TWITTER SENTIMENT ANALYSIS USING VADER ON PYTHON

Mrs. K. Bhagya Laxmi, Assistant Professor, Department of Computer Science and Engineering, Matrusri Engineering College, Saidabad, Hyderabad, Telangana

B. Yamini, CH. Rakshitha, D. Keerthi, B.E Scholars, Department of Computer Science and Engineering, Matrusri Engineering College, Saidabad, Hyderabad, Telangana

Abstract - The growing popularity of E-commerce, social medias, forums, blogs etc. created a new platform where anyone can discuss and exchange his/her views, ideas , suggestions and experience about any product or services. This trend accumulated a huge amount of user generated data on the web. If this content can be extracted and analyzed properly then it can act as a key factor in decision making. Twitter is one such a platform widely used by people to express their opinions and display sentiments on different occasions. But manual extraction and analysis of this content is an impossible task, as the content is unstructured in nature and it is written in natural language. This situation opened a new area of research called Opinion Mining and Sentiment Analysis. Opinion Mining and Sentiment Analysis is an extension of Data Mining that extracts and analyzes the unstructured data automatically.

Sentiment analysis is an approach to analyze data and retrieve sentiment that it embodies. Twitter sentiment analysis is an application of sentiment analysis on data from Twitter (tweets), in order to extract sentiments conveyed by the user. In the past decades, the research in this field has consistently grown. The reason behind this is the challenging format of the tweets which makes the processing difficult. The tweet format is very small which generates a whole new dimension of problems like use of slang, abbreviations etc.

The main motive of this project is to classify the polarity of the tweet where it is either positive or negative efficiently compared to already existing algorithms and provide the user with live graph of the topic, which gets updated every second in accordance with the tweets posted every second which makes decision making a very easy process by providing data in pictorial way. This project also presents a theoretical comparative analysis of various techniques to the hybrid technique used in this project.

1. INTRODUCTION

In the past few years, there has been a huge growth in the use of microblogging platforms. The increased popularity of E-commerce, social medias, forums, blogs etc. resulted in a huge accumulation of user generated data on the internet in the form of reviews, opinions and comments on different services, events and products and this trend is continually growing day by day. Both consumers and producers are beneficiaries of this content: consumers can consider others opinion and experience while taking decision about any product or services and producers can get clear idea about their product from the consumer point of view and thereby they can increase the quality of the product.

Twitter is one such a major micro-blogging website, having over 100 million users generating over 500 million tweets every day. With such large audience, Twitter has consistently attracted users to convey their opinions and perspective about any issue, brand, company or any other topic of interest. Due to this reason, Twitter is used as an informative source by many organizations, institutions and companies. On Twitter, users are allowed to share their opinions in the form of tweets, using only 140 characters. This leads to people compacting their statements by using slang, abbreviations, emoticons, short forms etc. Along with this, people convey their opinions by using sarcasm and polysemy. Hence it is justified to term the Twitter language as unstructured.

2. Literature Survey

Sentiment analysis, or opinion mining, is an active area of study in the field of natural language processing that analyzes people's opinions, sentiments, evaluations, attitudes, and emotions via the computational treatment of subjectivity in text.

2.1 Sentiment Lexicons

A substantial number of sentiment analysis approaches rely greatly on an underlying sentiment (or opinion) lexicon. A sentiment lexicon is a list of lexical features (e.g., words) which are generally labeled according to their semantic orientation as either positive or negative (Liu, 2010). Manually creating and validating such lists of opinion-bearing features, while being among the most robust methods for generating reliable sentiment lexicons, is also one of the most time-consuming. For this reason, much of the applied research leveraging sentiment analysis relies heavily on preexisting manually constructed lexicons. Because lexicons are so useful for sentiment analysis, we briefly provide an overview of several benchmarks. We first review three widely used lexicons (LIWC1, GI2, Hu-Liu043) in which words are categorized into binary classes (i.e., either positive or negative) according to their context free semantic orientation. We then describe three other lexicons (ANEW4, SentiWordNet5, and SenticNet6) in which words are associated with valence scores for sentiment intensity.

2.1.1 Semantic Orientation (Polarity-based) Lexicons

LIWC is text analysis software designed for studying the various emotional, cognitive, structural, and process components present in text samples. LIWC uses a proprietary dictionary of almost 4,500 words organized into one (or more) of 76 categories, including 905 words in two categories especially related to sentiment analysis (see Table):

LIWC Category	Examples	No. of Words
Positive Emotion	Love, nice, good, great	406
Negative Emotion	Hurt, ugly, sad, bad, worse	499

Table 2.1: Example words from two of LIWC's 76 categories. These two categories can be leveraged to construct a semantic orientation-based lexicon for sentiment analysis

LIWC is well-established and has been both internally and externally validated in a process spanning more than a decade of work by psychologists, sociologists, and linguists (Pennebaker et al., 2001; Pennebaker et al., 2007). Its pedigree and validation make LIWC an attractive option to researchers looking for a reliable lexicon to extract emotional or sentiment polarity from social media text. For example, LIWC's lexicon has been used to extract indications of political sentiment from tweets (Tumasjan, Sprenger, Sandner, & Welpe, 2010), predict the onset of depression in individuals based on text from social media (De Choudhury, Gamon, Counts, & Horvitz, 2013), characterize the emotional variability of pregnant mothers from Twitter posts (De Choudhury, Counts, & Horvitz, 2013), unobtrusively measure national happiness based on Facebook status updates (Kramer, 2010), and differentiating happy romantic couples from unhappy ones based on their instant message communications (Hancock, Landrigan, & Silver, 2007). However, as Hutto, Yardi, & Gilbert (2013) point out, despite its widespread use for assessing sentiment in social media text, LIWC does not include consideration for sentiment-bearing lexical items such as acronyms, initialisms, emoticons, or slang, which are known to be important for sentiment analysis of social text (Davidov, Tsur, & Rappoport, 2010). Also, LIWC is unable to account for differences in the sentiment *intensity* of words.

For example, "The food here is *exceptional*" conveys more positive intensity than "The food here is *okay*". A sentiment analysis tool using LIWC would score them equally (they each contain one positive term). Such distinctions are intuitively valuable for fine-grained sentiment analysis.

2.1.2 Sentiment Intensity (Valence-based) Lexicons

Many applications would benefit from being able to determine not just the binary polarity (positive versus negative), but also the *strength* of the sentiment expressed in text. Just how favorably or unfavorably do people feel about a new product, movie, or legislation bill? Analysts and researchers want (and need) to be able to recognize changes in sentiment *intensity* over time in order to detect when rhetoric is heating up or cooling down (Wilson, Wiebe, & Hwa, 2004). It stands to reason that having a general lexicon with strength valences would be beneficial. The Affective Norms for English Words (ANEW) lexicon provides a set of normative emotional ratings for 1,034 English words (Bradley & Lang, 1999). Unlike LIWC or GI, the words in ANEW have been ranked in terms of their pleasure, arousal, and dominance. ANEW words have an associated sentiment valence ranging from 1-9 (with a neutral midpoint at five), such that words with valence scores less than five are considered unpleasant/negative, and those with scores greater than five are considered pleasant/positive. For example, the valence for *betray* is 1.68, *bland* is 4.01, *dream* is 6.73, and *delight* is 8.26. These valences help researchers measure the intensity of expressed sentiment in microblogs (De Choudhury, Counts, et al., 2013; De Choudhury, Gamon, et al., 2013; Nielsen, 2011) – an important dimension beyond simple binary orientations of positive and negative. Nevertheless, as with LIWC and GI, the ANEW lexicon is also insensitive to common sentiment-relevant lexical features in social text.

3. OVERVIEW OF THESYSTEM

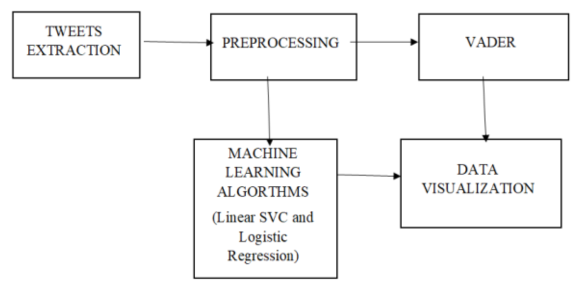

Fig 3.1 System Architecture

3.1 Problems with Existing System

Sentiment analysis played a great role in the area of researches done by many; there are many methods to carry out sentiment analysis. There are currently three different approaches to carry out sentiment analysis. In this section, let us look at each existing approach and the problems associated with them.

3.1.1 Machine Learning approach

Machine learning strategies work by training an algorithm with a training data set before applying it to the actual data set. Machine learning techniques first trains the algorithm with some particular inputs with known outputs so that later it can work with new unknown data. Again machine learning algorithms are of two types supervised and unsupervised learning. Supervised learning is the machine learning task of learning a function that maps an input to an output based on example input-output pairs. It infers a function from labelled training data consisting of a set of training examples.

Disadvantages

- The main disadvantage of existing machine learning models is their reliance on labelled data. It is extremely difficult to ensure that sufficient and correctly labelled data can be obtained.

- The other disadvantage is that the classifier trained on the texts in one domain in most cases does not work with other domains.
- Machine Learning approach needs enough time to let the algorithms learn and develop enough to fulfill their purpose with a considerable amount of accuracy and relevancy. It also needs massive resources to function. This can mean additional requirements of computer power for you.

3.1.2 Rule based approach

Rule based approach is used by defining various rules for getting the opinion, created by tokenizing each sentence in every document and then testing each token, or word, for its presence. If the word is there and has with a positive sentiment, a +1 rating was applied to it. Each post starts with a neutral score of zero, and was considered positive. If the final polarity score was greater than zero, or negative if the overall score was less than zero. After the output of rule based approach it will check or ask whether the output is correct or not. If the input sentence contains any word which is not present in the database which may help in the analysis of movie review, then such words are to be added to the database. This is supervised learning in which the system is trained to learn if any new input is given.

Disadvantages

- Lower recall.
- Difficult and tedious to list all the rules.
- Generating rules for a complex system is quite challenging and time consuming.
- Efficiency and accuracy depend the defining rules.

3.1.3 Lexicon based approach

Lexicon Based techniques work on an assumption that the collective polarity of a sentence or documents is the sum of polarities of the individual phrases or words. In the seminar ROMIP 2012 the lexicon based method proposed in [14] was used. This method is based on emotional research for sentiment analysis dictionaries for each domains. Next, each domain dictionary was replenished with appraisal words of appropriate training collection that have the highest weight, calculated by the method of RF (Relevance Frequency) [15]. The word-modifier changes (increases or decreases) the weight of the following appraisal word by a certain percentage. Word-negation shifts the weight of the following appraisal word by a certain offset: for positive words to decrease, for negative to increase. The procedure of the text sentiment classification was carried out as follows. First weights of all training texts the classified text is calculated. All the texts are placed into a one dimensional emotional space. The proportion of deletions was determined by the cross-validation method. Then the average weights of training texts for each sentiment class were found. The classified text was referred to the class which was located closer in the one dimensional emotional space.

Disadvantages

- Requires powerful linguistic resources which is not always available.

3.2 Proposed System

In the proposed system, we not only classify tweets using only VADER but also calculate confidence score of the tweets using VADER and machine learning algorithms LinearSVC and Logistic Regression, to provide better analysis. Further these sentiment scores are represented as graphs to provide better visualization.

4. IMPLEMENTATION

The current study consists of four phases. Phase one concerns the acquisition of Twitter data. Phase two focuses on the initial preprocessing work carried out to clean and remove irrelevant information from the tweets. Phase three deals with the use of the NLTK's VADER (Valence Aware Dictionary for Sentiment Reasoning) analyzer as well as the scoring method applied to the VADER results to assess its ability to classify tweets on a five-point scale. Phase four deals with the plotting of positivity, negativity, neutral, compound (overall) score of a tweet produced by the VADER.

As aforementioned that, the purpose of the data acquisition phase was to obtain Twitter data. The methods used to extract Twitter data allowed real-time access to publicly available raw tweets. To gather the data, we used TweePy API.

A tweet is a microblog message posted on Twitter. It is limited to 140 characters. Most tweets contain text and embed URLs, pictures, usernames, and emoticons. They also contain misspellings. Hence, a series of preprocessing steps were carried out to remove irrelevant information from the tweets. The reason is that the cleaner the data, the more suitable they are for mining and feature extraction, which leads to the improved accuracy of the results. The tweets were also preprocessed to eliminate duplicate tweets and retweets from the dataset. To preprocess these data, we used Python's Natural Language Toolkit (NLTK). Various functions of NLTK were used to convert the tweets to lowercase, remove stop words (i.e., words that do not express any meaning, such as is, a, the, he, them, etc.), tokenize the tweets into individual words or tokens, and stem the tweets. When the preprocessing steps are complete, the dataset was ready for sentiment classification.

In phase three, the sentiments expressed in the tweets were classified. VADER Sentiment Analyzer was applied to the dataset. VADER is a lexicon and rule-based sentiment analysis tool and a lexicon that is used to express sentiments in social media. VADER contains a systematically built sentiment lexicon, together with some syntactic rules to further improve the sentiment analysis. VADER was constructed especially for tweets, and contains both abbreviations and emojis. Emojis are emotional tokens, which are often used on the Internet. The rules in the application handle degree modifiers. Examples of these rules are:

- **Exclamation and interrogation marks**. These increase or decrease the sentiment intensity.
- **Capitalization:** If a sentiment laden word is capitalized while others are not, the sentiment intensity increases for this word.
- **Negators:** If a sentiment laden word is preceded by a negator such as "not", this reverts the sentiment. A positive sentiment turns negative and vice versa.
- **Booster Words**, such as "extremely", which if positioned before the word "good" will increase the sentiment of this word.

For the development of VADER, a combination of qualitative and quantitative methods are used to produce, and then empirically validate, a gold-standard sentiment lexicon that is especially attuned to microblog-like contexts. Next combined these lexical features with consideration for five generalizable rules that embody grammatical and syntactical conventions that humans use when expressing or emphasizing sentiment intensity. We find that incorporating these heuristics improves the accuracy of the sentiment analysis engine across several domain contexts (social media text, NY Times editorials, movie reviews, and product reviews). Interestingly, the VADER lexicon performs exceptionally well in the social media domain. The correlation coefficient shows that VADER (r = 0.881) performs as well as individual human raters (r = 0.888) at matching ground truth (aggregated group mean from 20 human raters for sentiment intensity of each tweet). Surprisingly, when we further inspect the classification accuracy, we see that VADER (F1 = 0.96) actually even outperforms individual human raters (F1 = 0.84) at correctly classifying the sentiment of tweets into positive, neutral, or negative classes.

VADER not only tells about the positivity and negativity scores but also tells about how positive or negative a sentiment is.

5. RESULTS

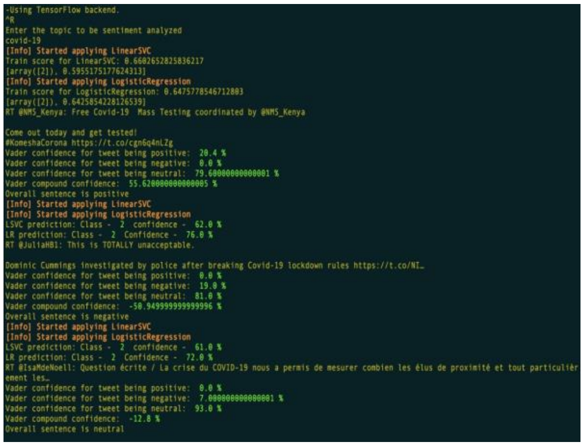

Fig 5.1: Output of sentiment scores of tweets

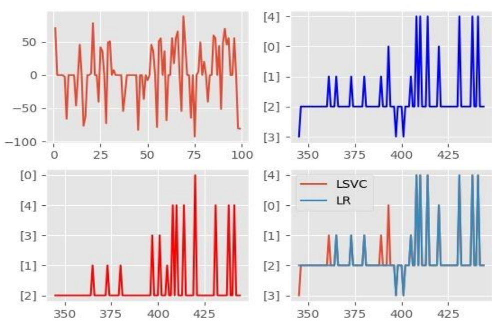

Fig 5.2: Graph of sentiment scores at instance 1

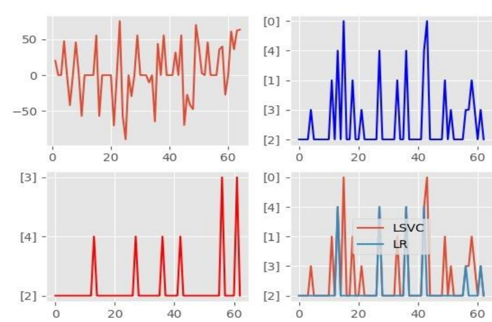

Fig 5.3: Graphs of sentiment scores at instance 2

6. CONCLUSION

VADER's is a simple lexicon and rule-based approach which not only gives polarity but also tell how positive or negative a sentence. However, when compared to sophisticated machine learning techniques, the simplicity of VADER carries several advantages. First, it is both quick and computationally economical without sacrificing accuracy. Running directly from a standard modern laptop computer with typical, moderate specifications (e.g., 3GHz processor and 6GB RAM), a corpus that takes a fraction of a second to analyze with VADER can take hours when using more complex models like SVM (if training is required) or tens of minutes if the model has been previously trained. Second, the lexicon and rules used by VADER are directly accessible, not hidden within a machine-access only black-box. VADER is therefore easily inspected, understood, extended or modified. By exposing both the lexicon and rule-based model, VADER makes the inner workings of the sentiment analysis engine more accessible (and thus, more interpretable) to a broader human audience beyond the computer science community. Sociologists, psychologists, marketing researchers, or linguists who are comfortable using LIWC should also be able to use VADER. Third, by utilizing a general (human-validated) sentiment lexicon and general rules related to grammar and syntax, VADER is at once both self-contained and domain agnostic – it does not require an extensive set of training data, yet it performs well in diverse domains. We stress that in no way do we intend to convey that complex or sophisticated techniques are in any wrong or bad. Instead we show that a simple, human-centric, interpretable, computationally efficient approach can produce high quality results – even outperforming individual human raters.

REFERENCES

- Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment analysis of Twitter data. In Proc. WLSM-11s.
- Akkaya, C., Wiebe, J., & Mihalcea, R. (2009). Subjectivity word sense disambiguation. In Proc. EMNLP-09.
- Baccianella, S., Esuli, A., & Sebastiani, F. (2010). SentiWordNet 3.0. In Proc. of LREC-10.
- Bradley, M. M., & Lang, P. J. (1999). Affective norms for English words (ANEW): Instruction manual and affective ratings.
- Vader: A parsimonious rule-based model for sentiment analysis of social media text, CJ Hutto, E Gilbert

- Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting Elections with Twitter:. In Proc. ICWSM-10.
- Turney, P. D., & Littman, M. L. (2003). Measuring praise and criticism. ACM Trans. Inf. Syst., 21(4), 315–346.
- Wang, H., Can, D., Kazemzadeh, A., Bar, F., & Narayanan, S. (2012). A system…real-time Twitter sentiment analysis. ACL-12.
- Wilson, T., Wiebe, J., & Hwa, R. (2004). Just how mad are you? finding strong and weak opinion clauses. In Proc. NCAI-04s.