Abstract

In machine learning classification and recognition are crucial tasks. Any object is recognized with the help of features associated with it. Among many features only some leads to classify object correctly. Feature selection is useful technique to detect such specific features. Feature selection is a process of selecting subset of features to reduce number of features (dimensionality reduction). Semi-supervised feature selection is difficult due to scarcity of labeled samples. Here constraint based approach is proposed to efficiently select features from semi-supervised data. Constraint based approach is selected as it incorporates supervised information in processing. In the absence of labels, features can be evaluated based on locality preserving ability. Hence for semi-supervised data, properties of both labeled and unlabeled data are combined to choose good features. Constraint based Laplacian score is used to find weight of features. To eliminate redundant features mutual information is calculated and graph
Semi-Supervised Feature Selection with Constraint Sets

based method is used to remove redundant features. Classification accuracy for different
dataset is measured to check performance of system.

References

and relief.
- Zhao Z.  and Liu H. , I.  C.  2007. Spectral feature selection for supervised and
unsupervised learning.
ACM transactions on knowledge discovery from data.
reduction.
- Benabdeslem K.  and Hindawi M. , 2011. Constrained laplacian score for
1080-1085.
partitional clustering algorithms.
- Peng H. , Long F.  and Ding C. , 2005. Feature selection basesed on mutual
information: criteria of max-dependency, max-relevance and min-redundancy.  IEEETrans.
- Hindawi M.  and Benabdeslem K. , 2014. Efficient semi-supervised feature selection:

partitional clustering algorithms, in Proc.  ECML/PKDD.
algorithms.

Index Terms

Computer Science  
Applied Mathematics
Keywords
Constraints  Feature Selection  Redundant  Relevant  Semi-supervised