

REDEFINING SENTIMENT ANALYSIS BASED ON RELATED CONCEPTS AND TASKS

Vishal M. Pandya
IQRA BCA College, Bharuch
Affiliated to VNSGU,
Bharuch, Gujarat, INDIA
vishalmpandya@gmail.com

Dr. Pravinchandra Bhathawala
Computer Science Department,
AURO University, Surat,
Gujarat, INDIA
pcb1010@yahoo.com

ABSTRACT

There is a fast development of Information and Communication Technology since last few years. The evolution of Web in to Social Web has made people, providers of the content rather than consumers, and has made deep changes into the lives of people and societies globally. In this new era Terabytes of information is generated on Web daily which is available to everyone. In early days the problem was to find the sources of information and now the problem is to find and retrieve relevant data from the data and information being flooded from Web every day. The challenge is to devise and imply methods to mine this data and dig out knowledge. The data which is unstructured and written in human language is difficult to process and mine comparing to the structured data (i.e. tables or databases). Natural Language Processing (NLP) comes to the picture here. NLP is a part of the research areas; Artificial Intelligence (AI) and Computational Linguistics. Many application areas of NLP are designed to deal with factual data. Today the Social Web has great influence on societies, people and market worldwide, as people are willingly sharing their knowledge, experience, thoughts and opinions through blogs, micro blogs, forums, review sites, and Social Network platforms like Facebook and Twitter. This influences nature of economy, social life and politics worldwide. ..Here we will first present an overview of the definitions given in the research literature related to NLP for opinion mining, sentiment analysis, and emotion detection. Subsequently definitions of the terms are given which are related to these tasks. Finally we propose an operational definition for Sentiment Analysis that is consistent with the definitions of the different terms related to Sentiment Analysis. The research question that is answered here is “How can Sentiment Analysis be defined in a correct way that can be used to define the related tasks and propose methods for them?”

Keywords- Sentiment Analysis, Opinion Mining, NLP (Natural Language Processing).

I. INTRODUCTION

Today our world is known as “Global Village” and the era is called “Information age”. Information is now a valuable trade factor. There is gigantic amount of data available on Web and it keeps flooding every day. The knowledge is to be mined from this data. Apart from factual data, opinionated data is of great importance and influence to the society, economy and politics. Sentiment Analysis or in broader perspective Opinion Mining transforms this data in knowledge. We also need this knowledge with characteristics of Accuracy, Reliability, and Timeliness. For such high quantity of data we need automatic processing

techniques. Natural Language Processing (NLP) is the field that manages automatic treatment of natural language in text or speech.

Sentiment Analysis and Opinion mining, although different terminologies have been used interchangeably, as they are considered by some authors to point to the same task (Pang and Lee, 2008). Emotion Detection is a closely related task which classifies text according to the emotion expressed. All these research areas are part of the wider field in Artificial Intelligence called Affective Computing (Picard, 1995).

NLP research community has shown growing interest in Sentiment Analysis in past few years. Good amount of research has been performed to develop computational methods for analyzing text to see if it contains opinions, sentiments, emotions or attitudes. There is a vast amount of research done in this field in a past few years but the terminology used in defining various tasks and concepts involved is not uniform across the research community. We aim to propose a set of definitions for the tasks and concepts used for Sentiment Analysis.

II. LITERATURE REVIEW

Most of the work in this field has concentrated on creating methods to discover whether a particular entity is regarded in a positive or negative manner by a specific source. This task is given many names like opinion mining, sentiment analysis, review mining, attitude analysis, appraisal attraction etc. Authors like Pang and Lee (2008) consider the terms “opinion mining” and “sentiment analysis” equivalent. At the same time these both terms are referred to the job of classifying text according to different criteria as shown below;

- The polarity of the sentiment expressed (positive/negative/neutral).
- Pros and cons (Kim and Hovy, 2006)
- Whether the candidate that text is talking about is likely or unlikely to win (Kim and Hovy, 2005)
- Whether the text includes good or bad news (Ku et al., 2005)
- Whether a person agrees or disagrees with a political debate (Balapur et al., 2009)
- Determining the polarity of the outcome (Niu et al., 2005)

As we see that there is high dynamics to the field of sentiment analysis and there is no uniformity and consistency in the terms used by different researchers to define Sentiment Analysis.

According to Pang and Lee (2003), Subjectivity Analysis performed prior to **Sentiment analysis** leads to better result in the latter. In this sense sentiment analysis continues with the classification of identified sentiment-containing text into two or three categories (positive, negative or positive, negative and neutral).

Opinion mining is a two-stage problem and that two different but complementary problems can be identified (Jijkoun et al., 2010):

- **Sentiment Extraction:** “given a set of textual documents, identify phrases, clauses, sentences or entire documents that express attitudes, and determine the polarity of these attitudes (kim and Hovy, 2004)
- **Sentiment Retrieval:** “given a topic (and possibly, a list of documents relevant to the topic), identify documents that express attitudes toward this topic (Ounis et al., 2007)”.

III. CONCEPTS OF SENTIMENT ANALYSIS

Subjectivity: In Philosophy Solomon (2005) defines subjectivity as the subjects and his/her perspective, feelings, beliefs, and desires. Wiebe (1994) proposed the definition of subjectivity as the “linguistic expression of somebody’s opinions, emotions, sentiments, evaluations, beliefs, and speculations”. Bin Liu (2010) defined subjective versus objective sentences in the Handbook of Natural Language Processing (2010) as follows: “An objective sentence expresses some factual information about the world, while a subjective sentence expresses some personal feelings or beliefs”.

Opinion: The Webster dictionary in the definition of Opinion gives following set of synonyms; opinion, view, belief, conviction, persuasion, and sentiment. But it is important to state the fact that these closely related terms have slightly different meanings.

Opinion means a conclusion thought out, yet open to dispute. It is (i) a view, judgment, or appraisal formed in the mind about particular matter. (ii) a belief stronger than impression and less stronger than positive knowledge. (iii) a former expression of judgment or advise by an expert.

Belief implies intellectual assent and often deliberate acceptance.

Conviction implies to a firmly and seriously held belief.

Persuasion suggests a belief based on evidence of its truth.

Sentiment suggests a settled opinion which reflects one’s feelings.

Feeling is defined as the conscious subjective experience of emotion. (Van den Bos, 2006)

These definitions suggest that there are different types of opinions and not all opinions are subjective. Also not all opinions have sentiment attached to them. An objective opinions is of the one like a doctor, and a subjective opinion is one that is based on personal criteria. According to Liu (2010):

- An **opinion** on a feature *f* is a positive or negative view, attitude, emotion or appraisal of *f* from an opinion holder.
- The **holder of an opinion** is the person or an organization that expresses the opinion.

- An **explicit opinion** on feature *f* is an opinion explicitly expressed on *f* in a subjective sentence.
- An **implicit opinion** on feature *f* is an opinion of *f* in an objective sentence.
- An **opinionated sentence** is a sentence that expresses explicit or implicit, positive or negative opinions. The sentence can be subjective or objective.
- **Emotions** are subjective feelings or thoughts.
- An **Attitude** is a “hypothetical construct that represents an individual’s degree of like or dislike for something. (Breckler and Wiggins, 1992).

IV. TASKS RELEVANT TO SENTIMENT ANALYSIS

The classification process of opinionated text is denoted by different terms; opinion orientation, sentiment polarity, polarity, sentiment orientation, polarity of opinion, and semantic orientation.

Sentiment Analysis as an NLP task is defined by Pang and Lee (2008) as “The binary classification task of labeling an opinionated document as expressing either an overall positive or an overall negative opinion is called sentiment polarity classification or polarity classification”.

Nasukawa and Yi (2003) entitled their paper, “Sentiment Analysis: Capturing favorability using natural language processing”.

Affective computing is a branch of Artificial Intelligence (AI) dealing with the designs of systems and devises that can recognize, interpret, and process human affect. The concept includes interdisciplinary work from computer science, psychology and cognitive science.

Emotion Detection and classification is the task of spotting linguistic expressions of emotions from text and classifying them in predefined categories/lables.

Subjectivity Analysis is the task of classifying contents into objective or subjective. Pang and Lee (2004) demonstrate that using subjectivity analysis as a first step towards opinion mining yields better results.

V. THE PROCESS

Based on the literature review and research, we identify following to be subtasks or process of Sentiment Analysis.

1) Creation of resources for Sentiment Analysis: Sentiment Analysis requires lexical resources which contain words and expressions that are subjective with value for polarity. To mine opinion from diverse type of text like from news, blogs, product reviews, micro blogs, tweets and comments, current lexical resources are not enough to capture specifics of opinions which may be direct, indirect or implicit. It is a necessary step to label corpora for the training and evolution of systems implementing sentiment analysis.

Following are several lexical resources that are created by researchers and are freely available for use in the task of opinion mining:

- WordNet Affect (Strapparava and Valitutti, 2004)
- SentiWordNet (Esuli and Sebastiani, 2006)
- Emotion tiggers (Balahur and Montoyo, 2008)
- Micro WNOp (Cerini et al., 2007)
- General Inquirer (Stone et al., 1966)

2) Extraction of opinion from text: The task of extracting opinions from text refers to the problem of spotting exact parts of text where a specifically sought opinion is presented on a specific given target.

3) Sentiment polarity classification: Classification of sentiment is done at various levels depending on the final use of the sentiment analysis task. For example if we are interested in finding out opinion people gave about a movie, the overall sentiment is enough to decide whether to see it or not. But if we are interested in buying a mobile phone, the general opinion given in its product review is not enough, because we might be more interested to know about its specific features like camera, processing power, looks and design etc.

VI. LEVEL OF ANALYSIS

Opinion mining required different methods and approaches based on the level of analysis that is required and the type of text to analyze. Different levels that are considered for sentiment analysis are:

- a) Sentiment Analysis at document level
- b) Sentiment Analysis at sentence level
- c) Feature level/based Sentiment Analysis

Feature level Sentiment Analysis which is also termed as Aspect Based Sentiment Analysis, involves a series of tasks:

Task-1: Identify and extract object feature/aspect that is commented upon by an opinion holder.

Task-2: Determine whether the opinions are positive, negative or neutral on the feature.

Task-3: Group feature synonyms.

Task-4: Present a result as a positive versus negative opinion on each feature.

VII. REDEFINING SENTIMENT AND SENTIMENT ANALYSIS

The operational definition proposed for the concept of **Sentiment** from the understanding given by research literatures, “Sentiment is a settled opinion reflective of one’s feelings”, and feelings as defined by Scherer (2005) “a single component of emotion, denoting the subjective experience process”.

Based on the literature review and research done in the field of Sentiment Analysis and/or Opinion Mining and the study of different concepts and tasks related to it, we propose an operational definition for Sentiment Analysis that may support our research and provide a base for proposing a model for generating influence and associability ratings for entities, as follows:

• “**Sentiment Analysis** is a research area that falls under the research fields of Affective Computing, Natural Language Processing and Linguistic Computing”

• **Sentiment Analysis** is a process of

(1) **Resource creation for Sentiment Analysis**: “To create or improve lexical resources and sentiment corpus to support the process of Sentiment Analysis”.

(2) **Sentiment Retrieval**: “To identify documents that expresses opinions, sentiments, emotions or attitudes towards a specific topic”. The result of which is opinionated text.

(3) **Sentiment Extraction and Polarity**: “From the opinionated text to identify phrases, clauses, sentences or entire document and classify them based on different criteria like to find polarity of sentiment (positive/negative/neutral), to find pros and cons of an object, to find likeliness to win/lose, to find agreement or disagreement to debate, or any other”.

VIII. SUGGESTIONS FOR FURTHER RESEARCH

Using the given definition as a base an effective feature based sentiment analysis model can be proposed to generate influence and associability rating for entities.

Also considering different concepts related to Sentiment Analysis a method can be proposed to identify implicit sentiment from subjective text.

REFERENCES

- [1] Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, Vol 2, Nr. 1-2, 2008.
- [2] Picard, R. (1995). Affective computing. Technical report, MIT Media Laboratory.
- [3] Kim, S.-M. and Hovy, E. (2004). Determining the sentiments of opinions.
- [4] Kim, S.-M. and Hovy, E. (2005). Automatic detection of opinion bearing words and sentences.
- [5] Kim, S.-M. and Hovy, E. (2005). Automatic identification of pro and con reasons in online reviews.
- [5] Kim, S.-M. and Hovy, E. (2005). Automatic identification of pro and con reasons in online reviews.
- [6] Ku, L.-W., Li, L.-Y., Wu, T.-H., and Chen, H.-H. (2005). Major topic detection and its application to opinion summarization.
- [7] Balahur, A., Kozareva, Z., and Montoyo, A. (2009). Determining the polarity and source of opinions expressed in political debates.
- [8] Niu, Y., Zhu, X., Li, J. and Hirst, G. (2005). Analysis of polarity information in medical text.
- [9] Pang, B. and Lee, L. (2003). Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales.
- [10] Jijkoun, V., de Rijke, M., Weerkamp, W. (2010). Generating Focused Topic-Specific Sentiment Lexicons.
- [11] Ounis, I., Macdonald, C., de Rijke, M., Mishne, G., and Soboroff, I. (2007). Overview of the TREK 2006 blog track.
- [12] Solomon, R. C. (2005). Subjectivity.
- [13] Wiebe, J. (1994). Tracking point of view in narrative.
- [14] Liu, B. (2010) Sentiment Analysis and Subjectivity. In Handbook of Natural Language Processing.
- [15] Van den Bos, G. (2006). APA Dictionary of Psychology.
- [16] Breckler, S. J. and Wiggins, E. C. (1992). On defining attitude and attitude theory: Once more with feeling.
- [17] Nasukawa T. and Yi, J. (2003). Sentiment Analysis: Capturing favorability using natural language processing.
- [18] Pang, B. and Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts.

- [19] Strapparava, C. and Valitutti, A. (2004). Wordnet-affect: An affective extension of Wordnet.
- [20] Esuli, A. and Sebastiani, F. (2006). Sentiwordnet: A publicly available resource for opinion mining.
- [21] Balahur, A. and Montoyo, A. (2008). Applying a culture dependent emotion triggers database for the text valence and emotion classification.
- [22] Cerini, S., Compagnoni, V., Demontis, A., Formentelli, M., and Gandini, G. (2007). Micro-wnop: A gold standard for the evaluation of auto-matically compiled lexical resources for opinion mining.
- [23] Stone, P., Dumphy, D. C., Smith, M. S., and Ogilvie, D. M. (1996). The General Inquirer: A Computer Approach to Content Analysis.
- [24] Scherer, K. (2005). What are emotions? And how can they be measured? *Social Science Information*, 3(44), 695-729.