

A STAR SPECTRUM OUTLIER MINING SYSTEM BASED ON SIMULATED ANNEALING

JIANGHUI CAI, JIFU ZHANG AND XUJUN ZHAO

School of Computer Science and Technology
Taiyuan University of Science and Technology
Taiyuan 030024, P. R. China
cjhj@sohu.com

Received July 2007; revised December 2007

ABSTRACT. *Seeking the unknown celestial body is one of main goals for the mankind exploring universe. Outlier mining is an effective way to find the spectrum data of unknown celestial body. Many recent algorithms have been proposed for outlier mining in order to find the outliers based on their relationship to the other points in the data. Many clustering algorithms are used to detect outliers as by-products of the clustering processes. The rational of using clustering algorithms to detect outliers is based on the understanding that outliers and cluster objects are mutually complementally. Unfortunately, in high-dimensional space, they fail to retain their effectiveness. Consequently, for high-dimensional data, the notion of finding meaningful outliers becomes substantially more complex and nonobvious. In this paper, we discuss new technology for outlier detection. A clustering model and algorithm is proposed and researched by using deterministic annealing. The model takes account of the interact ions of clusters; some models which were put forward previously are special cases of this one. Temperature parameter is introduced, and the clustering problem as a physical system is considered. Finding the optimal solution to clustering problem is transformed into simulating the equilibrium state of a physical system. The equilibrium state is simulated by solving a series of problems to minimize the free energy which varies with temperature. Finally, the ground state of the system is attained. That is the optimal solution of clustering problem. Using outlier mining as the way to analyze star spectrum data and VC++, Oracle9i as development tools, the outlier mining system on star spectrum data is designed and realized. The running results of the system show that it is feasible and valuable to apply this method to mine the outliers in star spectrum data.*

Keywords: Outliers, Star spectrum data, Clustering, Simulated annealing, Visual simulation

1. Introduction. The Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST) project is one of the National Major Scientific Projects undertaken by the Chinese Academy of Science. The project's budget is RMB 235 million yuan. It will bring Chinese astronomy into the 21century with a leading role in wide field spectroscopy and in the fields of large scale and large sample astronomy and astrophysics. LAMOST has been constructing by the National Astronomical Observatories for several years, and has been planned to set up in this year. Since LAMOST is planned to acquire ten-millions of spectra as a whole, it is of importance to import a novel way to deal with these very large high-dimensional data sets. Consequently, as an effective way to obtain scientific knowledge, data mining has been introduced into astronomical spectral data processing. Because the state or fact of knowing of chronometer to cosmic is Comparison finitude, one important mission is to discover the new special celestial body. It is worth studying