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
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


Index Terms

Computer Science

Algorithms

Keywords

Data Stream, conceptdrift, Genetic Algorithm, optimization.