

Identifying the Strength of Emotions in Relation with the Topic of Text Using Word Space

Mohsin Manshad Abbasi
Theoretical Foundation of Computer Sciences
Udmurt State University
Izhevsk, Russian Federation
mohsinmanshad@gmail.com

Anatoly Beltiukov
Theoretical Foundation of Computer Sciences
Udmurt State University
Izhevsk, Russian Federation
belt.udsu@mail.ru

Abstract—Emotion analysis from text is a topic of growing interest over recent years. It is because of the growth and availability of internet. The emotion analysis from text and the strength of the emotions in text plays an important role in understanding and predicting the future events. In this paper, we are identifying the relationship between the strength of emotions and the topic of the text. In our methodology, we extend the concept of Word space that we proposed in our previous works for analyzing emotions. In Word space distances between the words and their occurrences are measures. The emotion carry words with relatively high frequency and less distance between their occurrences are strong emotions. Whereas less frequent emotions that occurs far from, each other are considered less intense emotions in the text. We also made a comparison between the changes in intensity of emotions on same topics over period. The paper is divided into different section. In the methodology and conclusion section, the results of our research mentioned in detail.

Keywords—Word space, emotion, class, summarization, internet, blogs, communication, topic, strength.

I. INTRODUCTION

Emotion analysis from text recently got a lot of attention from the researchers. It is because of the evolution in internet that made human capable of communicating information and emotions quicker than ever. The emotion presented in text has various types and intensities. The major types are positive and negative emotions. The positive and negative emotions are further divided into subcategories based on their polarities, intensities and other characteristics as mentioned by William James [1], Robert Plutchik [2], Parrott [3] etc. The subcategories of emotions or secondary emotions are dependent on primary emotions [4].

The analysis of emotions plays an important role in identifying the future events such as stock market prices, election results, people respond to governmental policies etc. It also help to identify the people with similar interests, their feedback about a product or service, identifying hate groups over internet, etc. The emotions presented in text is a major source of advice over internet. People before buying a product, service, watching a movie, or traveling observes the comments of other people over internet.

In our work, we are using the method of Word space on the relevant blogs from internet over different periods. It is

observed that more people are expressing their emotions over internet with time. Similarly, the intensity and frequency of emotion also changes with time. It is observed that the topic plays an important role in expressing the emotions. About some topics, people are more expressive and moderate whereas about some topics people are less expressive. In addition, the global events and politics changes the people view about a specific topic and so the emotions and their intensities.

II. HISTORY OF EMOTION ANALYSIS

The development of the General Inquirer System (1966) (Stone, [5]) by Philip Stones in Harvard was the first milestone to identify textual emotions. The system was used to count the positive and negative emotional instances.

After this, the main contributions were from Jaynce Wiebe, Peter Turney and Vasileios Hatzivassiloglou during early 90's. Jaynce Wiebe in 1990 (Wiebe, [6]) defines the term "Subjectivity" for Information Retrieval research. Later on in the year of 1997 (Hatzivassiloglou et. al., [7]) identified the semantic orientation of adjectives in the text. After a few years Peter Turney (2002, [8]) came up with his revolutionary approach of Thumbs Up and Thumbs Down for positive and negative review classification. Pang (et. al., 2002, [9]) has suggested the building of sentiment lexicon manually for a domain. (Denecke, 2009, [10]) reported an interesting study on multiple domains to demonstrate the usefulness of the prior polarity scores from the emotional database called SentiWordNet. The preprocessing of text before analyzing emotions improves the performance and accuracy of results [11].

Clustering had been observed as a technique based on generalizations of graph partitioning that don't require pre specified ad hoc distance functions and is capable of automatically discovering document similarities or associations [12]. The logic formulas are used for defining the most common modalities for analyzing emotions from text. The logical properties of emotional modalities, the logic of emotional evaluations and the definition of various modalities enables the affective analysis of emotions [13].

The studies devoted to sentiment analysis in Russian language before 2011 are not very numerous. In 2009,

(Ermakov, 2009, [14]) proposed a sentiment analysis system extracting opinions about cars from a Russian blog community. The patterns were language and domain dependent, which means that patterns must capture the lexical, syntactic and stylistic features of the analyzed text. It was not possible to directly translate or map the English pattern base into Russian pattern [15].

In the beginning, emotion analysis from text in Russian language appears mainly in multilingual experiments. Zagibalov in (Zagibalov et al., 2010, [16]) compare corpora of reviews related to the same books in English and Russian. In (Steinberger et al., 2011, [17]), construction of general emotion vocabularies for several languages is described. Chetviorkin and Loukachevitch (2012, [18]) described, the generation of the Russian sentiments and emotional vocabulary for the generalized domain of products and services and so on.

In bilingual experiment on Russian and Romanian languages, it is observed that the word spelling can be considered as a word phonetic equivalent. This feature allowed limiting the search to letter-based representations [19]. For emotion identification and analysis in Russian language, most of the researchers are using Rule based techniques which is quite necessary as the rules are key to develop a grammar that can be used to develop stemmer, semantic and syntax analyzer [20]. It has been observe that the methods of machine learning performs better in text categorization and classification. The use of syntax for analyzing emotions from text written in Russian language recently gained popularity [21].

III. RELATED WORK

In the beginning of 1977 (Halliday and Hasan, [22]) classified lexical cohesion into two main categories: the reiteration and the collocation category. Reiteration category considers repetition, synonym, and hyponyms, while the collocation category deals with the co-occurrence between words in text document. After this, the major effort was the work on the surprising behavior of Distance Metrics in High Dimensional Space [23]. In this research, it was observed that the fractional distance metric provides better results from the theoretical and empirical perspective.

Edgar C. et al. 2001 [24] performed search in Spaces using the distance. They proposed a concept for working with general cases to model similarity with a distance function that satisfies the triangle inequality, and the set of objects.

Kruengkrai and Jaruskululchi determined the text title [25]. Their approach takes advantages of both the local and global properties of sentences. They used clusters of significant words within each sentence to calculate the local property of sentence and relations of all sentences in document to determine global property of text document. In 2012, Maryam Kiabod et al. [26] presented a technique for summarizing the text by identifying significant words from it. They used extractive method to select the subset of the sentences that contains main concept of the text. The

algorithms for preprocessing the text before analyzing the emotions improves the efficiency and effectiveness of the complete process.

In our methodology, we are using the concept of Word space in which we calculate the distance between the occurrence of the same emotion to identify its strength in the text. We also observed that the strength of the emotion is in relationship with the topic of the text. The common discussed topic over recent year are expected to have stronger emotions than less discussed topics.

IV. METHODOLOGY

The first phase of our work was to identify the topics currently in discussion over internet. For our experiment, we use blogs on similar topics posted during different periods and merge them under a single topic. The common text topic that we use in our research are extracted from reviews and blogs of people about product, science, future life, the current political situations, weapons, religion etc. Using our methodology we observed the change in emotions and their strength our period.

The algorithms that we are using to analyze emotions from text is the extension of our previous work where we use Word space as a measure to identify the strength of emotion and its frequency in the text [27]. Within this research, we are using Word space for comparing the change in emotions over the period.

To analyze the emotions and their strength from text we are using the mathematics from our previous work. We will perform calculation using several equations and formulas as explained below for identifying the strength of emotions in text.

In first equation, we have C_1 that is a numerical value represents the number of occurrence of a word expressing an emotion in text. For every word expressing an emotion in text, the number of its occurrence counted resulting in an array C carrying the occurrence count of all emotion-carrying words from the text. Similarly F_1 represents the frequency of occurrence of a word by dividing its word count C_1 with the total number of words T_m in the text and on applying it to all the members of Word count array C , we have a new array F that represent the frequency of all the emotion carrying words in the text.

$$C_1 = \text{Count (Word Occurrence)} \quad (1)$$

$$C = [C_1 \ C_2 \ C_3 \ C_4 \ \dots \ \dots \ \dots \ \dots \ \dots \ C_n] \quad (2)$$

$$F_1 = \left(\frac{C_1}{T_m} \times 100 \right) \% \quad (3)$$

$$F = [F_1 \ F_2 \ F_3 \ F_4 \ \dots \ \dots \ \dots \ \dots \ \dots \ F_n] \quad (4)$$

The next step is to collect the emotion carrying words in the form of an array W . The array W represents a list of all words expressing emotions. Here DI represents the average distance between the different occurrences of an emotion carrying word W_1 . On applying it to all words carrying emotion we have a list D that represents the average distances of all emotion expressing words. The strength S_1

of a word is directly proportional to its word count C_1 and inversely proportional to the average distance D_1 .

$$W = [W_1 W_2 W_3 W_4 \dots \dots \dots W_n] \quad (5)$$

$$D_1(W_1) = \frac{1}{N} \sum_{n=1}^N [D_n] \quad (6)$$

$$D = [D_1 D_2 D_3 D_4 \dots \dots \dots D_n] \quad (7)$$

$$S_1 \propto C_1 \quad (8)$$

$$S_1 \propto \frac{1}{D_1} \quad (9)$$

Before calculating the average distance occurrences, the maximum and minimum distance values are identified and then normalized. For this purpose the maximum value is divided by the total number of occurrences of that word and the minimum value is multiplied by it.

$$Norm = \max(D), \min(D) \quad (10)$$

$$Norm = \frac{\max(D)}{C}, \min(D) \times C \quad (11)$$

The strength of an emotion is the total occurrence of it C_i divided by the average distance of occurrence D and multiplied by the total number of emotions in the text T_E . On calculating the strength, average distance and occurrence, the strong, very strong and weak emotions can be identified from the text.

$$S(W) = \frac{\sum_{i=1}^n C_i}{D} \times T_E \quad (12)$$

V. EXPERIMENT & RESULTS

The objective of the above calculations is to identify the strength of emotions in text. We begin our research by identifying the text written on similar topics over internet and analyzed the change in emotions over a period of 15 years. We divide the period in three equal interval of time. Here we are mentioning the results of just two most searched and commented topics over internet. The first one is related with the emotions of people while answering a question about religion (Is religion good or bad for society?) while the second is about doing shopping online (Is online shopping safe or unsafe?).

The most common emotions expressed by people about religion over internet during 2005-2010, their frequency, average distance and strength is shown below in table 1.

The frequency and strength of emotions are calculated using the equations mentioned above in methodology section. It has been observed that the strength of positive emotions is more than the negative emotions. The positive emotions occurs more frequent and the distance between their occurrences is smaller than the negative emotions. The results observed during this period are then compared with the results observed during other periods mentioned below in the table 2.

TABLE I. SHOW MOST FREQUENT EMOTIONS IN TEXT FROM 2005-2010

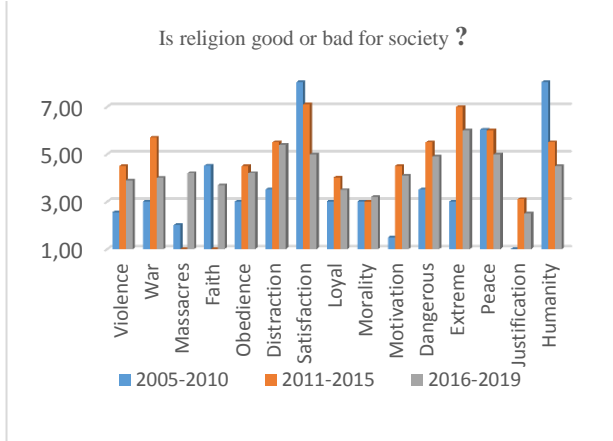
During 2005- 2010			
Emotion	Total Occurrence	AVG Distance	Strength
Violence	5	42.6	2.55
War	6	52.6	3.01
Massacres	4	39.2	2.01
Faith	9	77.3	4.52
Obedience	6	67.6	3.01
Distraction	7	45.7	3.51
Satisfaction	20	66	8.03
Loyal	6	52.6	3.01
Morality	6	95.8	3.01
Motivation	3	359	1.50
Dangerous	7	45.7	3.51
Extreme	6	95.8	3.01
Peace	12	223	6.03
Justification	2	27.5	1.005
Humanity	16	81	8.04

There is a remarkable increase in strength of some of the emotions and similarly remarkable decrease in the strength of some other emotions over the period. It expresses the change in the thinking of emotions of people over time.

TABLE II. SHOW COMPARISON BETWEEN STRENGTH OF EMOTIONS IN TEXT ON SIMILAR TOPIC FROM 2005 TO 2009

Is religion good or bad for society			
Year	2005-10	2011-15	2015-19
Emotion Type	Strength		
Violence	2.55	4.5	3.9
War	3.01	5.7	4.0
Massacres	2.01	4.0	4.2
Faith	4.52	4.0	3.7
Obedience	3.01	4.5	4.2
Distraction	3.51	5.5	5.4
Satisfaction	8.03	7.1	5.0
Loyal	3.01	4.02	3.5
Morality	3.01	3.0	3.2
Motivation	1.50	4.5	4.1
Dangerous	3.51	5.5	4.9
Extreme	3.01	7.0	6.0
Peace	6.03	6.0	5.0
Justification	1.005	3.1	2.5
Humanity	8.04	5.5	4.5

The table 2 describes the emotions and their strengths over different periods. It can be observed that in some emotions the change is remarkable while in others, it is a slight difference. During analysis, it is observed that the blogs and comments of people about this specific topic increased significantly over recent years. It is because of the growth and availability of internet, change in world politics and the events that occurred in last decade. The above analysis is also presented in the form of graph below;



Graph 1. Representing change in emotions of people over periods

The graph 1 above represents the emotions and the change in their strength over different periods represented by different colors. It is observed that some of the positive emotions such as humanity and satisfaction lost their strengths while some of the negative emotions such as war, violence gained it over recent years. As mentioned above the factors such as global politics, wars, extremism have affected the emotions of people about the topic.

TABLE III. SHOW MOST FREQUENT EMOTIONS IN TEXT RELATED TO ONLINE SHOPPING FROM 2016-2019

During 2016- 2019			
Is online shopping safe or unsafe?			
Emotion	Total Occurrence	AVG Distance	Strength
Safe	22	186	6.09
Secure	16	81	8.04
Good	12	203	4.03
Trust	9	77.3	4.52
Easiness	6	67.6	3.01
Hack	7	45.7	3.51
Trouble	6	95.8	3.01
Steal	6	52.6	3.01
Unsafe	2	27.5	1.005
Bad	3	359	1.50

The same methodology is applied on text from blogs related to online shopping. It is observed that the feedback about the products dramatically increased over the recent

years. It is because of the fact that more and more companies are providing their customers with the online feedback mechanisms and then use it for improving the quality of their products and services. Similarly, the people are using feedback of other people for making decision before buying a product or use a service. The results of the analysis is detailed below in table 3.

The analysis about online shopping is done on text available, between 2005 and 2019. For further elaboration, of the results of analysis, the periods are divided into three intervals from 2005 to 2010, 2011 to 2015 and 2016 to 2019. The quantity of the text increased over time. The online shopping emerged as a less cost and more available mechanism for buying products. The results are summarized in the form of a table below.

TABLE IV. SHOW COMPARISON BETWEEN THE STRENGTH OF EMOTIONS IN TEXT FROM 2005-2019

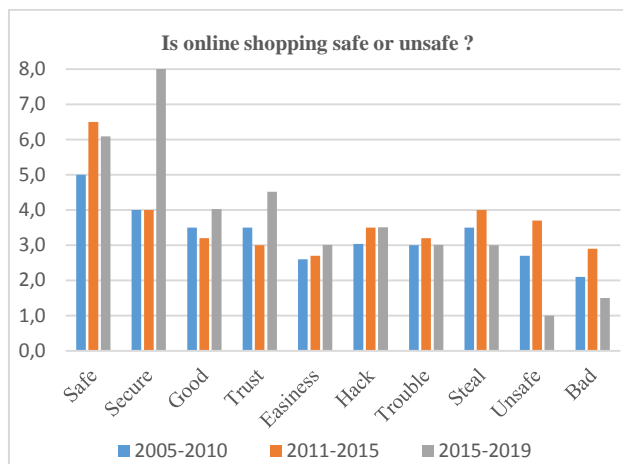
Is online shopping safe or unsafe?			
Year	2005-10	2011-15	2016-19
Emotion Type	Strength		
Safe	5.0	6.5	6.09
Secure	4.0	4.0	8.04
Good	3.5	3.2	4.03
Trust	3.5	3.0	4.52
Easiness	2.6	2.7	3.01
Hack	3.04	3.5	3.51
Trouble	3.0	3.2	3.01
Steal	3.5	4.0	3.01
Unsafe	2.7	3.7	1.005
Bad	2.1	2.9	1.50

In the above table 4, it can be observed that with time, the trust of people on buying products online has been increased. The positive emotions gained the popularity whereas the negative emotion lost their strength. The main reasons are the improvement in technology that made payment more safe and the efficient and safe mechanism that the seller provides to its customers for doing shopping online. It is also observed that the emotion expressed are soft emotions. It is because of the nature of the topic that enables people to express soft emotions. The result of the comparison is sketched below using the graph 2.

The above graph presents the variations in emotions over different period. Each color bar represent a specific period. In general, it is observed that positive emotions dominates the negative emotion over time. It expresses the positive change in attitude of people over time.

VI. CONCLUSION

Our methodology works effectively for analyzing and comparing emotions from text. It is also used to identify the quantity of text available on internet.



Graph 2 Representing change in emotions of people over periods

WordSpace effectively calculate the strength of each emotion in the text by using its frequency and distance between its occurrences. The mechanism is a straight forward and the results are appreciable. It has been observed during the analysis that the topic play an important role in the selection of the emotions by people. Similarly, the world politics and other circumstances affect the quantity of text about a specific topic online. During the analysis of both topics, we observed that the emotions expressed in first topic are relatively stronger compared to the emotion expressed in second topic. The emotions expressed by people are related to the topic of text. It seems that there is an inherent relationship between the emotions and the topic of text.

REFERENCES

- [1] James, William. The Principles of Psychology. Cosimo, Inc. Retrieved 11 June 2019 – via Google Books
- [2] W.Parrott. Emotions in Social Psychology. Key Readings in Social Psychology. Philadelphia: Psychology Press. ISBN 978-0863776830. 2001
- [3] R. Plutchik. The Nature of Emotions. Archived from the original on July 16, 2019.
- [4] M.M. Abbasi., A.P. Beltiukov, "Summarizing Emotions from Text Using Plutchik's Wheel of Emotions", Advances in Intelligent systems research, In Proceedings of the 7th Scientific Conference on Information Technologies for Intelligent Decision Making Support (ITIDS 2019), Atlantis press, vol. 166, 2019, pp. 291-294.
- [5] P.J. Stone, D.C. Dunphy, M.S.Smith. The General Inquirer: A Computer Approach to Content Analysis . MIT Press, Cambridge, 1966.
- [6] Wiebe, M. Janyce, "Identifying Subjectivity characters in Narrative", In Proceedings of the 13th International Conference on Computational Linguistics, Helsinki Morristown NJ, Association of Computational Linguistic, 1990, pp. 401-408.
- [7] H. Vasileios, R.M. Kathleen, "Predicting the Semantic Orientation of Adjectives" EACL '97, In Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics, Madrid, Spain, July 07-12, 1997.
- [8] D.T. Peter, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews", In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, 2002, pp. 417-424.
- [9] B. Pang, L. Lee L, "Thumbs up? Sentiment Classification using Machine Learning Techniques", In Proceedings of the Conference on Empirical Methods in Natural Language, Philadelphia, 2002, pp.79-86.
- [10] K. Denecke, "Are SentiWordNet scores suited for multi-domain sentiment classification?", In Proceedings of the 4th International Conference on Digital Information Management, University of Michigan, USA, 1-4 November 2009.
- [11] M.M Abbasi, A.P.Beltiukov, "Mekhanizm predvaritel'noy obrabotki teksta pered analizom nastroyeni", 6 Vserossiyskaya konferensiya? Informatsionnye tekhnologii intellektual'noy podderzhky prinytiya resheniy, Ufa – Stavropol, Rossiya, 2018.
- [12] D. Boley, M. Gini, R. Gross, "Partition based clustering for web document categorization". Elsevier Journal for Decision Support Systems, vol. 27, Issue 3, December 1999, pp. 329-341.
- [13] M.M. Abbasi, A.P. Beltiukov, "Logical analysis of emotions in text from natural language". Vestnik Udmurtskogo Universiteta.Matematika. Mekhanika. Komp'yuternye Nauki, vol. 29, Issue 1, 2019, pp. 106-116.
- [14] A. Ermakov, "Knowledge extraction from text and its processing: Current state and prospects", In Proceedings of the Computational Linguistics and Intellectual Technologies, 2009. pp. 50–55.
- [15] L. Pivovarova, R. Yangarber, "Adapting the PULS event extraction framework to analyze Russian text", In Proceedings of the 4th Biennial International Workshop on Balto-Slavic Natural Language Processing, Sofia, Bulgaria, 8-9 August 2013, pp. 100–109.
- [16] T. Zagibalov, Belyatskaya et al., "Comparable English-Russian Book Review Corpora for Sentiment Analysis", Edition 2010.In Proceedings fo the 1st Workshop on Computational Approaches to Subjective and Sentiment Analysis, Lisbon, Portugal pp.67-72.
- [17] J.Steinberger, P. Lenkova, M. Kabadjov, et al., "Multilingual Entity-Centered Sentiment Analysis Evaluated by Parallel Corpora", In Proceedings of Recent Advances in Natural Language Processing, Bulgaria, 12-14 September 2011, pp. 770–775.
- [18] I. Chetviorkin ,P. Braslavskiy, N. Loukachevich, "Sentiment Analysis Track at ROMIP 2011", Computational Linguistics and Intellectual Technologies, In Proceedings of the International Conference (Dialog 2012), Bekasovo, pp. 1–14.
- [19] M. Sokolova, V. Bobicev, "Classification of emotion words in Russian and Romanian languages", International Conference RANLP, Borovets, Bulgaria 2009, pp. 416–420.
- [20] M.M. Abbasi, A.P. Beltiukov, "Analysis of sentiment and emotion from text written in Russian language", In Proceedings of the 5th All Russian Conference on Information technology for intelligent decision making support (ITIDS), Ufa, Russian Federation, vol. 1, issue 1, page 42-47.
- [21] M.M. Abbasi, A.P. Beltiukov, "Analysis of emotions from text in Russian Language using syntactic methods" (In Russian), In Proceedings of the 7th International Science Conference. Information Technology and Systems, Khanty-Mansiysk , Russia, 2019, pp.137-142.
- [22] M. Halliday,R. Hasan. Cohesion in English. London: Longman, 1975.
- [23] C.C. Aggarwal, A. Hinneburg, D.A. Keim, "On the surprising behavior of distance metrics in high dimensional space", International Conference on Database Theory, 2001, pp. 420-434.
- [24] C. Edgar, N. Gonzalo , B.Y. Richardo, L.M. Jose, "Searching in Metric Spaces", ACM Computing Surveys, vol. 33, No. 3, September 2001, pp. 273–321.
- [25] C. Jaruskululchi, Kruengkrai, "Generic text summarization using local and global properties of sentences", IEEE/WIC International conference on web intelligence, October 2003, pp.13-16.
- [26] M. Kiabod, M. Naderi, S.M. Sharafi, "A novel method of significant words identification in text summarization", Journal of Emerging Technologies in Web Intelligence, vol. 4, issue 3, August 2012.
- [27] M.M. Abbasi, A.P. Beltiukov, "Analyzing emotions from text corpus using WordSpace", In Proceedings of the 20th international workshop on computer science and information technologies CSIT'2018, Bulgaria, Varna, 2018 vol. 3, issue 4, pp. 161-164.