

SENTIMENTAL CLASSIFICATION BASED ON KERNEL METHODS AND DOMAIN SEMANTIC ORIENTATION DICTIONARIES

CHANGQIN QUAN^{1,2}, FUJI REN² AND TINGTING HE¹

¹Department of Computer Science
Huazhong Normal University
Wuhan 430079, P. R. China
{ quanchqin; tthe }@mail.cnu.edu.cn

²Institute of Technology and Science
The University of Tokushima
2-1 Minamijosanjima, Tokushima city, Japan
{ quan-c; ren }@is.tokushima-u.ac.jp

Received January 2009; revised June 2009

ABSTRACT. *Kernel methods make use of the document information encoded in the inner-product between all pairs of document items, avoiding explicitly the computation of the feature vector for a given input, therefore they get considerable attention in classification tasks. In this paper, we focus our attention on the problem of sentimental classification based on three kernel methods: latent semantic kernel (LSK), polynomial kernel (PK), and Gaussian kernel (GK). It is well known that LSK has good performance in text classification, but it has relative low efficiency because of the process of the SVD decomposition, especially runs on large corpora. Our experiments demonstrate that PK has higher precision and efficiency compared with LSK and GK for the problem of sentimental classification. In particular, we compare the performances on different semantic orientation dictionaries, and find that the domain semantic orientation dictionaries can enhance the performance greatly. Also, our method can categorize the reviews with different degrees, such as 5-star, 4-star, . . . and 1-star by sorting the similarities between the reviews and the semantic orientation dictionaries. In our method, tagged corpus and certain rules are not necessary, so it is practical and has high efficiency.*

Keywords: Sentimental classification, Kernel methods, Semantic orientation dictionaries

1. Introduction. Sentimental classification is to classify a review according to the opinion polarity. This is similar to the task of text classification. However, sentimental classification is concerned with the opinion it expresses. Nowadays, a huge amount of information is available in the online documents such as web pages, newsgroup postings, BBS, blog and wiki. To decide automatically whether a given text expresses a positive or a negative opinion is very useful. For example, knowing a movie is good or bad can be determined by classifying the movie reviews, knowing the reputation of some products or brands can be specified by customers' review with different opinions. Also sentimental classification can be used on search engines, summarizing reviews, and so on. Automatic sentimental classification enables a fast and comprehensive investigation.

There has been extensive researches on sentimental classification. Generally, the fundamental step is to acquire the sentimental orientations of words, terms, phrases or collocations. For example, the word "beautiful" is positive while the word "waste" is negative. Semantic orientation dictionaries can be created in different ways. Hatzivassiloglou and McKeown [1] used textual conjunctions such as "fair and legitimate" or "simplistic but well-received" to get the clusters of similar and opposite words. Peter [2] introduced a