An Automated Error Detection for News Webpages of Chinese Portal

邱登裕, 李之中, 潘雅貞
Information Management
Computer Science and Informatics
leecc@chu.edu.tw

Abstract

There exists some news obviously classified into incorrect categories on Chinese webpage portals. This phenomenon is owing mainly the difficulty in automatically classifying Chinese news and the fact that news appearing on webpage portals is retrieved from numerous media sources. This study integrates genetic algorithms and multiclass support vector machine classifiers to construct an automated classification error detection approach for Chinese news classification. A genetic algorithm is utilized to select four feature thresholds used to obtain representative features/words of each class. The multi-class SVM classifier is then trained to construct an appropriate classifier to aid automated classification error detection. The experiment applies the proposed method to the Chinese news on Taiwan Yahoo!

Keyword：multi-class support vector machine, genetic algorithm, news classification error detection