IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

Generalized PCA Method and Its Application in Uncertainty Reasoning

BIN WU[1], XIAO YI[1*], AND DONG NING ZHAO[2*]
[1] Naval Aviation University, Yantai 264001 CHINA
[2] Shenzhen Vetose Technology Co. Ltd., Shenzhen 0755 CHINA

*Corresponding author: XIAO YI (e-mail: yxgx@sohu.com), DONG NING ZHAO(e-mail: 582101@qq.com).

ABSTRACT Principal Component Analysis (PCA) is an important mathematical dimension reduction method. In the process of uncertain reasoning, as the elements in the recognition framework increase, the evidence dimension increases exponentially, The computation required also increases exponentially, which greatly affects the application of uncertainty theory in practical engineering. To solve this problem, the PCA algorithm is used to reduce the dimension of evidence in uncertain reasoning. However, in evidence theory, the focal elements in evidence are not completely independent, which is the essential difference between evidence theory and probability theory. This is also the advantage of evidence theory in dealing with uncertain data. Due to the PCA algorithm cannot be used to reduce the dimension of evidence directly. This paper proposes a generalized PCA method and gives strict mathematical proof. The traditional PCA algorithm is a special case of the generalized PCA algorithm proposed in this paper. Finally, the application of a generalized PCA algorithm in uncertainty reasoning is given in this paper, with example results showing that the generalized PCA proposed can greatly reduce the computation required and obtain good evidence combination effect.

INDEX TERMS Generalized PCA, Evidence theory, Data dimensionality reduction.

I. INTRODUCTION

With the advancement of sensor development technology, various types of sensors are applied to achieve accurate determination of information such as target categories and states. However, the use of many sensors has produced problems such as uncertainty, conflict, and redundancy of information. Evidence theory[1] has a good effect as a solution to uncertain and conflict information, and it has been widely used in many fields, such as expert system, multi-source information fusion, target recognition, and military command system. However, with the use of many sensors and the improvement of the target feature recognition database, a complete detection will produce a large amount of data, which creates high demands on the processing capacity of the system. Accurate and fast processing of this information in the battlefield environment will bring great advantages to our operations. Therefore, how to deal with key information and abnormal information in massive data has potential significance.

In the expert system, the target discrimination is largely based on the feature information in the target library, and the target library has many categories so that the output data will show multi-dimensional characteristics. Most of the data belong to low-confidence evidence, but the key information and abnormal information are often hidden in them. Merging two or two corrections based on evidence theory does not show a time advantage in the large data environment. Therefore, this paper aims to study a method that can quickly screen abnormal information in high-dimensional data, providing convenience for evidence correction and fusion.

In many application fields, such as pattern recognition, semantic classification, text classification, and so on, it is usually high-dimensional data. In this case, dimension reduction is an effective method to process these data, and the most classical method is Principal Component Analysis (PCA). How to apply the PCA dimension reduction method to evidence theory is the focus of our research.

II. RELATED WORK

The PCA algorithm is mainly applied to image processing. Lkhagvdor *et al.*[2] proposed a hash tree PCA, which ensures that the distribution characteristics of the original data are not lost by sampling similar objects., and the number of objects is reduced by constructing a hash table for each sample set. Gopalakrishna *et al.* [3] proposed a distance-based distance similarity measure and a log-based Gabor-PCA method for low-resolution video, blurred video and fuzzy video sequence of moving targets in complex environments. Zyad [4] and so on applied the PCA and KPCA dimensionality reduction algorithms in the intrusion detection system and compared the advantages and disadvantages of the two types of algorithms in the intrusion detection. Xiao *et al.*[5] proposed a nonlinear PCA network for image classification, this method introduces nonlinear terms into the convolution filter by comparing with the traditional PCA, which ensures that the mapping depends not only on the globally distributed master but also depends on the location of a single entity, this method is applied to the non-linear distribution of data and maintains more original feature information. Nallammal *et al.* [6] analyzed four methods of PCA, linear discriminant analysis, independent component analysis and hidden Markov model in the field of face recognition comprehensively. Guo *et al.* [7] proposed a method based on PCA and support vector machine for power system transmission fault detection. reducing the signal dimension through PCA, and using the fault signal characteristics extracted from the signal to construct the support vector machine network. The identification method distinguishes the fault state and solves the large-scale network problem that cannot be solved by the conventional amplitude and frequency detection methods. Aloke *et al.* [8] discussed the application of PCA and KPCA in the field of high-spectral image dimensionality reduction and compared the two-dimensionality reduction methods in eliminating redundant spectral bands and reducing classification time. Senthilkumar *et al.* [9] extend one-dimensional PCA to two-dimensional, before feature extraction, image matrix does not need to be transformed into the vector, and the image matrix is directly used to construct covariance matrix. Anbang *et al.* [10] proposed a new image registration scheme, which combines 2DPCA and PCA to extract the features of the image set and input it into the feedforward neural network to provide translation, rotation, and scaling parameters. Tian *et al.* [11] proposed a combination of 2DPCA and SRC, which greatly improved the accuracy of image classification, and automatically obtained the optimal weight without parameters. Ma *et al.* [12]proposed a method of representing image based on PCA. Zhu *et al.* [13] designed an extreme learning machine based on PCA and Kernel for sonar image classification.

Besides in image processing, the PCA method is widely used in other fields. Dan *et al.*[14] turned big data into tiny data by K-means and PCA. Raphael *et al.*[15] proposed Kernel Hierarchical PCA for person re-identification. Markus *et al.* [16]proposed a method to solve the inverse problem with PCA. Pan *et al.*[17] proposed a method to solve streaming noisy data with high compression. Shravan *et al.*[18] designed the classification through PCA and One-class SVM to the text document. Marina *et al.*[19] proposed a method for visual speech recognition using PCA and LSTM. Zhang *et al.* [20] proposed a modified PCA-based approach for process monitoring. Bernardo *et al.* [21] proposed a hybrid subspace neural network to solve the low-resolution features in the visual data set by using PCA and the discriminant filter bank in the subspace method. Lou *et al.* [22] proposed a two-step PCA method that can represent the dynamic characteristics of process data and extract time-independent components. Leonard *et al.*[23] proposed the Hebbian PCA algorithm, which simplifies the existing network structure by removing intra-layer weights, reducing the weight required for training essentially. Tanu *et al.*[24] applied PCA to gesture recognition and verified its feasibility. Loisel *et al.* [25] studied the parameter recovery ability of PCA, RPCA, MDP and other methods in the absence of data. Neda *et al.* [26] used k-means and PCA to estimate the gold level in the deposit. Thierry *et al.* [27] applied RPCA to video processing, which utilized more space-time information than image processing. Rafferty *et al.* [28] and others use PCA to realize real-time detection and classification of events in the power system. Zhou *et al.* [29] proposed a new face recognition method, which extracts features and reduces the number of dimensions through PCA and uses logistic regression as a classifier.

Through the application of the above PCA algorithms, it is found that the PCA algorithms achieve good results in dimension reduction. Through analysis, it is not difficult to find that the data of the above analysis are independent of each other. If the PCA method is directly applied to evidence theory, the challenge is how to deal with the existence of attribute-related data, so this paper proposes a generalized PCA algorithm. It can effectively reduce the dimensionality of the data in which the attributes are related.

III. Proof process of PCA method and its existing problems

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2960261, IEEE Access

A. PCA method certification process

In the process of data analysis and processing, some feature information is redundant. By reducing the dimension of the feature data, the amount of data processing can be greatly reduced. The PCA method is a widely used dimension reduction method and has good theoretical support of matrix knowledge. In order to study the advantages and disadvantages of the PCA method better, the derivation process of the PCA method is given below.

In the article, all vectors are column vectors by default, and each data has n features, which are expressed as column vectors:

$$x=\left(x_1,x_2,x_3,\cdots,x_n\right)^T$$

The basic idea of the PCA method is to make the original data rotate in the new space to maximize the variance in the new dimension through orthogonal transformation. The variance represents the importance of information in this dimension., the characteristic dimension with smaller variance can be discarded Within the allowable error range to achieve the purpose of dimension reduction. Therefore, the problem is transformed into finding a set of orthogonal transformations to maximize the variance of the original data after transformation. The following is a modeling and analysis of the problem.

Assuming that the original data has N-dimensional characteristics, there are a total of P samples, forming the sample set,

$$X=\left[x_1,x_2,\cdots,x_p\right]=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Np} \end{bmatrix} \tag{1}$$

each row represents a feature dimension with a total of N features, and each column represents a sample with a total of p samples. X is a $N\times p$ matrix, x_i is a sample, and the dimension is $N\times 1$.

The following transformation matrix is constructed as

$$A=\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_M \end{bmatrix} \tag{2}$$

the dimension of A is $M\times N$, $M\le N$ so that the original feature of the N dimension is reduced to M-dimensional data. a_i represents a transformed vector with a dimension of $1\times N$.

In order to eliminate the influence of the dimension and the mean, the original data needs to be decentralized to make the expectation of the original data is 0, that is,

$$X'=X-\bar{X}=\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}-\begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_N \end{bmatrix} \tag{3}$$

where, $\bar{X}=\frac{1}{p}\sum_{i=1}^{p}x_i$, $\bar{x}_i=\frac{1}{p}\sum_{j=1}^{p}x_{ij}$, through analysis,

$$E\left(X'\right)=E\left(X-\bar{X}\right)=E\left(X\right)-E\left(\bar{X}\right)=0 \tag{4}$$

this achieves the purpose of decentralization.

So the transformed matrix Y is

$$Y=A\left(X-\bar{X}\right)=\begin{bmatrix} a_1\left(x_1-\bar{x}_1\right) & a_1\left(x_2-\bar{x}_2\right) & \cdots & a_1\left(x_p-\bar{x}_p\right) \\ a_2\left(x_1-\bar{x}_1\right) & a_2\left(x_2-\bar{x}_2\right) & \cdots & a_2\left(x_p-\bar{x}_p\right) \\ \vdots & \vdots & \ddots & \vdots \\ a_M\left(x_1-\bar{x}_1\right) & a_M\left(x_2-\bar{x}_2\right) & \cdots & a_M\left(x_p-\bar{x}_p\right) \end{bmatrix} \tag{5}$$

record the transformed $M\times p$ dimensional matrix Y as

$$Y=\begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_M \end{bmatrix}=\begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1p} \\ y_{21} & y_{22} & \cdots & y_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ y_{M1} & y_{M2} & \cdots & y_{Mp} \end{bmatrix} \tag{6}$$

Through the previous analysis, the basic idea of the PCA method is to find a set of orthogonal transforms to maximize the variance of the original data transformation. First, consider the first component y_1. The variance after the change is required to be the largest, and the constraint conditions are as follows:

$$\max\left(\sum_{i=1}^{p}\left(y_{1i}-\bar{y}_1\right)^2\right) \tag{7}$$

where, $\bar{y}_1=\frac{1}{p}\sum_{i=1}^{p}y_{1i}=\frac{1}{p}\sum_{i=1}^{p}\left(a_1\left(x_i-\bar{x}_i\right)\right)=\frac{a_1}{p}\sum_{i=1}^{p}\left(x_i-\bar{x}_i\right)=0$

so the constraint becomes

$$\begin{aligned} &\max\left(\sum_{i=1}^{p}y_{1i}{}^2\right) \\ &=\max\left(\sum_{i=1}^{p}\left(a_1\left(x_i-\bar{x}_i\right)\right)^2\right) \\ &=\max\left(\sum_{i=1}^{p}\left(a_1\left(x_i-\bar{x}_i\right)\left(x_i-\bar{x}_i\right)^T a_1{}^T\right)\right) \\ &=\max\left(a_1\left(\sum_{i=1}^{p}\left(x_i-\bar{x}_i\right)\left(x_i-\bar{x}_i\right)^T\right)a_1{}^T\right) \\ &=\max\left(a_1Ra_1{}^T\right) \end{aligned} \tag{8}$$

where, $R=\sum_{i=1}^{p}\left(x_i-\bar{x}_i\right)\left(x_i-\bar{x}_i\right)^T$ is the standard covariance matrix of the original data and a_1 is the transformation matrix to be solved, the constraint conditions are as follows:

$$a_1a_1{}^T=\left\|a_1\right\|=1 \tag{9}$$

therefore, the Lagrange equation is constructed to obtain

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2960261, IEEE Access

$$L(a_1) = a_1 R a_1^T - \lambda_1 (a_1 a_1^T - 1) \\ = a_1 (R - \lambda_1 I) a_1^T - \lambda_1 \tag{10}$$

where, I is the unit matrix, derivation of Formula (10) we can get

$$\frac{\partial L}{\partial a_1} = 2\left(R a_1^T - \lambda_1 a_1^T\right)^T = 0 \tag{11}$$

$$R a_1^T = \lambda_1 a_1^T \tag{12}$$

so R is the eigenvalue of the standard covariance matrix λ_1, and a_1 is the corresponding eigenvector, brought into the objective function we can get,

$$a_1 R a_1^T = a_1 \lambda_1 a_1^T = \lambda_1 a_1 a_1^T = \lambda_1 \tag{13}$$

therefore, to solve $\max(a_1 R a_1^T)$ is to solve $\max(\lambda_1)$, that is to solve the maximum eigenvalue of the standard covariance matrix R.

In the same way, a_2 can be solved. The difference is that the constraint condition a_2 is,

$$\begin{cases} a_2 a_2^T = 1 \\ a_1 a_2^T = 0 \end{cases} \tag{14}$$

finally, a_2 is also the corresponding feature vector, and the objective function is the feature value λ_2 corresponding to a_2. Since the largest feature value is λ_1, λ_2 can take the second-largest feature value. For the remaining steps, refer to Section IV.

B. Problems of PCA method

The biggest problem with the PCA method is that it cannot handle information with inconsistent degrees of confidence, which is the information with different weights. For example, the evidence theory to be studied below, in the evidence theory, if there are two elements $\{A, B\}$ in the recognition framework, then the focal element is $\{A, B, AB\}$, where $m(AB)$ is related to the elements $m(A)$ and $m(B)$. At this time, PCA method is not suitable. The following is a specific demonstration.

By analyzing the derivation process of PCA method, it is found that the method requires orthogonal transformation of the original data, so in the traditional Euclidean space, the distance between the transformed feature vectors remains unchanged, while the length of the feature vectors themselves remains unchanged, so the method does not consider whether the distance between the vectors before and after transformation is transformed.

However, when there is no correlation between each feature in the original data, there is no problem with this idea, and when the confidence (or weight) of the features in the feature vector is different, the method will have problems.

Take the evidence theory as an example. In the theory of evidence, the distance between evidence is an important research topic. Many experts and scholars put forward different distances of evidence, but there is not a recognized authoritative distance of evidence. The most widely used evidence distance is jousselme [30] evidence distance,

$$d_J(m_1, m_2) = \sqrt{\frac{1}{2}(m_1 - m_2)^T D_J (m_1 - m_2)} \tag{15}$$

This requires that the distance between the evidence before and after the transformation is the same, that is,

$$d_J(X'_1, X'_2) = d_J(AX'_1, AX'_2) \approx d_E(Y_1, Y_2) \tag{16}$$

where, $d_E(Y_1, Y_2) = \langle Y_1, Y_2 \rangle$.

That is to say, the distance of the original data before transformation should be equal to the distance of the transformed Euclidean within the allowable error range. However, it is obvious that the traditional PCA method does not meet this condition. In order to discuss the preciseness, the following is an example analysis.

In order to enhance the persuasiveness of the example, in addition to the jousselme evidence distance, we also add several evidence distances, including:

Euclidean distance: It takes BPA as the vector and directly calculates the distance between vectors. The definition of Euclidean distance is given below. Considering the basic probability assignment m_1, m_2 under the same recognition framework, and the function is given as follows

$$d_E = \sqrt{\frac{1}{2}\left(\sum_{A \in \Theta} (m_1(A) - m_2(A))^2\right)} \tag{17}$$

Tessem distance: Tessem is similar to Euclidean distance, but Tessem distance considers the distance between probability transformations. The probability transformation used is the most classic probability transformation.

$$BetP_m(A) = \sum_{B \in \Theta} \frac{|A \cap B|}{|B|} m(B) \tag{18}$$

The Tessem distance formula is given below,

$$d_T(m_1, m_2) = \max\{|BetP_1(A) - BetP_2(A)|\} \tag{19}$$

It can be seen that Tessem takes the probability conversion as the basic vector and calculates the infinite norm between the vectors.

Bhattacharyya distance: The product of the basic probability assignments considered by the Bhattacharyya distance. The product of different focal elements BPA is regarded as the distance between two pieces of evidence. The specific definition is as follows,

$$d_B(m_1, m_2) = \sqrt{1 - \sum_{A \in 2^\theta} \sqrt{m_1(A) m_2(A)}} \tag{20}$$

Distance based on membership function: The membership function is a concept in fuzzy mathematical theory. The distance given by the membership function is defined as follows,

$$d_F(m_1, m_2) = 1 - \frac{\sum_{A \in \theta} (\mu_1(A) \wedge \mu_2(A))}{\sum_{A \in \theta} (\mu_1(A) \vee \mu_2(A))} \tag{21}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2960261, IEEE Access

where, $\wedge, \vee$ represents disjunction and conjunction respectively, and function $\mu(\cdot)$ can be replaced with plausibility function or belief function.

C. Example analysis of traditional PCA method

Considering that the target recognition framework is $\{A, \ B, \ C\}$, the existing six sources give six pieces of evidence, as shown in Table 1.

Table.1 Source representation

	A	B	C	AB	AC	BC	ABC
1	0.55	0.1	0.1	0.1	0.1	0.05	0
2	0.6	0.1	0.05	0.1	0.1	0.05	0
3	0	0.65	0	0.15	0	0.2	0
4	0.5	0.1	0.2	0.1	0.05	0.05	0
5	0.65	0	0.1	0.1	0.1	0.05	0
6	0.7	0.05	0.05	0.1	0.1	0	0

The jousselme evidence distance between the first piece of evidence and the other 5 pieces of evidence is calculated separately, and the distance after PCA conversion is calculated as shown in Figure. 1.

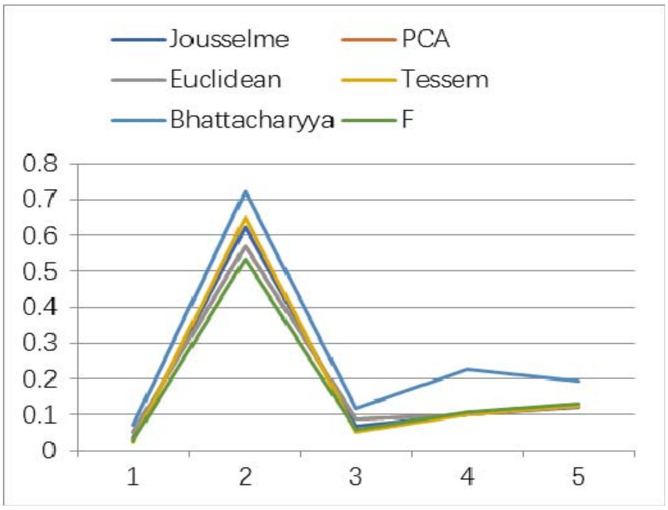

Figure.1 Comparison chart of evidence distance and PCA distance

It can be seen from Fig.1 that there is no positive correlation between the jousselme evidence distance and the distance after PCA conversion. The distance between the two distances is large or small, which is not in line with the actual situation. The reason for this phenomenon is that the calculation of the jousselme evidence distance takes into account the magnitude of each focal element in the evidence, and the PCA method does not consider the correlation between the focal elements in the calculation process, so the two distances are not positively correlated.

IV. Generalized PCA method and its proof process

In evidence theory, with the increase of elements in the identification framework, the calculation amount increases exponentially, so it is urgent to reduce the dimension of evidence. Therefore, this section will present an improved PCA method, the generalized PCA method. The derivation process of the generalized PCA method is given below.

It has been pointed out that the distance between the evidence before and after the transformation is required to be the same, that is,

$$d_J\left(X'_1, X'_2\right) = d_J\left(AX'_1, AX'_2\right) \approx d_E\left(Y_1, Y_2\right) \tag{22}$$

this requires the original data to be modified. And the correction matrix is B, and we can get

$$y_1^r = ABx_1 \tag{23}$$

where, y_1^r is a column of the transformed data Y, which is a sample, and it is required to be satisfied,

$$\begin{aligned}&\left(y_1^r - y_2^r\right)^T \left(y_1^r - y_2^r\right)\\ &= \left(ABx_1 - ABx_2\right)^T \left(ABx_1 - ABx_2\right)\\ &= \left(x_1 - x_2\right)^T B^T B\left(x_1 - x_2\right)\\ &= d_J\left(x_1, x_2\right)\end{aligned} \tag{24}$$

and it satisfies

$$\left(x_1 - x_2\right)^T B^T B\left(x_1 - x_2\right) = \frac{1}{2}\left(x_1 - x_2\right)^T D_J\left(x_1 - x_2\right) \tag{25}$$

So, $B^T B = \frac{1}{2} D_J$, where D_J is a matrix of $2^n \times 2^n$, any element d_{ij} in the matrix is defined as $d_{ij} = \frac{|A_i \cap A_j|}{|A_i \cup A_j|}$, and $|A|$ represents the potential of element A.

At this time, the constraints in PCA reasoning become

$$\begin{aligned}&\max\left(\sum_{i=1}^{p} y_{1i}^{\ 2}\right)\\ &= \max\left(\sum_{i=1}^{p}\left(a_1 B\left(x_i - \bar{x}_i\right)\right)^2\right)\\ &= \max\left(\sum_{i=1}^{p}\left(a_1 B\left(x_i - \bar{x}_i\right)\left(x_i - \bar{x}_i\right)^T B^T a_1^{\ T}\right)\right)\\ &= \max\left(a_1\left(\sum_{i=1}^{p} B\left(x_i - \bar{x}_i\right)\left(x_i - \bar{x}_i\right)^T B^T\right)a_1^{\ T}\right)\\ &= \max\left(a_1 B\left(\sum_{i=1}^{p}\left(x_i - \bar{x}_i\right)\left(x_i - \bar{x}_i\right)^T\right)B^T a_1^{\ T}\right)\\ &= \max\left(a_1 R' a_1^{\ T}\right)\\ &= \max\left(a_1 BRB^T a_1^{\ T}\right)\end{aligned} \tag{26}$$

where $R' = \left(\sum_{i=1}^{p} B\left(x_i - \bar{x}_i\right)\left(x_i - \bar{x}_i\right)^T B^T\right)$ is the quasi-standard covariance matrix of the original data, D_J is the weight coefficient in the jousselme evidence distance.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2960261, IEEE Access

More generally, when $R' = \sum_{i=1}^{p}\left(x_i - \overline{x}_i\right)\left(x_i - \overline{x}_i\right)^T$ is taken, where B is the weight coefficient of evidence distance, and when B is the identity matrix, the generalized PCA method is the traditional PCA algorithm. In order to solve the problem, and also considering that there is no recognized evidence distance in academia, $B = \sqrt{\frac{1}{2}D_J}$ is taken in this paper.

a_1 is the transformation matrix to be solved, and the constraint condition is,

$$a_1 a_1^T = \|a_1\| = 1 \tag{27}$$

In the same way, the Lagrange equation is constructed to obtain

$$\begin{aligned} L\left(a_1\right) &= a_1 R' a_1^T - \lambda_1\left(a_1 a_1^T - 1\right) \\ &= a_1\left(R' - \lambda_1 I\right)a_1^T - \lambda_1 \end{aligned} \tag{28}$$

derivation of Formula (23) we can get

$$\frac{\partial L}{\partial a_1} = 2\left(R' a_1^T - \lambda_1 a_1^T\right)^T = 0 \tag{29}$$

$$R' a_1^T = \lambda_1 a_1^T \tag{30}$$

Similarly, it is obtained that λ_1 is the eigenvalue of the standard covariance matrix R and a_1 is the corresponding eigenvector, which is brought into the objective function,

$$a_1 R' a_1^T = a_1 \lambda_1 a_1^T = \lambda_1 a_1 a_1^T = \lambda_1 \tag{31}$$

So, solving $\max\left(a_1 R' a_1^T\right)$ is to solve $\max\left(\lambda_1\right)$, that is solving the maximum eigenvalue R'.

In the same way, a_2 can be solved. The difference is that the constraint condition a_2 is,

$$\begin{cases} a_2 a_2^T = 1 \\ a_1 a_2^T = 0 \end{cases} \tag{32}$$

finally, a_2 is also the corresponding feature vector, and the objective function is the feature value λ_2 corresponding to a_2. Since the largest feature value is λ_1, λ_2 can take the second-largest feature value.

Through the above reasoning, the general steps of Generalized PCA algorithm can be obtained:

Step1: Solving the quasi covariance matrix $R' = \left(\sum_{i=1}^{p} B\left(x_i - \overline{x}_i\right)\left(x_i - \overline{x}_i\right)^T B^T\right)$;

Step2: Finding the eigenvalues of the quasi covariance matrix and arrange them in descending order;

Step3: Normalizing eigenvector;

Step4: Taking out the first few feature vectors for dimensionality reduction;

Step 5: Calculating dimensionality reduction results.

In order to illustrate the superiority of the generalized PCA method better, the following example analysis is given.

Considering that the target recognition framework is $\{A, B, C\}$, six pieces of evidence are given by six sources, as shown in Table 1. Now we calculate the jousselme evidence distance between the first evidence and the other five pieces of evidence, and the distance after the generalized PCA transformation. The jousselme evidence distance has been calculated in the previous example, and the distance after the generalized PCA transformation is calculated below.

Firstly, calculating D_J in the evidence distance of jousselme, as shown in Table 2.

Table.2 the value of D_J

	A	B	AB	C	AC	BC	ABC
A	1.000	0	0.500	0	0.500	0	0.333
B	0	1.000	0.500	0	0	0.500	0.333
AB	0.500	0.500	1.000	0	0.333	0.333	0.667
C	0	0	0	1.000	0.500	0.500	0.333
AC	0.500	0	0.333	0.500	1.000	0.333	0.667
BC	0	0.500	0.333	0.500	0.333	1.000	0.667
ABC	0.333	0.333	0.6667	0.333	0.667	0.667	1.000

Then calculating the class covariance matrix R' and bring it into the formula. The calculation results are shown in Table 3.

Table.3 Distance comparison table

Jousselme	0.0478	0.5675	0.0827	0.0979	0.1177
GPCA	0.0354	0.6232	0.0677	0.1	0.1208
Euclidean	0.05	0.5701	0.0866	0.1	0.1225
Tessem	0.025	0.65	0.05	0.1	0.125
F	0.0278	0.5361	0.0563	0.1081	0.1299

Draw a line graph of the calculation results, as shown in Figure 2,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2960261, IEEE Access

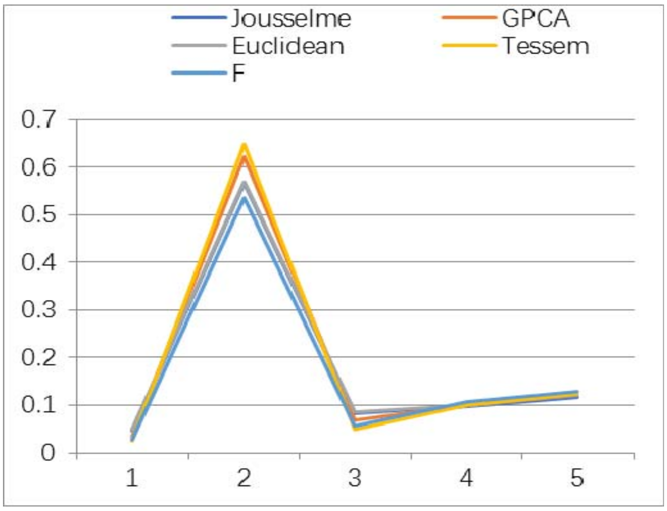

Figure.2 Comparison chart of evidence distances and GPCA distance

By analyzing the table and the calculation result graph, it can be obtained that the generalized PCA method can better preserve the relative distance between the evidence compared with the traditional PCA method so that the dimensionality reduction algorithm can be applied to the evidence theory.

V. Application of generalized PCA method in uncertainty reasoning

Now consider the application of the generalized PCA method in evidence theory. The fusion of information is divided into three levels: the data layer, the feature layer, and the decision layer. This example is based on the decision layer to fuse data. Assuming that there are some abnormal data among the data, that is there are large differences between some data and the other data, the traditional method calculates the distance between the two, and use the sum of the distance between one piece of evidence and all the evidence to represent the distance between this piece of evidence and the whole evidence set. If the distance is large, this piece of evidence is considered to be in conflict. However, the computation is large, and the distance between the two must be calculated once. If the dimension of evidence is large, the computation is bound to be large, which is necessary to reduce the dimension of evidence.

In this example, the generalized PCA is used to reduce the dimension of evidence, assuming that there are seven sources of evidence, and the evidence obtained is

$$m_1(A)=0.55, m_1(B)=0.1, m_1(AB)=0.1, m_1(C)=0.1,$$
$$m_1(AC)=0.1, m_1(BC)=0.05$$

$$m_2(A)=0.6, m_2(B)=0.1, m_2(AB)=0.05, m_2(C)=0.1,$$
$$m_2(AC)=0.1, m_2(BC)=0.05$$

$$m_3(A)=0, m_3(B)=0.65, m_3(AB)=0, m_3(C)=0.15,$$
$$m_3(AC)=0, m_3(BC)=0.2$$

$$m_4(A)=0.5, m_4(B)=0.1, m_4(AB)=0.2, m_4(C)=0.1,$$
$$m_4(AC)=0.05, m_4(BC)=0.05$$

$$m_5(A)=0.65, m_5(B)=0, m_5(AB)=0.1, m_5(C)=0.1,$$
$$m_5(AC)=0.1, m_5(BC)=0.05$$

$$m_6(A)=0.7, m_6(B)=0.05, m_6(AB)=0.05, m_6(C)=0.1,$$
$$m_6(AC)=0.1, m_6(BC)=0$$

$$m_7(A)=0, m_7(B)=0.55, m_7(AB)=0, m_7(C)=0.25,$$
$$m_7(AC)=0.2, m_7(BC)=0$$

According to the steps of the generalized PCA algorithm, the first step is to calculate the covariance S between evidence

$$S=\begin{bmatrix} 0.0899 & -0.0774 & 0.0118 & -0.0143 & 0.0137 & -0.0237 & 0 \\ -0.0774 & 0.0690 & -0.0122 & 0.0118 & -0.0116 & 0.0205 & 0 \\ 0.0118 & -0.0122 & 0.0049 & -0.0024 & 0.0013 & -0.0034 & 0 \\ -0.0143 & 0.0118 & -0.0024 & 0.0032 & -0.0021 & 0.0038 & 0 \\ 0.0137 & -0.0116 & 0.0013 & -0.0021 & 0.0023 & -0.0035 & 0 \\ -0.0237 & 0.0204 & -0.0034 & 0.0038 & -0.0035 & 0.0064 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

then calculating D_J in the jousselme distance, as shown in Table 4,

Table.4 value of D_J

	A	B	AB	C	AC	BC	ABC
A	1.000	0	0.5000	0	0.500	0	0.333
B	0	1.000	0.5000	0	0	0.500	0.333
AB	0.500	0.500	1.000	0	0.333	0.333	0.667
C	0	0	0	1.000	0.500	0.500	0.333
AC	0.500	0	0.333	0.500	1.000	0.333	0.667
BC	0	0.500	0.333	0.500	0.333	1.000	0.667
ABC	0.333	0.333	0.667	0.333	0.667	0.667	1.000

we can get B according to formula $B^T B = \frac{1}{2} D_J$,

$$B=\begin{bmatrix} 0.6552 & -0.0214 & 0.1779 & -0.0214 & 0.1779 & -0.0277 & 0.0762 \\ -0.0214 & 0.6552 & 0.1779 & -0.0214 & -0.0277 & 0.1779 & 0.0762 \\ 0.1779 & 0.1779 & 0.6091 & -0.0277 & 0.0696 & 0.0696 & 0.2352 \\ -0.0214 & -0.0214 & -0.0277 & 0.6552 & 0.1779 & 0.1779 & 0.0762 \\ 0.1779 & -0.0277 & 0.0696 & 0.1779 & 0.6091 & 0.0696 & 0.2352 \\ -0.0277 & 0.1779 & 0.0696 & 0.1779 & 0.0696 & 0.6091 & 0.2352 \\ 0.0762 & 0.0762 & 0.2352 & 0.0762 & 0.2352 & 0.2352 & 0.5627 \end{bmatrix}$$

then we can get quasi covariance matrix R' according to formula $R' = \left(\sum_{i=1}^{p} B\left(x_i - \bar{x}_i\right)\left(x_i - \bar{x}_i\right)^T B^T \right)$,

$$R'=\begin{bmatrix} 0.0445 & -0.0369 & 0.0065 & -0.0081 & 0.0112 & -0.0171 & 0.0003 \\ -0.0369 & 0.0312 & -0.0054 & 0.0062 & -0.0100 & 0.0147 & -0.0003 \\ 0.0065 & -0.0054 & 0.0018 & -0.0016 & 0.0011 & -0.0023 & 0.0002 \\ -0.0081 & 0.0062 & -0.0016 & 0.0020 & -0.0012 & 0.0025 & -0.0001 \\ 0.0112 & -0.0100 & 0.0011 & -0.0012 & 0.0041 & -0.0052 & 0.0000 \\ -0.0171 & 0.0147 & -0.0023 & 0.0025 & -0.0052 & 0.0073 & -0.0001 \\ 0.0003 & -0.0003 & 0.0002 & -0.0001 & 0.0000 & -0.0001 & 0.0001 \end{bmatrix}$$

The next step is the same as the classic PCA algorithm, which solves the eigenvalues and eigenvectors of the class covariance, and then selects the specific dimension according to the size of the principal component. Through calculation, the result after the final generalized PCA conversion is obtained. Here, the first four main dimensions of the generalized PCA conversion result are selected (according to the need of precision, the appropriate dimension is selected), the result we can get is shown in table.5

Table.5 result of 4 dimensions

-0.1685	0.0352	-0.0003	-0.0360
-0.1989	0.0303	0.0599	-0.0341
0.6275	0.0110	0.0182	0.0609
-0.1340	0.0896	-0.0975	-0.0280
-0.2992	0.0494	0.0182	-0.0756
-0.3137	0.0330	0.0765	-0.0092
0.4868	-0.2485	-0.0750	0.1220

in the table, each row represents one piece of evidence, and four columns indicate that the original feature dimension is reduced to four dimensions. The distance between the two pieces of evidence is calculated below.

Table.6 distance of two evidence

	m_1	m_2	m_3	m_4	m_5	m_6	m_7
m_1	0	0.048	0.568	0.0837	0.0989	0.117	0.520
m_2	0.048	0	0.589	0.1286	0.083	0.0839	0.543
m_3	0.567	0.589	0	0.551	0.6629	0.668	0.223
m_4	0.083	0.128	0.551	0	0.149	0.181	0.511
m_5	0.098	0.083	0.663	0.149	0	0.064	0.614
m_6	0.118	0.084	0.669	0.182	0.064	0	0.617
m_7	0.5120	0.543	0.223	0.511	0.614	0.617	0

Summing each column we can get

[1.4334 1.4751 3.2626 1.6037 1.6720 1.7333 3.0284]

Obviously, the 3rd and 7th pieces of evidence have a clear conflict compared with the evidence set, and the two pieces of evidence need to be amended or eliminated directly.

The following is the direct elimination of the conflict evidence, and the comparison results of the evidence combination, as shown in Table 7.

Table.7 Comparison table of dimension reduction fusion results

	without elimination		elimination	
	DS	**PCR6**	**DS**	**PCR6**
$m(A)$	0	0.609	0.9971	0.8930
$m(B)$	0.6296	0.259	0.0008	0.0210
$m(AB)$	0	0.0252	0.0000	0.0270
$m(C)$	0.3704	0.0535	0.0020	0.0331
$m(AC)$	0	0.0173	0	0.0194
$m(BC)$	0	0.0358	0	0.0066
$m(ABC)$	0	0	0	0

By analyzing the data in the table, we can get the following conclusions,

1. The generalized PCA method can well represent the conflict degree between pieces of evidence;
2. The generalized PCA method can be well applied in evidence theory;
3. DS evidence theory cannot deal with high conflict evidence well;
4. The generalized PCA method is an efficient dimensionality reduction algorithm. It not only retains the advantages of the classical PCA method in reducing computation but also can be applied in complex fields such as evidence theory. The classical PCA method is a special case of the generalized PCA method.

VI. SUMMARY

In this paper, a generalized PCA algorithm is proposed to solve data dimensionality reduction, and the classical PCA algorithm is strictly proved to be a special case. The generalized PCA algorithm is applied to the uncertainty evidence reasoning to solve the problem that the original PCA algorithm cannot deal with related attributes. When there is high-dimensional data, the algorithm proposed in this paper can well reduce the dimension, and then filter out the abnormal evidence through the reduced dimension data, and then the evidence can be corrected. The corrected evidence can be combined with the ideal result.

REFERENCES

[1] Zadeh, Lotfi A. "A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination." *Ai Magazine*, vol. 7, no. 2, pp. 85-90, 1986.

[2] Battulga, Lkhagvadorj, et al. "Hash-tree PCA: accelerating PCA with hash-based grouping." *The Journal of Supercomputing* pp. 1-17, 2019.

[3] Gopalakrishna, M. T, M. Ravishankar , and D. R. Rameshbabu . "Multiple Moving Object Recognitions in video based on Log Gabor-PCA Approach." *Computer Science*, vol.235, pp.93-100, 2014.

[4] Elkhadir, Zyad, K. Chougdali , and M. Benattou . *"Intrusion detection system using PCA and kernel PCA methods." Proceedings of the Mediterranean Conference on Information & Communication Technologies 2016*.

[5] Zhang, Xiao, and Youtian Du. "Nonlinear PCA Network for Image Classification." *pacific rim conference on multimedia*, pp. 449-457, 2016.

[6] Nallammal, N., and V. Radha. "Performance evaluation of face recognition based on PCA, LDA, ICA and hidden markov model." *international conference on data engineering*, pp.96-100, 2010.

[7] Guo, Yuanjun, Kang Li, and Xueqin Liu. "Fault Diagnosis for Power System Transmission Line Based on PCA and SVMs." *international conference on intelligent computing for sustainable energy and environment*, pp. 524-532, 2012.

[8] Datta, Aloke S. Ghosh , and A. Ghosh . *PCA, Kernel PCA and Dimensionality Reduction in Hyperspectral Images. Advances in Principal Component Analysis*, pp.19-46, 2018.

[9] Senthilkumar, R., and R. K. Gnanamurthy. "A Comparative Study of 2D PCA Face Recognition Method with Other Statistically Based Face Recognition Methods." *Journal of The Institution of Engineers*, vol. 97, no. 3, pp. 425-430, 2016.

[10] Xu, Anbang, Xin Jin, and Ping Guo. "Two-Dimensional PCA combined with PCA for neural network based image registration." *international conference on natural computation*, pp. 696-705, 2006.

[11] Tian C, Zhang Q, Zhang J, et al. "2D-PCA Representation and Sparse Representation for Image Recognition. " *Journal of Computational & Theoretical Nanoscience*, vol. 14, no. 1, p. 834, 2017.

[12] Zhongli, Ma, et al. "Image representation based PCA feature for image classification." *2017 IEEE International Conference on Mechatronics and Automation (ICMA)* IEEE, pp. 1121-1125, 2017.

[13] Zhu, Mingcui, et al. "PCA and Kernel-based extreme learning machine for side-scan sonar image classification." *ieee international underwater technology symposium*, pp. 1-4, 2017.

[14] Feldman, Dan, Melanie Schmidt, and Christian Sohler. "Turning big data into tiny data: constant-size coresets for k -means, PCA and projective clustering." *symposium on discrete algorithms*, pp. 1434-1453, 2013.

[15] Prates, Raphael Felipe, and William Robson Schwartz. "Kernel Hierarchical PCA for person re-identification." *international conference on pattern recognition*, pp. 2091-2096, 2016.

[16] Neumayer, Markus, et al. "PCA based state reduction for inverse problems using prior information." *Compel-the International Journal for Computation and Mathematics in Electrical and Electronic Engineering*, vol. 36, no. 5, pp. 1430-1441, 2017.

[17] Pan, Pingbo, et al. "Online compressed robust PCA." *international joint conference on neural network*, pp. 1041-1048, 2017.

[18] Kumar B S, Ravi V. "Text Document Classification with PCA and One-Class SVM" *Proceedings of the 5th International Conference on Frontiers in Intelligent Computing: Theory and Applications. Advances in Intelligent Systems and Computing*, vol 515, 2017.

[19] Zimmermann, Marina, et al. "Visual Speech Recognition Using PCA Networks and LSTMs in a Tandem GMM-HMM System." *asian conference on computer vision*, pp. 264-276, 2016.

[20] Zhang, Jingxin, Hao Chen, and Pinlong Cai. "A modified PCA-based approach for process monitoring." *chinese control and decision conference*, pp. 3011-3016, 2017.

[21] Gatto, Bernardo B., Lincon S. Souza, and Eulanda M Dos Santos. "A deep network model based on subspaces: A novel approach for image classification." *international conference on machine vision*, pp. 436-439, 2017.

[22] Lou, Zhijiang , J. Tuo , and Y. Wang . "Two-step principal component analysis for dynamic processes." *2017 6th International Symposium on Advanced Control of Industrial Processes (AdCONIP)* IEEE, pp.73-77, 2017.

[23] Johard, Leonard, et al. "Self-adaptive Node-Based PCA Encodings." *international conference on software engineering*, pp. 1-9, 2016.

[24] Srivastava, Tanu, et al. "Feasibility of Principal Component Analysis in hand gesture recognition system." *arXiv: Computer Vision and Pattern Recognition* (2017).

[25] Loisel, Sebastien, and Yoshio Takane. "Comparisons among several methods for handling missing data in principal component analysis (PCA)." *Advanced Data Analysis and Classification*, vol. 13, no. 2, pp. 495-518, 2019.

[26] Neda Mahvash Mohammadi, Ardeshir Hezarkhani, and Abbas Maghsoudi. "Application of K-means and PCA approaches to estimation of gold grade in Khooni district(central Iran)." *Acta Geochimica*, vol.37, no.1, pp.104-114, 2018.

[27] Thierry, Bouwmans , et al. "On the Applications of Robust PCA in Image and Video Processing." *Proceedings of the IEEE*, vol. 106, no.8, pp. 1427-1457, 2018.

[28] Rafferty, Mark , et al. "Real-Time Multiple Event Detection and Classification Using Moving Window PCA." *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2537-2548, 2016.

[29] Zhou, Changjun , et al. "Face recognition based on PCA and logistic regression analysis." *Optik - International Journal for Light and Electron Optics*, vol. 125, no. 20, pp. 5916-5919, 2014.

[30] Jousselme, Anne Laure , D. Grenier , and éloi Bossé. "A new distance between two bodies of evidence." *Information Fusion*, vol. 2, no. 2, pp. 91-101, 2001.

BIN WU received his B.S. degree in information and communication engineering from Naval Aeronautical Engineering Institute, Yantai, China, in 2013 and M.S. degree in Electronical Information and Engineering from Naval Aeronautical Engineering institute Yantai, China, in 2016, He is currently pursuing the Ph.D. degree in information and communication engineering at Naval Aviation University Yantai, China

His research interests in the area of wireless sensor networks and evidence theory, including tracking, security, and decision making.

XIAO YI received his B.S. degree in information and communication engineering from SHANDONG University, Jinan, China, in 1997 and M.S. degree in Electronical Information and Engineering from Naval Aeronautical Engineering Institute, Yantai, China, in 2001, he received Ph.D. degree in Electronical Information and Engineering from Naval Aeronautical Engineering institute Yantai, China, in 2004.

From 2012 to 2015, he was the Director of the Aviation Communication Teaching and Research Office. He is a professor in Naval Aviation University now. His research interests in WSN, Information Fusion, and Evidence theory.

DONGNING ZHAO received her B.S. degree in communication engineering from Nanjing Institute of Communication Engineering, Nanjing, China, in 2001 and M.S. degree in communication engineering from Nanjing Institute of Communication Engineering, Nanjing, China, in 2004, She received her Ph.D. degree in Signal and Information Processing from Shenzhen University in 2015.

Her research interests include information security, information hiding, and multimedia processing. She is currently working as a Vice General Manager with Shenzhen Vetose Technology Co. Ltd. China.