

Measuring human and Vader performance on sentiment analysis

Ajla Kirlić¹, Zeynep Orhan²

¹(American Univeristy in Bosnia and Herzegovina, Sarajevo, Bosnia and Herzegovina)

²(International Burch University, Sarajevo, Bosnia and Herzegovina)

ABSTRACT: Sentiment analysis was examined on Tweeter data and neutral polarity was excluded out of research. In this paper on Tweets were determined polarities in two ways: by group of ten people and also by Vader sentiment analysis. In total was examined 527 Tweets from 10 different companies. At the end obtained results were compared to see if there is significant similarity among the methodologies. Results showed that there is no significant difference among human and Vader sentiment.

KEYWORDS: human, polarity, sentiment analysis, Vader

I. INTRODUCTION

Lately topic of sentiment analysis becomes very popular for the researchers. There are many innovative products that are developed related to measurement of sentiment. Beside products there are many methods that help measuring sentiment including machine learning techniques and lexical-based approaches [1]. Up to now is uncertain which method is better for deciding about polarity. For that reason there is need to conduct comparison of methodologies. When comparison is done we can understand better disadvantages, advantages and limitations of methods [1]. The idea for this paper came from the study on Tweets and change on stock exchange prices. For owners and potential investors constantly following news, Tweets and price movements is essential in predicting future movements of prices [2]. It is adequate to say that connection between market and investor can be investigated and help to optimize strategy of investor. However, in demand to relate stock prices and Tweets about company, whole market or certain sector we need to determine whether the Tweets that we have are negative, positive. [3]. Usage of sentiment analysis allows deciding about nature of evaluation on text. Polarity perspective can help to decide and determine important impacts on stock prices of company [4].

Main difficulty while dealing with sentiment analysis is the withdrawal of forms from huge data collections and making conclusions out of them. Comparing the old-fashioned way of collecting information, today with help of technological systems that process is more efficient, more reliable and precise. If we want to gain some useful information, then we need to deal properly with data collections otherwise that will remain just data collection. A principally motivating example of study is that of the financial markets [2]. Nowadays it is easy to access information, since market competence relies upon the accessibility of information. Recent technological revolution is fluctuating dramatically the way how we look to all aspects of life. Interaction of people with technology is generating enormous informational datasets and documenting cooperative behavior [5].

In this paper we collected Tweets regarding 10 companies that will be mentioned in further text. Evaluation of Tweets was made regarding the human and Vader sentiment. In group of three people was deciding polarity of Tweets and results were compared to the Vader's sentiment using χ^2 test (Chi-square test) for association.

II. RELATED WORK

Since sentiment analysis is very popular topic in last decade, there are many papers published regarding that topic. We will mention some of the interesting studies regarding sentiment analysis. Interesting study was conducted by Hutto, Gilbert and Arbour [6] where was used Vader for deciding about polarity and strength of sentiments. Vader sentiment was compared with other sentiment classifiers. Study was done on amazon reviews, New York Times reviews and movie reviews. Data was used to help in determining accuracy of the classifiers. Bean [7] conducted a study where in tagging of Twitter keyword about satisfaction of airlines services was involved polarity and sentiment analysis. Sentiment detection algorithm was used in the Bean's research. As a result was concluded that sentiment analysis is giving quick and reliable insight and idea for improvement on customer satisfaction ratings. Research of Ma [8] classified articles from Reuter Corpus and Wall Street Journal as positive and negative with usage of Maximum Entropy and Naive Bayes classifier.

However study of Parikh and Movassate [9] involved sentiment analysis of Tweets. Researchers noticed that classification of Tweet sentiment requires different approach than product reviews. Sentiment of tweets was perceived by usage of Multinomial Bigram and Unigram Naïve Bayes. Two researchers Hashler and Nagar [10] introduced text mining that is automated and has based approach to sum up text from different sources and all together create News Corpus. Middle part (Corpus) is filtered until significant sentences and analyzed using Natural Language Processing techniques (NLP). In the research of these two authors as representation of sentiment is proposed negative and positive polarity. They collected news and made sentiment analysis by using open source packages. At the end they concluded that there is a strong correlation among movement of market price and time difference of News Sentiment. On the other hand Barbosa and Feng [11] in their research didn't explain how they collected data but they used predictions of polarity from three different websites to train a model. In total they collected 1000 tweets for training and 1000 tweets for testing. Their idea was also to include hash tags, links, re tweet together as prior polarity.

Vader (Valence Aware Dictionary for sentiment Reasoning) :Vader as a text sentiment analysis uses combining qualitative analysis and empirical authentication in terms of using human raters and it is introduced in 2014. It is sensitive to strength of emotion and polarity [12].Sentiment analysis using Vader relies on a dictionary which depending on emotion intensity is also known as sentiment score, maps lexical features. Score of a text is obtained aggregating intensity of each word that is in the text. In tweets beside words is possible to find emoticons, slangs or acronyms and with Vader there is ability to handle it. In this paper sentiment score is in the range between -1 and 1 and score is calculated by summing up the score of each Vader-dictionary registered word in the tweet [12].

Hutto [12] standardization is used

$$\frac{x}{\sqrt{x^2+\alpha}} \tag{1}$$

Here x is representing aggregation of sentiment scores of words in Tweet and α is standardization restriction that is agreed to be 15. It is obvious that how x is increasing it is getting closer to -1 or 1. Also it is known that Vader is best on short documents like in this case are Tweets [12].

In Table 1 is example of applying Vader sentiment score on Tweets:

Table 1. Representation of Vader sentiment score on Tweets

Text of Tweet	Vader Score	Vader Polarity
Amazon just made a major move into the restaurant food delivery business	0,1779	Positive
Sometimes I forget what I ordered from @amazon in the middle of the night	-0,2263	Negative
Buy Shares of General Motors Company \$GM on Weakness on Insider Selling	-0,1531	Negative
NO thanks Volkswagen! Take them back and disassemble them! Your poisonous lying and cheats precede you.	-0,8889	Negative
Free Amazon gift cards	0,967	Positive
There are NO hidden extras with #Citroen #vans – just brand-new, quality vehicles at discount prices	-0,4466	Negative
Ingersoll strike crippling production of one of General Motors' best-selling products, the Chevrolet Equinox	-0,128	Negative
Amazon customer support is always rubbish	0,5267	Positive
@amazon you can't manage order do not fool people	0,3412	Positive
I barely know how to use Microsoft... and I put that skill on my resumé. You can be sure I don't know how to use twitter	0,3182	Positive
It's frustrating that Microsoft is so obviously pushing to add ways to recoup what they paid	-0,4404	Negative
Three arrested after Greenpeace volunteers storm Volkswagen ship #Parkesandco	-0,4767	Negative

If Vader score >0.1 it was taken as positive, if it is <-0.1 then it was taken as negative, and if it was between -0.1 and 0.1 it is taken as neutral.

In Table 2 is comparison of human and Vader sentiment:.

Text Of Tweet	Vader Score	Vader Polarity	Human Polarity
Social media is a bummer lately. That's why I get my news from Amazon. Ooh! My new sheets will be delivered Monday!	-0,3578	Negative	Positive
Amazon just made a major move into the restaurant food delivery business	0,1779	Positive	Positive
Enter to #win a \$50 Amazon Gift Card from #PDangelico	0,5574	Positive	Positive
Free Amazon gift cards	0,967	Positive	Positive
There are NO hidden extras with #Citroen #vans – just brand-new, quality vehicles at discount prices	-0,4466	Negative	Positive
Ingersoll strike crippling production of one of General Motors' best-selling products, the Chevrolet Equinox	-0,128	Negative	Positive
Amazon customer support is always rubbish	0,5267	Positive	Negative
@amazon you can't manage order do not fool people	0,3412	Positive	Negative
I barely know how to use Microsoft... and I put that skill on my resumé. You can be sure I don't know how to use twitter	0,3182	Positive	Negative
It's frustrating that Microsoft is so obviously pushing to add ways to recoup what they paid	-0,4404	Negative	Negative
Three arrested after Greenpeace volunteers storm Volkswagen ship #Parkesandco	-0,4767	Negative	Negative
NO thanks Volkswagen! Take them back and disassemble them! Your poisonous lying and cheats precede you.	-0,8889	Negative	Negative

Table 2. Sample of comparison among Vader and human sentiment

III. METHODOLOGY AND RESULTS

Tweets were collected from 10 different companies: Amazon, Apple, Citroen, EBay, General Motors, IBM, Lukoil, Samsung, and Volkswagen. In total was collected 527 Tweets. Research was conducted among sentiment of group of people and sentiment of Vader. In research were included 10 people who were subjectively deciding about polarity of Tweet. If seven people said for particular Tweet that polarity is positive then that was accepted as a Tweet with positive polarity, otherwise if seven people said for Tweet that polarity is negative, Tweet was accepted as a Tweet with negative polarity. Ho: There is significant similarity among decision making of polarity between human and Vader. For the methodology was chosen Chi Square Test of Association, that is commonly used as a method of comparing proportions. In Fig. 1 it is visible how much positive and negative sentiment we have by human and how much by Vader sentiment analysis.

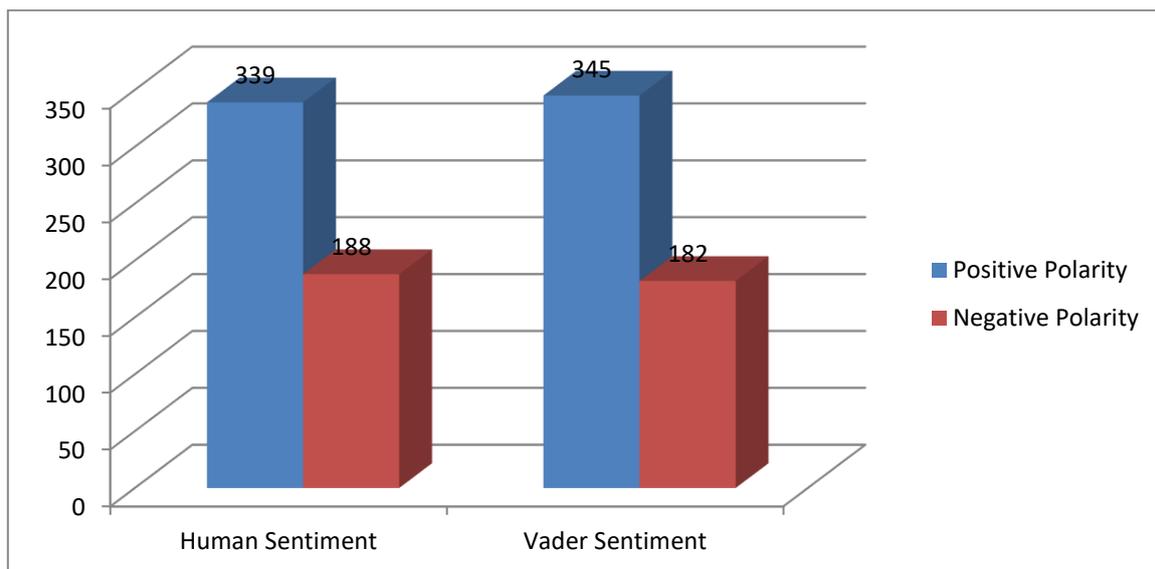


Figure 1. Representation of positive and negative sentiments

χ^2 (CHI) is calculated by using formula (2):

$$\chi^2(t, c) = \frac{N \times (AD - CB)^2}{(A+C) \times (B+D) \times (A+B) \times (C+D)} \quad (2)$$

$$N = A + B + C + D \quad (3)$$

Table 3. Representation of members in formula

$A = \#(t, c)$	$C = \#(-t, c)$
$B = \#(t, -c)$	$D = \#(-t, -c)$

When we do calculation part (4) we obtain that:

$$\chi^2 = 0,1499 \quad (4)$$

It is important to mention that in this research degree of freedom is 1. We used 0.05 as level of significance and we obtained that p-value = 0.698603.

Table 4. Representation of results

N	DF	Chi-Sq	P-Value
1054	1	0.1499	0.698603

In these results, the p-value is 0.698603. Because the p-value is greater than the significance level of 0.05, we fail to reject the null hypothesis. Therefore, we cannot conclude that the observed proportions are significantly different from the specified proportions.

IV. CONCLUSION

Although there are many research papers regarding the topic of sentiment analysis we hope that our research will make contribution in that field. In this paper we were comparing human and Vader sentiment and obtained that our null hypothesis failed to be rejected. Our research is making contribution to knowledge discovery in terms of comparative study of different approaches in sentiment analysis. Despite some limitations in our research like subjectivity in human decision of Tweet polarity it is believed that according to observed results there is no difference between Vader and human decision making. In future work we are hoping to examine how sentiment of Tweets is affecting financial market and to examine whether there are connections among polarity of Tweets and changes in stock prices. Additionally it is believed in future work if the threshold was changed that might be found some other insights among Vader and human sentiment.

REFERENCES:

- [1]. Filipe Nunes Ribeiro, Matheus Araújo, Pollyanna Gonçalves, Fabrício Benevenuto, Marcos André Gonçalves: *SentiBench - a benchmark comparison of state-of-the-practice sentiment analysis methods*(2016)
- [2]. Malkiel, B. G, Fama, E. F. Efficient capital markets: A review of theory and empirical work *Journal of Finance* 25, 383-417 (1970).
- [3]. Stefan Feuerriegel, Dirk Neumann: *Evaluation of News-Based Trading Strategies* (2012)
- [4]. Paul Tetlock, Maytal Saar and Sofus Macskassy: More than words: *Quantifying Language to Measure Firms' Fundamentals*, *Journal of Finance Forthcoming* (2011)
- [5]. King, G. *Ensuring the data-rich future of the social sciences*. *Science* 331, 720-721 (2011).
- [6]. Hutto, C.J. & Gilbert, E.E, Anne Arbour: *A Parsimonious Rule based Model for Sentiment Analysis of Social Media Text* (2014)
- [7]. J. Bean, R by example: *Mining Twitter for consumer attitudes towards airlines*, In *Boston Predictive Analytics Meetup Presentation*, (2011)
- [8]. Qicheng Ma, "Stock Price Prediction Using News Articles", *CS224N Final Report*, 2008.
- [9]. Ravi Parikh and Matin Movassate, *Sentiment Analysis of User-Generated Twitter Updates using Various Classification Techniques*, *CS224N Final Report*, (2009.)

- [10]. Anurag Nagar, Michael Hahsler, Using Text and Data Mining Techniques to extract Stock Market Sentiment from Live News Streams, *IPCSIT vol. XX (2012) IACSIT Press, Singapore*
- [11]. Luciano Barbosa and Junlan Feng. 2010. Robust sentiment detection on twitter from biased and noisy data. *Proceedings of the 23rd International Conference on Computational Linguistics: Posters, pages 36–44.*
- [12]. <http://datameetsmedia.com/vader-sentiment-analysis-explained/>