Spatio-Temporal Data Clustering Based on Type-2 Fuzzy Sets and Cloud Models

Kun Qin, Mengran Wu, Lingqiao Kong, Yao Liu
School of Remote Sensing Information Engineering, Wuhan University, Wuhan, China, 430079
Email: qinkun163@163.com

Abstract

The time series remote sensing data and meteorological satellite data offer new opportunities for understanding how the Earth is changing, for determining what factors cause these changes, and for predicting future changes. These time series spatial data are very huge. How to discover knowledge from these data is an urgent task. Clustering mining is an important approach of spatio-temporal data mining. It can explore the huge data space and discover the data clusters. Spatio-temporal data clustering becomes a kind of idea tools to explore huge data space of spatio-temporal data. There are many uncertainties in the huge spatio-temporal data, including fuzziness and randomness. For these uncertain problems, the spatio-temporal clustering methods with uncertainty are most suitable. Based on the theories of type-2 fuzzy set and cloud model, the paper analyzes the uncertainties of the membership values of FCM (fuzzy *C*-means), and proposes a FCM-type method based on cloud model. Take the time series SST data as example, the paper uses FCM, FCM based on type-2 fuzzy sets, and FCM based on cloud models to carry out the spatio-temporal clustering analysis, and analyzes the relationships between the clustering patterns and the climate events.

Firstly, the paper analyzes the uncertainties of FCM. FCM is a kind of objective function-based clustering methods, which are used to minimize the distance between a pattern and cluster prototypes. Conventional FCM method assigns a certain membership value for every pattern, which is not suitable in many real world applications.

Secondly, the paper introduces the principles of type-2 fuzzy sets and cloud models. In order to analyze the uncertainties of memberships, L.A.Zadeh proposes type-2

fuzzy sets in 1975; Deyi Li put forward cloud models in 1995. The concept of type-2 fuzzy sets is an extension of ordinary fuzzy sets, it can be characterized by a fuzzy membership function, i.e., the membership value for each element of this set is a fuzzy set. There are two kinds of type-2 fuzzy sets: interval type-2 fuzzy set whose secondary membership functions (MFs) are interval sets, and Gaussian type-2 fuzzy set whose secondary MFs are type-1 Gaussian MFs.

Cloud models use three numerical characters (*Ex*, *En*, *He*) to express a qualitative concept. *Ex* means the expected value; *En* means the entropy, which is determined by both the randomness and the fuzziness of the concept; *He* means the hyper-entropy, which is the uncertainty measurement of the entropy. Given the parameters *Ex*, *En*, *He* and the number of cloud drops *n*, cloud model can produce *n* cloud drops.The forward normal cloud generator includes the following steps: Generate a normally distributed random number $En_i^{'}$ with expectation *En* and variance He^2; Generate a normally distributed random number x_i with expectation *Ex* and variance $En_i^{'2}$; Calculate $\mu_i = \exp[-\frac{(x_i - Ex)^2}{2(En_i^{'})^2}]$; x_i with certainty degree μ_i is a cloud drop in the domain; Repeat the above steps until *n* cloud drops are generated.

Both type-2 fuzzy sets and cloud models deal with the uncertainties of membership values of pattern sets. Type-2 fuzzy sets mainly consider the fuzziness of memberships, use a fuzzy set to express the memberships; cloud models use random method to generate the membership values of pattern sets, and build the association between the fuzziness and randomness.

Thirdly, the paper propose a FCM-type clustering method based on cloud models. Hwang and Rhee proposes an uncertain fuzzy clustering, which extend a pattern set to interval type-2 fuzzy sets using two fuzzifiers m_1 and m_2 which create a footprint of uncertainty (FOU) for the fuzzifier *m*, and incorporate this interval type-2 fuzzy set into FCM. Some FCM-type clustering methods based on cloud models have been proposed. Qin *et al*. put forward a clustering method which uses cloud model to initialize the FCM, Yang *et al*. propose a method which introduces the normalization constraint, and use cloud models to express the clusters. Based on these researches, the paper put forward a clustering method which considers the membership values of a pattern as a cloud model including some discrete random points. We use a cloud model *Cloud*(*Ex*,*En*,*He*) to express *m*. for every cloud drop, there is a value of *m*. So, every sample x_i has several membership values. For each cloud, we set *En* =0.2, so *Ex* lies in [Ex-0.6, Ex+0.6], and let *Ex* increase 1 every time in the range of [1.1, 10]. So the values of *Ex* can fully cover the whole range [0, 10]. If the values are less than 0 or larger than 10, then let them be 0 or 10. For the method, the membership value of a pattern set is a random value which is a point of cloud model. Then carry out clustering analysis based on the basic procedure of FCM. In the last, use the cluster validity measurement techniques to define the best clustering results. In order to check the validity of the proposed approach, we use Gaussian random data with noise, and four datasets from UCI repository of machine learning databases to carry out clustering experiments based on FCM, FCM-type clustering method based on type-2 fuzzy sets, and the proposed method. The experiment results validate the proposed method.

Fourthly, the paper applies these clustering methods into the clustering analysis of time series SST data.The SST dataset includes the monthly data of 26 years, which cover the whole globe. SST data can be used to research the climate changes. The processing steps of time series SST data clustering include: Deal with the seasonality of data. For most earth scientists are only interested in non-seasonal patterns, the seasonal patterns in the data must be removed; Use the clustering methods to cluster the SST data in which the seasonal patterns have been deleted; Discover the associations among anomalies of SST data and climate events.

Keyword： type-2 fuzzy sets; cloud models; spatio-temporal clustering; sea surface temperature; climate change

References

Cheng H B, Tan P N. Semi-supervised Learning with Data Calibration for Long-Term Time Series Forecasting. Proc of the ACM SIGKDD International Conference on Data Mining, August, 2008.

Cheng H B, Tan P N, Potter C, Klooster S. Detection and Characterization of Anomalies in Multivariate Time Series. Proc of SIAM International Conference on Data Mining (SDM 2009), 2009.

Hwang C, Rhee F. Uncertain Fuzzy Clustering: Interval Type-2 Fuzzy Approach to C-Means, IEEE Transactions on Fuzzy Systems, 2007, 15(1):107-120.

Li D Y, Shi X M, Meng H J. Membership Clouds and Membership Cloud Generators. Journal of Computer Research and Development,1995,32(6): 15-20.

Li D Y, Du Y. Artificial Intelligence with Uncertainty, Chapman & Hall/CRC, 2007.

Karnik N N, Mendel J M. Centroid of a type-2 fuzzy set. Information Sciences,2001,(132):195-220.

Mendel J M, John R I. Type-2 Fuzzy Sets Made Simple. IEEE Transactions on Fuzzy Systems, 2002,10(2): 117-127.

Pedrycz W. Knowledge-Based Clustering: From Data to Information Granules, New York: John Wiley & Sons, 2005

Qin K, Xu M, Li D Y. Cloud Model Based Fuzzy C-Means Clustering and Its Application. Proceeding of SPIE, Geoinformatics 2007: Remotely Sensing Data and Information, Vol.6752,2007.

Steinbach M, Tan P N, Boriah S, Kumar V, Klooster S, and Potter C. The Application of Clustering to Earth Science Data: Progress and Challenges. Proceedings of the 2nd NASA Data Mining Workshop, May 2006.

Yang X, Xie W X, Huang J J. A C-Means Clustering Approach Based on Cloud Model, IEEE International Conference on Fuzzy Systems, pp.965-968,2008.

Zadeh LA.The concept of a linguistic variable and its application to approximate reasoning–1, Information Science, 1975,(8):199–249.