Abstract

Text categorization is the task of assigning text or documents into pre-specified classes or categories. For an improved classification of documents text-based learning needs to understand the context, like humans can decide the relevance of a text through the context associated with it, thus it is required to incorporate the context information with the text in machine learning for better classification accuracy. This can be achieved by using semantic information like part-of-speech tagging associated with the text. Thus the aim of this experimentation is to utilize this semantic information to select features which may provide better classification results. Different datasets are constructed with each different collection of features to gain an understanding about what is the best representation for text data depending on different types of classifiers.

References

- Ikonomakis M., Kotsiantis S. and Tampakas V.: "Text Classification Using
- Shen D, Sun J-T, Yang Q, Chen Z; Text Classification Improved through Multigram Models; at ACM Transactions at CIKM; Nov. 2006, Virginia, USA.
- Gayathri K, Marimuthu A; Text Document Pre-Processing with the KNN for Classification Using the SVM; Proceedings of 7th International Conference on Intelligent Systems and Control (ISCO 2013) IEEE.
- Zhixing Li, Zhongyang Xiong, Yufang Zhang, Chunyong Liu, Kuan Li; Fast text categorization using concise semantic analysis; Pattern Recognition letters (2011), Elsevier.
- Chagheri S, Calabretto S, Roussey C, Dumoulin C; Feature Vector Construction Combining Structure and Content for Document Classification; 6th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT), 2012 IEEE.
- Celik K; Gungor T; A comprehensive analysis of using semantic information in text categorization; International Symposium on Innovations in Intelligent Systems and Applications, 2013 IEEE.
- Yan Xu; A Study for Important Criteria of Feature Selection in Text Categorization; 2nd International Workshop on Intelligent Systems and Applications (ISA), 2010, IEEE.
- Wang Ziqiang, Qian Xu; Text Categorization Based on LDA and SVM; 2008
Feature Selection for Effective Text Classification using Semantic Information

International Conference on Computer Science and Engineering, IEEE.
- Jiang Xiao-yu, Fan Xiao-zhong, Chen Kang: "Chinese Text Classification Based on Summarization Technique", Third International Conference on Semantics, Knowledge and Grid, 2007 IEEE.
- Chen Z H, Huang L and Murphey Y Li: "Incremental Learning for Text Document Classification", International Joint Conference on Neural Networks, Orlando, Florida, USA, August 2007, IEEE.
- Qingxuan Chen, Dequan Zheng, Tiejun Zhao, Sheng Li: "A Fusion of Multiple Classifiers Approach Based on Reliability function for Text Categorization", Fifth International Conference on Fuzzy Systems and Knowledge Discovery, 2008 IEEE.

Index Terms
Feature Selection for Effective Text Classification using Semantic Information

Keywords

Context  POS tagging  semantic information  text categorization