ORIGINAL RESEARCH PAPER

Volume-7 | Issue-11 | November-2018 | PRINT ISSN No 2277 - 8179

INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH

FEATURE EXTRACTION BASED ON MULTI-OBJECTIVE LOCALLY LINEAR EMBEDDING

Computer Science

Han Baojin

Tianjin Polytechnic University School of Computer Science & Software Engineering, Tianjin 300387

ABSTRACT

Since most of the previous data feature extraction methods use continuous feature vectors to complete feature extraction, the biggest disadvantage of this method is that after dimensionality reduction, the data does not achieve the desired dimension or even greater than the previous dimension, and thus does not actually complete the function of dimensionality reduction. The multi-objective (moea / d) [1] method is combined with the local linear embedding method, which makes the dimensionality reduction of the data take into account the actual dimensionality reduction of the data while taking into account the classification accuracy. Theoretical and experimental results show that this algorithm can effectively reduce data dimension and improve classification accuracy.

KEYWORDS

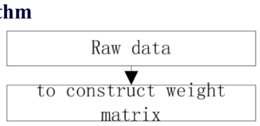
Data dimensionality reduction, feature extraction, moea / d

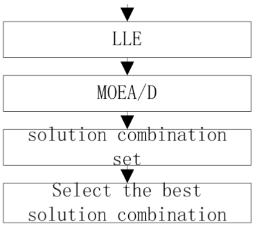
INTRODUCTION

The emergence of large-scale data has caused the formation of data disasters, so that people have fear of existing data, and make the real use of data can not be expanded, which has caused a large amount of waste of data resources, how to carry out data dimensionality reduction is an important task. Traditional methods, such as PCA[2] and ISOMAP[3], can effectively extract important feature data in global dimensionality reduction. However, in data dimension, there is always ambiguity that does not clearly set a limited dimension or give a range of values, which results in the selection of how many dimension data is appropriate. One difficulty is to solve the problem of feature extraction[4] only by the priority of eigenvalues. Sometimes the opposite will happen. The existence of noisy data makes the dimensionality reduction of data a key processing difficulty for these eigenvectors. Because of the influence of noisy data, the effect of dimensionality reduction[5] of data is unsatisfactory. Raw data is very important to extract feature data. In many previous literatures, there is a feature selection for the original data. But because the operation of the original data may affect the integrity of the later feature extraction, it is necessary to remove noise from the extracted feature extraction. In order to maintain the integrity of data features, local linear embedding is a better method of feature extraction, because it is a local, and the amount of calculation is much less than the global, so it has been applied in many feature extraction, such as face recognition system, voice extraction system. Because it may produce deviation in approximate close-range calculation, so it is particularly sensitive to the selection of nearest neighbor points. Therefore, this paper carries out many experiments on the size of nearest neighbor points (K value) in the preliminary processing to reduce the error caused by the nearest neighbor similarity. Compared with ISOMAP, this method of global selection is more sensitive. In order to make the data set too centralized, resulting in the discrete point of the separate data impact on the classification effect, this paper continues to use, previously used close point, close point weight relative reduction of the weight of the remote point is relatively larger, the most popular explanation is to close the distance, scattered close distance, the purpose of doing so is to make The data can be evenly dispersed among the points set as far as possible. Because the function of PCA in data processing is to maximize the difference of data, this method is very good for data with greater difference, but in this paper, clustering[6] is the main verification method, so we need to find the approximate data aggregation data set, relative to LE approximate data aggregation. Ideas, we need more approximate data, the purpose of doing this is to reduce the deviation between the data representation, so that it can be reduced from the source, many data because the single point similarity set is not enough, resulting in poor classification effect after data extraction.

specific algorithm description:

TABLE – 1 Algorithm





The algorithm first chooses the nearest neighbor K, then constructs the matrix weights, and then calculates the corresponding eigenvectors after the matrix weights are constructed. Finally, the eigenvector matrix is formed, and then the Pareto optimal solution is worked out by using MOEA / D method. Finally, the suitable eigenvector is formed. In this way, not only the classification accuracy is high, but also the optimized data dimension can be maximized, so that the data can be reduced to the optimal solution.

LLE algorithm

Specific description of LLE algorithm

1. calculate the nearest neighbor points of each sample point, that is, find the nearest neighbor set of the sample points; K
2. the local reconstruction weight matrix W of the sample points is calculated to minimize the reconstruction block error.
3. all the sample points are mapped to low dimensional space to achieve dimensionality reduction.

The concrete formula of LLE algorithm is:

a. Formula calculation of W weight obtained.

$$min\ \varepsilon(w) = \sum_{i=1}^{n} | x_i - \sum_{j=1}^{k} w_j^i x_{ij} |^2$$

Among them x_{ij}(j=1,2,…k) is the k nearest neighbors of w_j^i is the weight between x_i and x_{ij}, and the condition $\sum_{j=1}^{k} w_j^i=1$. ε(w)is the loss value.

b. Formula for finding specific eigenvectors:

$$min\ \varepsilon(y) = \sum_{i=1}^{n} \left| y_i - \sum_{j=1}^{k} w_j^i y_{ij} \right|^2$$

ε(y) is the loss function value, y_i is the output vector of x_i, y_{ij}(j=1,2,…k) is the k nearest neighbors of y_i, satisfying the following two conditions $\sum_{j=1}^{n} y_i=0.\frac{1}{n}\sum_{i=1}^{n} y_i y_i^T = I$ Where I is the unit matrix of m×m, where w_j^i can be stored in the sparse matrix w of n*n, where w_j^i is the weight value when x_j is the nearest neighbor of x_i, otherwise $w_j^i = 0$;

The influence of K value on dimension reduction results in LLE algorithm

a. Formula calculation of W weight obtained.

$$min\ \varepsilon(w) = \sum_{i=1}^{n} | x_i - \sum_{j=1}^{k} w_j^i x_{ij} |^2$$

Among them x_{ij}(j=1,2,…k) is the k nearest neighbors of w_j^iAmong them

x_{ij}(j=1,2,…k) is the k nearest neighbors of $\Sigma^{k}_{=1} w^{i}$=1. $\varepsilon(w)$ is the loss value.

b. Formula for finding specific eigenvectors:

$$min\ \varepsilon(y) = \sum_{i=1}^{n} \left| y_i - \sum_{j=1}^{k} w_j^i y_{ij} \right|^2$$

ε(y) is the loss function value, y_i is the output vector of x_i, y_{ij}(j=1,2,…k) is the k nearest neighbors of y_i, satisfying the following two conditions $\sum_{j=1}^{n} y_i$=0. $\frac{1}{n}\sum_{i=1}^{n} y_i y_i^T = I$. Where I is the unit matrix of m×m, where w_j^i can be stored in the sparse matrix w of n*n, where w_j^i is the weight value when x_j is the nearest neighbor of x_i, otherwise $w_j^i = 0$;

The influence of K value on dimension reduction results in LLE algorithm

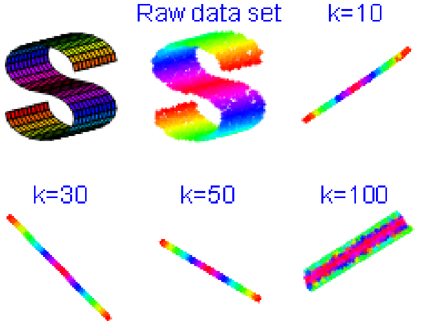


Figure 1

It can be seen from the graph that when k = 10, it shows that the projection of the data is not uniform after dimensionality reduction, so the effect on the data after dimensionality reduction is that it can not objectively reflect the real relative structure of the data, but increases the recognition ability of the data. When k = 50, it can be seen that the internal structure of the projected data has been deformed, so that the internal spatial structure of the original data can not be reflected, and some of the projected data has been significantly deformed. From k = 100, we can see that when k is too large, the relationship between the data is more blurred, thus forming a collapse, so that the data set should be issued out of a certain aspect of the data characteristics. When the k = 30 data projection is clearer, the relationship between the data is more clear, thus forming a better reflection of the projection of the data before the relationship between the data.

MOEA/D
algorithm description
Because the algorithm is based on the decomposition of tasks, there are a lot of decomposition methods in the previous data. This paper mainly runs the Chebyshev aggregation method to decompose tasks. The decomposition of tasks is better and the relative relationship is clearer. A large number of detailed steps can be found in reference []

optimization objectives
This article mainly optimizes the goal for classification accuracy and data dimension, because of the mutual restriction between classification accuracy and data dimension, to a great extent reflects this point, for example, when the data dimension increases, classification accuracy may show an upward state, but when the dimension decreases, on the contrary, how reasonable. The relationship between them has became a difficult point.

In this paper, the reduced dimension data set is used to generate vector combinations in the selection of MOEA/D. After these feature vectors are generated, K-means is used as clustering method to classify the selected feature vectors, and the classification accuracy function is set up.

ACC=$\frac{1}{n}\sum_{i=1}^{n} \mathrm{P}(\boldsymbol{k_i}, \boldsymbol{map(m_i)})$

Among them, k_i and m_i are the class labels which are regenerated after feature selection and those actually labeled in the original data set. Map () is the optimal mapping function, and Hungarian algorithm (Hungarian algorithm) is used to match the class labels generated after feature selection and the actual labels in the original data set Map () is the optimal mapping function, and Hungarian algorithm (Hungarian algorithm) is used to match the class labels generated after feature selection and the actual labels in the original data set. P ***(k_i, m_i)*** is the indicator function, P ***(k_i, m_i)*** is the indicator function, when ***k_i, m_i*** is 1, when **$ck_i \neq m_i$** is 0, ACC is the classification accuracy.

Objective function: min ***g=α_1*ACC+α_2*H (α_1+α_2=1)*** ***α_1*** and ***α_2*** are random production numbers, ACC is correct classification, H is the number of data dimensions, G is the minimum value of data.

experimental part
In this paper, MOEA/D is used, the mutation probability is 0.3, the crossover probability is 0.8, the population size is 20, the coding length is its characteristic dimension, and the maximum number of iterations is 30. Leave the two decimal part. The LLE algorithm is used to extract features before the experiment, and then the extracted feature set is put into the MOEA / D algorithm. In the experiment, nine data sets are selected from the UCI database, and their attributes are listed in Table 1 below. By choosing the classical algorithm to make a comparison, which the comparison algorithm KPCA [6], LE [7], ISOMAP, LTSA [9], MDS [10], PCA.

Data set information:
Table 2 (data set, sample number, attribute dimension, positive class number, negative class number)

data set	Total	Attribute dimension	Positive class number	Negative class number
brest	277	9	81	196
german	1000	24	300	700
heart	270	13	120	150
ionospher	351	34	126	225
liver	345	6	145	200
Iris	270	14	151	119
sonar	208	60	111	97
vote	435	16	267	168
wpbc	569	30	357	212

Table 3 (Data Set ISOMAP, KPCA, LE, Algorithms, LTSA, MDS, PCA Notes: Under Data Set, For Classification Correctness (Average Dimension)

Algorithm / data set	breast	german	heart	ionosphere
ISOMAP	0.69 (5.16)	0.62 (11.9)	0.56 -7	0.67 -16.66
KPCA	0.68 (6.1)	0.6414 -12.57	0.67 (2.56)	0.63 -14.87
LE	0.68 (5.6)	0.69 (13.97)	0.63 -7	0.69 -17.03
? ? ? ?	0.68 (4 .66	0.70 (13.03)	0.67 (7.3)	0.75 -17.37
LTSA	0.698 3 (5)	0.65 (12.63)	0.71 (6.33)	0.72 -18
MDS	0.69 (5.2)	0.67 (12.67)	0.59 (6.97)	0.68 (18.7)
PCA	0.63 (4.26)	0.61 (12.8)	0.58 -8.36	0.71 -17.96

From the experiment, we can see that the proposed algorithm is better than most algorithms in average performance, but it does not perform well in some data sets, such as heart set and LTSA algorithm, which shows that this data set may have a good performance in sorting, and its data set is better. Plus is suitable for sorting algorithm processing. The performance of MDS algorithm in Sonar data sets is remarkable, which shows that multiple choices are helpful to the improvement of the algorithm. The algorithm should be improved greatly in multiple choices, while the global algorithm is relatively trivial in these data sets.

Table 4(Data Set ISOMAP, KPCA, LE, Algorithms, LTSA, MDS, PCA Notes: Under Data Set, For Classification Correctness (Average Dimension)

Algorithm / data set	liver	Iris	Sonar	vote	wpbc
ISOMAP	0.55 (3.5)	0.59 (7.03)	0.74 (27.3)	0.78 (5.43)	0.59 (18.33)
KPCA	0.57 (3.23)	0.68 (7.16)	0.53 (27.2)	0.61 (4.77)	0.55 (15.03)
LE	0.57 (3.53)	0.628 -7.3	0.65 (25.3)	0.75 (4.83	0.73 (16.03)
? ? ? ?	0.60 (3.53)	0.71 (7.77)	0.66 (29.4)	0.77 (4.67	0.75 (16.53)
LTSA	0.57 (3.73)	0.71 (6.63)	0.55 -27.1	0.63 -4.7	0.74 (15.57)
MDS	0.54 (3.4)	0.59 (5.57)	0.83 (32.3)	0.79 -4.8	0.62 -15.26
PCA	0.55 -4.23	0.59 -8.06	0.76 -28.4	0.74 -5.4	0.61 -14.03

CONCLUSIONS

We can see that the effect of LLE algorithm on some data sets is obvious. Especially for some small sample data sets, but the algorithm still has many shortcomings, such as, for processing similar data sets and nonlinear data sets, these aspects should be strengthened learning, and for data noise anti-jamming ability to be strengthened, for data-intensive data, this should be added more. Reinforcement learning.

REFERENCES:

[1] Harada K, Hiwa S, Hiroyasu T. Adaptive weight vector assignment method for MOEA/D[C]// Computational Intelligence. IEEE, 2018:1-9.

[2] Callegari C, Gazzarrini L, Giordano S, et al. Improving PCA based anomaly detection by using multiple time scale analysis and Kullback–Leibler divergence[J]. International Journal of Communication Systems, 2015, 27(10):1731-1751.

[3] Singh K P, Bhai R, Mishra V, et al. Localization in wireless sensor network using LLE-ISOMAP algorithm[C]// TENCON 2017 - 2017 IEEE Region 10 Conference. IEEE, 2017:393-397.

[4] Zhang Y, Lu L. Affine invariant feature extraction based on the shape of local support region[C]// Pattern Recognition and Computer Vision. 2018:3.

[5] Swarna M, Sowmya V, Soman K P. Effect of Dimensionality Reduction on Sparsity Based Hyperspectral Unmixing[C]// International Conference on Soft Computing and Pattern Recognition. Springer, Cham, 2016:429-439.

[6] Wu X, Zhu J, Wu B, et al. Discrimination of tea varieties using FTIR spectroscopy and allied Gustafson-Kessel clustering[J]. Computers & Electronics in Agriculture, 2018, 147:64-69.