

Research on the Management Mechanism of Multidimensional Data Organization for ZY-1 02C Satellite

Ken CHEN¹, Ping LIAO¹, Naixue ZHANG¹, Yuchuan WANG¹, Pengfei XIAO¹, Fang Wang², Fang MIAO³

¹Sichuan Institute of Land Planning and Survey, 610045, Chengdu, China

²College of Computer Science and Technology Southwest University for Nationalities, 610041, Chengdu, China

³Spatial Information Technology Institute, Chengdu, 610059 China

email: chenken@foxmail.com

Keywords: ZY-1 02C Satellite; Multidimensional Model; Clustering Classification; Land Resources Management

Abstract. The development of China-made satellites has begun to take shape. It will form an earth observation data system, combined with optics and microwave, with high, medium and low resolution at the same time. Particularly, the application of ZY-1 02C satellite, with strong dynamic observation capabilities, leads the way, providing a new platform for our land management work. Reasonable data classification in terms of data attributes and effective organization of data in free state is the primary task to solve the problem of how to use land resources data efficiently for service. Therefore, the establishment of multidimensional data coordinated organization management is the foundation for data services and data collaboration. It can strengthen supervision by technical means, standardize management and operation, and also enhance the most stringent farmland protection and intensive land use, by studying the management mechanism of multidimensional data organization for ZY-1 02C satellite, which reflects the spirits of active service and tracking service from land departments, providing theoretical support for the construction of land Integrated monitoring platform.

Introduction

With the launch of resources satellite series, remote sensing satellite series, environment satellite series, as well as other small domestic satellites and the application of its products, the development of China-made satellites has begun to take shape. It will form an earth observation data system, combined with optics and microwave, with high, medium and low resolution at the same time. In particular, the ZY-1 02C satellite, China's first high-resolution satellite, launched successfully on December 22, 2011, stands out. Taking full advantages of high-resolution remote sensing data acquired by ZY-1 02C satellite, it has great significance to carry out image processing, and study land use thematic information extraction technology, to enrich our land resources remote sensing application system, etc. With unceasing expansion of remote sensing application on depth, width and utilization, China's domestic satellite has been widely used in agriculture, forestry, land, water conservancy, urban construction, environment, mapping, transportation, meteorology, oceanography, earth science, and many other fields [1].

Multidimensional Data Classification Model

(1) General introduction to the Multidimensional Data Classification Model

Satellite data are becoming key data resources for the scientific research of Earth System and social and economic life [2]. To efficiently utilize the land data resources for application service, first of all, reasonable data classification in terms of data attributes and multidimensional data organization, effectively organizing data in free state, are the basis of data service and data collaboration. Any data can be classified and analyzed in any dimension, thereby confirming the

location in cloud storage; Meanwhile, the data at any location in cloud can be confirmed which type it belongs to, such as dimension of process, structure, time, industry, space, etc. The data attribute, definition and distribution have sound multidimensional descriptions in the Multidimensional Data Classification Model, having fundamental research significance for the image data application of domestic satellite. Multidimensional Data Classification Model as shown in figure 1: 1) Structural Dimension, including structuralization, non-structuralization, semi-structuralization, etc.; 2) Process Dimension, including original data, intermediate results, final results, etc.; 3) Industrial Dimension, including dynamic surveillance, prime farmland, mineral resource, geological hazard, etc.; 4) Spatial Dimension, including two-dimensional/three dimensional data, vector/raster data, etc.; 5) Temporal Dimension, including temporary data, offline data, near-line data, online data, etc.

(2) Features of Multidimensional Data Classification Model

The features of Multidimensional Data Classification Model are as follows:

1) To be Theme-Oriented

Traditional data classification model applies to transaction process and assignment, and data from each system separate respectively; While the data from multidimensional data classification model are theme and architecture oriented, organized by a certain theme, which are the key aspects of data analysis and decision making with multidimensional model, such as land utilization, farmland protection, vegetation coverage, mineral reserves, etc..

2) To be Sharable

The data from multidimensional data classification model are for shared service and application, which are stored dispersedly after marking the data, and processed, collected, organized through unified data exchange standard. The redundancy mechanism allows for the inconsistency of the original data and integrity of the data information, such as name switching, variables measurement, encoding structure, data attributes, etc.

3) To be Dynamic

Under the big data environment, data usually need to be updated in real time, and modified in time according to requirements. The data from multidimensional model are mainly for decision making, multidimensional display, data inquiry, etc. Once the data are stored in cloud through data registration center, they are permanently retained and redundantly backed up. Users can carry out regular update, maintenance, etc. through data rights management.

4) To be Real-Time

Through dispersed storage of data, information service can be provided online, near-line and offline. There is element of time in the multidimensional data classification model, that is, full life cycle properties. The information provided is related to time. The data information covers the data generating, storage, management, sharing, destruction, invalidity, etc. during data classification.

Data Clustering Management for ZY-1 02C Satellite Oriented

On the basis of multidimensional data classification model of ZY-1 02C, effectiveness of data collaboration increased, while complexity of data type reduced, start from temporal, industrial and spatial dimension, and combine with clustering algorithm for categorical data, to improve data organization and management capability. Classification method based on thresholds is a basic method, belonging to hard classification method, simple and fast. But under the big data environment, it is unsuitable for multidimensional data classification. The common classification methods are K-means Clustering (KC) [3] [4], Fuzzy C-means Clustering (FCM) [5] [6], etc. KC belongs to hard C-means algorithm. FCM is the most well-known fuzzy clustering algorithm, brought forward by Bezdek, belonging to C-means algorithm. It is impossible to be sure about the number of clusters and each initial cluster center, which is the greatest weakness of FCM. Essentially, FCM is local search algorithm. With improper selection of initial values, it not only needs more iterations, but also converges to local optimum [7] [8], which strongly influences clustering effects.

Aiming at the above issues, by combining KC with FCM, K-means Clustering & Fuzzy C-means Clustering (KC-FCM) is put forward, in which KC is used to optimize the selection problem of

initial clustering center for FCM, and FCM to optimize the given clustering center. Thus fuzzy clustering partition is accomplished. Combining with fuzzy reasoning method, clustering classification is finally achieved.

Given data set $X = \{x_1, x_2, \dots, x_n\}$, $x_k \in R^p$, let $P(X)$ be the power set of X , KC initial clustering center for X is a family of sets $A_i \in P(X) | 1 \leq i \leq c$ (c is positive integer), that is, $\{A_1, A_2, \dots, A_c\}$ put X into c class, in which each A_i is a category. It meets:

$$\begin{cases} \cup_{i=1}^c A_i = X \\ A_i \cap A_j = \emptyset (1 \leq i \neq j \leq c) \end{cases} \quad (1)$$

KC can be represented by the membership function of elements x_k in A_i . x_k belongs to A_i , $u_{ik} = 1$, otherwise $u_{ik} = 0$. Accordingly, there is a given value of u_{ik} , a KC partition of a sole can X be confirmed, vice versa. Meanwhile, u_{ik} should meet 3 conditions:

$$u_{ik} \in \{0,1\}, 1 \leq i \leq c, 1 \leq k \leq n \quad (2)$$

$$\sum_{i=1}^c u_{ik} = 1, \forall k \in \{1,2, \dots, n\} \quad (3)$$

$$0 < \sum_{k=1}^n u_{ik} < n, \forall i \in \{1,2, \dots, c\} \quad (4)$$

Matrix $U_{c \times n}$ is consisted of elements u_{ik} , hence, matrix form of initial clustering center according to KC is obtained. The definition is as follows:

Definition 1 initial clustering center according to KC: Let $X = \{x_1, x_2, \dots, x_n\}$ be any set, V_{cn} is real $c \times n$ matrix, in the set of $U = [u_{ik}]_{c \times n}$, U meets formulas from (2) to (4), c is integer, and $n > c \geq 2$, thus KC classification space of X is set:

$$M_c = \{U | U \in V_{cn}\} \quad (5)$$

Definition 2 optimized classification according to FCM: Let the meanings of X , V_{cn} and c be same with definition 1, therein, u_{ik} is the membership of A_i , which meets formula (3), thus FCM space of X is set:

$$M_{fc} = \{U \in V_{cn} | u_{ik} \in [0,1], 1 \leq i \leq c, 1 \leq k \leq n\} \quad (6)$$

Through total error sum of squares of objective function, combining with M_c or M_{fc} , the optimum objective function is :

$$J_w(U,V) = \sum_{k=1}^n \sum_{i=1}^c u_{ik} \|x_k - v_i\|^2 \quad (7)$$

In the formula, $U \in M_c$ or M_{fc} , $V = (v_1, x_2, \dots, v_c)$, v_i is the center of class A_i :

$$v_i = \sum_{k=1}^n u_{ik} x_k / \sum_{k=1}^n u_{ik} \quad (8)$$

To make objective function reach local minimum value, and all points are close to clustering center according to KC, thus $U = [u_{ik}] \in M_{fc}$ and $V = (v_1, x_2, \dots, v_c) (v_i \in R^p)$ are found, objective function (9) is the minimum:

$$J_m(U,V) = \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^m \|x_k - v_i\|^2 \quad (9)$$

Therein, $m \in (1, \infty]$ is weighted index, based on the establishment of the necessary condition of minimization matter, combine KC with FCM:

Theorem 1 Let $X = \{x_1, x_2, \dots, x_n\}$, $x_k \in R^p$ be the given data set, assume $c \in \{2, 3, \dots, n-1\}$ and $m \in (1, \infty)$, and for all conditions of $1 \leq i \leq c$ and $1 \leq k \leq n$, $\|x_k - v_i\| \neq 0$, only when:

$$u_{ik} = 1 / \sum_{j=1}^c \left(\frac{\|x_k - v_i\|}{\|x_k - v_j\|} \right)^{2/m-1}, \quad 1 \leq i \leq c, \quad 1 \leq k \leq n \quad (10)$$

$$v_i = [\sum_{k=1}^n (u_{ik})^m x_k] / [\sum_{k=1}^n (u_{ik})^m], \quad 1 \leq i \leq c \quad (11)$$

$U = [u_{ik}] \in M_{fc}$ and $V = (v_1, v_2, \dots, v_c) (v_i \in R^p)$ are the local minimum of $J_m(U, V)$.

Support Vector Machine (SVM) is a general linear classification algorithm, having simple structure and strong generalization ability. Through non-linear transformation, the input space is transformed to a high-dimensional Eigen space to find out the optimum linear classification. For a group of identified mode (x, y) , therein, x is observation data, y is identity, α represents decision function parameter, ϕ is a mapping from original data space to high-dimensional Eigen space, b is deviation, the optimum is approximated to be :

$$f(x, \alpha) = \omega \cdot \phi(x) + b \quad (12)$$

Therein, $X = (x^1, x^2, \dots, x^d)$ is input vector, X_1, X_2, \dots, X_s are support vectors, $K(X_i, X)$ is kernel function, and $\omega_i = \alpha_i y_i$ are weights. After calculation, the output Y is:

$$Y = \text{sgn}[\sum_{i=1}^s \alpha_i y_i K(X_i, X) + b] \quad (13)$$

Through the verification of theoretical analysis and experimental results of programming, compared with traditional FCM, it is achievable to converge to global optimum with less iterations with the above methods. Combining with SVM, better workouts are obtained, and precision and efficiency of data classification under big data environment can be effectively improved, less complex and adaptive for the data environment of ZY-1 02C. The method can be applied to data classification system and data organization of ZY-1 02C.

Conclusion

With the rapid development of the domestic satellite data, series of satellites with strong dynamic observational capability, provide a new platform for our land management work. By the research of classification method of multidimensional model, it provides an effective support for follow-up work of land resources for ZY-1 02C satellite data:

(1) Products of ZY-1 02C satellite data, with features of multi-resources, mass, heterogeneity and complexity of applications, etc., raise a higher demand for sharing data products;

(2) In view of the data product standards system has not been refined, it needs necessary, scientific and technical support for sharing data products through the establishment of production standards, product distribution and information services standards, data management practices and application system construction specifications;

(3) Data oriented Architecture provides a technical reference for applications based on ZY-1 02C satellite through the establishment of data registration centers for sharing data services, combined with land-related applications.

In conclusion, by studying the management mechanism of multidimensional data organization for ZY-1 02Csatellite, it can strengthen supervision by technical means, standardize management and operation, and also enhance the most stringent farmland protection and intensive land use, which reflects the spirits of active service and tracking service of land department, providing technical support and policy advice for urban development zone, as well as theoretical support for the construction of land Integrated monitoring platform.

Acknowledgement

In this paper, the research was sponsored by the Nature Science Foundation (Project No. 61071121) and the Fundamental Research Funds for the Central Universities of Southwest University for Nationalities (Project No. 11NZYTD09).

References

- [1] Tianhe Chi. The Rapid Development of Remote Sensing Spatial Information Industry [J]. Journal of Chinese Scientific and Technological Achievements, 2008 (12): 21-23.
- [2] RuiLiang Pu, Peng Gong. Hyperspectral Remote Sensing and its Application [M]. Beijing: Higher Education Press, 2001.
- [3] Xu R, Wunsch D. Survey of Clustering Algorithms [J]. IEEE Transactions Neural Networks, 2005, 16(3): 645-678.
- [4] Chang Dongxia, Zhang Xianda, Zheng Changwen. A Genetic Algorithm with Gene Rearrangement for K-Means Clustering [J]. Pattern Recognition, 2009, 42(7): 1210-1222.
- [5] Bezdek JC, Ehrlich R, Full W FCM: the Fuzzy C-means Clustering Algorithm [J]. Computers and Geosciences, 1984, 10(2-3): 191-203.
- [6] Jie Yu, Peihuang Guo, Pinxiang Chen. Remote Sensing Image Classification Based on Improved Fuzzy C-means [J]. Geo-Spatial Information Science, 2008, 11(2): 90-94.
- [7] Kun Qin, Min Xu. Based on Cloud Model and FCM Clustering of Remote Sensing Image Segmentation Method [J]. Earth Information Science, 2008, 10 (3): 302-307.
- [8] Xiangyang Wang Chunhua Wang. Based on the Characteristics of the Divergence of Adaptive FCM Image Segmentation Algorithm [J]. Chinese Journal of Image and Graphics, 2008, 13 (5): 906-910.