Abstract

Mining data streams is concerned with extracting knowledge structures represented in models and patterns in high-speed streams of information. It raises new problems for the data mining community in terms of how to mine continuous high-speed data items that you can only have one look. The increasing focus of applications that generate and receive data streams stimulates the need for online data stream analysis tools. Recently, mining data streams with concept drifts has become an important and challenging task for a wide range of applications such as target marketing, network intrusion detection, credit card fraud protection, etc. Clustering and classification ensemble learning is a frequently used tool for building prediction models from data streams, due to its fundamental nature of managing large volumes of stream data. These both are the tools which help in improving the performance of mining systems. In order to improve the accuracy and error rate of traditional ensemble models, we propose a new ensemble model which combines both classifiers and clusters together and utilizes genetic algorithms for mining data streams. The main reason for using genetic algorithms along with clustering and classification is its high ability to solve optimization.
An Efficient Approach to Enhance Classifier and Cluster Ensembles Using Genetic algorithms for Mining Drifting Data Streams

References

- B Peng Zhang, Xingquan Zhu, Jianlong Tan, Li Guo. 2007. Classifier and Cluster Ensembles for Mining Concept Drifting Data Streams. QCIS Centre, Faculty of Eng. & IT, Univ. of Technology.
- Mohamed Medhat Gaber, Arkady Zaslavsky and Shonali Krishnaswamy. Mining Data Streams: A Review. Centre for Distributed Systems and Software Engineering, Monash University, 900 Dandenong Rd, Caulfield East, VIC3145, Australia.
- C. Shafer, R. Agrawal, and M. Mehta. 1996 Sprint: A scalable parallel classifier for data mining. In VLDB.
- J. Gao, W. Fan, J. Han. On appropriate assumptions to mine data streams: analysis and practice, 2003 In Proc. of ICDM.
- S. W. Mahfoud. 1993. Simple analytical models of genetic algorithms for multimodal function optimization. Proc. of 5th ICGA.

Index Terms

Computer Science

Keywords

Classifier And Cluster Ensembles Using Genetic Algorithms genetic Algorithms Based Cluster And Classifier Ensembles For Mining Drifting Data Streams
An Efficient Approach to Enhance Classifier and Cluster Ensembles Using Genetic algorithms for Mining Data Streams

Mining Data Streams