Abstract

Feature selection involves the process of identifying the most useful feature's subset which produces compatible results similar to original set of feature. Efficiency and effectiveness are the two measures to evaluate feature selection algorithm. The time to find the cluster concerns to efficiency, while effectiveness is concerned to quality of subset feature. With these criteria, fast clustering algorithm was proposed and experimented in two steps. Features are divided into cluster in first step and followed by selection representative feature related to the target class from each cluster. Fast algorithm has the probability of producing a useful and independent feature subset. Performance of this algorithm is evaluated against several
Feature Subset Selection Algorithm for High-Dimensional Data by using FAST Clustering Approach

selection algorithms (FCBF, Relief, and CFs) and it outperforms the other algorithm. The result analyzed from 35 real world dataset (image, microarray, text data) proves not only that FAST produces smaller subset but also improves the performance.

References

- Liu, H. , Motoda, H. and Yu, L.  2004 &quot;Selective Sampling Approach to Active Feature Selection,&quot; Artificial Intelligence, vol. 159, nos. 1/2, pp. 49-74.
- Souza, J.  2004 &quot;Feature Selection with a General Hybrid Algorithm,&quot; PhD dissertation, Univ. of Ottawa.
- Das, S.  2001 &quot;Filters, Wrappers and a Boosting-Based Hybrid for Feature Selection,&quot; Proc. 18th Int&apos;Conf. Machine Learning, pp. 74-81.
Feature Subset Selection Algorithm for High-Dimensional Data by using FAST Clustering Approach


Index Terms

Computer Science
Algorithms

Keywords

Feature Selection Filter Method Feature Clustering Graph-based Clustering.