Acta Technica *62* No. 1B/2017, 381–390

High-dimensional data express model based on tensor[1]

Zhang Jing[2,4], Guo XinChang[3]

Abstract. Taking into account the requirements of four-dimensional field data with high-dimension analysis and expression, the model of high-dimensional data structure analysis and dynamic expression based on tensor was constructed, the tensor definition, basic operators and tensor decomposition method were briefly introduced, and the process of multi-dimensional characteristic analysis based on tensor decomposition was given in this paper. Then, according to the multi-dimensional fusion feature, the high-dimensional data was organized and expressed, and the unified organization and storage method of high-dimensional data was designed. At the same time, by using the tensor decomposition method, the analysis and dynamic reconfiguration of high-dimensional data structure characteristics in different dimensions were realized, and the multi-dimensional data with high-dimensional analysis model and feature-driven high-dimensional linkage data based on tensor were established to exhibit the strategy. In addition, the experiments were verified by the grid data of the India ocean satellite RUEB, so that the multi-dimensional perspective, subset extraction, contour surface rendering and spatial data representation function based on tensor were achieved. Finally, the analysis and extraction of FARP event time type and space type were succeed by using tensor decomposition, and the data representation driven with time, longitude and latitude coefficient was realized. The examples verified that this method can exhibit the spatial and temporal patterns and the dynamic evolution characteristics of FARP event better, and the multi-dimensional perspective in the evolution process of FARP spatial can be realized.

Key words. Tensor decomposition, the four dimensional field, feature-driven, data representation.

1. Introduction

The analysis and expression of high-dimensional data features are the main contents of high-dimensional data expression. Due to the continuous development of

[1]This work was supported by the Sponsored by a project of National Social Funds named "a Research on China-Japan Maritime Crisis Monitoring and Warning" (No. 16BGJ028).

[2]Basic courses Department, Qingdao Ocean Shipping Mariners College, Qingdao, Shandong, 266071, China

[3]Ocean University of China, Qingdao, Shandong, 266071, China

[4]Corresponding author

the network observation, a large number of remote sensing images, satellite and radar images have been obtained. All the images have the characteristics of large scale and huge amount [1–2]. Due to the large scale and high-dimensional data, the significant structure of the data can be extracted, the direct data observation is the identification approach of the solutions and rules, and the high-dimensional statistical method is generally used. This method is based on matrix analysis and statistics as the theoretical basis, which can be widely used in the analysis of space and time. The disunity of the multi-dimensional operation may lead to the asymmetry in time-space dimension. General geoscience shorthand software uses the mentioned method to express the data in three-dimensional space, instead of utilizing the underlying mathematical basis for multidimensional data expression. The general high-dimensional dynamic and analytic expression method is different, which can generate an impact on the current analysis of the geographical space and time. According to the different requirements of the data analysis, new support data can be found to analyze the space and time better, which is a useful solution for high-dimensional data analysis. The data analysis is used to extend the traditional vector and data. The analytic method of tensor decomposition is used to analysis the data center, which is conducive to ensure the form of the structure and reveal the relationship among different data dimensions, so that more sophisticated data analysis method can be proposed. This method has been widely used in plenty of fields. The general analysis methods are analyzed on the basis of tensor analysis, however, there is no the comprehensive analysis of tensor analysis in the aspects of data organization and characteristic analysis. In this paper, the tensor structure was taken as the basis to discuss the data organization and storage of the tensor [3–5]. By using this decomposition method, the field data was analyzed and the multi-dimensional space-time form was represented. The results show that this study is favorable for the validation of satellite data in the India Ocean.

2. The multi-dimensional solution of high-dimensional data features based on tensor

2.1. Tensor and tensor operation

Tensor is an expansion of the vector quantity matrix, and is a kind of expression model for multi-dimensional data to express the physical properties of multiple forms, so it can be used to express and deal with the data of multi-dimensional points. An M-step tensor can be written as $A \in R^{L_1 \times L_2 \times \ldots \times L_M}$. Here, L_i means the dimension of the ith order. A lot of operational symbols can be used in the tensor product and Kronecker product, which can be applied in different algebraic tensors and the earthquakes. The symbolic significance in different conditions are identical. Generally, a bilinear calculation is required to conduct and expressed with $\otimes$. The main difference is the identical restructuring method and structure reset of the tensors [12]. Assuming that U and V are two tensors, the tensor product can be expressed as

$$U \otimes V = \begin{bmatrix} u_{11}V & u_{12}V & L \\ u_{21}V & u_{22}V & \\ M & \ddots & O \end{bmatrix} = \begin{bmatrix} u_{11}u_{11} & u_{11}u_{12} & L & u_{12}u_{11} & u_{12}u_{12} \\ u_{11}u_{21} & u_{11}u_{22} & & u_{12}u_{21} & u_{12}u_{22} \\ M & & & & \\ u_{21}u_{11} & u_{21}u_{12} & O & & \\ u_{21}u_{21} & u_{21}u_{22} & & & \\ M & & & & \end{bmatrix} . \quad (1)$$

2.2. QRB tensor decomposition

Tensor decomposition is mainly extended by the method of component analysis, which can make data analysis of high order tensor through the low order tensor. The singular value decomposition method is used in this method, and the tensors are calculated by least square method. The tensors can be separated through the high dimensional space, and then the reliability test can be carried out.

$$\begin{aligned} \sigma_i = & \max_{\substack{\| \psi \|_s = 1 \\ \| \varphi \|_v = 1 \\ \| \phi \|_t = 1}} Y. . (\psi \otimes \varphi \otimes \phi) \\ & = Y.. (\psi_i \otimes \varphi_i \otimes \phi_i) . \end{aligned} \quad (2)$$

Here, σ_i is the i characteristic value of the tensor as well as the peacekeeping vector. The operator can be shown as indentation operation, which can be called the accumulation of tensor or the decomposition of tensor coefficient and tensor. Then, dimensionality based on the tensor decomposition can be constructed.

$$\begin{aligned} Y.. (\psi \otimes \varphi \otimes \phi) &= (Y..\psi) .. (\varphi \otimes \phi) \\ &= (Y..\varphi) .. (\psi \otimes \phi) \\ &= (Y..\phi) .. (\psi \otimes \phi) . \end{aligned} \quad (3)$$

For k-dimensional tensor, the tensor decomposition model can be constructed with the dimensionality $k-1, k-2, \cdots, k-n, \cdots, 1$.

2.3. Multi-dimensional feature reconstruction based on tensor decomposition

Generally, tensor decomposition is mainly to carry out the structure characteristics analysis of high-dimensional data, which is conducive to the data reconstruction for multi-dimensions. Taking the four-dimensional field data as an example, the sum of tensors can be used in this four-dimensional field data, so that the tensors can obtain the coefficient information of direction axis and the 3 directions. Then, the process matrix can be constructed, and different dimension coefficient can represent the information in practical works, so that the characteristics of spatial and temporal process in different dimensions can be obtained. Finally, the process of dimension perspective can be conducted on the basis of specific dimensions. For example, by constructing kinds of spaces, the specific combination of time and space can be dis-

tributed through different visual angles. The common dimensions are the tensor products in two-order tensors of a tensor and two dimensions, which can be used to represent the space-time evolution process. A large number of decomposition results guarantee the correlation of different dimension structures, which is beneficial to the performance of dynamic data in different dimensions, and it is also in favor of the analysis and expression of the unit structure. However, the dimensionality of data may be at the expense of partial data and information, which is good for high-dimensional data to be mapped into a low-dimensional space. Because there are some losses of the original information, the essential premise of the original data needs to be kept, so as to express the high-dimensional data in low-dimensional better.

3. Characteristic data representation of high-dimensional data

3.1. High-dimensional data analysis and data expression process based on tensor

Figure 1 shows the decomposition conditions of tensor data between the two-order tensors. This process can be used in the management and analysis of the original data. The data have better decomposition, so that the data coefficients can be obtained, which is not only beneficial for the data reconstruction to obtain the data with physical meaning, but also good for the analysis of space-time and import parameters of external time to find the data meaning in different dimensions, so that the dimension fusion of the tensor data organization can be conducted, and the different fusion results can be reconstructed. Finally, the observation and design pattern can be gained to express the data for simple structures. Each of the flow chart is compatible with a variety of multi-dimensional data, which can not only prompt a variety of data organization and the unified expression of the process, but also has a variety of dimension and analysis functions.

3.2. High-dimensional data organization based on tensor

Based on the increment, the current four-dimensional data has been constructed by the multi-dimensional array to make the object retrieval and object reorganization and analysis process. This is conducive to the expression of four-dimensional data object, analysis of the history value, and has significant meanings to the construction of geographic models. It can carry on the reorganization of the market according to the storage mode. It can analyze and express the data by using tensors, and the multidimensional spatial data can be analyzed. It is possible to gain a better effect, but there may be more of the value of irregular boundary and missing, so the estimation of all kinds of parameter is required. Generally, based on the stereoscopic space-time model, a variety of model data can be stored through the positive structure, including all of the core components and the tensor storage structure. Therefore, the thought data and the attribute of tensors can be analyzed,

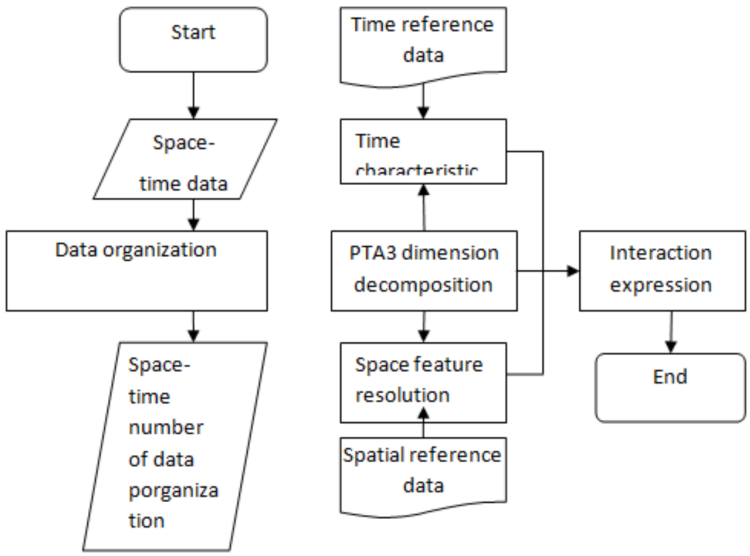

Fig. 1. High-dimensional data analysis and dynamic expression process

and the market can be described. According to the methods of the last operation tensor and a variety of data operation plans, the decomposition of tensor and tensor data analysis, multi-dimensional tensors perspective in the same group of tensors of these components, the goal of object modeling can be achieved, which is favor of the retrieval of capital market data. Based on this, kinds of strategies can be constructed, so that the structural characteristics and perspective target can be obtained easily, which is conducive to the management and operation of the relevant data, reorganization and calculation of all the data, and the analysis and function expansion of the follow-up data.

3.3. Data representation of feature driven

The goal of geographic data is to reveal all kinds of news in organizations, which is conducive to analyze the indicators, and is helpful to achieve the current data expression. The system is good for the interpretation and extraction of the data through the method of dimension reduction. This data obtained should be in the method of matrix operation, which has a great effect on the high-dimensional data. The data analysis should be linkage and the tensor data should have various fusion characteristics, so as to promote the combination and dispatch among different data. According to the integrality and reconstruction of the tensor decomposition, there are internal linkage and perspective. Since the characteristics of this idea can be obtained, as shown in the scheme, the tensor form organization data can perform the dimension reorganization and penetration, which is conducive to the integration of dimensions. By adding the geography index, the signal can be obtained from high level data. According to the different analytical results, different expression

strategies can be developed. According to the characteristics of data and function, all the result data can be used to carry out the time and space data election. Then the observer pattern is set to express the dynamic state and characteristics in different dimensions, so as to provide a better environment for data analysis, which is conducive to the analysis to the solution of the geoscience problem.

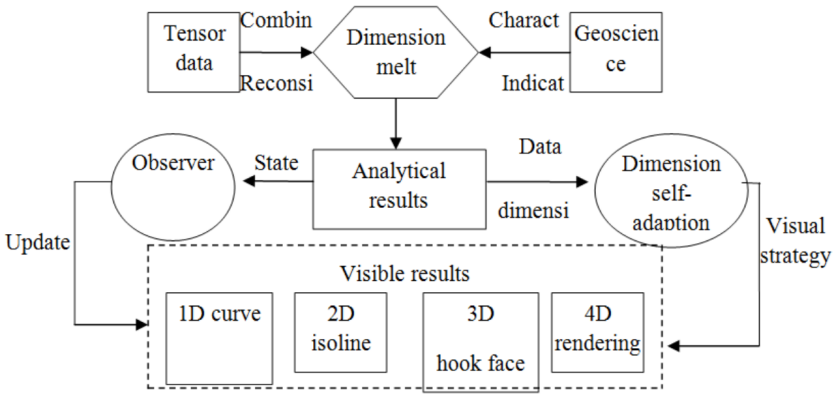

Fig. 2. Scheme of data representation for feature driven

4. Experiment and verification

4.1. Original data

The BQHTP not only can release the mixture of T/P and Jason4 in India ocean grid data, but also can perform all kinds of experimental verifications. The original data is the NetCDF format with the spatial resolution of 1.4×1.4, the time scale being from 1995 to 2003, and the spatial range being from 15 degrees north latitude to 15 degrees south latitude, 150 degrees east longitude to 150 degrees west longitude. In addition, there are 92 time slices, 21 time dimensions and 421 longitude intervals, which are compared with the standard indicators. This system is conducive to the decomposition of the tensor and representation of the data, and the import and convert of data are good for the unified investigation of various tensors, so as to draw all kinds of VTK pictures. Figure 3 shows the perspective four-dimensional field data along the time direction.

4.2. Multi-dimensional analysis of data characteristics

Since the sea surface is treated average monthly processing, abnormal high-dimensional field data may occur in the sea surface. Each of the tensor has its own contribution rate to the data, so that the fusion refactoring of the solved results can be performed, and the time change graphs can be obtained. The MEI index can be introduced to express the events' degree, which is conducive to the final analysis

Fig. 3. Data representation based on tensor data

of the characteristics. As shown in Fig. 4, the main tensor can decompose the space and time, so as to obtain the time factor, which has better correlation. Based on this effect, the general FARP evolution has differences and shows the obvious difference of latitude in space. Through a comprehensive comparison of reconstruction at different time, it can be found that the sea space changes with time; and the two images indicate the sea location, the amplitude and distribution, respectively.

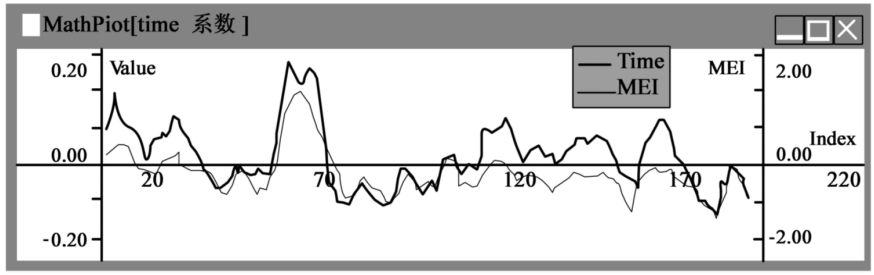

Fig. 4. Space and time decomposition of the main tensor

4.3. Dynamic data representation with feature driven

The three coefficients can be obtained by using tensor decomposition and the data can be expressed. This method is favorable of the perspective different vectors and different dimensions. Through the decomposition of the vector data, each data has a unique dimension basic training. The two-dimensional data can support a two-order tensor to express this three-dimensional data, and both of the two have relationship through the data sequence. When various sequence data changes, the temperature of the tensor will appear a variety of changes. The decomposition can be constructed through the varying process in latitude of the sea level changing and

the specific perspective, which is favorable to construct and discuss the parameters structure, so as to implement the dynamic data expression. Based on this, the exploration and analysis can be carried out.

5. Conclusion

The expression of multidimensional data is an important direction for the development of the subject. Through the analysis of the tensor structure, various data models can be established, which is conducive to the establishment of a unified expression and calculation framework. Taking into account the characteristics of high dimension, new solutions can be proposed to analyze the dynamic expression and analysis method. Through the study in this paper, it can be seen that the tensor structure is good for the multi-dimensional statistics and calculation, the tensor structure established is conducive to the plenty decomposition and the decomposition of the feature data, and the data expression can also be realized. At the same time, according to India ocean satellite data, the original pattern and dynamic characteristics can be reproduced. The results of this paper are conducive to a better use of the tensor structure extension, which is supportive to the intrinsic computer. Based on the construction method of the tensor, the significant of the high-dimensional data analysis and expression can be analyzed. In addition, the tensor analysis is unceasing developing and progressing, which is conducive to go deep into the comprehensive use of related fields. In this research, the related demands were analyzed on the basis of the existing problems, and the calculation process was optimized by using tensor operation characteristics, so as to analyze and deconstruct the irregular data through various multi-dimensional data representation methods, instead of the data values.

References

[1] J. Luo, C. Zhou, Z. Sehn, X. Jang, C. Qiao, Q. Chen, D. Ming: *Theoretic and methodological review on sensor information tupu computation.* Journal of Geo-Information Science *11* (2009), No. 05, 664–669.

[2] G. Andrienko, N. Andrienko, U. Demsar, D. Dransch, J. Dykes, S. I. Fabrikant, M. Jern, M. J. Kraak, H. Schumann, C. Tominski: *Space, time and visual analytics.* International Journal of Geographical Information Science *24* (2010), No. 10, 1577–1600.

[3] H. Su, Y. Wen, M. Chen, H. Tao, J. Shen: *VGEs-oriented multi-sourced heterogeneous spatial data integration.* Proc. Geoinformatics 2008 and Joint Conference on GIS and Built Environment: Advanced Spatial Data Models and Analyses, 28–29 June 2008, Guangzhou, China, paper 71460D.

[4] A. H. Monahan, J. C. Fyfe, M. P. H. Ambaum, D. B. Stephenson, G. R. North: *Empirical orthogonal functions: The medium is the message.* American Meteorological Society, Journal of Climate *22* (2009), No. 24, 6501–6514.

[5] L. Zhang, C. Yang, D. Liu, Y. Ren, X. Rui: *A web-mapping system for real-time visualization of the global terrain.* Computers & Geosciences *31* (2005), No. 3, 343–352.

[6] J. Hu, Z. Ma, H. Wu, M. Pan: *3D visualization of massive terrain data based on*

grid partition. Journal of Computer Aided Design & Computer Graphics *16* (2004), No. 8, 1164–1168.

[7] Y. Xia, R. Mo, J. Zhang, Q. Tian, Z. Chang: *Design process saving and reconstruction based on knowledge*. Journal of Computer Aided Design & Computer Graphics *17* (2005), No. 12, 2702–2708.

[8] Y. G. Ren, G. Yu: *Research and development of the data visualization techniques*. Computer Science *31* (2004), No. 12, 92–96

[9] A. Zeli: *The financial distress indicators trend in Italy: An analysis of medium-size enterprises*. Eurasian Economic Review *4* (2014), No. 2, 199–221.

[10] L. D. Lathauwer, B. D. Moor, J. Vandewalle: *On the best Rank-1 and Rank-(R1 ,R2 ,. . .,RN) approximation of higher-order tensors*. Siam Journal on Matrix Analysis & Applications *21*, (2006), No. 4, 1324–1342.

[11] M. Bachmayr, W. Dahmen: *Adaptive near-optimal rank tensor approximation for high-dimensional operator equations*. Foundations of Computational Mathematics *15* (2015), No. 4, 839–898.

[12] D. Leibovici, R. Sabatier: *A singular value decomposition of ak-way array for a principal component analysis of multiway data, PTA-k*. Linear Algebra and its Applications *269* (1998), No. 1, 307–329.

[13] P. Drineas, M. V. Mahoney: *A randomized algorithm for a tensor-based generalization of the singular value decomposition*. Linear Algebra and its Applications *420* (2007), Nos. 2–3, 553–571.

Received June 29, 2017