Abstract

Many studies in data mining have proposed a new learning called semi-Supervised. Such type of learning combines unlabeled and labeled data which are hard to obtain. However, in unsupervised methods, the only unlabeled data are used. The problem of significance and the effectiveness of semi-supervised clustering results is becoming of main importance. This paper pursues the thesis that much greater accuracy can be achieved in such clustering by improving the similarity computing. Hence, we introduce a new approach of semi-supervised clustering using an innovative new homogeneity measure of generated clusters. Our experimental results demonstrate significantly improved accuracy as a result.

References

A New Homogeneity Inter-Clusters Measure in Semi-Supervised Clustering

- KiriWagstaff, Claire Cardie, Seth Rogers, Stefan Schrödl, "Constrained K-means Clustering with Background Knowledge"; ICML &apos;01 Proceedings of the Eighteenth International Conference on Machine Learning, 2001

Index Terms

Computer Science Artificial Intelligence
Keywords
Semi-supervised clustering distance computation homogeneity measure