Keyword-Based Sentiment Mining using Twitter

M. Baumgarten, Department of Computing & Engineering, University of Ulster, Newtownabbey, UK

M. D. Mulvenna, Department of Computing & Engineering, University of Ulster, Newtownabbey, UK

N. Rooney, Department of Computing & Engineering, University of Ulster, Newtownabbey, UK

J. Reid, RepKnight, Belfast, UK

ABSTRACT

Big Data are the new frontier for businesses and governments alike. Dealing with big data and extracting valuable and actionable knowledge from it poses one of the biggest challenges in computing and, simultaneously, provides one of the greatest opportunities for business, government and society alike. The content produced by the social media community and in particular the micro blogging community reflects one of the most opinion- and knowledge-rich, real-time accessible, expressive and diverse data sources, both in terms of content itself as well as context related knowledge such as user profiles including user relations. Harnessing the embedded knowledge and in particular the underlying opinion about certain topics and gaining a deeper understanding of the overall context will provide new opportunities in the inclusion of user opinions and preferences. This paper discusses a keyword-based classifier for short message based sentiment mining. It outlines a simple classification mechanism that has the potential to be extended to include additional sentiment dimensions. Eventually, this could provide a deeper understanding about user preferences, which in turn could actively and in almost real time influence further development activities or marketing campaigns.

Keywords: Big Data, Data Sources, Keyword-Based Classifier, Micro Blogging Community, Sentiment Dimensions, Short Message Based Sentiment Mining, Social Media Community, User Relations

INTRODUCTION

Today’s connected society is characterized by the way people share information and by how such information affects the community as a whole. This is particular relevant when such information reflects the opinion of individuals about other individuals, companies, products, specific product features, etc. Arguably, Twitter is one of the most popular platforms for publishing opinions and other information to a global audience. In general, such platforms enable the networked community to easily express likes or dislikes, to convey personal feelings or moods, to comment about events or activities of other individuals, to publish news about general
topics or themselves. Individually, they can be seen as simple statements that have limited or no value at all. However, collectively they reflect a powerful mechanism that can actively influence other people’s behavior. For instance, if a large number of people approve of a given product by expressing their satisfaction with it then other people may be more likely to buy this particular product, which is referred to as online word-of-mouth branding. Vice versa, people may choose not to buy a certain product if a substantial amount of people has commented negatively about it. Equally compelling in this context is how individuals think about celebrities for which public opinion can be considered a valuable commodity. Disturbingly in both cases is that it is often irrelevant if the publicized opinions are based on actual facts or if they are based on unfounded information, hearsay, harassment, etc.

Moreover, individual comments are often simply aggregated into positive, negative and sometimes also into neutral categories that supposedly reflect the public’s opinion. Such a simplification is not only problematic but it often leads to false interpretations because expressed sentiments seldom refers to a single general topic but generally to a highly specific context that is defined by multiple factors including time of occurrence, topicality, related topics, demographics, etc. Analyzing the sentiment individuals have towards a given context or person could not only help to assess the reputation of the context or person concerned but could also influence future decisions and actions. However, extracting the sentiment from relatively short and often slang-based messages such as tweets is a non-trivial task. Similar, correlating such sentiments to specific aspects, properties or even complex personal profiles is equally challenging.

The following reviews the foundations and challenges of short message based sentiment analysis, briefly reviews the state of the art in this area and also prototypes a keyword based classifier which categorizes individual tweets into distinct groups that reflect different sentiments. Over time, this can be used to reason about the changing opinion a given context has within the connected community. Towards the end of this paper, a performance study is presented and future work is discussed before concluding remarks are given.

FOUNDATIONS AND CHALLENGES OF SENTIMENT ANALYSIS

Based on Wikipedia, sentiment analysis is defined as an “application of natural language processing, computational linguistics and text analytics” with the objective to identify and to “extract subjective information” from various content. In other words, sentiment analysis refers to the problem of extracting the opinion or emotion a person has about a given context, product, person, etc. While personal opinion can be expressed in various forms such as written, verbal as well as visual, it is, for the scope of this work, assumed that it is expressed or can be converted into short text fragments such as typically published via various micro blogging platforms or topic specific forums. Micro blogging platforms in general but Twitter in particular have become the platform of choice for the real-time publishing of personal messages that relate to products, celebrities, personal or public events or other contexts of interest. On an individual level they simply reflect the personal statement of a single person about a specific topic at a given moment in time. However, collectively they are expressive enough to reason about the social reputation a person or product may have at any moment in time and equally important in real-time. The most basic sentiment analysis is to determine the polarity a given message has in relation to a given topic. That is if the underlying sentiment is positive, negative or neutral. Other alignments such as angeriness, sadness, happiness, etc. reflect more complex dimensions in this area of research that are more challenging to detect. One of the popular methods to determine the polarity is referred to as keyword spotting in which a given text is classified according to the occurrence
Related Content

Providing an Attractive Environment for People to Engage in Health Activities: Serving with Landscape
[www.igi-global.com/article/providing-an-attractive-environment-for-people-to-engage-in-health-activities/176586?camid=4v1a](www.igi-global.com/article/providing-an-attractive-environment-for-people-to-engage-in-health-activities/176586?camid=4v1a)

Computational Intelligence for Functional Testing
[www.igi-global.com/chapter/computational-intelligence-functional-testing/36450?camid=4v1a](www.igi-global.com/chapter/computational-intelligence-functional-testing/36450?camid=4v1a)

Mining E-Mail Messages: Uncovering Interaction Patterns and Processes using E-Mail Logs
[www.igi-global.com/article/mining-mail-messages/2437?camid=4v1a](www.igi-global.com/article/mining-mail-messages/2437?camid=4v1a)
A Novel Approach of Restoration of Digital Images Degraded by Impulse Noise
www.igi-global.com/chapter/a-novel-approach-of-restoration-of-digital-images-degraded-by-impulse-noise/178392?camid=4v1a