



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 8, August 2014

An Efficient Surveillances of Products Based on Opinion Mining

Meenambigai B

Assistant Professor, Dept. of Computer Science, Bharathidhasan College of Arts and Science, Bharathiyar University,
Coimbatore, India

ABSTRACT: Sentiment Analysis is a *Natural Language Processing* and *Information Extraction* task that aims to obtain writer's feelings expressed in positive or negative comments, questions and requests, by analysing a large numbers of documents. Generally speaking, sentiment analysis aims to determine the attitude of a speaker or a writer with respect to some topic or the overall totality of a document. In recent years, the exponential increase in the Internet usage and exchange of public opinion is the driving force behind Sentiment Analysis today. The Web is a huge repository of structured and unstructured data. The analysis of this data to extract latent public opinion and sentiment is a challenging task.

Sentiment analysis is a technique to classify people's opinion in product reviews, blogs or social networks. It has different usages and has received much attention from researchers. In this study, we are interested in product feature based emoticons in sentiment analysis. In other words, we are more interested in identifying the opinion polarities (positive, neutral or negative) expressed on product features. This is termed as the product feature based sentiment analysis. Sentiment Analysis can be performed on both supervised and unsupervised dataset.

Sentiment Analysis identifies the phrases and emoticons in a text that bears some sentiment. The sentiment can be *objective facts* or *subjective opinions*. It is necessary to distinguish between the two. It identifies the polarity and degree of the sentiment. Sentiments are classified as *objective* (facts), *positive* (denotes a state of happiness, bliss or satisfaction on part of the writer) or *negative* (denotes a state of sorrow, dejection or disappointment on part of the writer). The sentiments can further be given a score based on their *degree* of positivity, negativity or neutral. Whenever emoticons are used, their associated sentiment dominates the sentiment conveyed by text and forms a good proxy for intended sentiments.

KEYWORDS: Blog Data, Feature Extraction, Emoticons, Type of opinion

I. INTRODUCTION

Large datasets are available on-line today, they can be numerical or text file and they can be structured, semi-structured or non-structured. Approaches and technique to apply and extract useful information from these data have been the major focuses of many researchers. Many different information retrieval techniques and tools have been proposed according to different data types. Sentiment analysis, also known as opinion mining, is to identify and extract subjective information in source materials, which can be positive, neutral, or negative. Researchers in sentiment analysis have focused mainly on two problems—detecting whether the text is subjective or objective, and determining whether the subjective text is positive or negative. The techniques relied on two main approaches: supervised and unsupervised classifications based on machine learning. The Sentiment Orientation opinion polarity calculations are token based calculations, which calculate the polarity (positive or negative) scores of words based on a small list of words with prior-polarities.

In sentiment analysis, many weighting values have been used for feature values such as term frequency (TF), term presence, term frequency-inverse document frequency (tf-idf). Based on the analysis above, we conducted both unsupervised and supervised learning on product feature based sentiment analysis. The first experiment was



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 8, August 2014

unsupervised learning that calculated sentiment score for each product feature by applying different linguistic rules and constraints. Product features used in this experiment were extracted from unsupervised learning experiment.

In face-to-face communication, sentiment can often be deduced from visual cues like smiling. However, in plain-text computer-mediated communication, such visual cues are lost. Over the years, people have embraced the usage of so-called emoticons as an alternative to face-to-face visual cues in computer-mediated communication like virtual utterances of opinions. In this light, we define emoticons as visual cues used in texts to replace normal visual cues like smiling to express, stress, or disambiguate one's sentiment. Emoticons are typically made up of typographical symbols such as “:.”, “=”, “-”, “)”, or “(” and commonly represent facial expressions. Emoticons can be read either sideways, like “:-)” (a sad face), or normally, like “(^)” (a happy face).

II. RELATED WORK

SENTIMENT ANALYSIS FOR TEXT

Several systems have been built which attempt to quantify opinion from product reviews. Their results show that the machine learning techniques perform better than simple counting methods. They achieve an accuracy of polarity classification of roughly 83%. They identify which sentences in a review are of subjective character to improve sentiment analysis. They do not make this distinction in this system, because we feel that that both fact and opinion contribute to the public sentiment about news entities. They focus on identifying the orientation of sentiment expressions and determining the target of these sentiments. Shallow parsing identifies the target and the sentiment expression; the latter is evaluated and associated with the target. Our system also analyses local sentiments but aims to be quicker and cruder: we charge sentiment to all entities the same sentence as instead of a specific target. They follow-up by employing a feature-term extractor. For given item, the feature extractor identifies parts or attributes of that item, e.g., battery and lens are features of a camera.

EMOTICONS

The nonverbal cues are deemed important indicators for people in order to understand the intentions and emotions of whoever they are communicating with. Translating these findings to computer-mediated communication does hence not seem too far-fetched, if it were not for the fact that plain-text computer-mediated communication does not leave much room for nonverbal cues. However, users of computer-mediated communication have found ways to overcome the lack of personal contact by using emoticons. The first emoticon was used on September 19, 1982 by professor Scott Fahlman in a message on the computer science bulletin board of Carnegie Mellon University. In his message, Fahlman proposed to use “:-)” and “:-)” to distinguish jokes from more serious matters, respectively. It did not take long before the phenomenon of emoticons had spread to a much larger community. People started sending yells, hugs, and kisses by using graphical symbols formed by characters found on a typical keyboard. Thus, nonverbal cues have emerged in computer-mediated communication. These cues are however conceptually different from nonverbal cues in face-to-face communication like laughing and weeping are often referred to as involuntary ways of expressing oneself in face-to-face communication, whereas the use of their respective equivalents “:-)” and “:-)” in computer-mediated communications is intentional. As such, emoticons enable people to indicate subtle mood changes, to signal irony, sarcasm, and jokes, and to express, stress, or disambiguate their (intended) sentiment, perhaps even more than nonverbal cues in face-to-face communication can. Therefore, harvesting information from emoticons appears to be a viable strategy to improve the state-of-the-art of sentiment analysis. Yet, the question is not so much whether, but rather how we should account for emoticons when analysing a text for sentiment.

III. PROPOSED ALGORITHM

The main problem of the existing approach is about matching the attribute values. Since the values are texts composed of many words and are often noisy, accepting only the exact matches is misleading. It is clearly wrong to distinguish between ‘Desktop’ and ‘Desktop Computer’. The problem becomes worse when we try to use an attribute like ‘product description’ which is sometimes composed of full sentences. In this model, each extracted term is treated as an independent attribute. To cope with this model, we redefine the value of an attribute as a set of terms.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 8, August 2014

$$v_i = \{t_{i1}, t_{i2}, \dots, t_{im}\}$$

Where t is a term produced. Then we can reasonably assume that

$$P(a_i = v_i | c_j) = \prod_k P(t_{ik} \text{ appears in } a_i | c_j) = \prod_k \frac{n(c_j, a_i, t_{ik})}{n(c_j, a_i)}$$

Where j, i, k, n, c, a, t denotes the number of occurrences of t_{ik} in a_i of the catalogs that belong to class c_j . Similarly, $(c_j | j)$ in c denotes the sum of frequencies of all terms in a_i of the catalogs that belong to class c_j . n, a denotes the total number of terms appearing in attribute a_i . When none of the terms in training data matches t_{ik} , t appears in a c_j . The classifier is reformulated as follows.

The text below the nodes represents values of an input catalog. Fig.(1) is the structure after extension. In Fig.(1), the terms are generated from each attribute. Using the training data, we count all the parameters. They are stored in tables and retrieved when the classification is performed.

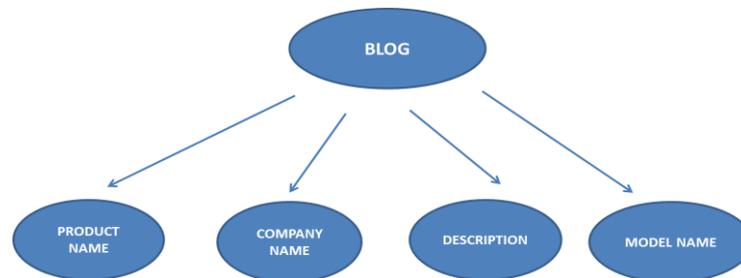


Fig.(1)

Given a product review containing multiple features and varied opinions, the objective is to extract expressions of opinion describing a target feature and classify it as positive or negative. The objectives can be summarized is:

- **Extract all the features from the given review**

In the absence of any prior information about the domain of the review (in the form of untagged or tagged data belonging to that domain), this will give a list of potential features in that review which needs to be pruned to obtain the exact features.

- **Extract opinion words referring to the target feature**

The opinion words are adjectives like hate, love. A naive method, like extracting the opinion words closest to the target feature, does not work so well when the sentence has multiple features and distributed emotions. In the example above, pathetic and not bad are the opinion expressions referring to battery life and multimedia features respectively.

- **Classify the extracted opinion words as positive, negative or neutral**

Each word is retrieved from the blog data and it is compared with the Bag-of-Words(BoW) here it is two files consisting of positive and negative words and if a match occurs then the corresponding count is increased. Finally, if the positive count is more than the blog comment is declared as positive comment else if the negative count is more than the blog comment is declared as negative comment. If both the count are equal then the blog comment is declared as neutral comment.

IV. PSEUDO CODE

- Step 1: Retrieve the blog post from social network.
- Step 2: Extract all the features from the given review
- Step 3: Now split the features into text and emoticons separately.
- Step 4: Extract opinion words referring to the target feature
- Step 5: Classify the extracted opinion words as positive, negative or neutral.
- Step 6: Now, Calculate the total number of positive and negative words.
- Step 7: Repeat from step 2 to step 6 for all text.
- Step 8: End.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 8, August 2014

V. SIMULATION RESULTS

After analysing the sentiment type for various products of different companies, all the result are stored in database. Once we click the button “Sentiment Analysis” the result is displayed in a tabular format with the count of positive, negative and neutral count along with total comments retrieved for analysis.

Query	Positive	Negative	Neutral	Total
Sony Digital camera	12	5		17
Nokia C5	23	5	6	34
iPad - Apple	12	6	1	19
Nokia lumia 520	45	2	5	52
Nokia lumia 720	22	16	6	44
Nokia lumia 920	41	20		61
Nokia lumia 1020	16	20	7	43
Samsung ATIV S	18	11	10	39
Samsung Galaxy Pocket	19	13	12	44
Samsung Galaxy S III	33	15	23	71
Nokia 3020	44	22	14	80
LG Optimus L3 E400	12	4	1	17
LG Optimus L5 E615	10	15		25
LG Optimus Sol E730	9	13	4	26
LG Optimus Vu II Price	15	12	3	30

VI. CONCLUSION AND FUTURE WORK

Sentiment analysis deals with the classification of text based on the sentiments they contain. It focuses on a typical sentiment analysis model consisting of three core steps, namely data preparation, review analysis and sentiment classification and describes representative techniques involved in those steps. Sentiment analysis is an emerging research area in text mining and computational linguistics and has attracted considerable research attention in the past few years. Sentiment analysis has been used to support business and customer decision making by assisting users to explore customer opinions on products that they are interested in recently discussed the potential use of sentiment analysis. Future research shall explore sophisticated methods for opinion and product feature extraction, as well as new classification models that can address the ordered labels property in rating inference. Applications that utilize results from sentiment analysis are expected to emerge in the near future.

As our results are very promising, we envisage several directions for future work. First, we would like to further explore and exploit the interplay of emoticons and text, for instance in cases when emoticons are used to intensify sentiment that is already conveyed by the text. Another possible direction for future research includes applying our results in a multilingual context and thus investigating how robust our approach is across languages. Additionally, future research could be focused on other collections of texts in order to verify our findings in, e.g., specific case studies. Last, we would like to exploit structural and semantic aspects of text in order to identify important and less important text spans in emoticon-based sentiment analysis.

The key contribution of our work lies in our analysis of the role that emoticons typically play in conveying a text’s overall sentiment. The results indicate that people typically use emoticons in natural language text in order to express, stress, or disambiguate their sentiment in particular text segments, thus rendering them potentially better local proxies for people’s intended overall sentiment than textual cues.



ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 8, August 2014

REFERENCES.

- [1] Alexander Pak, Patrick Paroubek. 2010. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. Proceedings of LREC 2010.
- [2] Alekh Agarwal and Pushpak Bhattacharyya, Sentiment Analysis: A New Approach for Effective Use of Linguistic Knowledge and Exploiting Similarities in a Set of Documents to be Classified, International Conference on Natural Language Processing (ICON 05), IIT Kanpur, India, December, 2005.
- [3] Alistair Kennedy and Diana Inkpen, Sentiment classification of movie and product reviews using contextual valence shifters, Computational Intelligence, 22(2):110–125, 2006.
- [4] Dave.D, Lawrence.A, Pennock.D, Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews, Proceedings of International World Wide Web Conference, 2003.

BIOGRAPHY

Meenambigai B is an assistant professor in Department of Computer Science, Bharathidasan College of arts and science, Bharathiyar University, Coimbatore. She received Master of Computer Application (MCA) degree in 2005 from Gobi arts and Science College, Gobichettipalayam, India. Her research interest is in Data Mining.