yields a 1-D embedding of X. Considering that the interval $[t_1, t_n]$ may be different in different permutations, it adjusts the interval $[t_1, t_n]$ to $[0, 1]$ at last. That is, the map $h : X \to [0, 1]$ gives a 1-D embedding of X.

When the starting point in (2.1) is different, the resulting 1-D embedding or the shortest path is different. Assume there are s different paths and mappings $h_i, 1 \le i \le s$, the vector $\mathbf{h} = [h_1, h_2, \cdots, h_s]$ gives a 1-D multi-embedding of X.

2.2. *The binary classifier based on 1-D multi-embedding*

Let $h : X \to [0, 1]$ be a single 1-D embedding associated with the permutation P such that $h(PX) = T = \{t_1, t_2, \cdots, t_n\}$ with $t_1 = 0, t_n = 1$. Assume that f is a binary classifier on X, whose values on the subset $X_0 \subset X$ are known. For $x \in X_0$, $f(x) = +1$ or $f(x) = -1$, where $+1$ and -1 denote the labels of two different classes respectively. The 1-D embedding h converts the classifier f on the original high dimensional data X to a classifier $\tilde{f}$ on 1-D interval [0.1] such that

$$\tilde{f} = f \circ (P^{-1} \circ h^{-1}).$$

In this framework, $\tilde{f}$ will be learned instead of the target function f. Note that the values of $\tilde{f}$ are known on the subset $T_0 = h(PX_0)$. Hence, a simple 1-D learning algorithm can solve the task and obtain an approximation of $\tilde{f}$, such as polynomial interpolation methods on interval $[0, 1]$. Due to the lost of information in the embedding from high dimensional space to 1-D space, a binary classifier built on a single 1-D embedding is quite weak. To overcome the problem, it employs 1-D multi-embedding.[17]

Assume that a 1-D multi-embedding $\mathbf{h} = [h_1, h_2, \cdots, h_s]$ of X has been given. For each embedding h_i, a univariate function $\tilde{f}_i = f \circ (P_i^{-1} \circ h_i^{-1}), 1 \le i \le s$ can be constructed. Assume that $T^i = \{t_1^i, t_2^i, \cdots, t_n^i\}$ with $t_1^i = 0, t_n^i = 1$ is the image of X under the embedding $h_i : T^i = h_i(P_i X)$, and the labeled set X_0 is mapped onto $T_0^i = h_i(P_i X_0)$. Then each $\tilde{f}_i$ yields an approximation of f in the following way:

$$f_i^* = \tilde{f}_i \circ (h_i \circ P_i). \tag{2.3}$$

The vector $\mathbf{f}^* = [f_1^*, f_2^*, \cdots, f_s^*]$ gives s different approximations of the binary classifier f.

To improve the the quality of the approximations, an *autodidactic interpolation scheme* is constructed to enlarge the labeled data set X_0 in Ref. 17. Algorithm 1 shows the detail of the scheme on 1-D multi-embedding and gives the decision function vector $\mathbf{f}^* = [f_1^*, f_2^*, \cdots, f_s^*]$.

Algorithm 1 Autodidactic interpolation scheme

Input: Data set $X = \{x_i\}_{i=1}^n$ with labeled subset $X_0 \subset X$.
Output: The decision function vector $\mathbf{f}^* = [f_1^*, f_2^*, \cdots, f_s^*]$.

Step 1 Construct the multi-embedding $\mathbf{h} = [h_1, h_2, \cdots, h_s]$ of X.
Step 2 **While** 1 **Do**

- **Step 2.1** $T^i \leftarrow h_i(P_i X);\quad T_0^i \leftarrow h_i(P_i X_0), 1 \leq i \leq s$.
- **Step 2.2** Construct $\mathbf{f}^* = [f_1^*, f_2^*, \cdots, f_s^*]$ using interpolation operator.
- **Step 2.3** Update feasible confident set X_{conf}, and set $X_1 = X_0 \bigcup X_{\text{conf}}$. If $X_0 \neq X_1$, set $X_0 = X_1$; else if $X_0 = X_1$, break.

End While

Step 3 Construct $\mathbf{f}^* = [f_1^*, f_2^*, \cdots, f_s^*]$ using **Step 2.1** and **Step 2.2**.

Remark 2.1. *In algorithm 1, the feasible confident set is constructed as follows:* $X_{\text{conf}} \subset X_{\text{f}}$ *satisfies* $|X_{\text{conf}}| = p|X_{\text{f}}|, 0 < p < 1$, *where* $X_{\text{f}} = \{x \in X \backslash X_0; \sigma(x) \leq \epsilon\}$ *with* $\sigma(x) = \frac{1}{s}\sum_{i=1}^s \sqrt{(\text{sgn}(f_i^*(x)) - \frac{1}{s}\sum_{j=1}^s \text{sgn}(f_j^*(x)))^2}$ *and* $\epsilon > 0$ *is a constant.* The signum function, denoted $\text{sgn}(x)$, is defined as $\text{sgn}(x) = 1$ if $x > 0$, $= -1$ otherwise.

2.3. *Multi-category classification based on 1-D multi-embedding*

We employ two popular methods, one-versus-rest and one-versus-one[8] to solve the multi-class classification problem based on a series of binary classifiers. Let k be the number of classes.

2.3.1. *One-versus-rest method based on 1-D multi-embedding*

The main idea of one-versus-rest method is to construct k binary classifiers, where the j-th binary classifier is learned based on the assumption that all of labeled samples in the j-th class have positive labels while all other labeled samples have negative ones.

In the one-versus-rest multi-category classification case, k binary decision function vectors $\mathbf{f}^1, \mathbf{f}^2, \cdots, \mathbf{f}^k$ will be constructed respectively as described in subsection

2.2, where $\mathbf{f}^j = [f_1^j, f_2^j, \cdots, f_s^j]$ gives s approximations of the binary classifier between j-th class and the others, $1 \leq j \leq k$. For each sample x, let

$$p^j(x) = \frac{1}{s}\sum_{i=1}^{s} \mathrm{sgn}(f_i^j(x)) \tag{2.4}$$

be the binary decision function.

Based on these k decision function vectors, $p^1(x), p^2(x), \cdots, p^k(x)$ are constructed. For the sample $x \in X$, we denote

$$M = \{m : p^m(x) = \max_{1 \leq j \leq k} p^j(x), 1 \leq m \leq k\} \tag{2.5}$$

as a alternative label set. If $|M| = 1$, we label x by the only label in M. In particular, if $\exists 1 \leq m_1, m_2 \leq k$, satisfy $p^{m_1}(x) = p^{m_2}(x) = \max\limits_{1 \leq j \leq k} p^j(x)$, we compute the average

$$w^j(x) = \frac{1}{s}\sum_{i=1}^{s} f_i^j(x) \tag{2.6}$$

as the weight that x is labeled as j. If $w^{m_1}(x) > w^{m_2}(x)$, we assign x to the class m_1, else if $w^{m_1}(x) < w^{m_2}(x)$, we assign x to the class m_2. In case that $w^{m_1}(x) = w^{m_2}(x)$, we randomly draw one of them as the class label of x. Algorithm 2 shows the one-versus-rest method based on 1-D multi-embedding.

Algorithm 2 One-versus-rest method based on 1-D embedding

Input: k binary decision vectors $\mathbf{f}^1, \mathbf{f}^2, \cdots, \mathbf{f}^k$, where $\mathbf{f}^j = [f_1^j, f_2^j, \cdots, f_s^j]$.
Output: The multi-classifier L.

Step 1 Calculate $[p^1(x), p^2(x), \cdots, p^k(x)]$:

$$p^j(x) \leftarrow \frac{1}{s}\sum_{i=1}^{s} \mathrm{sgn}(f_i^j(x)),\ 1 \leq j \leq k.$$

Step 2 Update the alternative label set of the sample x by (2.5).
Step 3 If $|M| = 1$, then set

$$L(x) \leftarrow \arg\max_{1 \leq j \leq k} p^j(x);$$

else calculate $w^j(x) \leftarrow \frac{1}{s}\sum_{i=1}^{s} f_i^j(x), j \in M$, and set

$$L(x) \leftarrow \arg\max_{j \in M} w^j(x).$$

2.3.2. *One-versus-one method based on 1-D multi-embedding*

Algorithm 3 One-versus-one method based on 1-D embedding

Input: $\frac{k(k-1)}{2}$ binary decision function vectors $\mathbf{f}^{i,j},\ 1 \le i < j \le k$.
Output: The multi-classifier L.

Step 1 Calculate $\frac{k(k-1)}{2}$ binary decision function:

$$g^{i,j}(x) \leftarrow \frac{1}{s}\sum_{n=1}^{s} \mathrm{sgn}(f_n^{i,j}(x)),\ 1 \le i < j \le k.$$

Step 2 For $1 \le i < j \le k$,

$$\hat{g}_{i,j}(x) = \begin{cases} i, & g_{i,j}(x) > 0, \\ j, & g_{i,j}(x) < 0. \end{cases}$$

Step 3 Update the alternative label set by

$$M \leftarrow \{m : |S_m| = \max_{1\le l\le k} |S_l|, 1 \le m \le k\},$$

where $S_l = \{g_{i,j}(x) :\ \hat{g}_{i,j}(x) = l\}, 1 \le l \le k$.

Step 4 If $|M| = 1$, then set

$$L(x) \leftarrow \arg\max_{1\le l\le k} |S_l|;$$

else calculate $w_m(x) = \frac{1}{|S_m|}\sum_{g_{i,j}\in S_m} |g_{i,j}(x)|,\ m \in M$, and set

$$L(x) \leftarrow \arg\max_{m\in M}\ w_m(x).$$

The one-versus-one method first constructs binary classifiers between any two different classes, and then uses voting method to generate the final multi-category classification decision function.

Picking up the samples belonging to i-th and j-th classes, the decision function vector $\mathbf{f}^{i,j} = [f_1^{i,j}, f_2^{i,j}, \cdots, f_s^{i,j}]$ is constructed using the method in subsection 2.2, where $f_m^{i,j}(1 \le m \le s)$ is a approximation of binary classifier between i-th class and j-th class. Assume samples in i-th class have positive labels while the samples in j-th class have negative ones, the binary decision function $g_{i,j}$ can be constructed using the method described in one-versus-rest method, which is represented as

$$g_{i,j}(x) = \frac{1}{s}\sum_{m=1}^{s} \mathrm{sgn}(f_m^{i,j}(x)). \tag{2.7}$$

Hence, $g_{i,j}(x) \in [-1, 1]$.

After all $\frac{k(k-1)}{2}$ functions $g_{i,j}(1 \le i < j \le k)$ have been constructed, each sample has been labeled $\frac{k(k-1)}{2}$ times. The class of the sample x is given by the majority voting rule. More precisely, we define

$$\hat{g}_{i,j}(x) = \begin{cases} i, & g_{i,j}(x) > 0, \\ j, & g_{i,j}(x) < 0, \end{cases} \tag{2.8}$$

and divide the $\frac{k(k-1)}{2}$ functions $g_{i,j}$ into k sets $S_1, \cdots, S_k$ such that $S_l = \{g_{i,j} : \hat{g}_{i,j}(x) = l\}$.

Let c_l be the cardinality of S_l: $c_l = |S_l|$. If $c_{\tilde{m}}$ is the unique largest number among $c_l, 1 \le l \le k$, then we assign x to class $\tilde{m}$. Otherwise, say $c_{m_1} = c_{m_2} = \max\{c_l, 1 \le l \le k\}$, we compute $w_{m_1}(x) = \frac{1}{|S_{m_1}|}\sum_{g_{i,j}\in S_{m_1}} |g_{i,j}(x)|$ and $w_{m_2}(x) = \frac{1}{|S_{m_2}|}\sum_{g_{i,j}\in S_{m_2}} |g_{i,j}(x)|$. If $w_{m_1}(x) > w_{m_2}(x)$, then we assign x to the class m_1, else if $w_{m_1}(x) < w_{m_2}(x)$, then we assign x to the class m_2. In case that $w_{m_1}(x) = w_{m_2}(x)$, then we assign x to either the class m_1 or the class m_2 at random. In a straightforward way, we may apply the voting rule for the case that there are more than 2 subsets achieving the maximal cardinality.

3. Experimental results

The performance of the proposed method is evaluated on two well-known handwritten digit data sets, MNIST,[11] USPS[10] and a Yale face[7] data set. In MNIST, USPS and Yale, a digit or face is originally represented as a $c \times c$ matrix, where $c = 28$ for MNIST, $c = 16$ for USPS, and $c = 100$ for Yale. To reduce the shift-variance, a distance between $x = (x_{i,j})_{i,j=1}^{c}$ and $y = (y_{i,j})_{i,j=1}^{c}$ is defined by (3.1).[17]

$$d(x, y) = \min_{|i'-i|\le 1, |j'-j|\le 1} \sqrt{\sum_{i=2}^{c-1}\sum_{j=2}^{c-1}(x_{i,j} - y_{i',j'})^2} \tag{3.1}$$

In our experiments, a digit is represented as a $c \times c = c^2$ dimensional vector, so the data set can be rewritten as $X = \{x_i\}_{i=1}^{n}$ where x_i is a c^2 dimensional vector. In data set X, only a subset X_0 is labeled.

3.1. *Classification on MNIST*

In this section, the sample subsets of MNIST handwritten digits are chosen according to Ref. 17. For digits in $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 0\}$, 200 samples for each digit are randomly selected so that $|X| = 2000$ and $X \subset R^{784}$. In each experiment, we choose $2, 4, \cdots, 20$ samples from each digit as the labeled samples randomly, therefore, the initial labeled set X_0 has the size $20, 40, \cdots, 200$, respectively.

We first consider the one-versus-rest classification method. Let the initial labeled samples selected from 1-th class be labeled by $+1$ and other labeled samples be labeled by -1. In the experiment, we construct the iteration scheme enlarging the labeled set to improve the algorithm, and the ratio of digits in Class $+1$ to digits in Class -1 is kept to be $r = 1 : 9$ in each updated labeled set. In the experiment, we employ $s = 3$ in iteration scheme and $s = 10$ in the last step in algorithm 1.

By the means of step 2.1 and step 2.2 in algorithm 1, the function vector $\mathbf{f}^* = [f_1^*, f_2^*, \cdots, f_s^*]$ is obtained based on the labeled set X_0. In the experiment, X_f is divided into two subsets as follows:

$$X_f^+ = \cap_{i=1}^s \{x \in X \backslash X_0; f_i^*(x) > 0\}, \quad X_f^- = \cap_{i=1}^s \{x \in X \backslash X_0; f_i^*(x) < 0\}.$$

Given a constant $p = 0.5$, we select $X_{conf}^+ \subset X_f^+$ and $X_{conf}^- \subset X_f^-$ satisfying:

(1) $|X_{conf}^+| \leq p|X_f^+|, |X_{conf}^-| \leq p|X_f^-|$;
(2) $|X_{conf}^+| : |X_{conf}^-| = r$.

The feasible confident set is constructed as $X_{conf} = X_{conf}^+ \bigcup X_{conf}^-$ and the labeled set is updated as $X_0 = X_0 \bigcup X_{conf}$. In the last iteration, the decision function vector $\mathbf{f}^1$ is given.

Repeating the above procedures, let initial labeled samples selected from m-th class $(2 \leq m \leq k)$ be labeled by $+1$ and other labeled samples be labeled by -1. The decision function vector $\mathbf{f}^m (2 \leq m \leq k)$ is constructed by interpolation based on 1-D multi-embedding.

When k binary decision function vectors $\mathbf{f}^1, \mathbf{f}^2, \cdots, \mathbf{f}^k$ are constructed, k binary decision functions $p^1(x), p^2(x), \cdots, p^k(x)$ are constructed by (2.4), a sample x_i is labeled to the class corresponding to maximal decision function value. If more than one class achieve the maximal value, we compare their weights to decide the label of the sample as mentioned in algorithm 2.

In the one-versus-one method, let the initial labeled samples selected from i-th class and j-th class be labeled by $+1$ and -1, respectively, where $1 \leq i < j \leq k$. According to algorithm 3, $\frac{k(k-1)}{2}$ binary decision functions can be constructed, then each sample is labeled $\frac{k(k-1)}{2}$ times. Based on the majority voting method, the final label for each sample can be obtained.

In the experiments, our 1-D multi-embedding multi-category classification method is compared with SVM,[3] Laplacian SVM,[1] random forest,[2] KNN method and adaboost algorithm.[6] The one-versus-rest (1-v-r) and one-versus-one (1-v-1) 1-D multi-embedding multi-category classification methods are compared with the corresponding 1-v-r SVM and Laplacian SVM, 1-v-1 SVM and Laplacian SVM, respectively.

Fig. 1 shows the average accuracy of different algorithms on MNIST. It clearly shows that our method gives the best results using different number of labeled samples per class. Especially, when the number of labeled samples per class is 2, the accuracy of other algorithms is about 60% while our algorithm almost correctly classifies all the samples. Except for 1-v-r 1-D multi-embedding multi-category classification method, the 1-v-r Laplacian SVM provides the best results because it uses both the labeled and unlabeled samples. Fig. 2 shows the average accuracies of our proposed 1-v-1 1-D multi-embedding multi-category classification method, 1-v-1 SVM, 1-v-1 Laplacian SVM, random forest, KNN method and adaboost algorithm. Similar to Fig. 1, our method shows the best classification performance even if the labeled set is extremely small.

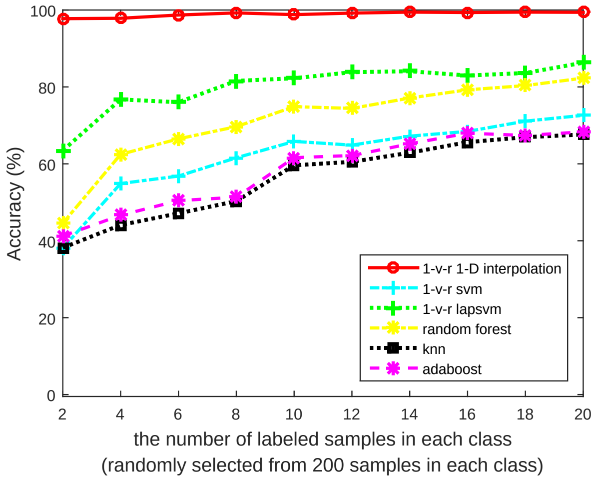

Fig. 1. Comparison of 1-v-r 1-D multi-embedding multi-category classification method and other methods on MNIST.

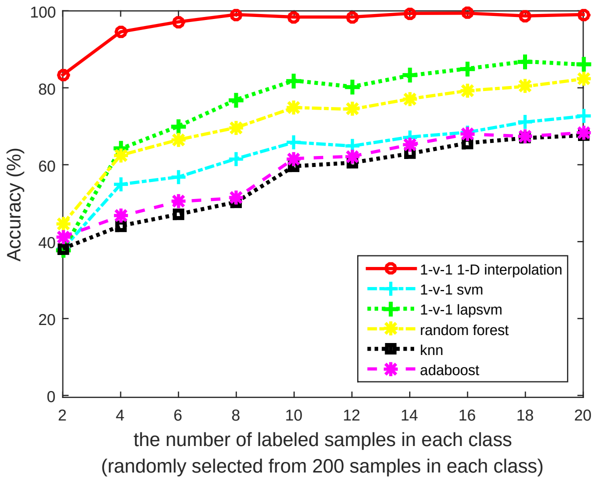

Fig. 2. Comparison of 1-v-1 1-D multi-embedding multi-category classification method and other methods on MNIST.

3.2. *Classification on USPS*

For digits in $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 0\}$, 200 samples for each digit are randomly selected so that $|X| = 2000$ and $X \subset R^{256}$. In each experiment, choose $2, 4, \cdots, 20$ samples from each digit as the labeled samples randomly, therefore, the initial labeled set X_0 has the size $20, 40, \cdots, 200$, respectively.

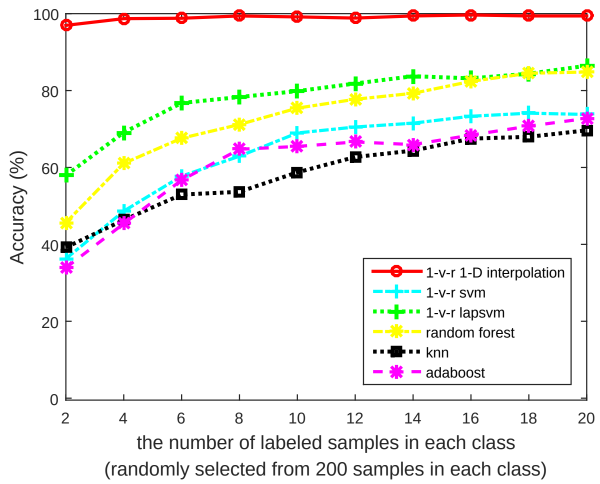

Fig. 3. Comparison of 1-v-r 1-D multi-embedding multi-category classification method and other methods on USPS.

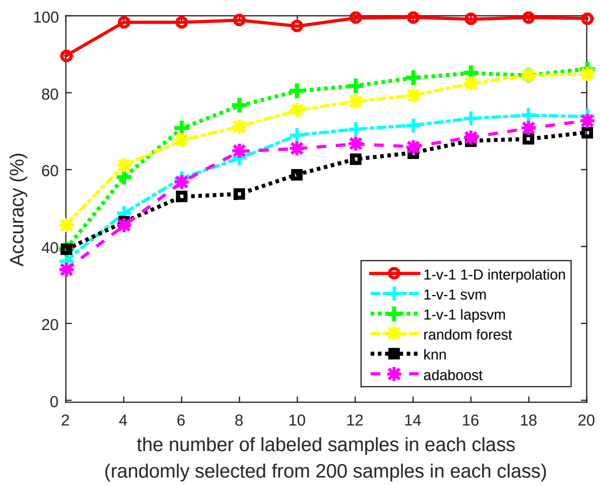

Fig. 4. Comparison of 1-v-1 1-D multi-embedding multi-category classification method and other methods on USPS.

Figs. 3 and 4 show the average accuracy of different methods on USPS data. It can be seen that our method shows consistent better results than other methods even in the case of extremely limited labeled samples.

3.3. *Classification on Yale face data*

In this experiment, we use the proposed 1-D multi-embedding multi-category classification methods to classify the face images in Yale face data set. The Yale data set has 15 classes, and each class contains 11 pictures. The size of each face image is 100×100, so the data $X \subset R^{10,000}$. We randomly select $1, 2, 3, 4, 5$ samples from each class to form the labeled set. Fig. 5 and Fig. 6 show the average accuracy of different methods. The proposed method provides the best results in all cases, which demonstrates that the 1-D multi-embedding method is very effective for the classification of high dimensional data.

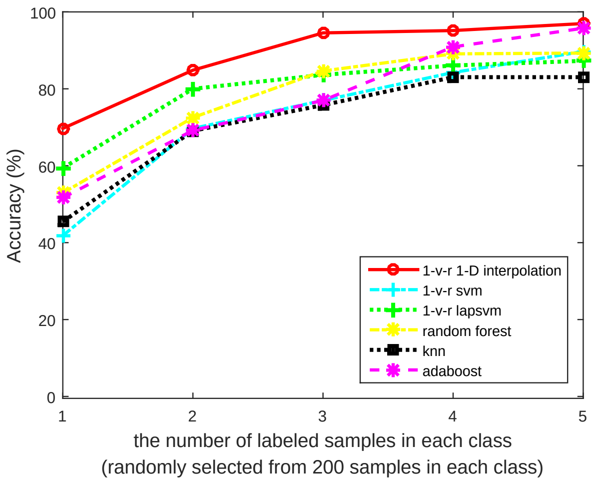

Fig. 5. Comparison of 1-v-r 1-D multi-embedding multi-category classification method and other methods for face recognition on Yale face data.

4. Conclusion

In this paper, a multi-category classification method based on 1-D embedding of the data has been proposed. The 1-D multi-embedding method transfers the complex high dimensional classification problem to a simple 1-D interpolation classification problem. In the 1-D space, the binary classifier is represented as a univariate function. By means of one-versus-rest and one-versus-one techniques, a multi-category classifier is generated on the basis of several binary classifiers on 1-D multi-embeddings. Experimental results have shown that the 1-D multi-category classification method is very effective to handle high dimensional data with only limited labeled samples.

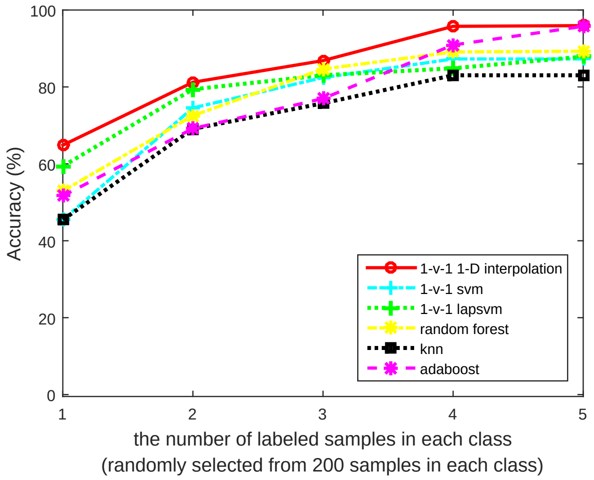

Fig. 6. Comparison of 1-v-r 1-D multi-embedding multi-category classification method and other methods for face recognition on Yale face data.

Acknowledgements

This work was in part supported by the National Natural Science Foundation of China under Grant 11371007.

References

1. M. Belkin, P. Niyogi and V. Sindhwani, Manifold regularization: a geometric framework for learning from labeled and unlabeled examples, *Journal of Machine Leraning Research*, **7** (3) (2006) 2399–2434.
2. L. Breiman, Random forests, *Machine Learning*, **45** (1) (2001) 5–32.
3. C. Chang and C. Lin, LIBSVM: a library for support vector machines, *ACM Trans. Intelligent Systems and Technology*, **2** (3) (2011), Article No. 27 (27 pages).
4. H. Chen and L. Li, On the rate of convergence for multi-category classification based on convex losses, *Science in China Series A: Mathematics*, **50** (11) (2007) 1529–1536.
5. H. Chen and L. Li, Semisupervised multicategory classification with imperfect model, *IEEE Trans. Neural Networks*, **20** (10) (2009) 1594–1693.
6. Y. Freund and R. Schapire, A decision theoretic generalization of on-line learning and an application to boosting, *Journal of Computer and System Sciences*, **55** (1) (1997) 119–139.
7. A. S. Georghiades, P. N. Belhumeur and D. J. Kriegman, From few to many: illumination cone models for face recognition under variable lighting and pose, *IEEE Trans. Pattern Anal. Mach. Intelligence*, **23** (6) (2001) 643–660.
8. C. Hsu and C. Lin, A comparison of methods for multi-class support vector machines, *IEEE Trans. Neural Networks*, **13** (2) (2002) 415–425.
9. G. F. Hughes, On the mean accuracy of statistical pattern recognizers, *IEEE Trans. Inf. Theory*, **14** (1) (1968) 55–63.
10. J. J. Hull, A database for handwritten text recognition research, *IEEE Trans. Pattern*

Anal. Mach. Intell., **16** (5) (1994) 550–554.
11. Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, **86** (11) (1998) 2278–2324.
12. G. Ou and Y. L. Murphey, Multi-class pattern classification using neural networks, *Pattern Recognition*, **40** (1) (2007) 4–18.
13. I. Rish, An empirical study of the naive Bayes classifier, *Journal of Universal Computer Science*, **1** (2) (2001).
14. S. Roweis and L. Saul, Nonlinear dimensionality reduction by locally linear embedding, *Science*, **290** (5500) (2000) 2323–2326.
15. J. Sulzmann, J. Furnkranz and E. Hullermeier, On Pairwise Naive Bayes Classifiers, *Machine Learning*, **4701** (2007) 371–381.
16. J. B. Tenenbaum, V. de Silva and J. C. Langford, A global geometric framework for nonlinear dimensionality reduction, *Science*, **290** (5500) (2000) 1959–1966.
17. J. Wang, Semi-Supervised Learning Using Multiple One-Dimensional Embedding-Based Adaptive Interpolation, in this special issue of *International Journal of Wavelets, Multiresolution and Information Processing*, (2015)
18. J. Wang, Semi-Supervised Learning Using Ensembles of Multiple 1D-Embedding-Based Label Boosting, in this special issue of *International Journal of Wavelets, Multiresolution and Information Processing*, (2015)
19. K. Q. Weinberger and L. K. Saul, Distance metric learning for large margin nearest neighbor classification, *Journal of Machine Learning Research*, **10** (1) (2009) 207–244.
20. X. Zhu, *Semi-Supervised Learning Literature Survey, Technical Report 1530, Computer Sciences, University of Wisconsin-Madison*, (2005).