Information Sciences 394–395 (2017) 38–52

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

A method for multi-class sentiment classification based on an improved one-vs-one (OVO) strategy and the support vector machine (SVM) algorithm

Yang Liu[a], Jian-Wu Bi[a], Zhi-Ping Fan[a,b,*]

[a] Department of Information Management and Decision Sciences, School of Business Administration, Northeastern University, Shenyang 110167, China
[b] State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang 110819, China

ARTICLE INFO

Article history:
Received 26 September 2016
Revised 19 January 2017
Accepted 6 February 2017
Available online 10 February 2017

Keywords:
Multi-class sentiment classification
One-vs-one (OVO) strategy
Support vector machine (SVM) algorithm
Experimental study

ABSTRACT

Multi-class sentiment classification is a valuable research topic with extensive applications; however, studies in the field remain relatively scarce. In the present paper, a method for multi-class sentiment classification based on an improved one-vs-one (OVO) strategy and the support vector machine (SVM) algorithm is proposed. First, an improved OVO strategy is proposed wherein the relative competence weight of each binary classifier is determined according to the K nearest neighbors and the class center of each class in the training sample set concerning the binary classifier. A method for multi-class sentiment classification is proposed based on this improved OVO strategy and the SVM algorithm. After converting the training texts into term feature vectors, the important features (terms) for multi-class sentiment classification are selected using the information gain (IG) algorithm. A binary SVM classifier is then trained on the training feature vectors of each pair of sentiment classes. To identify the sentiment class of a test text, a confidence score matrix of multiple SVM classifiers is constructed based on the results of multiple SVM classifiers. Using this score matrix, the sentiment class of the test text can be determined using the improved OVO strategy. The results of our experimental studies show that the performance of the proposed method is significantly better than that of the existing methods for multi-class sentiment classification.

1. Introduction

With the rapid development of information technology, an increasingly greater number of people are surfing the web. There are more than 3 billion internet users worldwide, and the number of internet users continues to grow rapidly [19]. Internet users typically express their opinions on social issues and share their experiences about products or services by posting online texts to various websites [32,48]. These online texts are valuable for the improvement of government, company and consumer decision making. Governments can make sounder public decisions by analyzing their citizens' online texts on social issues [3], companies can identify product weaknesses and forecast market demand by analyzing online product reviews [6,22,25] and consumers can make suitable purchasing decisions by identifying the sentiment orientation

* Corresponding author at: Department of Information Management and Decision Sciences, School of Business Administration, Northeastern University, Shenyang 110167, China. Tel.:+86-24-8368 7753; Fax: +86-24-2389 1569.
E-mail addresses: liuy@mail.neu.edu.cn (Y. Liu), jianwubi@126.com (J.-W. Bi), zpfan@mail.neu.edu.cn (Z.-P. Fan).

http://dx.doi.org/10.1016/j.ins.2017.02.016
0020-0255/