Abstract

Due to rapid growth of on-line information, text classification has become one of key technique for handling and organizing text data. One of the reasons to build taxonomy of documents is to make it easier to find relevant documents, content filtering and topic tracking. LS-SVM is the classifier, used in this paper for efficient classification of text documents. Text data is normally
high-dimensional characteristic, to reduce the high-dimensionality also possible with SVM. In this paper we are improving classification accuracy and dimensionality reduction of a large text data by Least Square Support Vector Machines along with Singular Value Decomposition.

Reference

- A Practical Guide to Support Vector Classification Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin.
- Robert Burbidge, Bernard Buxton “An Introduction to Support Vector Machines for Data Mining “
- Jaiwai, Han, Michelia Kamber Data Mining concepts and Techniques, Morgan Kaufmann publisher 2000.
- D. S´anchez, M.J. Mart´ın-Bautista, I. Blanco C. Justicia de la Torre “Text Knowledge Mining: An Alternative to Text Data Mining” IEEE International Conference on Data Mining Workshops 2008.
- Dustin Boswell “Introduction to Support Vector Machines” August 6, 2002

Index Terms

Computer Science

Information Retrieval
Key words
Text classification        Least-Square Support Vector Machines
Singular Value Decomposition