A Statistical Approach to Star Rating Classification of Sentiment

Alexander Hogenboom, Ferry Boon, and Flavius Frasincar

Erasmus University Rotterdam, P.O. Box 1738, NL-3000 DR, Rotterdam, The Netherlands
{hogenboom,frasincar}@ese.eur.nl, ferry.boon@gmail.com

Abstract. Automated analysis of the ever-increasing amount of reviews available through the Web can enable businesses to identify why people like or dislike (aspects of) products or brands, yet to this end, a reliable indication of the intended sentiment of reviews is of crucial importance. This sentiment is typically quantified in universal star ratings, which are not always available. We propose and compare the performance of several statistical methods of automatically classifying star ratings of reviews represented by means of a binary vector representation, with features signaling the presence of sentiment-carrying words. A nearest neighbor classifier maximizes recall, whereas a naïve Bayes classifier excels in terms of precision, accuracy, and the root mean squared error of the assigned number of stars.

Keywords: Sentiment analysis, star ratings, nearest neighbor, naïve Bayes.

1 Introduction

The Web as it exists today encompasses a vast and ever-increasing amount of user-generated content. Popular Web sites like Twitter, Blogger, or Epinions enable anyone to write and publish short messages, blog posts, or reviews about anything at any time. Today's typical Web user exhibits a hunger for and reliance upon on-line advice and recommendations, yet in the wealth of user-generated content, explicit information on user opinions is often hard to find, confusing, or overwhelming [11]. Nevertheless, user-generated content does contain traces of people's sentiment. As recent estimates indicate that twenty percent of all tweets [6] and one third of all blog posts [8] discuss products or brands, automated information monitoring tools for consumer sentiment are crucial for today's businesses.

For such information systems, reviews form an important source of information for, e.g., marketing and reputation management. In reviews, users describe their experiences with a particular brand or product, while implicitly or explicitly expressing what they do or do not like about the subject of their respective reviews. The overall verdict of a review can typically be classified by means of universal star ratings, where the number of stars reflects the extent to which a reviewer intends to convey positive sentiment with respect to the review's subject. Such star classes,

J. Casillas et al. (Eds.): Management Intelligent Systems, AISC 171, pp. 251–260.
springerlink.com

typically five, are defined on an ordinal scale, e.g., a piece of text that is assigned five stars is considered to be more positive than a four-star piece of text.

Star ratings can enable the extraction of valuable information from the multitude of available reviews, as they can facilitate analyses of, e.g., which aspects of an arbitrary product are mentioned in what context in reviews associated with particular ratings. Sentiment analysis techniques can be used to this end. Some of such techniques focus on identifying the subjectivity or objectivity of a text, whereas other techniques aim to determine the polarity of natural language text.

Typical sentiment analysis approaches involve scanning a text for cues signaling subjectivity or polarity, e.g., words, parts of words, or other (latent) features of natural language text, typically in statistics-based machine learning approaches. The use of sentiment lexicons – lists of words and their associated sentiment, possibly differentiated by Part-of-Speech (POS) and/or meaning [1] – has gained attention in recent research endeavors [2, 3]. Such lexicon-based methods have been shown to have a more robust performance across domains and texts than pure machine learning approaches [14]. Additionally, lexicon-based methods allow for intuitive ways of incorporating deep linguistic analysis into the sentiment analysis, for instance by accounting for structural or semantic aspects of text, but this comes at a cost of significant decreases in processing speed with respect to statistical approaches [2].

In order to be able to (semi-)automatically analyze user-generated content for clues as to, e.g., why people like or dislike (aspects of) products or brands, or how different aspects of products contribute to the overall user experience, a reliable indication of intended sentiment associated with this content is of crucial importance. Some Web sites offer users the possibility to assign scores to their reviews in order to express their intended sentiment, but such scores are not always available. For instance, opinionated blog posts or tweets are not typically assigned scores by their respective authors in order to signal their intended sentiment. Therefore, a major challenge is to automatically determine the star rating associated with reviews based on cues in the actual natural language content.

In this light, we propose and compare several statistical methods for classifying the star rating of reviews. In our current endeavors, we aim to contribute to combining the accuracy and processing speed benefits of statistics-based sentiment analysis approaches with the robustness of lexicon-based approaches.

The remainder of this paper is structured as follows. First, we discuss related work on sentiment analysis in Sect. 2. Then, we propose several statistics-based approaches to star rating classification of the sentiment associated with reviews in Sect. 3. An evaluation of our methods is presented in Sect. 4. Last, we conclude and propose directions for future work in Sect. 5.

2 Sentiment Analysis

The research area of sentiment analysis is related to natural language processing, computational linguistics, and text mining. The main goal of sentiment analysis is