Rutkowski, Leszek; Jaworski, Maciej; Pietruczuk, Lena; Duda, Piotr
The CART decision tree for mining data streams. (English) Zbl 1339.68229

Summary: One of the most popular tools for mining data streams are decision trees. In this paper we propose a new algorithm, which is based on the commonly known CART algorithm. The most important task in constructing decision trees for data streams is to determine the best attribute to make a split in the considered node. To solve this problem we apply the Gaussian approximation. The presented algorithm allows to obtain high accuracy of classification, with a short processing time. The main result of this paper is the theorem showing that the best attribute computed in considered node according to the available data sample is the same, with some high probability, as the attribute derived from the whole data stream.

MSC:
68T05 Learning and adaptive systems in artificial intelligence

Keywords:
data stream; decision tree; CART; Gini index; Gaussian approximation

Software:
UCI-ml; C4.5

Full Text: DOI

References:
[21] Larose, D., Discovering Knowledge in Data, An Introduction to Data Mining, (2005), Wiley \& Sons · Zbl 1095.68643


[26] Quinlan, J., Learning Efficient Classification Procedures and Their Application to Chess end Games, Learning Efficient Classification Procedures and Their Application to Chess end Games, (1983), Morgan Kaufman San Francisco, CA

[27] Quinlan, J., C4.5: programs for machine learning, (1993), Morgan Kaufman San Francisco, CA


This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.