

LEARNING TO CLASSIFY SEMANTIC ORIENTATION ON ON-LINE DOCUMENT

LEI YU¹, XIANGYANG LIU², FUJI REN^{1,3} AND PEILIN JIANG^{1,4}

¹Graduate School of Advanced Science Technology Education
The University of Tokushima
Tokushima, 7708506, Japan
{yulei; ren; jiang}@is.tokushima-u.ac.jp

²Computer Science Department
Shangqiu Normal University, P. R. China
sqsfxyly@hotmail.com

³Beijing University of Posts and Telecommunications, P. R. China

⁴Institute of Artificial Intelligence and Robotics
Xi'an Jiaotong University
Xi'an, P. R. China

Received July 2008; revised December 2008

ABSTRACT. *As a result of advance in Internet technology, automatic text sentiment classification for a large amount of on-line documents in the form of surveys or called reviews becomes attractive. The task of sentiment classification is to construct an effective classifier with the knowledge data of vocabularies' semantic meaning and the relationships between the vocabularies to determine the sentiment orientation of a document. In this paper, one method combining HowNet knowledge base with a robust supervised sentiment classifier is proposed. It computes semantic similarity of characteristic words and phrases by using HowNet. Sentiment features of text are divided into characteristic words and phrases, and they adopt the positive and negative terms as features of sentiment classifier. Finally in the experiment, evaluation results show the effectiveness of our method.*

Keywords: Sentiment classification, Text semantic orientation, Opinion mining

1. **Introduction.** Being increasingly popular, the Internet greatly changes our lives. A large amount of information is available in on-line documents such as web pages, news-group postings, and personal web-blog. Among these myriad types of available on-line documents, the sentiment or called opinions which people express towards a subject is very useful. For example, the reputation of businessman or their competitors' products or brands is very important for product [17, 21] development, marketing and consumer relationship management. Traditionally, most of companies conduct consumer surveys or called questionnaire for this purpose. Well-designed surveys can provide some quality estimations, but manual analysis of positive and negative opinions for these documents is costly and impractical task for us.

The evaluative character of a vocabulary or text is called its semantic orientation [18]. Semantic orientation of a word or a text is a linguistic term which refers to the direction in which a concept or an opinion is interpreted. For example 'active' has a positive orientation, 'not active' has a negative orientation. This term has been used to some extent co-opted by research in some area of sentiment or affect analysis, where the purpose is trying to report on either the emotional direction of report, or whether the report is positive or negative with respect to some topics.