Cascaded Dimensionality Reduction Method and Its Application in Spectral Classification

Huifen Dong¹ and Jingchang Pan²,*
¹,²School of Mechanical, Electrical & Information Engineering, Shandong University, China

*Corresponding author email: pjc@sdu.edu.cn

Abstract. The classification of the high-dimensional spectral is one of the important study domains in the astronomy. However, the curse of dimension problem restrains the performance of the methods to classify the spectral data. In this paper, the cascaded dimensionality reduction, combining with the virtues of the principal component analysis and t-distributed stochastic neighbour embedding, is conducted to improve the performance of classification methods for spectral data. In the cascaded dimensionality reduction, the PCA is employed to pre-reduce dimensions of spectral data for reducing redundant information, under the constraint of preserving the information integrity as far as possible; T-SNE highlights the differences among the samples with different labels, and outputs target results after the dimension reduction. The support vector machine in conjunction with the cascaded dimensionality reduction is applied to classify the spectral data, and its performance is compared with the PCA based SVM and T-SNE based SVM. Experimental results demonstrate that the cascaded dimensionality reduction assists the SVM obtaining better performance than PCA and T-SNE.

Keywords: Cascaded dimensionality reduction; T-SNE; LAMOST.

1. Introduction
Large Sky Area Multi-Object Fiber Spectroscopy Telescope (LAMOST) can produce over ten-thousands spectral records every night, which provides substantial information to explore mysteries of universe. Although the machine learning methods, such as the decision tree, the neural network and the support vector machine [1], have obtained exciting results for spectral classification tasks, the curse of dimension problem of high-dimensional spectral data restrains the performance of classification methods.

The dimension reduction methods have brought light to reduce the impact of high dimensional data on classification methods. Among these methods, the principal component analysis (PCA) is widely applied to spectral classification problems, as its simple implementation and insensitive to noise [2]. Also, the information integrity of spectral data can be protected as far as possible by controlling the contribution rates in PCA when reducing the dimension. However, it is hard for PCA to simultaneously save the local and global structures contained in spectral data. Thus, compared with original data, sometimes, the performance of classification methods is decreased after the spectral data reduced dimensions by PCA. t-distributed stochastic neighbour embedding (T-SNE) can capture local and global structures of data, and highlights the different among samples with different tags [3], which can decrease the complexity of classifying spectral data. Nevertheless, the high time complexity for the high-dimensional data restrains the applications of T-SNE.

The above observations have derived us to conduct the cascaded dimensionality reduction to improve the performance of the classification method for high dimensional spectral data. The cascaded
dimensionality reduction (CDR) can synthesize the advantages of PCA and T-SNE. It includes the pre-reducing dimension (PD) and the extracting difference (ED) operations. The PD is employed to reduce the high-dimensional spectral data with the limitation that the information integrity contained in spectral data is saved as far as possible; the ED can highlight the difference among spectral records with different labels, and displays the target spectral data after the dimension reduction. The contributions of this study are listed as follows:

- The CDR is conducted to deal with the high-dimensional spectral data.
- The CDR is employed to assist the support vector machine (SVM) to improve the performance for classifying the spectral data.

The structure of this study is arranged as follows. Section 2 reviews some work for dimensionality reduction and classification of the high-dimensional spectral data. Section 3 introduces the methodology in details. Section 4 displays the experiments, and analyses the results. Section 5 concludes this study, and proposes the future works.

2. Related Works
In this subsection, we given the review of these works, but is by no means complete. The linear dimension reduction is widely applied in this field. Kheirdastan et al. [4] applied the PCA as feature extraction method, and investigated the performance of probabilistic neural network, K-means and SVM for classifying the stellar spectra. Tang et al. [5] applied fisher discriminant analysis (FDA) to extract the features based on the fusing information in galaxy spectral data, and obtained superior classification performance. The non-linear dimension reduction method has obtained lots of attention, because it can capture the non-linear structure of the original data. Liu [6] employed the locality preserving projections (LPP) as dimension reduction method to preserve the local structure of the spectral data and then the SVM is used to classify spectral data. Pan et al. [7] employed the Locally linear embedding (LLE) to assist the classification method to improve the performance in spectral classification. Wang et al. [8] proposed a spectral feature extraction algorithm for astronomical spectra, which uses the deep neural network to extract features from different levels.

3. Methodology
In this section, the process of dimensionality reduction for spectral data by using cascaded dimensionality reduction is introduced in details. The cascaded dimensionality reduction includes two operations, the pre-reducing dimension (PRD) and extracting difference (ED). The PRD is employed to decrease computational complexity for the ED operation. During PRD operation, the PCA is used to reduce the number of dimensions of the original spectral dataset, and produce the pre-reduced dataset. In this process, the redundant information of the spectral is removed under the condition that valuable information is saved as much as possible. In the ED operation, for the pre-reduced dataset, the T-SNE is adopted to highlight the difference of samples with different labels. In this process, the number of dimensions of pre-reduced dataset are reduced to the target number.

3.1. Pre-reducing Dimension
In PRD, the PCA is used to pre-reduce the dimension of spectral dataset \( X = \{X^1, \ldots, X^i, \ldots, X^m\} \), where \( m \) represents the number of spectra. \( X^i = (x_{i1}, x_{i2}, \ldots, x_{iN})^T \) is the \( i \)th record in \( X \). \( N \) is equal to the number of dimensions of \( X^i \). For the records of possibly correlated variables (dimensions) are converted into a set of linearly uncorrelated principal components by employing orthogonal transformation technique. In this process, the original data is projected to \( K \) bases, which can save the largest variance of different dimensions, and remain valuable information as far as possible. Normally, \( K \) is smaller than \( N \), thus, the dimensions of spectral data are reduced to \( K \) from \( N \).

For obtaining suitable the \( K \), the contribution rate of the first \( m \) principal components is defined in Eq. (1). The valuable information contained in original spectral dataset is saved by controlling the value of \( \alpha \). The pre-reducing dimension operation is able to remove the redundant information of original data, and save the valuable information as far as possible. After pre-reducing dimension operation, the pre-
reduced spectral dataset $Y$ is obtained. The computation complexity for the extracting differentiation operation is reduced by using the pre-reduced dataset.

$$\sum_{i=1}^{m} \lambda_i (\sum_{j=1}^{p} \lambda_j)^{-1} \geq \alpha\% \quad (1)$$

### 3.2. Extracting Difference

For the pre-reduced spectral dataset $Y$, although the redundant information is reduced, the difference between heterogeneous records (samples) are not highlighted. T-SNE is employed to extract difference between heterogeneous spectral samples, and outputs the goal spectral dataset. T-SNE assumes that the distribution of original data is subject to the gaussian distribution, and the distribution of dimensions reduced data obeys a general t-distribution, which can make the distance between clusters with larger distance [3]. Thus, the goal spectral dataset obtained by T-SNE can highlight the difference between heterogeneous spectral samples. The process of extracting difference consists of two steps: firstly, T-SNE constructs a probability distribution among high-dimensional spectral samples, so that similar samples have higher probability of being selected, meanwhile, non-similar samples have lower probability of being selected. Secondly, T-SNE constructs the probability distribution of these spectral samples in a low-dimensional space, making the two probability distributions as similar as possible.

Suppose the two spectral samples in the high dimensional space are $Y_i$ and $Y_j$ respectively. $Y_i$ selects $Y_j$ to be its neighbour with the conditional probability $p(j|i)$ defined in Eq. (2). If $Y_j$ is closer to $Y_i$, $p(j|i)$ is larger.

$$p(j|i) = \frac{\exp(-\|Y_i - Y_j\|^2/2\sigma_i^2)}{\sum_{k\neq i} \exp(-\|Y_i - Y_k\|^2/2\sigma_i^2)} \quad (2)$$

Suppose the mapping points of high-dimensional data points $Y_i$ and $Y_j$ respectively are $Z_i$ and $Z_j$ in low-dimensional space. The conditional probability of low-dimensional space is represented by $q(j|i)$, defined in Eq. (3).

$$q(j|i) = \frac{\exp(-\|Z_i - Z_j\|^2)}{\sum_{k\neq i} \exp(-\|Z_i - Z_k\|^2)} \quad (3)$$

Construct joint probability distributions $p$ and $q$ in high-dimensional space and low-dimensional space respectively, for $Y_i$ and $Y_j$, there always are $p(j|i) = q(i|j)$, $q(i|j) = q(j|i)$. $p(ij)$ is defined in Eq. (4).

$$p_{ij} = \frac{p_{ij} + p_{ji}}{2m} \quad (4)$$

The cost function $F$, defined in Eq. (5), is written using the KL. The gradient descent method is employed as the optimization method. For $ith$ sample $Z_i$ in the low-dimensional space, its gradient is defined in Eq. (6). The updated formula of $Z_i$ expressed as Eq. (7), where the $r$ is the learning rate.

$$F = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \quad (5)$$

$$\frac{\partial F}{\partial Z_i} = 4 \sum_j (p_{ij} - q_{ij}) (Z_i - Z_j) (1 + \|Z_i - Z_j\|^2)^{-1} \quad (6)$$

$$Z_i^{t+1} = Z_i^t + r \frac{\partial F}{\partial Z_i} + \gamma (Z_i^t - Z_i^{t-1}) \quad (7)$$

For spectral dataset $Z = \{Z_1, \ldots, Z_i, \ldots, Z_m\}$, which is the representation of pre-reduced spectral dataset $Y$ in the low dimensional space, is the target result of dimension reduction.

### 3.3. Classification

The spectral dataset $z$ is used as the training data to train the SVM. For $ith$ spectral record $Z_i$ in dataset $Z$, the predicted label that is produced by SVM is viewed as the label of $X_i$ in the spectral dataset $X$. 

3
4. Experiments

4.1. Data Preprocessing
In experiments, the spectral data is from LAMOST spectral survey, data release 3, including 1500 M giant spectrums and 1500 M dwarf spectrums. Figure 1. displays the comparison between the M giant and M dwarf. In the spectral dataset, each sample contains 3952 dimensions, in which, the first dimension is the label of this sample. In order to avoid the effect of different scales between dimensions, the spectral dataset was normalized.

![Figure 1. Comparison between the M giant and M dwarf.](image)

4.2. Performance of Dimensional Reduction
In this subsection, the performance of CDR for the spectral dataset was compared with PCA, ISOMAP, T-SNE and LLE. Figure 2. displays the results of these algorithms in two-dimensional space. The parameters of each algorithm were carefully tuned to achieve a fair comparison by trial and error. For the ISOMAP, the number of neighbours, \( N_e \), was selected from range \([21, 61]\). For the T-SNE, the iteration number was set as 1000. Its perplexity ranges from \([5, 50]\), learning rate \( r \) is equal to 200. For the LLE, the neighbours number ranges from \([21, 61]\). For the CDR, the contribution rate was selected from \([98, 100]\). The other parameters are same with the T-SNE. By observing Figure 2., we can find that the CDR displays more better results for the spectral dataset compared with the PCA, T-SNE, ISOMAP and LLE.

![Figure 2. Performance of the dimension reduction for PCA, T-SNE, ISOMAP, LLE and CDR](image)

|                       | Rand Index | Mutual Information |
|-----------------------|------------|--------------------|
| PCA+K-means           | 0.661      | 0.577              |
| LLE+K-means           | 0.652      | 0.543              |
| CDR+K-means           | **0.742**  | **0.642**          |
| T-SNE+K-means         | 0.729      | 0.624              |
| ISOMAP+K-means        | 0.651      | 0.571              |

To validate the capability of the CDR, for the spectral dataset, the K-means method was employed to cluster the outputs of PCA, ISOMAP, LLE, T-SNE and CDR. The indicators, rand index (RI) and mutual information (MI), are used to evaluate the quality of clusters. Table. 1 shows the values of the RI and MI. It can be observed that the CDR+K-means obtains the best results both RI and MI compared with
other methods. Therefore, CDR is the more appropriate choice than other methods in reducing the dimensions of the spectral data.

4.3. Spectral Classification
In this subsection, the capability of CDR allying the classification method for classifying the spectral data was investigated. Due to widely used in spectral classification, the SVM was selected as the classification method. The performance of CDR+SVM was compared with PCA+SVM, ISOMAP+SVM, LLE+SVM and T-SNE+SVM. Figure 3 and 4 display the training accuracy and generalization accuracy of SVM combined with 5 dimensionality reduction for classifying spectral dataset. The CDR+SVM shows the better result compared with the PCA+SVM, LLE+SVM and ISOMAP+SVM. This phenomenon demonstrates that the extracting difference operation of CDR, which is used to highlight the different between heterogeneous samples, can assist the SVM to improve performance for classifying the spectral data. The CDR+SVM performs same performance with the T-SNE+SVM, when the number of dimensions of the spectral dataset is reduced to 2 dimensional. However, CDR+SVM achieves best result compared with the T-SNE+SVM, when the number of dimensions of the spectral dataset is respectively reduced to 1 and 3. These results reflects that the CDR+SVM has better capability than the PCA+SVM, ISOMAP+SVM, LLE+SVM and T-SNE+SVM.

![Figure 3. Training accuracy.](image1)
![Figure 4. Generalization accuracy.](image2)

5. Conclusion
In this study, the CDR has been conducted to promote the performance of the classification methods in classifying the spectral data. The performance of CDR is compared with 4 dimensionality reduction methods in spectral data. The cluster performance demonstrates the superior of CDR. Moreover, the capability of CDR + SVM for classifying the spectral data is better than other comparison methods. These experiments reveal that CDR obtains promising performance for reducing the dimension of spectral data, and improves the ability of SVM for classifying the spectral data. In the future, we will use more spectral data to mine the potential of the CDR.

Acknowledgements
This work was financially supported by the National Natural Science Foundation of China (No. U1931209, No. 11873037).

References
[1] Tan W, Sun L, Yang F, et al. Study on bruising degree classification of apples using hyperspectral imaging and GS-SVM[J]. Optik, 2018, 154: 581-592.
[2] Karamizadeh S, Abdullah S M, Manaf A A, et al. An overview of principal component analysis[J]. Journal of Signal and Information Processing, 2013, 4(3B): 173-175.
[3] Maaten L, Hinton G. Visualizing data using t-SNE[J]. Journal of machine learning research, 2008, 9(Nov): 2579-2605.
[4] Kheirdestan S, Bazarghan M. SDSS-DR12 bulk stellar spectral classification: Artificial neural networks approach[J]. Astrophysics and Space Science, 2016, 361(9): 304.
[5] Tang L, Peng S, Bi Y, et al. A new method combining LDA and PLS for dimension reduction[J]. Plos One, 2014, 9(5).
[6] Zhong-bao L. Stellar Spectral Classification with Locality Preserving Projections and Support Vector Machine[J]. Journal of Astrophysics and Astronomy, 2016, 37(2): 12.

[7] Yude B, Jingchang P, Bin J, et al. Stellar Spectral Subclass Classification Based on Locally Linear Embedding[J]. Publications of the Astronomical Society of Japan, 2013, 65(4).

[8] Wang K, Guo P, Luo A L. A new automated spectral feature extraction method and its application in spectral classification and defective spectra recovery[J]. Monthly Notices of the Royal Astronomical Society, 2017, 465(4): 4311-4324.