

A classification approach of news web pages from multi-media sources at
Chinese entry website-Taiwan Yahoo! as an example

邱登裕, 李之中, 潘雅真

Information Management

Computer Science and Informatics

chiuden@chu.edu.tw

Abstract

There exists numerous news obviously classified into incorrect categories on Chinese web pages portals. For example, the news dated Aug 13, 2008 with title of “An 78-year-old man completed his Bachelor’s degree” is classified incorrectly into class politics at Taiwan Yahoo website. This phenomenon is owing to mainly the difficulty in automatically classifying Chinese news and the fact that news appearing on web page portals are retrieved from numerous various media sources having various categories. We utilize genetic algorithm to select four feature thresholds used to obtain representative features of each class, and to construct the vector space model for each document. The multi-class SVM classifier is then trained to construct an appropriate classifier to perform automatic classification error detection of Chinese news classification

Keyword : multi-class support vector machine; genetic algorithm