URN etd-0407109-183725-2083
Author Jen wen yan
Author's Email Address No Public.
Statistics This thesis had been viewed 50 times. Download 20 times.
Department
Year 2004
Semester 1
Degree Master
Type of Document Master's Thesis
Language zh-TW.Big5 Chinese
Title A Study of Using Data Mining Techniques for Text Classification
Date of Defense 0000-00-00
Page Count 70

Keyword
Data Mining
Text Mining
Classification Analysis
Data Mining
knowledge
data
news

Abstract
In the wake of rapid information upgrade in the current society, to boost competitive edge, we must be able to present, apply, and receive data in an effective, organized, and accurate approach. This study aims at improving quality on text classification via informational and technical support. Data mining technology, currently the most popular application approach, retrieves useful knowledge and information from a vast amount of data. The technology has now been widely applied in journalism and search engines. This paper is going to seek the most proper classification principle out of news data by classification technology, one of the data mining technologies, and evaluate Reuters-21578 portfolio, work on the drawbacks of traditional classification approaches, and enhance text classification quality.
The content of Reuters-21578 is the source for analysis. We classify news tile data based on the primary classification items. During the analysis process, we collect a huge amount of key words so as to find out the number of kinds words that articles are composed. Next, a ID3-tree of policies is made with classification analysis technology to realize the proper categories that key word items characteristics belong to, the research analyzes the aspects of frequency and weighted location for an enhanced accuracy of classification results, In this thesis, the first evaluating criterion is using the values of precision, recall to verify the effectiveness, it may be offer reference contributing to the automation of text classification.
Advisory Committee Chui Cheng Chen - advisor

Files etd-0407109-183725-2083-001.pdf

indicate access worldwide
Date of Submission 0000-00-00