IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received November 9, 2018, accepted November 19, 2018, date of publication November 26, 2018, date of current version December 27, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2883460

Distributed Dimensionality Reconstruction Algorithm for High Dimensional Data in Internet of Brain Things

YIMIN ZHOU, HUANG HANG, AND LING TIAN
School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China
Corresponding author: Yimin Zhou (yiminzhou@uestc.edu.cn)

This work was supported by the Foundation of Science and Technology Department, Sichuan Province, under Grant 2017JZ0032.

ABSTRACT There are a large number of high-dimensional data in the Internet of Brain Things, and the data is reduced from the high-dimensional to low-dimensional to maintain the similarity between the data, thereby effectively ensuring the operating speed of the Internet of Brain Things. In the traditional method, the distributed dimensionality reduction reconstruction algorithm of the High dimensional data has poor dimensionality reduction effect and serious data distortion after reconstruction. A method for dimensionality reconstruction of high-dimensional data in the Internet of Brain Things is proposed. By using the algorithm of linear discriminant analysis, the projection matrix of high dimensional data is constructed to solve it. According to the solution results, using improved implicit variable model to establish a high dimensional large data dimensionality reduction model for the Internet of Brain Things. The fitness value of data after dimensionality reduction is calculated by quantum immune clonal algorithm, and the optimal individual and optimal solution are determined. The data reconfiguration is realized through the optimal solution marshalling. Experimental results show that the proposed algorithm can effectively improve the dimensionality reduction of High dimensional data in the Internet of Brain Things. After reconstruction, the reconstructed data retain accurate data information, the reliability of reconstructed data is high, and the computational complexity is not high, the need for small storage space, and the advantages of strong promotion ability.

INDEX TERMS High dimensional data, distributed, dimensionality reduction, reconstruction.

I. INTRODUCTION

An Internet of Brain Things is a multi-hop network formed by thousands of thousands of energy-constrained micro-sensors with environmental awareness, data storage and computing capabilities, and wireless communication capabilities randomly deployed in the target area. The purpose is to realize the informationization of the physical object in the target area and transmit this information to the user to realize the monitoring and control of the target area. The Internet of Brain Things integrates wireless communication technology, sensing technology and embedded technology to promote the interaction and integration of the objective world and the information world. There is a large amount of data in the operation of Internet of Brain Things, which constitute a large data set. Big data usually contains a large amount of unstructured data, and big data requires more real-time analysis. The rapid development of computer technology has made Internet of Brain Things data acquisition increasingly easy, and the data objects of interest have become increasingly complex. Academic and industry needs for data processing and analytics are also more urgent [1]. As an important part of the data set, high-dimensional data in the Internet of Brain Things plays an important role in the normal operation of the Internet of Brain Things, which can effectively guarantee the running speed of the Internet of Brain Things. High-dimensional data analysis and processing in Internet of Brain Things has become a hot topic in the field of computer science and technology research and industry [2]–[3]. How to get information describing the essential characteristics of high-dimensional data from Internet of Brain Things is a basic problem in information science field [4]. High-dimensional data in Internet of Brain Things has a nonlinear structure, it makes it difficult to extract the high-dimensional data information in the Internet of Brain Things effectively by the traditional data extraction method, which affects the utilization effect of High dimensional data in the Internet of Brain Things [5]. To solve this problem, it is necessary to perform effective dimensionality reconstruction processing on high-dimensional data in the Internet of Brain Things. With the increase of the number of high-dimensional large data in Internet of Brain Things, the research on High dimensional data has been paid more and more attention by relevant experts and scholars, and some mature theories and applications have been produced [6].

Reference [7] proposed a data reduction and reconstruction algorithm based on local constraint dictionary learning. By reconstructing the local internal manifolds of some potential mark points and embedding the training data with the unknown data into the internal manifolds, the intrinsic geometric features of the data were maintained. But this algorithm did not process data, which affected the effect of data reconstruction. Reference [8] proposed a data reconstruction algorithm based on local structure preserving. A relational graph was constructed based on least squares under nonnegative constraints and the local neighbor geometric information was described. In order to integrate the global structure information with the local structure information, a new model selection method was designed to estimate the parameters of the model and realize the reconstruction of data dimensionality reduction. But the algorithm could cause data loss and affect data dimensionality reduction. Reference [9] proposed a data dimensionality reduction and reconstruction algorithm based on the fusion of global and local discriminant information. In the algorithm, KPCA algorithm was used to effectively reduce the correlation of data sets, eliminate the redundant attributes, and retain the global nonlinear information of the original data to the greatest extent. The OLSDA algorithm was used to fully excavate the information of the local manifold structure of the data, and extract the essential features with the high discriminant and low dimension, and realize the reconstruction of the data dimensionality reduction. But the algorithm did not recognize the attributes of data, resulting in poor data reconstruction. Reference [10] proposed a regularized semi supervised algorithm for data dimensionality reduction and isometric mapping. Using tag samples from training samples, K Unicom graph (K-CG) was built, to get the geodesic distance between approximate samples, and replace the original data points with vector features. The kernel matrix was computed through the geodesic distance, and the semi supervised regularization method was used instead of the multidimensional scaling analysis (MDS) algorithm to process vector characteristics. The objective function was constructed by regularized regression model, and the explicit mapping of low dimensional representation was achieved to realize data dimensionality reduction and reconstruction. But the algorithm had a poor effect on data dimensionality reduction. Reference [11] proposed a data dimensionality reduction and reconstruction algorithm combining PCA and information entropy. Using the sparsity of SIFT eigenvector matrix, the extracted SIFT eigenvectors were used as training data, and the PCA projection was used to generate the PCA matrix group of feature vector. According to the relationship between the energy level of PCA matrix and the information entropy of data, PCA template was selected and the relation of ''dimension-information entropy'' was fitted. PCA template and the relation of ''dimension-information entropy'' were loaded, and the adaptive dimensionality reduction algorithm based on information entropy of sample data was built. But this algorithm had poor data recognition effect and it affected the effect of data reconstruction.

In order to solve the above problems, a distributed dimensionality reduction and reconstruction algorithm for high dimensional data based on the combination of latent variable model and the quantum immune clone algorithm is proposed. The whole structure is arranged as follows:

(1) The linear discriminant analysis algorithm is used to construct high dimensional data projection matrix and achieve high dimensional data processing.

(2) The latent variables model is optimized, and the dimensionality reduction model of high dimensional data is constructed, to optimize model parameters and achieve dimensionality reduction of high dimensional data.

(3) Quantum immune clonal algorithm is used to calculate the fitness of data after dimensionality reduction, to determine the optimal individuals and optimal solutions, and achieve data reconstruction through combinatorial grouping.

(4) Experimental results and analysis. Through the experimental data, the practical application effect of distributed dimensionality reduction and reconstruction algorithm for high dimensional data based on the combination of latent variable model and quantum immune clone algorithm is demonstrated.

(5) The prospect is put forward based on the existing research.

II. MATERIAL AND METHODS

The primary design goal of Internet of Brain Things is to provide high quality of service and high-dimensional data dimensionality reduction. Therefore, the dimensionality reconstruction of high-dimensional data in the Internet of Brain Things is performed, and the high-dimensional data utilization effect of the Internet of Brain Things is enhanced.

A. A HIGH-DIMENSIONAL DATA PROCESSING ALGORITHM BASED ON LINEAR DISCRIMINANT ANALYSIS

For high-dimensional data in the Internet of Brain Things, the manifold structure information of data itself is very important. In the process of dimensionality reduction, keeping the manifold structure information of data can greatly improve the effect of subsequent processing. Through the linear discriminant analysis (LDA) [12], processing high dimensional data in the Internet of Brain Things can better maintain the popular structure information of the data.

LDA algorithm can be regarded as the linearization of Laplasse feature mapping, in the Internet of Brain Things, High dimensional data processing and other fields have caused widespread concern. For a given data set

$X = [x_1, x_2, \cdots, x_n]$, the realization process of the LDA algorithm for high dimensional data is as follows:

1) CONSTRUCTION OF ADJACENT GRAPH

G represents the high dimensional data of the G dimensions. If x_i and x_j Xare very similar, the data is similar between the dimension i and j. This data is usually called similar data, which implements the description of the similarity between the data. For the two sample data of x_i and x_j, the similarity of them can be measured in the following two ways.

(1) If $\left\|x_i - x_j\right\|^2 \leq \varepsilon$, x_i and x_j have edge connection, and ε represents neighborhood.

(2) If x_i is the k neighbors of x_j or x_j is the k neighbors of x_i, then x_i and x_j have edge connection, and k represents neighbor.

2) SELECTION OF THE WEIGHT OF THE EDGE CONNECTION

If W is the weight matrix of the edge connection, the general requirement is that W is sparse and symmetric. The elements in W are defined by the thermonuclear function, and the dimension number i and j are connected.

$$W_{ij} = e^{-\frac{\left\|x_i - x_j\right\|^2}{t}} \tag{1}$$

Wherein, e represents the weight of the characteristic attributes of the data.

3) THE FOLLOWING GENERALIZED EIGENVALUE IS SOLVED THROUGH FEATURE MAPPING

$$XLX^T = \lambda XDX^T W_{ij} \tag{2}$$

Where, D is the diagonal matrix, and the diagonal elements $D_{ii} = \sum_j W_{ij}$ and $L = D - W$ represents laplacian matrix, λ represents the mapping coefficients, and T indicates the matrix exponents.

The eigenvectors corresponding to the d eigenvalues obtained by the formula (2) are $a_1, a_2, \cdots, a_d$, and the low-dimensional embedding y_i of the high dimensional data x_i in the Internet of Brain Things is:

$$y_i = A^T x_i \tag{3}$$

$A = [a_1, a_2, \cdots, a_d]$ is the LDA projection matrix [13].

Linear Discrimination for Local Maintenance of high dimensional data in Internet of Brain Things, classical LDA is used to transform the objective function into the least squares problem of matrix [14].

$$\arg\min_A \left\|X^T A - Y\right\|_F^2 \tag{4}$$

Where, X represents the data matrix after data centralization, and Y represents data output. For the convenience of use, the formula (4) is rewritten to:

$$\min_A \left\|XA^T - Y\right\|_F^2 \tag{5}$$

Setting $Z = A^T X$, each column z_i of Z can be regarded as a low dimensional representation after A to be projected. In order to make the projection matrix A to have local preserving property, it needs to keep the similarity between the data points after projection. According to the analysis of LDA, a good projection matrix should make the following formula as small as possible.

$$\min_{Z=[z_1,\cdots,z_n]} \sum_{ij} \left\|z_i - z_j\right\| W_{ij} \tag{6}$$

Where, W_{ij} is data weight matrix induced by data matrix X, and its construction method is similar to the data structure in LDA algorithm. After the simplification, the formula (6) can be written as:

$$\min_Z tr(ZWZ^T) \tag{7}$$

Where, $tr(\cdot)$ represents the trace of any matrix, that is the sum of the diagonal elements of a matrix. According to the above analysis, $z_i = A^T x_i$, x_i is the data after centralization. The formula (7) can be further written as:

$$\min_Z tr(A^T XWX^T A) \tag{8}$$

By combining the formula (8) with the formula (5), a linear discriminant analysis model with local preserving of matrix's least squares form can be expressed as

$$A^* = \min_A \{\left\|A^T X - Y\right\|_F^2 + \lambda tr(A^T XWX^T A) + \beta \left\|A\right\|_F^2 \tag{9}$$

In the upper formula, the first one is the least squares form of classical LDA, which makes the projection matrix fully discriminative. The second one is manifold regularization term, which is used to characterize the low dimensional manifold structure of high dimensional data. The last one is used to describe the complexity of the model. β is the regularized parameter.

$F(A) = \left\|A^T X - Y\right\|_F^2 + \lambda tr(A^T XLX^T A) + \beta \left\|A\right\|_F^2$ is set. When A is the optimal solution of upper formula, then the derivative of $F(A)$ to A should be zero, i.e.:

$$\frac{\partial F(A)}{\partial A} = X(X^T A - Y^T) + \lambda XLX^T A + \beta A = 0 \tag{10}$$

Then A has a closed solution as:

$$A = (XX^T + \lambda XLX^T + \beta A)^{-1} XY^T \tag{11}$$

In practical applications, the computational complexity of the theoretical solution is very large, especially when the dimension of the data is far larger than the number of the samples, the calculation complexity of the matrix inversion is high, and the algorithm needs to be improved.

Set $m \times n$ be the size of block diagonal matrix of W. According to the above, the solution of A is improved to get:

$$A = \frac{1}{\beta}(I_{m\times n} + X\bar{L}X^T)^{-1} XY^T \tag{12}$$

Where, $I_{m\times n}$ is the unit matrix with the size of $m \times n$, $\bar{L} = \frac{\lambda}{\beta}L + I_\beta$, $I_\beta = \frac{1}{\beta}I_{m\times n}$. Supposing that there is a relationship as the next formula.

$$(I + AB)^{-1}A = A(I + BA)^{-1} \tag{13}$$

Then, A can be transformed into:

$$A = \frac{1}{\beta}X(I + \bar{L}X^TXA)^{-1}Y^T \tag{14}$$

Through the above discussion, the computation complexity of high-dimensional data processing is reduced by using matrix inversion operation. But matrix inversion is not recommended in practice. A fast algorithm for solving A is proposed, which can avoid direct matrix inversion. In order to avoid direct matrix inversion, we use primitive closed form to set up $\tilde{X} = XX^T + \lambda XLX^T + \beta I$, $\tilde{Y} = XY^T$ and $A = \tilde{X}^{-1}\tilde{Y}$, and there is:

$$\tilde{X}A = \tilde{Y} \tag{15}$$

Assuming $A = [a_1, a_2, \cdots, a_d]$, the upper formula is equivalent to the following d systems of linear equations.

$$\tilde{X}a_i = y_i \tag{16}$$

In the upper formula, $i = 1, 2, \cdots, d$. By solving the above d linear equations, it should avoid matrix inversion and get the best A. According to the optimal solution, the high dimensional data processing in the Internet of Brain Things is realized, and the High dimensional data quality in the Internet of Brain Things is improved, and the reliability after data dimensionality reconstruction is ensured.

B. DISTRIBUTED DIMENSIONALITY REDUCTION OF HIGH DIMENSIONAL DATA IN INTERNET OF BRAIN THINGS BASED ON HIDDEN VARIABLE MODEL

The Supervised Latent Linear Gaussian Process Latent Variable Model (SGPLVM) is a supervised dimensionality reduction method based on the latent variable model, which has the best performance of the present [15]. However, there are two shortcomings in the model: the problem of the use of annotated information and the complexity of the algorithm. In order to overcome these shortcomings, a new nonlinear supervised dimensionality reduction method based on latent variable model is proposed, that is, the Supervised Latent Linear Gaussian Process Latent Variable Model (SLLGPLVM), and expand it to achieve high dimensional data dimensionality reduction in Internet of Brain Things [16].

Given a high dimensional data set in an Internet of Brain Things $D = \{(x_i, y_i)\}_{i=1}^N$, where, x_i represent a sample data of $X \in R^D$ dimensional input space $X \in R^D$. $y_i \in R^q$ represents a data sample in q dimensional output space. Set the dimension $q = 1$ of the output space, and each pair of observation data (x_i, y_i) has a p dimensional latent variable $z_i \in R^p$, and the whole observation data and latent variables are expressed in the following way:

$$X = [x_1, x_2, \cdots, x_N] \tag{17}$$

$$Y = [y_1, y_2, \cdots, y_N] \tag{18}$$

$$Z = [z_1, z_2, \cdots, z_N] \tag{19}$$

The goal of dimensionality reduction is to find a low dimensional manifold embedded in high-dimensional observation space. High-dimensional observations include input data X and output tag Y. The definition of SLLGPLVM is as follows:

$$y = g(z) + \varepsilon = g(Wx) + \varepsilon \tag{20}$$

Where, function g represents the mapping from latent space to output space. Using the similar skills in SGPLVM, supposing that g is the Gauss process, then there is.

$$P(g) = N(g|0, X^TA) \tag{21}$$

In the regression task, that is, Y is a continuous real number, then the noise item in SLLGPLVM can be assumed to obey the Gauss distribution, that is, $P(\varepsilon|\sigma^2) = N(\varepsilon|0, \beta^{-1})$, then the likelihood of the data can be further written as:

$$P(Y|g) = \prod_{n=1}^{N} N(y_n|g(z_n), \beta^{-1}I) \tag{22}$$

According to Bias's theory, the unknown variable g is marginalized to obtain the following marginal likelihood.

$$\begin{aligned} P(Y) &= \int P(Y|g)P(g)dg \\ &= \frac{1}{2\pi^{n/2}|K|^{1/2}}\exp\{-\frac{1}{2}(X^TALX)\} \end{aligned} \tag{23}$$

In the upper formula, $K = K_{Z,Z} + \beta^{-1}I$, $K_{Z,Z}$ represent the kernel matrix.

In regression task, the output data Y is continuous real number. In order to optimize SLLGPLVM in such data sets, the conjugate gradient method (SCG) is used to maximize the marginal distribution $P(Y)$, and the optimal (W, β) is determined [17]–[20].

$$\begin{aligned} (W, \beta) &= \arg\max_{W,\theta} L = \arg\max_{W,\theta} \log P(Y) \\ &= \arg\max_{W,\theta} \left\{\frac{1}{2}\log|K| + \frac{1}{2}tr(K^{-1}YY^T)\right\} \end{aligned} \tag{24}$$

To express the conciseness, it can write the target function as:

$$L = \frac{1}{2}\log|K| + \frac{1}{2}tr(K^{-1}YY^T) \tag{25}$$

The key to use the optimization method of SCG based gradient is to calculate partial derivatives of optimization parameters, that is, it need to deduce $\frac{\partial L}{\partial W}$ and $\frac{\partial L}{\partial \beta}$.

In order to solve the partial derivative of latent matrix W, it needs to note that the existence of linear relation $z_n = Wx_n$, then the following way can be used to calculate the derivative.

$$\frac{\partial L}{\partial W} = \sum_{n=1}^{N}\frac{\partial L}{\partial z_n} \otimes x_n^T \tag{26}$$

Where, $\otimes$ represents the external product between $\frac{\partial L}{\partial z_n} = \left[\frac{\partial L}{\partial z_{1n}}, \cdots, \frac{\partial L}{\partial z_{pn}}\right]^T$ and $x_n = [x_{1n}, \cdots, x_{Dn}]^T$.

According to the chain rule of derivatives, it can further get the following conclusions.

$$\frac{\partial L}{\partial Z} = \frac{\partial L}{\partial K}\frac{\partial K}{\partial Z} \tag{27}$$

The derivative of the kernel matrix is:

$$\frac{\partial L}{\partial K} = K^{-1}YY^TK^{-1} - K^{-1} \tag{28}$$

To solve the partial derivative of kernel parameters, it can use similar way to calculate [21].

$$\frac{\partial L}{\partial \beta} = \frac{\partial L}{\partial K}\frac{\partial K}{\partial \beta} \tag{29}$$

Once the partial derivative $\frac{\partial L}{\partial W}$ and $\frac{\partial L}{\partial \beta}$ are obtained, it can directly use the SCG algorithm to maximize the objective function L and iterate to optimize the parameter (W, β).

For classification tasks, namely, the data set $D = \{(x_i, y_i)\}_{i=1}^N$ is given, each input sample point x_i is D vector, while output y_i is a discrete category tag.

According to the idea of Gauss process classification model, a continuous latent function $f(z)$ and logical function $\sigma(z)$ are introduced. Where $z = Wx$. SLLGPLVM is applied to classification problems, and a GP priori is applied to latent function f. Assuming that the likelihood is $\pi(z) = p(y = +1|z) = \sigma(f(z))$. The Gauss distribution $q(f|Z, Y)$ is used to approximate the posterior distribution $p(f|Z, Y)$ in Laplasse approximation algorithm [22].

$$p(f|Z, Y) = \frac{p(Y|f)p(f|Z, \beta)}{p(Y|Z, \beta)} \tag{30}$$

Similar to the Gauss process classification model, SLL-GPLVM optimizes the model parameter (W, β) by maximizing the marginal likelihood $p(Y|Z, \beta)$, so that the model derivation is more concise, and the marginal likelihood can be simplified as:

$$\begin{aligned} L &= \log p(Y|Z, \beta) \\ &= -\frac{1}{2}\hat{f}^TK^{-1}\hat{f} + \log p(Y|\hat{f}) - \frac{1}{2}\log|B| \end{aligned} \tag{31}$$

Where, $B = I + U^{\frac{1}{2}}KAU$, $U = -\log p(Y|\hat{f})$, $\hat{f}$ is the largest posterior found by Newton method.

Based on the above mentioned, the partial derivative function can be expressed as:

$$\begin{aligned} \frac{\partial L}{\partial \beta} &= \frac{1}{2}\hat{f}^TK^{-1}\frac{\partial K}{\partial \beta}K^{-1}\hat{f} - \frac{1}{2}tr((U^{-1}+K)^{-1}\frac{\partial K}{\partial \beta}) \\ &\quad + \sum_{i=1}^{n}\frac{\partial \log q(Y|Z, \beta)}{\partial \hat{f}_i} \end{aligned} \tag{32}$$

Based on the above mentioned, z=1 $\frac{\partial \log q(Y|Z,\beta)}{\partial \hat{f}_i} = -\frac{1}{2}[(K^{-1}+U)^{-1}]\frac{\log p(Y|\hat{f})}{\hat{f}}$。

Where, $\frac{\partial \log q(Y|Z,\beta)}{\partial \hat{f}_i} = -\frac{1}{2}[(K^{-1}+U)^{-1}]\frac{\log p(Y|\hat{f})}{\hat{f}}$.

Through the self-continuous relationship $\hat{f} = K\log p(Y|\hat{f})$, it can further get the following:

$$\frac{\partial \hat{f}}{\partial \beta} = (I + KU)^{-1}\frac{\partial K}{\partial \beta}\log(Y|\hat{f}) \tag{33}$$

In order to calculate the partial derivative $\frac{\partial L}{\partial W}$ of the mapping matrix W, it needs to calculate the partial derivative $\frac{\partial L}{\partial K}$ of the kernel matrix, that is:

$$\begin{aligned} \frac{\partial L}{\partial K} &= \frac{1}{2}K^{-1}\hat{f}^TK^{-1} - \frac{1}{2}(I + U^{\frac{1}{2}}KU)^{-1}U \\ &\quad + \sum_{i=1}^{n}\frac{\partial \log q(Y|Z, \beta)}{\partial \hat{f}}\frac{\partial \hat{f}}{\partial K} \end{aligned} \tag{34}$$

In the upper formula, $\frac{\partial \hat{f}}{\partial K}$ can be calculated by the formula (35).

$$\frac{\partial \hat{f}}{\partial K} = \frac{\partial K_i \log p(Y|\hat{f})}{\partial K} = T^i - K_i\sum_{i=1}^{n}U\frac{\partial \hat{f}}{\partial K_{mn}} \tag{35}$$

Where, $\frac{\partial \hat{f}}{\partial K_{mn}} = (\frac{\partial \hat{f}_1}{\partial K_{mn}}, \cdots, \frac{\partial \hat{f}_N}{\partial K_{mn}})^T$ and K_{mn} represents the partial derivatives of every sample data.

For $i = 1, 2, \cdots, N$, the upper formula can be rewritten as:

$$\frac{\partial \hat{f}}{\partial K_{mn}} = A - KU\frac{\partial \hat{f}}{\partial K_{mn}} \tag{36}$$

$\frac{\partial \hat{f}}{\partial K}$ is substituted to, get:

$$\frac{\partial \hat{f}}{\partial K_{mn}} = (I + KU)^{-1}A \tag{37}$$

Once the derivative $\frac{\partial \hat{f}}{\partial K_{mn}}$ of all the sample points is obtained, it can rearrange every partial derivative $\frac{\partial \hat{f}}{\partial K_{mn}}$ to get the partial derivative $\frac{\partial \hat{f}}{\partial K}$ of the required kernel matrix.

By using the chain rule based on derivative, it can calculate the partial derivative of latent variable Z by mixing $\frac{\partial L}{\partial K}$ and $\frac{\partial K}{\partial Z}$, and get the derivative $\frac{\partial L}{\partial W}$ of the mapping matrix.

Based on the above discussion, it can calculate the partial derivatives of optimization parameters, and complete the distributed dimensionality reduction model based on partial derivatives [23].

C. HIGH DIMENSIONAL DATA RECONSTRUCTION ALGORITHM IN INTERNET OF BRAIN THINGS BASED ON QUANTUM IMMUNE CLONING ALGORITHM

High dimensional data after dimension reduction in Internet of Brain Things, data packet dropout may occur. In order to improve the problem, quantum immune clonal algorithm is used to reconstruct high dimensional data after dimension reduction. Quantum immune clonal algorithm is based on quantum computation and genetic algorithm. The encoding of antibodies is quantum bit encoding. The specific process is described as follows.

After dimensionality reduction, the status of qubit in an antibody is uncertain, and it can be 0 or 1, so the data state can be expressed as:

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle \tag{38}$$

Where, α and β represent the two plural of the corresponding state occurrence probability, and their relationship is $\alpha^2 + \beta^2 = 1$.

The antibody with m qubit bits in the data can be described by the following formula.

$$q_i^t = \begin{bmatrix} \alpha_1^t & \alpha_2^t & \cdots & \alpha_m^t \\ \beta_1^t & \beta_2^t & \cdots & \beta_m^t \end{bmatrix} \tag{39}$$

Where t represent the generation of population. m is the amount of data. The size of quantum population of n is represented as $Q(t) = \{q_1^t, q_2^t, \cdots, q_n^t\}$, and $Q(t)$ it is the solution space for high-dimensional data reconstruction in Internet of Brain Things after dimension reduction. The qubit in previous formula is made real encoding, and the encoding rules are set as follows:

$$X_i = 0.5[rangeMax(1+\alpha_i) + rangeMin(1-\alpha_i)] \tag{40}$$

Where, $i = 1, 2, \cdots, k$ represents the size of data after dimensionality reduction, and α_i represents the magnitude of the current data.

After encoding the real number, the weights of the number of non-zero elements in the current frame is set to 0.

The population before cloning is $Q(t) = \{q_1^t, q_2^t, \cdots, q_n^t\}$, and the population after clone operation is $\tilde{Q}(t) = \{Q(t), c(t)\}$. Where, $c(t)$ is the antibody's subsets produced by cloning. The sparseness of each antibody in the population is the same as that of the original population, and only the location is changed. The population size of $c(t)$ is to use roulette method, and its clone size can be expressed as:

$$m_i = ceil(n_c \times \frac{f(q_i)}{\sum_{k=1}^{N} f(q_i)}) \tag{41}$$

Where, m_i is the clonal sizes of the ith antibody in the population, n_c is a constant which is related with the clonal size but larger than the population size N, and $f(q_i)$ is the fitness of the ith antibody [24].

In quantum cloning immune algorithm, quantum antibody realizes update operation through quantum revolving door and global crossover. The quantum revolving door is used as follows:

$$U(\theta) = \begin{bmatrix} -\cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \tag{42}$$

Where, θ represents the rotation angle, which can be expressed as $\theta = k \times \frac{gen}{\max gen}$. Among them, $k \in [0.01\pi, 0.15\pi]$ is a constant, gen represents the generation number of current evolutionary. Using the above formula, the optimal solution and the optimal individual for the fitness value are determined and the data reconstruction model can be expressed as:

$$M = \frac{U(\theta) \cdot f(q_i)}{m_i} \tag{43}$$

According to the above discussion, high dimensional data reconstruction in Internet of Brain Things based on quantum immune cloning algorithm is obtained. The specific implementation process is shown in Figure 1.

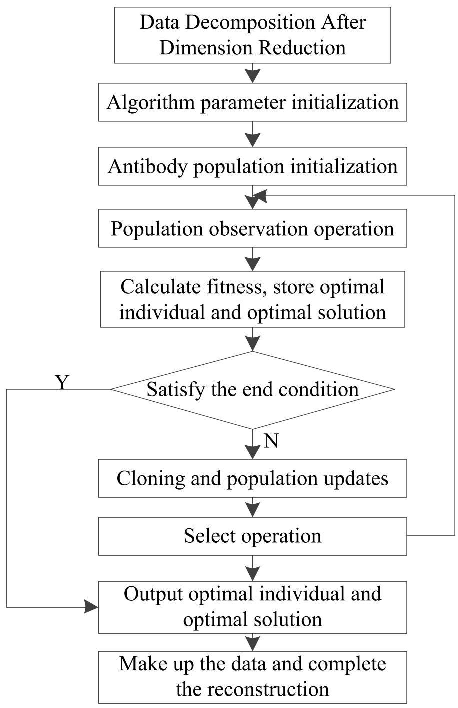

FIGURE 1. **Realization of High dimensional data reconstruction process in Internet of Brain Things based on quantum immune cloning algorithm.**

III. RESULTS

In order to prove the validity and feasibility of the distribution reconstruction algorithm for high dimensional data based on the latent variable model and the quantum immune clone algorithm, an experiment is carried out. The data set used in the experiment is shown in Figure 2. The dimension is a constant, and its unit is c.

In order to verify the ability of this algorithm to solve the incremental dimension reduction problem and to deal with large-scale data sets, the following experiments are carried out on the classic dimensionality reduction data set "Swiss-roll", "Swissholl" and "Twopeaks".

2000 points are randomly sampled from three data sets as a high-dimensional data set X. The data reduction algorithm based on latent variable model, data reduction algorithm based on ISOMAP and the data reduction algorithm based on LTSA are used to calculate the embedding Y of data set X on low dimensional manifold, and the data set (X, Y) is divided into two equal size of data sets, that is (X_1, Y_1) and (X_2, Y_2). Where, (X_1, Y_1) is as the training samples to train the

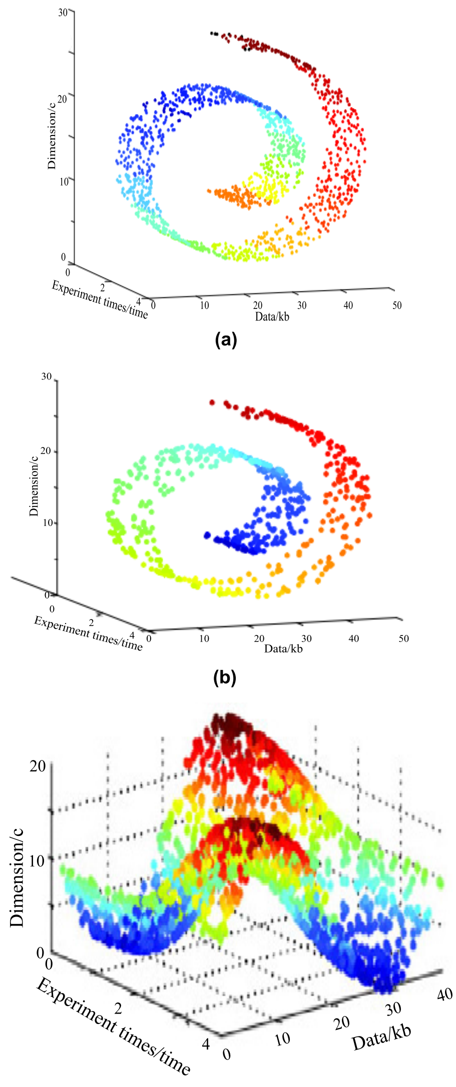

FIGURE 2. Experimental data set. (a) "Swiss Roll." (b) "Swissholl." (c) "Twopeaks."

high-dimensional data, and (X_2, Y_2) is as test samples to test the data dimensionality reduction algorithm based on latent variable model.

In order to analyze the effect of data size X on dimensionality reduction, it is set $X = 50, 100, 200, 300, 400$. The data dimensionality reduction algorithms based on latent variable model, ISOMAP and LTSA are used to reduce the dimension of " Swissroll " dataset, and the results of dimensionality reduction are shown in Table 1. Where, X is a constant and its unit is o, the result of dimensionality reduction and the embedding dimension result of low dimension manifold are constant by using different algorithms, and the unit is e.

From Table 1, we can see that the dimensionality reduction results obtained by the algorithm based on latent variable model are most close to the embedding results. According to the above discussion, the results of data dimensionality reduction algorithm based on latent variable model is the best. The data dimensionality reduction algorithm based on latent variable model is applied to process data before dimensionality reduction, which improves the quality of data and guarantees the result of dimensionality reduction.

TABLE 1. The results of dimensionality reduction for swissroll dataset.

X/o	Embedded results /e	Latent variable mc algorithm/e	ISOMAP algori	LTSA/ algo
50	57	56	50	52
100	68	69	64	65
200	81	80	83	78
300	95	94	91	93
400	103	104	99	105

Data dimensionality reduction algorithm based on latent variable model, data dimensionality reduction algorithm based on ISOMAP and data dimensionality reduction algorithm based on LTSA are used to carry out the data dimensionality reduction for the Swissroll dataset, and the performance of them is compared. The results are shown in Figure 3.

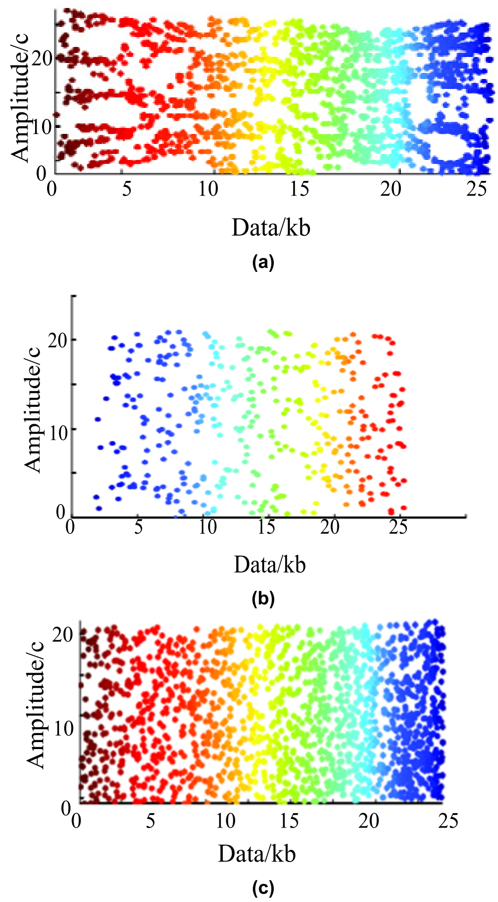

FIGURE 3. Comparison of dimensionality reduction results of different algorithms. (a) ISOMAP algorithm. (b) LTSA algorithm. (c) Latent variable model algorithm.

According to Figure 3, the data dimensionality reduction algorithm based on the latent variable model can better reduce the dimension of a single point, a small amount of data set, and even a large scale data set. It can solve the incremental

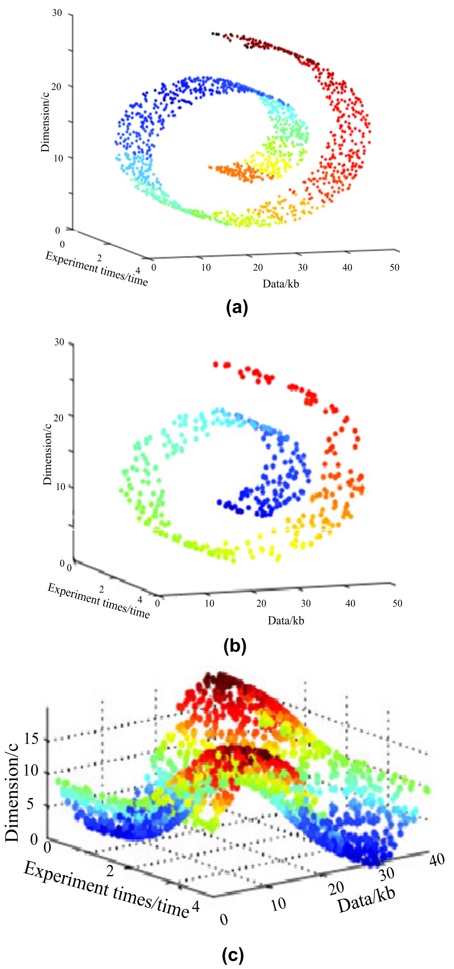

FIGURE 4. Results of data reconfiguration. (a) "Swissroll." (b) "Swissholl." (c) "Twopeaks."

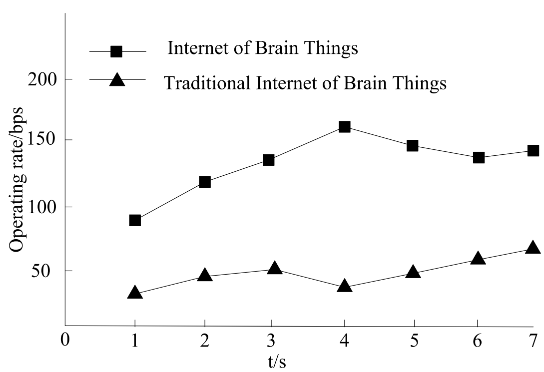

FIGURE 5. Comparison between Internet of Brain Things and traditional networks.

dimensionality problem and deal with the large-scale data sets well, and it needs the low computational complexity, the small storage space and has strong promotion ability.

In order to verify the effectiveness of the data reconstruction algorithm after dimensionality reduction based on quantum immune cloning algorithm, three data sets of Swissroll, Swissholl and Twopeaks are reconstructed, and the obtained results are shown in Figure 4.

Analysis of Figure 5 above shows that the operating rate of the traditional network is 68 bps/s and the lowest is 28 bps/s. The operating speed of the Internet of Brain Things is 156 bps/s and the lowest is 85 bps/s. The Internet of Brain Things has a relatively stable rate growth with less fluctuations. According to the above experiments, the operating speed of the Internet of Brain Things is significantly higher than that of the traditional network. The validity of the data dimensionality reconstruction method is verified. The operating speed of the Internet of Brain Things is much higher.

IV. CONCLUSIONS

With the development of computer technology, the popularity of personal computers and the Internet, and the deepening application of information technology, high-dimensional data in Internet of Brain Things has become the main focus of data analysis and processing, and the processing of high-dimensional data in Internet of Brain Things has become an academic and the urgent needs of the industry, is also a hot topic in the field of statistics and computer science and technology related research. The dimensionality reduction of high-dimensional data in Internet of Brain Things has very important research significance and practical value. Aiming at the poor dimensionality reduction effect of the high dimensional data distributed dimensionality reconstruction algorithm in current Internet of Brain Things, and the data distortion problem after reconstruction, an Internet of Brain Things based on the combination of hidden variable model and quantum immune cloning algorithm is proposed. Dimensional data dimensionality reduction reconstruction algorithm, experimental results show that the proposed algorithm has better dimensionality reduction performance and higher reliability of reconstructed data, but there are still some problems to be further researched and explored.

(1) The performance evaluation standard of the dimensionality reduction and reconstruction algorithm. Different algorithms evaluate the algorithm through different evaluation criteria, and other evaluation criteria are needed to evaluate the dimensionality reduction effect of different algorithms in the same data set, which involves the quantitative analysis of data dimensionality reduction. Although the algorithm has established explicit mapping, it is not clear which standard represents the internal structure of data set.

(2) High-dimensional data dimensionality reconstruction algorithm in Internet of Brain Things does improve the dimensionality reduction effect, but on the premise of keeping the accuracy of recognition rate, how to reduce the complexity of the algorithm and improve the speed of calculation is a direction for future research. It makes the algorithm not only stay in theoretical simulation, but also apply data dimensionality reduction and reconstruction algorithm to solve practical problems. Next, we will study the data dimensionality reduction algorithm and transplant it to the

hardware platform, and further study the problems existing in the application.

ACKNOWLEDGEMENT

The authors declare that there are no conflicts of interest regarding the publication of this article.

REFERENCES

[1] W. Xu, Y. Cui, Z. Li, and J. Lin, "A decentralized reconstruction algorithm for distributed compressed sensing," *Wireless Pers. Commun.*, vol. 96, no. 4, pp. 6175–6182, 2017.

[2] X. Wang, M. Wang, and Y. Gu, "A distributed tracking algorithm for reconstruction of graph signals," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 4, pp. 728–740, Jun. 2015.

[3] H. Fu, Z. Li, Z. Liu, and Z. Wang, "Research on big data digging of hot topics about recycled water use on micro-blog based on particle swarm optimization," *Sustainability*, vol. 10, no. 7, p. 2488, 2018.

[4] G. Guerrini, E. Landi, K. Peiffer, and A. Fortunato, "Dry grinding of gears for sustainable automotive transmission production," *J. Cleaner Prod.*, vol. 176, pp. 76–88, Mar. 2018.

[5] A. M. da Fonseca and C. M. Ylmaz, "The eigenvalues of some anti-tridiagonal Hankel matrices," *Kuwait J. Sci.*, vol. 45, no. 1, pp. 1–6, 2018.

[6] S. Wei, C. Wang, and Z. Li, "Bayes estimation of Lomax distribution parameter in the composite LINEX loss of symmetry," *J. Interdiscipl. Math.*, vol. 20, no. 5, pp. 1277–1287, 2017.

[7] L. Kang, H. L. Du, H. Zhang, and W. L. Ma, "Systematic research on the application of steel slag resources under the background of big data," *Complexity*, vol. 2018, Oct. 2018, Art. no. 6703908.

[8] R. Li and M. M. Taylor, "The first zagreb index and some Hamiltonian properties of the line graph of a graph," *J. Discrete Math. Sci. Cryptogr.*, vol. 20, no. 2, pp. 445–451, 2017.

[9] S. Sambathkumar *et al.*, "Host induced genetic variation in legume pod borer, Maruca vitrata," *J. Environ. Biol.*, vol. 38, no. 6, pp. 1281–1291, 2017.

[10] W. Gao, M. R. Farahani, A. Aslam, and S. Hosamani, "Distance learning techniques for ontology similarity measuring and ontology mapping," *Cluster Comput.*, vol. 20, no. 2, pp. 959–968, 2017.

[11] S. Ge, Z. Liu, R. Li, Y. Furuta, and W. Peng, "Desulphurization characteristics of bamboo charcoal from sulfur solution," *Saudi J. Biol. Sci.*, vol. 24, no. 1, pp. 127–131, 2017.

[12] J. Wu, L. Zhang, S. Yin, H. Wang, J. Wang, and J. Yuan, "Differential diagnosis model of hypocellular myelodysplastic syndrome and aplastic anemia based on the medical big data platform," *Complexity*, vol. 2018, Nov. 2018, Art. no. 4824350.

[13] D. Huang and J. Zhang, "Global smooth solutions for the nonlinear schrödinger equation with magnetic effect," *Discrete Continuous Dyn. Syst. S*, vol. 9, no. 6, pp. 1753–1773, 2016.

[14] S. Jiang, M. Lian, C. Lu, Q. Gu, S. Ruan, and X. Xie, "Ensemble prediction algorithm of anomaly monitoring based on big data analysis platform of open-pit mine slope," *Complexity*, vol. 2018, Aug. 2018, Art. no. 1048756.

[15] V. Petlusova, P. Petlus, and J. Hresko, "The effect of agricultural land use on soil erosion processes at upland landscape in Slovakia," *J. Environ. Biol.*, vol. 38, no. 5SI, pp. 999–1007, 2017.

[16] Y. Han, J. Li, X.-L. Yang, W.-X. Liu, and Y.-Z. Zhang, "Dynamic prediction research of silicon content in hot metal driven by big data in blast furnace smelting process under Hadoop cloud platform," *Complexity*, vol. 2018, Oct. 2018, Art. no. 8079697.

[17] L. B. Si and H. Y. Qiao, "Performance of financial expenditure in China's basic science and math education: Panel data analysis based on CCR model and BBC model," *Eurasia J. Math. Sci. Technol. Educ.*, vol. 13, no. 8, pp. 5217–5224, 2017.

[18] A. Yang, Y. Li, F. Kong, G. Wang, and E. Chen, "Security control redundancy allocation technology and security keys based on Internet of Things," *IEEE Access*, vol. 6, pp. 50187–50196, 2018.

[19] J.-H. Cheng and S.-F. Liu, "A study of innovative product marketing strategies for technological SMEs," *J. Interdiscipl. Math.*, vol. 20, no. 1, pp. 319–337, 2017.

[20] F. Liu, Y. Liu, D. Jin, X. Jia, and T. Wang, "Research on workshop-based positioning technology based on Internet of Things in big data background," *Complexity*, vol. 2018, Oct. 2018, Art. no. 7875460.

[21] A.-M. Yang, X.-L. Yang, J.-C. Chang, B. Bai, F.-B. Kong, and Q.-B. Ran, "Research on a fusion scheme of cellular network and wireless sensor for cyber physical social systems," *IEEE Access*, vol. 6, pp. 18786–18794, 2018.

[22] M. Wang, D. Q. Zhang, J. Su, J. W. Dong, and S. K. Tan, "Assessing hydrological effects and performance of low impact development practices based on future scenarios modeling," *J. Cleaner Prod.*, vol. 179, pp. 12–23, Apr. 2018.

[23] Y. Yang, M. Zhong, H. Yao, F. Yu, X. Fu, and O. Postolache, "Internet of Things for smart ports: Technologies and challenges," *IEEE Instrum. Meas. Mag.*, vol. 21, no. 1, pp. 34–43, Feb. 2018.

[24] Z.-G. Li, C. Han, and D.-M. Wei, "Empirical research on the relationship between natural gas consumption and economic growth in the northeast asia," *Energy Environ.*, vol. 29, no. 2, pp. 216–231, 2018.

YIMIN ZHOU received the B.S., M.S., and Ph.D. degrees in computer science from the School of Computer Science and Engineering, University of Electronic Science and Technology of China (UESTC) in 2003, 2006, and 2009, respectively. He was a joint Ph.D. student with the University of Central Arkansas, Conway, AR, USA, from 2007 to 2009. He was a Visiting Scholar at Georgia State University and the University of California at Santa Barbara, Santa Barbara, CA, USA, in 2013 and 2017, respectively. He has been a post Ph.D. major in signal processing since 2013. He is currently an Associate Professor at UESTC. He owns 12 granted parents and over 20 proposals adopted to the standards like HEVC, MPEG, IVC, and AVS. His research interests include big data and video streaming.

HUANG HANG was born in Hubei, China, in 1995. He is currently pursuing the bachelor's degree with the School of Computer Science and Engineering, University of Electronic Science and Technology of China. His research direction is digital image processing and data analysis. He was a recipient of the National Inspirational Scholarship, the People's Scholarship, and the People's Scholarship in 2016, 2017, 2018, respectively. He received the First Prize of the International ICIP Competition and the Third Prize of the National Competition "Image Compression Processing Competition Based on Deep Learning" in 2017.

LING TIAN received the B.S., M.S., and Ph.D. degrees from the School of Computer Science and Engineering, University of Electronic Science and Technology of China (UESTC) in 2003, 2006, and 2010, respectively. She has been a Visiting Scholar at Georgia State University, USA, in 2013. She is currently an Associate Professor with UESTC. She has edited two books and holds over 10 China patents. She has contributed over 10 technology proposals to the standardizations, such as China Audio and Video Standard and China Cloud Computing Standard. Her research interests include big data with cloud computing, image processing, and smart city.