Abstract

Data mining techniques have been widely used to find new patterns and knowledge from large amounts of data. While Bayesian models were widely used in the early days, more advanced machine learning methods, such as artificial neural networks and support vector machines, have been applied in recent years when these techniques are used in different areas. The problem of text mining has gained increasing attention in recent years because of the large amounts of text data, which are created in a variety of social network, web, and other information-centric applications. Unstructured data is the easiest form of data which can be created in any application scenario. As a result, there has been a tremendous need to design methods and algorithms which can effectively process a wide variety of text applications. This paper will provide an overview of the different methods and algorithms which are common in the text domain, with a particular focus on mining methods.

References

A Survey on Effective Classification for Text Mining using one-class SVM

- Yang Zhang. One-class Classification of Text Streams with Concept Drift. 2008 IEEE International Conference on Data Mining Workshops.
- Gabriel Pui Cheong Fung, Jeffrey X. Yu, Member, IEEE Computer Society, Hongjun Lu, and Philip S. Yu, Fellow, IEEE. Text Classification without Negative Examples Revisited. IEEE Transactions on Knowledge and Data Engineering, vol. 18, no. 1, January 2006
- Z. Jiawei Han and Micheline Kamber. Data mining concepts and techniques book referred third edition 2010
- Data Mining: Concepts and Techniques Second Edition Jiawei Han University of Illinois at Urbana-Champaign Micheline Kamber 2006.

Index Terms

Computer Science
Information Sciences

Keywords
Data mining text mining text categorization partially supervised learning labelling unlabelled data
feature selection
information filtering
SVM