Abstract

Data Stream Mining is the process of extracting knowledge structures from continuous, rapid data records. In these applications, the main goal is to predict the class or value of new instances in the data stream given some knowledge about the class membership or values of previous instances in the data stream. Machine learning techniques can be used to learn this prediction task from labeled examples in an automated fashion. In many applications which are in non-stationary environments, the distribution underlying the instances or the rules underlying their labeling may change over time, i.e., the class or the target value to be predicted may change over time. This problem is referred to as Concept drift[8]. Evolutionary Computations like Genetic Algorithm is a strong rule based classification algorithm which is used for mining static small data sets and inefficient for large data streams. Evolutionary Algorithms are one of the population optimization techniques done by calculating fitness evaluation measures using gene reproduction, crossover, mutation and selection of the individual gene mechanisms. If the Genetic Algorithm can be made scalable and adaptable by reducing its I/O intensity, it will become an efficient and effective tool for mining large data sets like data streams. In this paper a
scalable and adaptable online genetic algorithm is proposed to mine classification rules for the
data streams with concept drifts. The results of the proposed method are comparable with the
other standard methods which are used for mining the data streams.

References

1. Periasamy Vivekanandan and Raju Nedunchezhian, “Mining data streams with concept
Springer, October 2011.
3. Wang H, “Mining Concept-Drifting Data Streams”, IBM T.J. Watson Research Center,
Classification Rules in Data Mining”, International Journal of Computer Applications
(0975-8887), Vol. 41-No. 18, March 2012.
classifiers”, In Proceedings of the 9th ACM SIGKDD international conference on knowledge
6. Syed Shaheena and Shaik Habeeb, “Classification Rule Discovery Using Genetic
Algorithm-Based Approach”, NIMRA Institute, Department of CSE, IJCTT Journal, Vol. 4, Issue
8, pp 2710-2715, August 2013.
7. E Padmalatha, C R K Reddy and Padmaja B Rani. Article: Ensemble Classification for
Streams"In the proceedings of the Fifth International Conference on Information Science and

Index Terms

Computer Science Algorithms

Keywords

Data Stream, conceptdrift, Genetic Algorithm,optimization.