Abstract

Stream data mining is the process of excerpting knowledge structure from large, continuous data. For stream data, various techniques are proposed for preparing the data for data mining task. In recent years stream data have become a growing area for the researcher, but there are many issues occurring in classifying these data due to erroneous and noisy data. Change of trend in the data periodically produces major challenge for data miners. This research concentrates on incremental missing value replacement for stream data. The proposed method generates the value for the missing data considering the data type and data distribution. It also considers the concept drift in the data stream. The method is applied to different datasets and promising results derived.
References

- Data preprocessing. CCSU http://www.cs.ccsu.edu/~markov/ccsu_courses/DataMining-3.html
- Cw. flek. cvut. cz/lib/exe/fetch.php/sources/ac4m33sad/2_tutorial.pdf.
- Tony Finch, "Incremental calculation of weighted mean and variance"; February 2009.
- Lukasz A. Kurgan, Member, IEEE, and Krzysztof J. Cios, "CAIM Discretization Algorithm"; Senior Member, IEEE, IEEE transactions on knowledge and data engineering,
Incremental Missing Value Replacement Techniques for Stream Data

vol. 16, no. 2, february 2004
- UCI repository dataset, "http://archive.ics.uci.edu/ml/"
- Weka tool "http://www.cs.waikato.ac.nz/ml/weka/Vote data/"
- Contact lance data.

Index Terms

Computer Science
Databases

Keywords

Skewness  Mean  Median  Standard deviation  Discretization.