Abstract

Many data mining techniques have been proposed for mining useful patterns in text documents. However, how to effectively use and update discovered patterns is still an open research issue, especially in the field of text mining. This survey paper is based on effective classification of streamed data for text mining by PNLH & one-class classification SVM for text contained audit, we consider the problem of one-class classification of text streams with respect to concept drift where a large volume of documents arrives at a high speed and with change of user interests and data distribution. In this case, only a small number of positively labelled documents is available for training. And text classification without negative examples revisit, by this we propose a labelling heuristic called PNLH to tackle this problem. PNLH aims at extracting high quality positive examples and negative examples from U and our survey can be used on top of any existing classifiers.
An Effective Supervised Streamed Text Classification Approach for Mining Positive and Negative Examples

An Effective Supervised Streamed Text Classification Approach for Mining Positive and Negative Examples

- S. Wu, C. Yang, and J. Zhou, &quot;Clustering-training for data stream mining,&quot; Sixth IEEE International Conference of Data Mining Workshops, pages 653–656, 2006.
- X. Zhu, P. Zhang, X. Lin, and S. Y., &quot;Active Learning from Data Streams,&quot; Proceedings of the Sixth International Conference on Data Mining, (ICDM'06), 2007.
- X. Jeffrey member, &quot;Text classification without negative examples revisited,&quot; IEEE computer society 2008.
- Z. Zhang Yang, &quot;One-class classification of text streams with concept drift,&quot; University of Queensland Australia, 2008.
- Z. Jiawei Han and Micheline Kamber, &quot;Data mining concepts and techniques,&quot; third edition, 2010.
- Ning Zhong, Yuefeng Li, &quot;Effective Pattern Discovery for Text Mining,&quot; IEEE Transactions on Knowledge and data engineering vol. 24, No. 1 January 2012.

Index Terms

Computer Science
Artificial Intelligence

Keywords
An Effective Supervised Streamed Text Classification Approach for Mining Positive and Negative Examples

Text mining  text categorization  partially supervised learning  labelling unlabelled data  pattern mining

information filtering