International Mathematical Forum, Vol. 12, 2017, no. 3, 111 - 118
HIKARI Ltd, www.m-hikari.com
https://doi.org/10.12988/imf.2017.610133

Multidimensional Scaling Analysis Based on Attribute Reduction of Bivariate Mutual Information

Gao-Hui Peng

North China University of Water Resources and Electric Power
Zhengzhou 450011, Henan Province, China

Xiang-Xiang Wang

North China University of Water Resources and Electric Power
Zhengzhou 450011, Henan Province, China

Yuan-Zhang Zhang

North China University of Water Resources and Electric Power
Zhengzhou 450011, Henan Province, China

Abstract

An improved MDA is proposed to improve the operation efficiency of dimensionality reduction algorithm, it includes two steps: First simplify the data by attribute reduction of bivariate mutual information, and then dispose the simplified data by the original MDA. Compared with the original, the improved MDA is good at keeping structural relationships between the data points, and at the same time can be applied on ultra high dimensional data with more redundant attributes, because of its lower computational complexity. Experimental results show that the improved MDA is effective.

Keywords: data dimensionality reduction; attribute reduction; entropy; mutual information

1 Introduction

In information age, the rapid development technologies of data collection and data storage bring us various types of big data, and the dimensions of some data usually can reach hundreds of thousands. Thus, scholars begin to pay more attention to how to use the potential information effectively from these data. Currently, the common method to reduce the dimensionality is linear dimensionality reduction, mainly including Principal Component Analysis (PCA) [1], Linear Discriminant Analysis (LDA) [2], Locally Linear Embedding (LLE) [3] and so on. However, linear dimensionality reduction still has some problems about computational complexity, and even some certain limitations on dimensions. For instance, PCA algorithm may not be feasible to calculate the feature vector of very high-dimensional data [4]; the largest dimension limitation of LDA algorithm is k-1(where k is the category number) [4], etc. In practical application, the high-dimensional database may have many shortcomings, containing many attributes which are redundant and unnecessary in the process of rule discovery [5]. This not only increases the algorithm computational complexity but even limits the realization of the algorithm, so in dealing with high dimensional data with more redundancy attributes, we can remove the redundant attributes first and then use the dimension reduction algorithm, the algorithm can effectively improve the arithmetic speed.

Based on the two-dimensional mutual information theory, the paper proposes the attribute reduction algorithm and acquires the improved Multidimensional Scaling Analysis by studying Multidimensional Scaling Analysis (MDA) [6], we carried dimensionality reductions by these two methods respectively, and the results are compared to analyze the feasibility of the improved algorithm through examples.

2 Algorithmic research

2.1. Multidimensional Scaling Analysis(MDA)

Assume there is a set of high-dimensional array $p = [p_1, p_2, p_3,..., p_n]'$, $p_i \in R^D$, where $p_i = [x_{i1}, x_{i2}, x_{i3},..., x_{iD}]'$, principle of MDA is as follows:

Make each point do dot product with any point, we acquire a computing matrix $\overline{p} = (\overline{p}_{i,j})_{n\times n}$ order n by n, where

$$\overline{p}_{i,j} = p_i p_j = (x_{i1}x_{j1} + x_{i2}x_{j2} + ...x_{iD}x_{jD})$$

Matrix $\overline{p}$ is disposed by the eigenvalue analysis, let λ_i be the eigenvalue, let $(\alpha_{i1}, \alpha_{i2},..., \alpha_{in})$ be the eigenvector of λ_i, then we can obtain a new array after reduction of dimensionality, and the coordinate of point i is:

$$L_i = (\lambda_1^{1/2}\alpha_{1i}, \lambda_2^{1/2}\alpha_{2i},..., \lambda_d^{1/2}\alpha_{di})$$

Where d is the number of larger eigenvalues of matrix $\overline{p}$.

2.2. Multidimensional scaling analysis based on attribute reduction of bivariate mutual information

Mutual information is a measure of the amount of overlapping information between random variables, two-dimensional mutual information refers to the mutual information between two properties, its computation complexity is much more lower than conditional mutual information between multiple properties, therefore, the attribute reduction of bivariate mutual information is put forward, and applied to the MDA algorithm, the algorithm theory is as follows:

1) Assume there are n attributes recorded as $[D_1, D_2, ..., D_n]$, entropy of each attribute is calculated respectively and arranged in descending, the result is as follows: $[H_a \geq H_b \geq, ..., \geq H_c]$,where

$$D_i = [d_{1i}, d_{2i}, ..., d_{mi}]', \; H_i = H(D_i) = -\sum_{d_i} p(d_i)\log p(d_i)$$

2) Relation formulas between mutual information and entropy are as follows:

$$I(X;Y) = H(X) - H(X \mid Y) \tag{1}$$

$$I(X;X) = H(X) - H(X \mid X) = H(X) \tag{2}$$

$$H(X \mid Y) = \sum_y p(y)H(X \mid y) = -\sum_y p(y)\sum_x p(x \mid y)\log p(x \mid y) \tag{3}$$

It can be seen that mutual information $I(X;Y)$ represents the intersecting part of the information of X and Y, in other words, the larger the mutual information is , the more overlapping information X and Y contain, so the following formula is introduced to filter the property:

$$Q_{ij} = \frac{I(D_i;D_j)}{H(D_i)} = \frac{H(D_i) - H(D_i \mid D_j)}{H(D_i)} \geq K(K \geq 85\%) \tag{4}$$

According to the formula (1) and (2), When X approaches Y, $Q = \frac{I(X;Y)}{H(X)}$ approaches 1, This means that the more similar information X and Y contain, the larger the value of Q is. Hence, If property D_i and D_j satisfy formula (4), it indicates that the most information of D_j is covered with the information of D_i, then we can delete D_j, and continue to test the next property until all are tested completely.

3) Through above two steps we eliminate the unnecessary attributes and finally regard the remaining attributes as the simplified attributes.

4) Dealing with the simplified data by the original MDA.

3 Application Example

There are ten sample points, each sample has four attributes, and the corresponding datas are shown in Table 1.

Table 1The data of sample points

serial number	X_1	X_2	X_3	X_4
1	1	6	3	7
2	2	6	4	7
3	1	7	4	7
4	3	5	5	4
5	4	4	6	3
6	4	7	8	6
7	3	1	9	1
8	5	5	10	4
9	8	8	11	9
10	9	9	12	9

1) The steps of attribute reduction are as follows:

①Use the partition method to calculate the information entropy of each attribute:

Let $\max_i$ and $\min_i$ be the maximum and minimum value of the property X_i respectively, d is the interval number (where d is 5), and $k_i = \frac{\max_i - \min_i}{d}$ is the length of the each interval, the intervals are as follows: $[\min_i, \min_i + k_i)$ 、 $[\min_i + k_i, \min_i + 2k_i)$ 、 ... 、 $[\min_i + (d-1)k_i, \max_i]$. By calculating the frequency of each interval, we acquire the information entropy of X_i.

The entropies of four properties obtained by above method are as follows: $H(X_1) = 1.8464$, $H(X_2) = 2.1219$, $H(X_3) = 2.2465$, $H(X_4) = 1.8464$.

② According to $H(X_3) > H(X_2) > H(X_1) = H(X_4)$, we start from the attribute X_3, and successively evaluate the values of the mutual information and Q , results are shown in Table 2:

Table 2 The values of mutual information and Q

Attributes (X_i, X_j)	$I(X_i; X_j)$	Q_{ij}
(X_3, X_2)	1.7220	0.77
(X_3, X_1)	1.6465	0.73
(X_3, X_4)	1.6464	0.73
(X_2, X_1)	1.319	0.62
(X_2, X_4)	1.8465	0.87

In Table 2, $Q_{24} = 0.87 > 0.85$, it indicates the main information of X_4 is covered by X_2, then we can delete X_4, finally, $[X_1, X_{2,} X_3]$ is the simplified attribute group.

2) Use MDA algorithm to process the original data and the simplified respectively, results are shown in Table 3.

Table 3 Results of dimensionality reduction

	Results of dimensionality reduction of the original data		Results of dimensionality reduction of the simplified data	
	X_1	Y_1	X_2	Y_2
1	8.584959281	-4.543810977	5.859817147	-3.346831025
2	9.575769198	-3.602009928	7.009518711	-2.599024824
3	9.69583322	-4.398612483	7.124560868	-3.717481949
4	8.614874502	-0.506662011	7.614063240	-1.012884780
5	8.627746593	1.453414524	8.218607769	0.573255265
6	12.80119908	-0.409735355	11.293252246	-1.006380428

Table 3 (Continued): Results of dimensionality reduction

7	7.669284609	5.382157394	8.311781842	4.211182276
8	12.45595285	3.164454954	12.072226425	1.885776380
9	18.09632386	0.608083687	15.717628853	0.678827609
10	19.57818821	1.099244942	17.412487452	0.588299965

Attribute-groups $[X_1, Y_1]$ and $[X_2, Y_2]$ denote respectively the mapping of the original data and the simplified data in two-dimensional space through DMA algorithm. In order to analysis the result more precisely, the distribution of the two data sets are given in Figure 1.

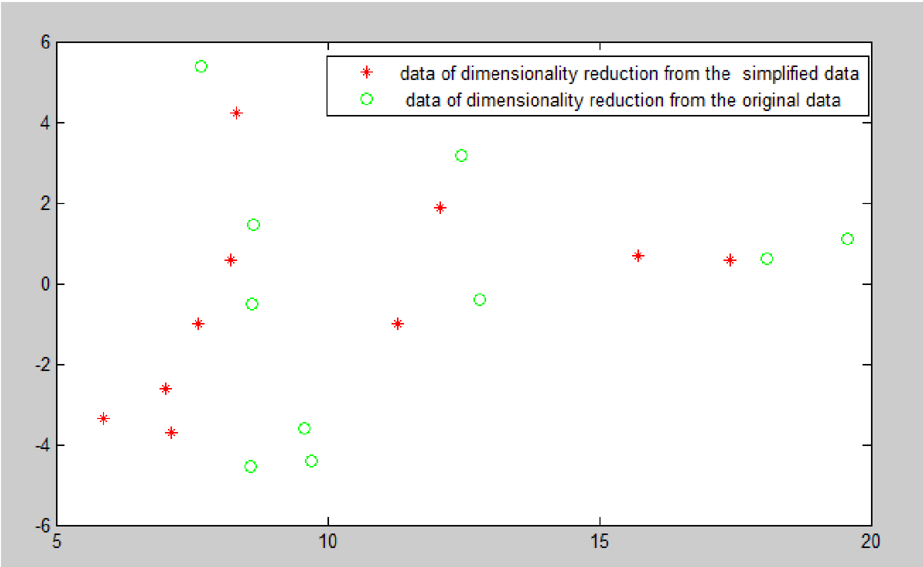

Figure 1 The distribution of the two data sets

In Figure 1, the distributions of the two data sets are almost the same, MDA algorithm can better keep the structure relationships between the original high-dimensional data points accurately [6], moreover, the structural relationship between the simplified data points is very close to the structural relationship between the original. So the improved MDA algorithm is an effective dimensionality reduction method.

4 Conclusions

By using the points of low-dimensional space to represent the high dimensional data to keep the inner structure of the original data in the process of the dimensionality reduction. The paper addresses the improved MDA by combining the attribute reduction of bivariate mutual information with MDA algorithm, we found that the distribution of the results of the improved algorithm and the original were very similar, the improved MDA is faster in disposing high dimensional data with more redundant attributes, because the attribute reduction of two-dimensional mutual information simplifies the data by scanning the database and few operations of multiplication. Currently, most of the algorithms in dimensionality reduction cannot be applied very well for the limitations of the complexity, besides, attribute reduction techniques have been applied to training, classification, and machine learning field in reducing the complexity effectively. Combinations of attribute reduction and dimension reduction are useful in improving dimensionality reduction rate, However, how to better improve the precision and the speed of attribute reduction are the focus of future research.

Acknowledgements. This work was supported by Colleges and universities in Henan province experimental teaching demonstration center construction projects - mathematics and information science experiment center, and also supported by Program for Science & Technology Innovation Talents in Universities of Henan Province (15HASTIT046), Science and technology project of Henan Province (152102110095) and Key scientific research project of Henan Province universities (15A570008).

References

[1] I.T. Jolliffe, *Principal Component Analysis*, Springer-Verlag, New York, 1986. https://doi.org/10.1007/978-1-4757-1904-8

[2] R.A. Fisher, The statistical utilization of multiple measurements, *Annals of Eugenics*, **8** (1938), 376-386. https://doi.org/10.1111/j.1469-1809.1938.tb02189.x

[3] S.T. Roweis and L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, *Science*, **290** (2000), 2323-2326. https://doi.org/10.1126/science.290.5500.2323

[4] X.T. Wu and D.Q. Yan, Analysis and Research on Method of Data Dimensionality Reduction, *Application Research of Computers*, **26** (2009), 2833-2835.

[5] Y. Wu, *Research on Attribute Reduction Methods based on Information Emtropy*, Changchun University of Science and Technology, Changchun, 2004.

[6] J.X. He, J.H. Yan and Y.Y. Fan, Dimensionality Reduction of Large Volu-

mes of Data Analysis, *Aerospace Electronic Warfare*, **30** (2014), 58-59.

Received: November 1, 2016; Published: January 19, 2017