

Conflict analysis for Turkish debates using text mining and text segmentation techniques

Efe Buyuk, Information Technologies MSc Program, Bahcesehir University, 34349 Besiktas, Istanbul, Turkey.

Adem Karahoca*, Department of Software Engineering, Bahcesehir University, 34349 Besiktas, Istanbul, Turkey.

Dilek Karahoca, Health Sciences Faculty, Health Management, Bahcesehir University, 34349 Besiktas, Istanbul, Turkey.

Suggested Citation:

Buyuk, E., Karahoca, A. & Karahoca, D. (2016). Conflict analysis for Turkish debates using text mining and text segmentation techniques. *Global Journal of Computer Sciences: Theory and Research*. 6(2), 19-25.

Received July 21, 2016; revised September 14, 2016; accepted November 05, 2016.

Selection and peer review under responsibility of Prof. Dr. Dogan Ibrahim, Near East University, Cyprus.

©2016 SciencePark Research, Organization & Counseling. All rights reserved.

Abstract

Conflict Analysis is one of the most challenging issues in the world that many organizations and governments try to carry out perfectly. It is crucial to have a correct analysis to prepare a resolution for a problem. Thus, this study paper focuses on the ways that a software program can detect the reasons of arguments in a debate. The examples of debate dialogs are chosen from Turkish language because there is not much research in this area with this language. Moreover, the techniques which are applied in this work can also be applied to other languages, because a sentiment word dictionary is used and sentiments are almost the same in every language. This is a prepared dictionary from SentiWordNet with all the sentiment points for English words. It is translated and extended for the Turkish language. Furthermore, both machine learning and lexicon-based approaches are implemented in order to increase the diversity of results. This paper aims to show that languages can be processed in a technical manner and meanings can be extracted from sentences to understand the reasons of arguments. Likewise, the main contribution of this study is that conflict analysis for Turkish debates can be applied with the techniques which are examined here and they are also suitable for other languages.

Keywords: text mining, opinion mining, sentiment analysis, machine learning.

*ADDRESS FOR CORRESPONDENCE: **Adem Karahoca**, Department of Software Engineering, Bahcesehir University, 34349 Besiktas, Istanbul, Turkey. E-mail address: adem.karahoca@eng.bau.edu.tr / Tel.: +90-212-381-0560

1. Introduction

This master study aims to show that standings of people for a specific debate or discussion can be understood by the help of sentiment analysis. This means that when there are two opposite opinions for a topic, we can find who is positive or negative about the related debate. In the light of this information, it is clear to understand why this type of work is necessary in our today's world. It is important because there are lots of debates and some of them have huge history that people cannot find a common ground in order to solve the problem. The more a debate takes time, the less chance there is to solve it because the total amount of words and sentences that are spoken increase constantly. That is why we need software programs to examine debates and find a common ground for all parties. Furthermore, when it comes to understanding this master study, it is necessary for readers to know or get familiar with the following topics beforehand that they are text mining, opinion mining, sentiment analysis and machine learning. This work uses Turkish as language because there is not much work related with it. Most of work are done for English or other European languages. To sum up, this work's purpose is to make contribution for the technique of solving Turkish debates. In the future, people can create a more accurate mechanism to detect people's standings as positive and negative and make suggestions to them like showing the common ground. In the following sections of this master study, the chapters of result, discussion and conclusion can give more idea about the output of this work and future thoughts.

2. Background to The Proposed Research

There are two main types of textual information on the Web: Facts and Opinions. Current search engines search for facts that they assume that information is true. Search engines do not search for opinions. Opinions are hard to express with topic keywords. Current search ranking strategy is not appropriate for opinion retrieval/search.

2.1. What is Sentiment Analysis?

Sentiment analysis's past is more than most of people would think. According to Ahmad, the origin of the sentiment analysis lies in the mid-twentieth century political science and pioneers were Harold Lasswell and Philip Stone. They developed content-analysis systems for analyzing speeches by politicians and manifesto of political parties [1].

Sentiment Analysis is the process of determining whether a piece of writing is positive, negative or neutral. As also indicated in a research, the purpose of sentiment analysis is to determine the overall opinion of the user review in terms of positive and negative (Mandal and Kar and Das and Panigrahi 2015). It's also known as opinion mining, deriving the opinion or attitude of a speaker. A common use case for this technology is to discover how people feel about a particular topic.

According to Liu, people's sentiments, opinions, attitudes, and emotions are for the computational study in sentiment analysis. This interesting problem is crucial in our daily business and society life. This area is a big challenge for interested researchers especially in opinion analysis and social media analysis (2015).

Additionally, there is another interesting study which was conducted by Steels, Kaplan, McIntyre and Looveren at Paris at the Sony Computer Science Laboratory in Paris. This study is important because it forms the basis of sentiment analysis. According to their study, they focused on why language has evolved (2000).

Last but not least, what is more important about Sentiment Analysis is that you can even learn why people think the idea is good or bad, by extracting the exact words that indicate why people did or didn't like it. For example, in the experiment of analyzing Twitter's data, when people mention about a food which they tried and said that it is too salty, then it shows as a common theme and you immediately have a better idea of why consumers aren't happy.

2.2. Problems with Sentiment Analysis

Sentiment analysis has problems in its own and it actually deals with the evaluation type of opinions or opinions which imply positive or negative sentiments but opinions itself is a broad concept and not all researchers agree with a set of emotions (Liu 2012). Moreover, there are common formulations of classification problems in sentiment analysis and opinion mining and also ranking problems (Pang and Lee 2008,). For example, in this study ranking problem is very popular because there is not a ready database for all the Turkish words that the software could use. However, the SentiWordNet has a database of English words which has been created for a long time period. Although there is not a ready Turkish sentiment words' database, it is being created with the one that is coded with this master study.

Furthermore, emerging new ideas for an existing opinion is also a problem in sentiment analysis. When this happens, the new information is not aggregated to an existing opinion for the given entity (Tkalčič and De Carolis and De Gemmis and Odić and Košir 2016). According to the researchers mentioned, there is an experiment with Twitter data and the example is related with Nexus4 smartphone. Opinions about this device is formed by a set of posts expressing sentiments. However, opinions about this device may change when there is a new bug or technical problem. Sentiment analysis's techniques should therefore spot opinion changes for these entities. Moreover, the main of goal of sentiment analysis especially in product reviews on the Internet is to rank the entire review or separate into different categories or strength polarities like strongly-positive, weakly-positive, fair, weakly-negative and strongly-negative (Management Association 2013). However, reviews are generally a combination of affirmative and negative opinions.

2.3. Areas to Use Sentiment Analysis

There are many areas to use sentiment analysis. If there is a text related area, then probably sentiment analysis can be applied to it. As Singh and Husain explained in their study, when there is a decision making moment, both for an individual or organizational level, then search of other's opinion becomes important (2014). For example, nowadays blogs are a big part of our life and according to Balijepalli, the texts in these blogs generally represents the sentiments of people. There are many topics there including sports, politics, business, entertainment and etc. He also emphasizes that a good sentiment detection tool is necessary for us in order to make better analysis (2008).

Moreover, according to a research, the subject areas are also retail, call centers, financial institutions and telecommunications (Ishikiryama and Miro and Gomes 2015). In addition to that, many of the text analysis techniques can be applied to diverse subject areas like marketing, politics and medicine. In order to obtain the best possible results, a certain degree of domain-subject knowledge is required (Kagan and Rossini and Sapounas 2013). That is why in this master study, firstly a Turkish sentiment database for the words is created and the language's properties are studied in order to determine which part of them can be applied to the software.

Furthermore, sentiment analysis can also be used to analyze organizations in social networks. When this analysis is completed, then you can understand that which areas in your organization is performing well and what areas may need some more work (Pierson 2015). In addition to that, there is a very related research and study area that local governments use Twitter data for sentiment analysis and they try to understand their people's opinions by doing that. One example about this is that this kinds of analysis focused on the capital city local governments and a sample of six Sydney suburban local governments. These governments were chosen because their social media use was significant enough to produce results (Sobaci, 2015).

2.4. Why Does Sentiment Analysis Become Popular?

There are lots of reasons to answer this question but the first answer which is more imaginable is that people want to know what the other people think. Besides that, the below factors are determined by Pang and Lee (2008);

- i. The rise of machine learning methods in natural language processing and information retrieval
- ii. The availability of datasets for machine learning algorithms
- iii. Realization of the fascinating intellectual challenges and commercial and intelligence applications that the area offers.

3. Identifying The Scope of Work

The topic which I have chosen is very difficult and that is why my work will be experimental. It is my hope that I can make some contributions on this field.

It is important to describe what I am going to try to solve in my study. That is why I have to draw some lines and set some rules for inputs which in this case will be Turkish debates. Every text will show a basic Turkish debate that basically two parties' arguments will be there. At the end, the output will be the sentence sentiment points of the both parties. The following can be a good template example as a result of conflict analysis;

Table 3.1. Illustration of Result Screen

Topic: Here written the topic (discussion)

Causes of Party A (Stand: Positive)	Causes of Party B (Stand: Negative)
Sentence Sentiment Point	Sentence Sentiment Point

4. Research Questions to Be Investigated

- i. Can a software program find arguments or reasons of a specific thought by specifying some key words?
- ii. Can a software program detect the familiarities between two different arguments by using words' sentiment points?
- iii. Can we understand one word's different meanings or sentiments in one sentence by using words' sentiment points?

5. Expected Inputs and Outputs

For the sake of my experimental research work in this study, I will keep it simple and give the topic and stand statuses of both parties before expecting an analysis. By doing so, it will be a bit easier to specify the causes during getting the related texts as inputs. To make it clear, let's give a simple example;

- i. Given Topic: "Bazı suç vakalarında ölüm cezasının uygulanması" (Execution of death punishment for some criminal cases)
- ii. Given the Stand Status of Party A: ['Merve', 'İdam cezası olmalı.', 1.0] (First column: Name; Second column: Sentence; Third column: Sentence Sentiment Point)
- iii. Given the Stand Status of Party B: ['Ali', 'İdam cezası olmamalı.', -1.0] (First column: Name; Second column: Sentence; Third column: Sentence Sentiment Point)

6. Results

6.1. How Will the Proposed Research Contribute to the Research Outputs and Research Culture in the Research Area?

My study aims to show that we as humans can solve all our conflicts among us. This can be done by formulating our languages. All languages have already their rules and we call them as grammars. These grammars can be formalized and calculated to understand what people say and mean in the context of a

dialog. Most of the time, people have difficulties for explaining their opinions during a discussion. It is because when there is a dispute or conflict between two people or parties, not only the thoughts people try to explain but also the feelings they should suppress to be more clear against each other. Because of these feelings' interruptions, machines can help us to understand and transmit the ideas of people.

6.2. Main Features

There are two main features of the software and these two main features also have some sub-features in order to make the user experience high and more logical and useful.

Two main features:

- i. Search for the sentiment points of a Turkish word
 - a. View all the sentiment points of negative and positive
 - b. Add a new word to database if the word cannot be found
- ii. Analyze a given text's each sentence to see their sentence sentiment points
 - a. List the attendees' names which are got involved in the given text, discussion
 - b. Find the ngram points of each sentences
 - c. Return as a list for the given text that includes attendees' names, their sentences and sentences' sentiment points
 - d. Add a new word to database if the word in the text cannot be found in the database

7. Findings

The main purpose of this study is to make contribution to sentiment analysis for Turkish language. When it comes to code the software, only the document analyzing was thought to complete. However, in the process of time it was understood that word searching functionality should be added to this software. In any way, word searching function was a part of document analysis but additionally separating them created better user experience.

Moreover, it is interesting that Marcus asked on his study that the way the sentiment analysis is performed and why a statement is characterized as positive, negative or neutral (2014). This characterization is important because otherwise we cannot specify the sentences' states and what people think about.

Table 7.1. Comparison between test results and true sentence states

True Sentence State	Test Result		Total
	Positive (+)	Negative (-)	
Positive (+)	33	8	41
Negative (-)	30	39	69
Total	63	47	110

Given the numbers on above table, results are shown below;

- i. Sensitivity is $33 / 41 = 0,8$ or 80 %
- ii. Specificity is $39 / 69 = 0,56$ or 56 %
- iii. False positive rate is $30 / 69 = 0,43$ or 43 %
- iv. False negative rate is $8 / 41 = 0,19$ or 19 %
- v. Positive predictive value is $33 / 63 = 0,52$ or 52 %

vi. Negative predictive value is $39 / 47 = 0,82$ or 82 %

8. Conclusion

This study's aim is to show that text mining and sentiment analysis techniques can be applied to Turkish language and by doing that positive results can be generated. It is the expected that some positive results were obtained. In the software which is coded for this study purpose there are some pre-defined documents. Those documents have Turkish sentences with Turkish sentence rules. Every document has its own specific rule in it and every first sentence in those documents carries that rules. Every second sentences there are the answers for the first sentences.

Results of the study is given in the part of results in this paper. Moreover, in the part of findings, measurement techniques and some highlights are shared. Furthermore, in the part of discussion, both theoretical and practical implications are given to readers. All the results show that when word's positive and negative scores are correctly types, then the sentences' sentiment points are calculated correctly. The important point here is that all of the words' sentiment points should be found beforehand and made ready for giving sentences. The words' sentiment points should not be based on one or two person's judgements. Therefore, the number of participants who will complete the missing words and their sentiment points should be increased as much as possible.

Lastly, text analysis is one of the most popular subject in computer science nowadays. In some languages like English, there are very good software programs with very accurate, consistent and correct results. However, this field could not be applied to Turkish language much enough. There is not much study like there are in English. One of the main purposes of this project is to increase the momentum of current studies of related topics in this area and to attract attention for this area from curious students. It is the hope that this study can be helpful or inspirational for other students.

References

- [1] Ahmad, K. *Affective computing and sentiment analysis: emotions, metaphor and terminology*. Springer Science & Business Media, 2011, 45, pp. viii.
- [2] Ertav, B. *Turkce dilbilgisi notlari: ygs – lys'ler icin ders notlari*, 2014, pp 20.
- [3] Balijepalli, S. *Information retrieval based on sentiment analysis*. VDM Publishin, 2008.
- [4] Agarwal, B. and Mittal, N. *Prominent feature extraction for sentiment analysis*. Springer, 2015, pp. 13.
- [5] Gokalp, M. *Turk dili bilgisi*. Sark Press, 1976, pp 157.
- [6] Kagan, V., Rossini, E. and Sapounas, D. *Sentiment analysis for ptsd signals*. Springer Science & Business Media, 2013, pp. 23.
- [7] Akdeniz, S. *Turkcede yeni gelismeler*. Ozek Press, 1973, pp 56.
- [8] Tkalcic M., De Carolis, B., De Gemmis, M., Odic, A. and Kosir, A. Emotions and personality in personalized services: models, evaluation and applications. *Springer*, 2016, pp. 134.
- [9] Gulsevin, G., & Boz E. *Turkce nin cagdas sorunlari*. Divan Press, 2006, pp 78.
- [10] Liu, B. *Sentiment analysis and opinion mining*. Morgan & Claypool Publishers, 2012, pp 22.
- [11] Stuckenschmidt, H. and Jannach, D. E-commerce and web technologies: 16th international conference on electronic commerce and web Technologies, ec-web 2015. *Valencia: Springer*, 2015, 239, pp. 109.
- [12] Liu, B. *Sentiment analysis: mining opinions, sentiments, and emotions*. Cambridge University Press, 2015
- [13] Pang, B. and Lee, L. *Opinion mining and sentiment analysis*. Now Publishers Inc, 2008, pp 26.
- [14] Coleman, S. and Freelon, D. *Handbook of digital politics*. Edward Elgar Publishing, 2015, pp. 465.
- [15] Ozaydin, U. *Temiz turkce kilavuzu: iletisim guzeldir*. Nubuto Limited Press, 2015, pp 29.
- [16] Mandal, D., Kar, R., Das, S. and Panigrahi, B. K. Intelligent computing and applications: proceedings of the international conference on ica. *Springer*, 2015, 343, pp. 368.
- [17] Bhattacharjee, A. and Fitzgerald, B. *Shaping the future of ict research: methods and approaches: ifip wg 8.2 working conference*. Tampa: *Springer*, 2012, 13-14, pp. 3.
- [18] Management Association Information Resources. *Enterprise resource planning: concepts, methodologies, tools, and applications: concepts, methodologies, tools, and applications*. IGI Global, 2013, pp. 1505.
- [19] Goker, O. *Uygulamali turkce bilgileri*. Ankara: The Ministry of Education, 1997, 3, pp 95.

- [20] Pierson, L. *Data science for dummies*. John Wiley & Sons, 2015, pp. 277.
- [21] Sobaci, M. Z. *Social media and local governments: theory and practice*. Springer, 2015, pp. 121.
- [22] Kutlu, M. *Türk dili ve edebiyati ansiklopedisi: devirler, isimler, eserler, terimler*, Dergah Press, 1977, 7, pp 119.
- [23] Marcus, A. Design, User Experience, and Usability: User Experience Design for Diverse Interaction Platforms and Environments: *Third International Conference, DUXU 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014, Proceeding*. Springer, 2014, 2, pp. 487.
- [24] Ishikiriya, C. S., Miro, D. and Gomes, C. F. S. *Text Mining Business Intelligence: a small sample of what words can say*. *Procedia Computer Science*, 2015, 55, 261 – 267.
- [25] Medhat, W., Hassan, A. and Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 2015, 5, 1093 – 1113.
- [26] Singh, P. K. and Husain, M. S. Methodological Study of Opinion Mining and Sentiment Analysis Techniques. *International Journal on Soft Computing (IJSC)*, 2014, 5 (1).
- [27] Cambria, E., Schuller, B., Xia, Y. and Havasi, C. *New Avenues in Opinion Mining and Sentiment Analysis*. IEEE Computer Society, 2013.
- [28] Pang, B. and Lee, L. *Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval*, 2008, 2 (1-2), pp. 4 - 5.
- [29] Istanbul University Faculty of Literature. *Türk dili ve edebiyati dergisi*. Istanbul: Istanbul University Faculty of Literature Press, 2006, 34, pp. 44.
- [30] Türk Dil Kurumu. *Türk dili araştırmaları yaylığı 1. bolum*. Turkish Language Association, 2006, pp. 80.
- [31] Balijepalli, S. *Blogvox2: a modular domain independent sentiment analysis system* (Thesis for M.S. Degree). Maryland: University of Maryland.
- [32] Demirel, O. *Yapıcı Olumlu Anlamca Olumsuz Cümleler Hakkında Bir Değerlendirme. A research about positive and negative sentences*. Istanbul: Istanbul University, 2009
- [33] Demirel, O. *Türkiye Türkçesinde Olumsuz Cümle Yapıları* (Thesis for M.A).Degree. Istanbul: İnönü University, 2002
- [34] İlhan, N. *Türkçede Olumsuzluk*. A research about the negative sentences in Turkish. Elazığ: Fırat University, 2007.
- [35] Ucan, A. *Automatic Sentiment Dictionary Translation and Using in Sentiment Analysis* (Thesis for M.A. Degree). Ankara: Hacettepe University, 2014.
- [36] Blunsom, P. *Hidden Markov Models*. Melbourne: The University of Melbourne, 2014.
- [37] Steels, L., Kaplan, F., McIntyre, A. and Van Looveren, J. *Crucial factors in the origins of word-meaning*. Paris: Sony Computer Science Laboratory, 2000.
- [38] Learn Turkish with Manisa Turkish. Basic Grammar. Istanbul: SIGN, 2016. Available from: <http://www.turkishlanguage.co.uk/>
- [39] SentiWordNet database, 23 July 2016. Available from: <http://sentiwordnet.isti.cnr.it/>
- [40] Solutions to Diagnostic Testing Practice Exercises, 2016. <http://sites.stat.psu.edu/~lSimon/stat250/sp00/solutions/misc/diagtests.html>
- [41] Positivism Research Philosophy, 31 August 2016. Available from: <http://research-methodology.net/research-philosophy/positivism/>
- [42] Deductive Approach, 2016. Available from: <http://research-methodology.net/research-methodology/research-approach/deductive-approach-2/>
- [43] Inductive Approach, 2016. Available from: <http://research-methodology.net/research-methodology/research-approach/inductive-approach-2/>
- [44] Research Design, 2016. Available from: <http://research-methodology.net/research-methodology/research-design/>
- [45] Research Methods, 2016. Available from: <http://research-methodology.net/research-methods/>
- [46] Convenience Sampling, 2016. Available from: <http://research-methodology.net/sampling-in-primary-data-collection/convenience-sampling/>