Knowledge Assisted Visualization for Imbalanced Data Clustering

Abstract

The common challenge which is faced by much of the data clustering techniques is data complexity, which leads to many issues such as overlapping, lack of representative data and class imbalance. This may deteriorates the clustering process. The situation gets worse when the class imbalance is very high. To cluster such imbalanced data sets, better understandings of the dataset and efficient clustering algorithms are required. This could be achieved by integrating suitable domain intelligence into the clustering process. In this paper, Knowledge Assisted Visualization framework is proposed for imbalanced data clustering and validation. The proposed Knowledge Assisted Visualization framework integrates an efficient visual
clustering framework with suitable domain intelligence acquired from domain experts and users into clustering process. An experimental analysis is carried out over a wide range of highly imbalanced data sets. Experiments demonstrate that the proposed method works well with imbalanced dataset and eases the cluster identification and validation in an effective way.

References

- Barbara, D., Chen, P., "Using the fractal dimension to cluster dataset," KDD&amp;apos;00 proceedings of the sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 260-264.
- Jeatrakul P., Wong K. W., Fung C. C., Takama Y., "Misclassification Analysis
Knowledge Assisted Visualization for Imbalanced Data Clustering


- Keim D. A. , Hans-Peter, Kriegel, &quot;Visualization Techniques for Mining Large Databases: A Comparison," IEEE Transactions on Knowledge and Data Engineering, Vol. 8(6), pp. 923-938, 1996.

Index Terms

Computer Science Information Technology

Keywords