NEWS STORY ANALYSIS WITH CREDIBILITY ASSESSMENT BY OPINION MINING

A Thesis Submitted to
the Graduate School of Engineering and Sciences of
İzmir Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

in Computer Engineering

by
Erhan SEZERER

July 2015
İZMİR

We approve the thesis of **Erhan SEZERER**

Examining Committee Members:

Asst. Prof. Dr. Selma Tekir
Department of Computer Engineering, İzmir Institute of Technology

Asst. Prof. Dr. Mustafa Özuysal
Department of Computer Engineering, İzmir Institute of Technology

Asst. Prof. Dr. Zerrin Işık
Department of Computer Engineering, Dokuz Eylül University

13 July 2015

Asst. Prof. Dr. Selma Tekir
Supervisor, Department of Computer Engineering
İzmir Institute of Technology

Prof. Dr. Halis Püskülcü
Head of the Department of
Computer Engineering

Prof. Dr. Bilge Karaçalı
Dean of the Graduate School of
Engineering and Sciences

ACKNOWLEDGMENTS

I would like to thank my advisor Asst. Prof. Dr. Selma Tekir for her guidance. She was very helpful, patient and positive throughout the process. Furthermore, I am grateful for her efforts and ideas she put into this project to help me accomplish this task.

Special thanks to my colleagues Mustafa Toprak, Furkan Eren Uzyıldırım and Deniz Eraslan for helping me to prepare the training data. Their collaboration was very valuable.

I would like to thank Asst. Prof. Dr. Mustafa Özuysal and Asst. Prof. Dr. Zerrin Işık, for accepting to be a part of my thesis defence jury and sparing their valuable time to read and evaluate my thesis and provide me with feedbacks.

Also I would like to thank all of my friends and family for their support. They supported me in every aspect and motivated me to do better.

ABSTRACT

NEWS STORY ANALYSIS WITH CREDIBILITY ASSESSMENT BY OPINION MINING

With the growing influence of media and the popularity and widespread use of social networks, credibility of the news sources became an important subject that needs more attention. The biggest problem of finding credible sources is, instead of giving every aspect of the incident, news sources tend to accept one of the parties' idea as a whole while rejecting every other ideas, or even worse, they focus on only one side of the incident and ignoring the rest.

Credibility is defined as "The quality of believable and trustworthy". The notion of trustworthiness can further be decomposed into components like bias, fairness, factual/opinionated, etc. In this thesis, credibility is measured using the fact/opinion ratio of the articles. Two methods, which are the traditional Naive Bayes method and the Relativistic method, are proposed. The intuition of relativistic method comes from the theory of relativity where the sentiment of the articles is determined relatively to the ordinary context used by people in daily speech.

We have tested our methods on four different types of data, hand-written articles, editorials, New York Times articles and Reuters articles, and aimed to show that our proposed models are able to differentiate the sentiments in the articles. In the experimental work, we provided a detailed evaluation of the results.

ÖZET

FİKİR MADENCİLİĞİ TEKNİKLERİYLE GÜVENİLİRLİK DEĞERLENDİRMESİ VE HABER HİKAYESİ ANALİZİ

Sosyal ağların yaygınlaşması ve popülerleşmesi ve medyanın etkisinin giderek artmasıyla birlikte, haber kaynaklarının güvenilirliği, üzerinde durulması gereken önemli bir konu haline gelmiştir. Güvenilir kaynakları bulmaktaki en büyük sorun, haber kaynaklarının haberin tüm yönlerini vermek yerine bir fikri kabul edip diğerlerini reddetmesi ya da daha kötüsü, tek bir fikrin savunulup diğerlerinin tamamen görmezden gelinmesidir.

Güvenilirlik birçok kaynakta "inanılabilirlik ölçüsü" olarak tanımlanmıştır. İnanılabilirlik kavramı taraflılık, adillik gerçeğe-dayalılık/fikre-dayalılık olarak daha da alt başlıklara bölünebilir. Bu tezin kapsamında, güvenilirlik ölçümü gerçek/fikir oranı kullanılarak yapılmıştır. Geleneksel Naive Bayes ve göreliliksel yöntemleri kullanan iki yöntem önerilmiştir. Göreliliksel yöntem kavramı, haber makalelerindeki fikir ve duyguların genel ve günlük konuşmalardan göreli olarak ne kadar farklı olduğuyla anlaşılabileceğini iddia edecek şekilde görelilik kuramından esinlenilmiştir.

Yöntemlerimizi dört çeşit veri (elle yaratılmış makaleler, New York Times'ın başyazıları, New York Times haber makaleleri ve Reuters haber makaleleri) üzerinde test edilmiş ve yöntemlerimizin gerçek ve fikirleri ayırt edebildiği gösterilmiştir. Deneysel değerlendirme bölümünde elde edilen sonuçlar detaylı olarak açıklanmıştır.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

POS . . . . Part-of-Speech

NYT . . . . New York Times

CDF . . . . Cumulative Distribution Function

CHAPTER 1

INTRODUCTION

The concept of credibility can be associated with the accuracy of the given information in general. Ensuring the accuracy involves verifying the correctness of information. The verification task is partially performed through the use of information source validation. Thus, credibility is addressed as a characteristic of information (object) as well as the source (subject) providing it.

Merriam-Webster describes credibility as "the quality of being believed or accepted as true, real, or honest"[1] while the American Heritage Dictionary describes it as "The definition of credibility is the quality of being trustworthy or believable"[2]. Although it seems like all the sources agree upon the general definition of "quality of being believed", thus creating a well-defined non-problematic understanding of the credibility, from a research perspective it is rather problematic. That problem resides in the use of the ambiguous term "quality". Main issues are;

- How can we qualify credibility in a news source?
- What methods or metrics should we use to analyze its "quality of believable"?

These are the research questions that should be answered first if we would like to base our research on the concept of credibility.

With a further research into the concept, Gaziano and McGrath (1986) find out that credibility of news source depends on two main components: trustworthiness and expertise. Also, there are many more minor components such as fairness, bias, moral, factual or opinionated, and accuracy. Main focus of this research will be on the main component trustworthiness since the expertise is not an easy task to calculate using traditional text mining methods.

Although this broader definition of credibility seems fitting, most of the time perception of truth and trustworthiness of common people is hardly the same. The trustworthiness of a source or information requires it to be independent of specific perceptions. For example, these two examples of news sources can reveal the severity of the problem:

[1] "Credibility" (2015)
[2] "Credibility" (2013)

> (1) … "Palestinian children arrested by (Israeli) military and police are systematically subject to degrading treatment, and often to acts of torture, are interrogated in Hebrew, a language they did not understand, and sign confessions in Hebrew in order to be released," it said in a report … Most Palestinian children arrested are accused of having thrown stones, an offence which can carry a penalty of up to 20 years in prison, the committee said. Israeli soldiers had testified to the often arbitrary nature of the arrests, it said… Israeli soldiers had used Palestinian children to enter potentially dangerous buildings before them and to stand in front of military vehicles to deter stone-throwing, it said. "Almost all those using children as human shields and informants have remained unpunished and the soldiers convicted for having forced at gunpoint a nine-year-old child to search bags suspected of containing explosives only received a suspended sentence of three months and were demoted,"it said[3].

> (2) … Militant rockets can be seen launching from crowded neighborhoods, near apartment buildings, schools and hotels. Hamas fighters have set traps for Israeli soldiers in civilian homes and stored weapons in mosques and schools. Tunnels have been dug beneath private property … "Hamas uses schools, residential buildings, mosques and hospitals to fire rockets at Israeli civilians," Prime Minister Benjamin Netanyahu told his Canadian counterpart in a call over the weekend, according to a statement from Mr. Netanyahu's office. "Hamas uses innocent civilians as a human shield for terrorist activity"[4].

As it can be seen from the articles, two sources can accuse different parties of war. Two main questions arise then: which one should be trusted by the people reading them? or are they both telling the truth? Israelis who read the first article, which is published by the Reuters in June 2013, would hardly find it trustworthy on the other hand it is expected them to find the second one, which is written by New York Times in July 2014, trustworthy. Same result can be expected when it comes to Palestinians. Vice versa they would find the first one trustworthy and the second one is not.

Since people are eager to believe and trust the sources that can help or agree to their cause, we need to dig deeper if we want to find a general meaning of credibility free from the perspectives of certain groups or individuals. To find that general meaning of credibility ,we have to find how objective and free of any kind of opinion, negative or positive, that news source is. The more emotion exists in an article the more opinionated it can be, and the more opinionated an article, the better chance that it is a subjective article that tries to mislead or convince the people reading it by either exaggerating the goodness of one side and the bad side of the other, or worse, completely ignoring an aspect of the story. Thus, a credible source should make a clear distinction between facts, mysteries and judgements.

[3]Nebehay (2013)

[4]Barnard and Rudoren (2014)

The objectivity of news is a credibility factor on its own, which is confirmed by the study of Gaziano and McGrath (1986), through the use of the sub-factors of "does or does not separate fact and opinion" and "is factual or opinionated" in news credibility measurement.

Objectivity has become a hot topic of debate, since the beginning of the 90's, where the mass media peeked to its highest level with the new technologies like web 2.0 and followed by the social media. Identifying the objective, therefore trustworthy news sources is lot harder now considering the millions of pieces of information that flows between individuals in social media and created in the news sources that spawn or vanish each day.

For addressing the aforementioned problem of identifying the objectivity, two models are used in the scope of this thesis: naive bayes classification and the relativistic approach. Different from the other scientific works on this topic, we do not build the naive bayes using any kind of training data, instead, using the information taken from the knowledge-bases, we applied naive bayes directly to the articles. Second method, the relativistic approach, is an implementation inspired from the theory of relativity. In this method we tried to create a reference point for opinions and evaluate the articles using the deviation of scores from that reference point.

Using the traditional naive bayes method and our novel approach, the relativistic method, we showed that both methods work very well in case of finding the polarity of single sentences, but when it comes to comparing the articles or even more the entire news sources in terms of subjectivity our method works better than the traditional naive bayes approach. We will also prove our hypothesis about the sentiment levels of four different kind of articles (hand-written emotional texts, editorials, New York Times articles and Reuters articles).

1.1. Organisation of Thesis

In this thesis, this objectivity problem has been addressed in a two different way with knowledge-based systems using various methods of natural language processing and sentiment analysis. Chapter 2 gives the definition of the sentiment analysis and some background information on the different methods of sentimental analysis used on the news sources. Chapter 3 introduces our model and methodology for a solution on the stated problem. Here, we also give the information about the dataset we used, and the type of dictionaries that is used to process natural language processing on our dataset. Furthermore, our two proposed models are introduced with details. Chapter 4 give their results from both our models and also comparisons and comments on the effectiveness of them. We conclude this chapter by humbly giving some recommendations and directions and some possible issues on future work and research on this subject.

CHAPTER 2

BACKGROUND ON SENTIMENT ANALYSIS

The main objective of sentiment analysis is to find in what emotional state the author of the given text is, using the data processing and indexing methods with the natural language processing support. The search of emotional state can be applied to different areas for finding different aspects of emotions. It can be applied to a product review to learn the level of satisfaction of the customer that bought the product, or it can be applied to a comment written about a trend topic by a user on twitter or a blog to learn whether that person is in favor of the situation or not. Last, but not the least important, area of practice is the news articles where we would like to learn how objective and clear from any kind of affection an author is in his writing.

Depending on these areas of practice, the results are expected to be different in terms of emotional state. In the area of product review, the aim is to find the polarity of the author as negative or positive. Other details are usually not necessary unless you would like to know why the customer does not like the product. The situation is not the same when it comes to the tweets and blogs. Especially if it is a comment on a social issue, the Twitter and blog texts are mostly examined in more details. Here an applied sentiment analysis is expected to give the results that express the emotions of the author as “angry”, “happy”, “frustrated”, “content”, etc. In the domain of news articles, instead of the type of emotions, the existence of it is searched. The main goal while analyzing journal texts is to find objectivity/subjectivity of the author. The reason of this search is to separate the texts which are affected by some school of thought or the texts written in some perspectives, from the texts which are written without any agenda.

Sentiment analysis, or opinion mining, has three different models when it comes to granularity:

- Document-based sentimental analysis: Document-based models treat the whole document as a single entity and retrieve the general emotion of the text. Works best with the texts where it is known that there is only one subject and one emotion that belongs to that subject.

- Sentence-based sentimental analysis: Based on the fact that a document may contain more than one type of opinion, namely negative-positive or sad-happy, sentence-based models divide the text into sentences and apply the sentiment analysis on every individual sentence independently. Sentence-based methods are built to exploit this and retrieve all of the opinions in text. This model usually works best with opinion mining on tweets or such short media, since the text is already small and may contain too many opinions and/or emotions in different sentences and it is subject to a considerable amount of emotional change between sentences.
- Aspect-based sentiment analysis: Aspect-based models enhance the sentence-based models by trying to learn the aspects of the text before applying sentiment analysis. These models extract the different subjects that are mentioned in the text first, then apply the opinion mining procedures to each of them independently. Being the most state-of-the-art model, aspect-based sentiment analysis is able to extract both the subject and the emotions that belong to them in a desired and effective way.

2.1. Related Work

Several methods have been proposed on the area of sentiment analysis. Pang and Lee (2008) and Liu (2012) describe, what sentiment analysis is and what are those methods, in general. Using crowdsourcing, Mejova et al. (2014) examines the effect of controversial topics on sentiment analysis. According to them, controversial topics tend to have negative sentiment throughout the text. Takala et al. (2014) uses inner-annotator agreement metrics to evaluate economy related news-articles with the assertion that they have different sentiments from other type of articles. Karamibekr and Ghorbani (2012) examines the difference between sentiment analysis on products and sentiment analysis on social issues. It is proposed that verbs are more important than other type of words when it comes to a topic related to a social issue. Balahur et al. (2010) tries to separate good and bad news from good and bad sentiment and proposes a method to get the aspect of the news from the perspective of the reader, author and text. Instead of marking the text with just positive and negative scores, Strapparava and Mihalcea (2008) proposes a novel approach where they use the hourglass of emotions (Cambria et al. (2011)) to try to find emotions in text. Pang et al. (2002) applies three methods for finding sentiments: bayesian classifiers, support vector machine and maximum entropy classification and claims that they are not performing better than traditional topic-based categorisation. Lim and Buntine (2014) uses Latent Dirichlet Allocation-based model and leverages hashtags and sentiment lexicons (hashtags, emoticons and mentions) to find opinion on products. SenticNet Cambria et al. (2014) on the other hand uses methods such as ConceptNet, AffectiveSpace and hourglass of emotions to find common sense concepts.

The idea of defining credibility as fact/opinion ratio in the news is first described by Gaziano and McGrath (1986). Since then, there have been some research done on the sentiment analysis using the same approach. Yu and Hatzivassiloglou (2003) proposed a bayesian classification on both sentence and document levels for solving fact/opinion ratio. A similar approach is proposed by Lin et al. (2006). It is focused on finding perspectives from which the articles are written, rather than just finding opinions or facts.

CHAPTER 3

SENTIMENT ANALYSIS IN THE NEWS

In this chapter, details of the proposed models are explained. As can be seen from the Figure 3.1, our model can be divided in three main parts.

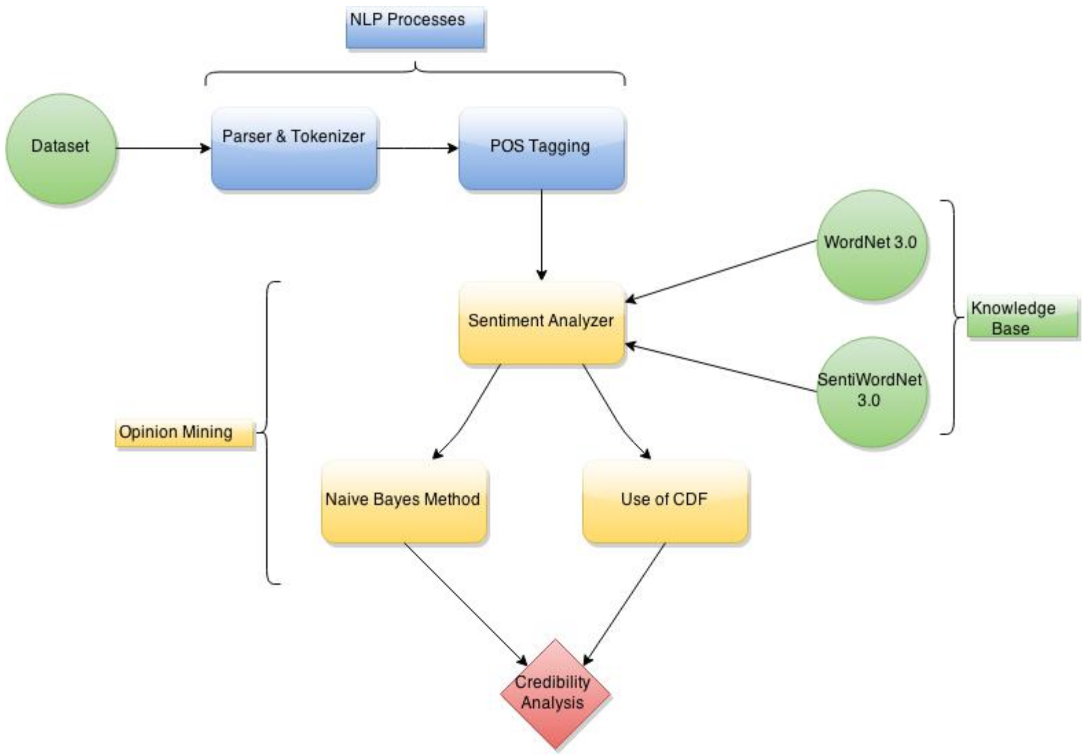

Figure 3.1. Model

In sections 3.1 and 3.2, we have explained the first part of this model (indicated with green) which consists of the dataset and the knowledge-base we use. Second part, which is indicated with blue colours, is the Natural Language Processing part which consists of parsing, tokenizing and part-of-speech (POS) tagging the articles in order to prepare our data for a sentiment analysis evaluation. This part is explained in section 3.3. Finally the third and the last part (indicated with yellow) contains the information of our proposed sentiment analysis methods. First of these methods uses Naive Bayes to calculate the probability of an article to be either subjective or objective, while the second one uses the cumulative distribution function of the knowledge-base to calculate the probabilities. They are explained in the sections 3.4.1 and 3.4.2 respectively.

3.1. Dataset

We have worked on two different datasets for sanity test and the actual tests. The use of test data for sanity tests was necessary for understanding whether our methods are working properly and they are able to extract the correct sentiment from the text. In a way, they worked as a ground-truth of emotional texts. Only after working on the sanity test data and making sure that our method works properly, we begin working on our eventual data, the test data.

Our sanity test data consists of two different type of articles. We used 30 editorial articles from New York Times that are manually tagged as either positive or negative. Using news articles for training would not have been helpful to us since they are expected to be written in an objective way. Editorial articles on the other hand are known for being subjective and we tried to exploit this property to understand whether our methods are in a good position to judge the objectivity of news articles. But the first sanity test data we used was the hand-written notes we created that have extreme levels of emotions and subjective sentence. These small notes, about 10 texts created by us, have either love or hatred sentences.

Our test data consists of the articles from two different sources: The New York Times Sandhaus (2008) and Reuters (Rose et al. (2002)). The New York Times (NYT) corpus contains approximately 1.800.000 articles, published in New York Times between January 1st. 1987 and June 19th. 2007. The Reuters corpus has around 800.000 articles, published between July 20th, 1996 and July 19th,1997.

The reasons behind this dataset selections are, first, those two news sources are one of the most popular and influential sources in English, second, they have a distinguishing property that we will try to verify with our tests. While The New York Times is a newspaper (both online and paper), Reuters is a news agency, and because of this Reuters is expected to be more free from opinion of the authors. On the other hand, New York Times, although it can be very objective too, is more likely to have opinions. This assumption comes from the fact that, the aim of Reuters and such agencies is to provide the information to other newspapers, and to do so, it should be, and in most of the cases it actually is, more fact based and pure. Newspapers, on the other hand, have a particular audience to address, so they tend to have more emotions and opinion on articles to conform with the general view of their audience. After validating our methods with the training articles, our aim is to reach a conclusion that agrees with our assumptions on our dataset.

By using these two types of training data with test data itself, we aimed to create an environment with four levels of objectiveness: non-article notes , editorials , newspaper articles and news agency articles where the objectivity increases from left to right. The goal of this thesis is to be able to find and differentiate those four levels from each other in terms of objectivity.

3.2. Sentiment Analysis And Knowledge-Base

Since we needed an understanding of what can be an opinion in an article, we have used two sources as a knowledge-base in our proposed methods: WordNet (Miller (1995)) and SentiWordNet (Baccianella et al. (2010)), both are versions 3.0. WordNet is a database of English words that contains information about the definition and category, or more specifically part-of-speech, of words. It contains more than 200.000 words that exist in English. SentiWordNet is a dictionary that is built on the the WordNet. SentiWordNet uses the words from the WordNet and scores them according to their probability of having a negative or positive sentiment. Around 117.000 words are scored in the range of 0-1 that corresponds to the probability of that word being positive, negative or objective. The sum of all three probabilities are add up to 1.

The SentiWordNet dictionary does not just give one score-set for every unique word, instead it provides different scores for every different meaning of a word. For different POS categories, different scores exist for a word. Furthermore, it is possible for a word to have two different score for the same POS if it has more than two meanings in different contexts. For example the word "mean" has 16 different scores; 1 for noun, 7 for verb, 8 for adjective (See Figure 3.3 for the word "mean" in lines: 4983, 12340, 51074 where the 8 digit integer next to it corresponds to WordNet id and the next two numbers correspond to scores). When it is used in a context as "of no value or worth" it has 0.5

```
114684 mealy_sage%1:20:00:: 12865239 1 0
114685 mealybug%1:05:00:: 02250822 1 0
114686 mealymouthed%5:00:00:indirect:02 00768927 1 0
114687 mean%1:09:00:: 06023969 1 10
114688 mean%2:31:00:: 00708538 4 27
114689 mean%2:31:01:: 00730052 6 8
114690 mean%2:31:05:: 00708840 7 0
114691 mean%2:32:01:: 00955148 1 93
114692 mean%2:32:03:: 00931852 3 43
114693 mean%2:42:00:: 02635189 2 89
114694 mean%2:42:03:: 02742482 5 9
114695 mean%5:00:00:contemptible:00 00905039 8 0
114696 mean%5:00:00:ignoble:01 01589650 3 3
114697 mean%5:00:00:nasty:00 01587787 2 6
114698 mean%5:00:00:normal:01 01594146 1 16
114699 mean%5:00:00:poor:03 02025718 5 0
114700 mean%5:00:00:skilled:00 02227663 4 0
114701 mean%5:00:01:stingy:00 01112969 7 0
114702 mean%5:00:02:stingy:00 01113807 6 0
114703 mean_deviation%1:09:00:: 06023476 1 0
114704 mean_deviation_from_the_mean%1:09:00:: 06023476 1 0
114705 mean_distance%1:07:00:: 05084982 1 0
```

Figure 3.2. The word "mean" in WordNet

probability of being negative and 0.125 probability of being positive, on the other hand if it is used in a context as "excellent" , "skilled" (he has a mean forehand) it has 1 as a positive probability. It should be noted that in both of these examples, the word "mean" is a verb. If it used as a noun where it means "an average of n numbers computed by adding some function of the numbers and dividing by some function of n", it has zero probability of being either positive or negative, it is a completely objective word. Both SentiWordNet and WordNet dictionaries consist of four type of POS tags: noun, verb, adverb, adjective.

```
4981 a  00904548  0      0.75   contemptible#1  deserving of contempt or scorn
4982 a  00904745  0      0.875  scurvy#1 scummy#1 miserable#3 low-down#1 low#6 abject#1 of the most contemptible kind; "abject
     cowardice"; "a low stunt to pull"; "a low-down sneak"; "his miserable treatment of his family"; "You miserable skunk!"; "a scummy
     rabble"; "a scurvy trick"
4983 a  00905039  0.125  0.5    mean#8 bastardly#2   of no value or worth; "I was caught in the bastardly traffic"
4984 a  00905181  0.333  0.667  pitiful#1 pitiable#1 pathetic#2 inspiring mixed contempt and pity; "their efforts were pathetic";
     "pitiable lack of character"; "pitiful exhibition of cowardice"
4985 a  00905386  0.625  0      ethical#2   conforming to accepted standards of social or professional behavior; "an ethical
     lawyer"; "ethical medical practice"; "an ethical problem"; "had no ethical objection to drinking"; "Ours is a world of nuclear giants and
     ethical infants"- Omar N. Bradley
12339 a  02227485  0.5    0    hot#10  performed or performing with unusually great skill and daring and energy; "a hot
      drummer"; "he's hot tonight"
12340 a  02227663  1      0    mean#4  excellent; "famous for a mean backhand"
12341 a  02227772  0.625  0    sure-handed#1   proficient and confident in performance; "promising playwrights...sure-handed
      enough to turn out top-drawer scripts"
12342 a  02227946  0      0    technical#4 expert#2    of or relating to or requiring special knowledge to be understood;
      "technical terminology"; "a technical report"; "technical language"
51072 n  06023675  0  0  mode#6 modal_value#1    the most frequent value of a random variable
51073 n  06023821  0  0  median_value#1 median#1 the value below which 50% of the cases fall
51074 n  06023969  0  0  mean_value#1 mean#1     an average of n numbers computed by adding some function of the numbers
      and dividing by some function of n
51075 n  06024230  0  0  first_moment#1 expected_value#1 expectation#4 arithmetic_mean#1 the sum of the values of a random
      variable divided by the number of values
51076 n  06024431  0  0  geometric_mean#1        the mean of n numbers expressed as the n-th root of their product
51077 n  06024576  0  0  harmonic_mean#1 the mean of n numbers expressed as the reciprocal of the arithmetic mean of the
      reciprocals of the numbers
```

Figure 3.3. The word "mean" in SentiWordNet

Since understanding the concept in which the word is written is not within the scope of this thesis, we simply take the average value of the scores if a word has more than one meaning with the same POS. For example, if we have the word "mean" in an article as a verb, instead of finding the particular meaning of it, we take the average of the scores where the word is "mean" and the POS is "verb". Same thing will be applied if it is noun, then we would have to search for scores where the word is "mean" and the POS is "noun". But we still have to find which POS does a word belongs to, which is explained in the next section.

3.3. Part-of-Speech Tagging

With the help of NYT's own document parser and a similar version we wrote for Reuters, we parsed the documents. For performing POS tagging we used the Stanford POS Tagger of Stanford NLP Group (Toutanova et al. (2003)). It is a maximum-entropy (CMM) part-of-speech tagger for many languages, including English, which produces the results in Penn Treebank tag set (Marcus et al. (1994)) [1]. Tagged articles are stored in different files with each word in an article is labeled with a part-of-speech tag, separated from the word with an underscore ('_'), thus producing the word/tag pairs that will be used extensively in our calculations. In Figures 3.4 and 3.5, you can see an example output of the POS Tagger that is applied to the articles given in the Chapter 1. Figure 3.4 shows the output of the article from New York Times while the Figure 3.5 shows the output of the article from Reuters. The Stanford POS tagger does not require the text to be tokenized, but the output of it needs one. So, we use the whitespace tokenizer of the same tool.

As it can be seen from the Penn Treebank tag set in Appendix, there are more than just one type of noun, i.e. plural noun, singular noun, singular proper noun, etc. But our knowledge-base of SentiWordNet dictionary only has one type of noun. Same issue repeats itself with the adjectives, adverbs and verbs. For this reason, every word/tag pair is grouped into one of these word/tag pairs: word/noun, word/adjective, word/verb, word/adverb. After this conversion, some of the tags are left out, i.e. interjections, prepositions, foreign words, etc. This reduction works as some kind of stop words removal for our thesis. Since they are not assumed to carry any kind of opinion, they do not exist in SentiWordNet, and neither do in the scoring part of our thesis.

[1] See Appendix A and Appendix B for details

```
traps_NNS for_IN Israeli_JJ soldiers_NNS in_IN civilian_JJ homes_NNS and_CC stored_VBD weapons_NNS in_IN mosques_NNS and_CC schools_NNS ._.
Tunnels_NNS have_VBP been_VBN dug_VBN beneath_IN private_JJ property_NN ._. With_IN international_JJ condemnation_NN rising_VBG over_IN
the_DT death_NN toll_NN in_IN Gaza_NNP exceeding_VBG 650_CD in_IN the_DT warâ_NN $_$ s_NNS 16th_JJ day_NN ,_, Israel_NNP points_VBZ to_TO
its_PRP$ adversariesâ_JJ $_$ practice_NN of_IN embedding_NN forces_NNS throughout_IN the_DT crowded_VBN ,_, impoverished_VBN coastal_JJ
enclave_NN of_IN 1.7_CD million_CD people_NNS ._. â_RB $_$ Hamas_NNP uses_VBZ schools_NNS ,_, residential_JJ buildings_NNS ,_, mosques_NNS
and_CC hospitals_NNS to_TO fire_VB rockets_NNS at_IN Israeli_JJ civilians_NNS ,_, â_VB $_$ Prime_NNP Minister_NNP Benjamin_NNP Netanyahu_NNP
told_VBD his_PRP$ Canadian_JJ counterpart_NN in_IN a_DT call_NN over_IN the_DT weekend_NN ,_, according_VBG to_TO a_DT statement_NN from_IN
Mr._NNP Netanyahuâ_NNP $_$ s_NNS office_NN ._. â_RB $_$ Hamas_NNP uses_VBZ innocent_JJ civilians_NNS as_IN a_DT human_JJ shield_NN for_IN
terrorist_JJ activity.â_NN $_$ Two_CD brothers_NNS on_IN Wednesday_NNP grieving_VBG their_PRP$ father_NN ,_, killed_VBN in_IN shelling_VBG
in_IN Gaza_NNP ._. Polls_NNS find_VBP Israel_NNP is_VBZ losing_VBG backing_NN in_IN many_JJ countries_NNS ,_, though_IN Americans_NNPS
are_VBP supportive_JJ ._. Nothing_NN is_VBZ ever_RB so_RB clear_JJ in_IN the_DT complex_NN and_CC often_RB brutal_JJ calculus_NN of_IN
urban_JJ warfare_NN ._. There_EX is_VBZ no_DT evidence_NN that_IN Hamas_NNP and_CC other_JJ militants_NNS force_VBP civilians_NNS to_TO
stay_VB in_IN areas_NNS that_WDT are_VBP under_IN attack_NN ._. But_CC it_PRP is_VBZ indisputable_JJ that_IN Gaza_NNP militants_NNS
operate_VBP in_IN civilian_JJ areas_NNS ,_, draw_NN return_NN fire_NN to_TO civilian_JJ structures_NNS ,_, and_CC on_IN some_DT level_NN
benefit_NN in_IN the_DT diplomatic_JJ arena_NN from_IN the_DT rising_VBG casualties_NNS ._. They_PRP also_RB have_VBP at_IN times_NNS
encouraged_VBD residents_NNS not_RB to_TO flee_VB their_PRP$ homes_NNS when_WRB alerted_VBN by_IN Israel_NNP to_TO a_DT pending_VBG strike_NN
and_CC ,_, having_VBG prepared_VBN extensively_RB for_IN war_NN ,_, did_VBD not_RB build_VB civilian_JJ bomb_NN shelters_NNS ._.
Israel_NNP ,_, for_IN its_PRP$ part_NN ,_, says_VBZ it_PRP takes_VBZ precautions_NNS to_TO avoid_VB killing_VBG civilians_NNS ,_, but_CC
has_VBZ also_RB accepted_VBN as_IN inevitable_JJ that_IN there_EX will_MD be_VB large_JJ numbers_NNS of_IN civilian_JJ casualties_NNS
when_WRB it_PRP strikes_VBZ at_IN certain_JJ targets_NNS ,_, like_IN Hamas_NNP members_NNS in_IN their_PRP$ houses_NNS ,_, or_CC
offices_NNS ,_, or_CC mosques_NNS ._. â_RB $_$ Itâ_JJ $_$ s_VBZ a_DT bit_NN of_IN a_DT fluid_NN concept_NN ,_, â_VB $_$ said_VBD Bill_NNP
Van_NNP Esveld_NNP ,_, a_DT lawyer_NN and_CC senior_JJ researcher_NN in_IN the_DT Middle_NNP East_NNP division_NN of_IN Human_NNP Rights_NNP
Watch_NNP ._. â_RB $_$ If_IN you_PRP have_VBP any_DT choice_NN in_IN the_DT matter_NN ,_, you_PRP should_MD not_RB be_VB fighting_VBG from_IN
an_DT apartment_NN building_NN full_JJ of_IN civilians.â_JJ $_$ Some_DT of_IN Hamasâ_NNP $_$ s_NNS tactics_NNS mirror_VBP those_DT of_IN
other_JJ insurgent_JJ groups_NNS ,_, like_IN the_DT Irish_JJ Republican_JJ Army_NNP and_CC Nicaraguan_JJ contras_NNS ._. â_RB $_$ Hamas_NNP
knows_VBZ that_IN it_PRP works_VBZ to_TO its_PRP$ advantage_NN ,_, politically_RB and_CC diplomatically_RB ,_, as_IN the_DT civilian_JJ
death_NN toll_NN mounts_VBZ ,_, there_EX is_VBZ increasing_VBG pressure_NN to_TO end_VB the_DT war_NN immediately_RB ,_, and_CC what_WP
that_WDT typically_RB entails_VBZ ,_, if_IN past_NN is_VBZ precedent_NN ,_, is_VBZ making_VBG some_DT concessions_NNS to_TO Hamas_NNP ,_,
â_VBD $_$ said_VBD Nathan_NNP Thrall_NNP ,_, co-author_NN of_IN a_DT recent_JJ International_NNP Crisis_NNP Group_NNP report_NN on_IN
Gaza_NNP ._. â_RB $_$ Hamas_NNP deliberately_RB buries_VBZ the_DT weapons_NNS in_IN populated_JJ civilian_JJ areas_NNS hoping_VBG that_DT
will_MD reduce_VB the_DT chance_NN that_IN those_DT weapons_NNS will_MD be_VB taken_VBN out_RP from_IN the_DT air_NN â_NN $_$ ''_'' thatâ_IN
$_$ s_VBZ a_DT pretty_RB clear-cut_JJ case.â_NN $_$ Experts_NNS in_IN international_JJ law_NN say_VBP that_IN even_RB as_IN Hamas_NNP is_VBZ
legally_RB obligated_VBN to_TO minimize_VB its_PRP$ operations_NNS near_IN civilians_NNS â_VBP $_$ ''_'' and_CC is_VBZ committing_VBG a_DT
war_NN crime_NN by_IN firing_NN rockets_NNS indiscriminately_RB â_VBP $_$ ''_'' so_RB ,_, too_RB ,_, is_VBZ Israel_NNP obligated_VBN to_TO
```

Figure 3.4. An Example Output of POS Tagger - 1

```
human_JJ rights_NNS body_NN accused_VBD Israeli_JJ forces_NNS on_IN Thursday_NNP of_IN mistreating_VBG Palestinian_JJ children_NNS ,_,
including_VBG by_IN torturing_VBG those_DT in_IN custody_NN and_CC using_VBG others_NNS as_IN human_JJ shields_NNS ._. Palestinian_JJ
children_NNS in_IN the_DT Gaza_NNP and_CC the_DT West_NNP Bank_NNP ,_, captured_VBN by_IN Israel_NNP in_IN the_DT 1967_CD war_NN ,_, are_VBP
routinely_RB denied_VBN registration_NN of_IN their_PRP$ birth_NN and_CC access_NN to_TO health_NN care_NN ,_, decent_JJ schools_NNS and_CC
clean_JJ water_NN ,_, the_DT U.N._NNP Committee_NNP on_IN the_DT Rights_NNS of_IN the_DT Child_NNP said_VBD ._. ``_`` Palestinian_JJ
children_NNS arrested_VBN by_IN -LRB-_-LRB- Israeli_JJ -RRB-_-RRB- military_JJ and_CC police_NNS are_VBP systematically_RB subject_JJ to_TO
degrading_JJ treatment_NN ,_, and_CC often_RB to_TO acts_NNS of_IN torture_VBP ,_, are_VBP interrogated_VBN in_IN Hebrew_NNP ,_, a_DT
language_NN they_PRP did_VBD not_RB understand_VB ,_, and_CC sign_NN confessions_NNS in_IN Hebrew_NNP in_IN order_NN to_TO be_VB
released_VBN ,_, ''_'' it_PRP said_VBD in_IN a_DT report.The_NNP Israeli_NNP Foreign_NNP Ministry_NNP said_VBD it_PRP had_VBD responded_VBN
to_TO a_DT report_NN by_IN the_DT U.N._NNP children_NNS 's_POS agency_NN UNICEF_NNP in_IN March_NNP on_IN ill-treatment_NN of_IN
Palestinian_JJ minors_NNS and_CC questioned_VBD whether_IN the_DT U.N._NNP committee_NN 's_POS investigation_NN covered_VBD new_JJ
ground_NN ._. ``_`` If_IN someone_NN simply_RB wants_VBZ to_TO magnify_VB their_PRP$ political_JJ bias_NN and_CC political_JJ bashing_VBG
of_IN Israel_NNP not_RB based_VBN on_IN a_DT new_JJ report_NN ,_, on_IN work_NN on_IN the_DT ground_NN ,_, but_CC simply_RB recycling_VBG
old_JJ stuff_NN ,_, there_EX is_VBZ no_DT importance_NN in_IN that_DT ,_, ''_'' spokesman_NN Yigal_NNP Palmor_NNP said_VBD ._. The_DT
report_NN by_IN the_DT U.N._NNP Committee_NNP on_IN the_DT Rights_NNS of_IN the_DT Child_NNP acknowledged_VBD Israel_NNP 's_POS national_JJ
security_NN concerns_NNS and_CC noted_VBD that_DT children_NNS on_IN both_DT sides_NNS of_IN the_DT conflict_NN continue_VBP to_TO be_VB
killed_VBN and_CC wounded_VBN ,_, but_CC that_IN more_JJR casualties_NNS are_VBP Palestinian_JJ ._. Most_JJS Palestinian_JJ children_NNS
arrested_VBN are_VBP accused_VBN of_IN having_VBG thrown_VBN stones_NNS ,_, an_DT offence_NN which_WDT can_MD carry_VB a_DT penalty_NN of_IN
up_RB to_TO 20_CD years_NNS in_IN prison_NN ,_, the_DT committee_NN said_VBD ._. Israeli_JJ soldiers_NNS had_VBD testified_VBN to_TO the_DT
often_RB arbitrary_JJ nature_NN of_IN the_DT arrests_NNS ,_, it_PRP said_VBD ._. The_DT watchdog_NN 's_POS 18_CD independent_JJ experts_NNS
examined_VBN Israel_NNP 's_POS record_NN of_IN compliance_NN with_IN a_DT 1990_CD treaty_NN as_IN part_NN of_IN its_PRP$ regular_JJ review_NN
of_IN a_DT pact_NN signed_VBN by_IN all_DT nations_NNS except_IN Somalia_NNP and_CC the_DT United_NNP States_NNPS ._. An_DT Israeli_JJ
delegation_NN attended_VBD the_DT session_NN ._. The_DT U.N._NNP committee_NN regretted_VBD Israel_NNP 's_POS ``_`` persistent_JJ refusal_NN
''_'' to_TO respond_VB to_TO requests_NNS for_IN information_NN on_IN children_NNS in_IN the_DT Palestinian_JJ territories_NNS and_CC
occupied_VBN Syrian_JJ Golan_NNP Heights_NNP since_IN the_DT last_JJ review_NN in_IN 2002_CD ._. ``_`` DISPROPORTIONATE_NN ''_'' ``_``
Hundreds_NNS of_IN Palestinian_JJ children_NNS have_VBP been_VBN killed_VBN and_CC thousands_NNS injured_VBN over_IN the_DT reporting_NN
period_NN as_IN a_DT result_NN of_IN the_DT state_NN party_NN military_JJ operations_NNS ,_, especially_RB in_IN Gaza_NNP where_WRB the_DT
state_NN party_NN proceeded_VBD to_TO -LRB-_-LRB- conduct_NN -RRB-_-RRB- air_NN and_CC naval_JJ strikes_NNS on_IN densely_RB populated_JJ
areas_NNS with_IN a_DT significant_JJ presence_NN of_IN children_NNS ,_, thus_RB disregarding_VBG the_DT principles_NNS of_IN
proportionality_NN and_CC distinction_NN ,_, ''_'' the_DT report_NN said_VBD ._. Israel_NNP battled_VBD a_DT Palestinian_JJ uprising_NN
during_IN part_NN of_IN the_DT 10-year_JJ period_NN examined_VBN by_IN the_DT committee_NN ._. It_PRP withdrew_VBD its_PRP$ troops_NNS and_CC
settlers_NNS from_IN the_DT Gaza_NNP Strip_NNP in_IN 2006_CD ,_, but_CC still_RB blockades_VBZ the_DT Hamas-run_JJ enclave_NN ,_, from_IN
where_WRB Palestinian_JJ militants_NNS have_VBP sometimes_RB fired_VBN rockets_NNS into_IN Israel_NNP ._. During_IN the_DT 10-year_JJ
period_NN ,_, an_DT estimated_VBN 7,000_CD Palestinian_JJ children_NNS aged_JJ 12_CD to_TO 17_CD ,_, but_CC some_DT as_IN young_JJ as_IN
nine_CD ,_, had_VBD been_VBN arrested_VBN ,_, interrogated_VBN and_CC detained_VBN ,_, the_DT U.N._NNP report_NN said_VBD ._. Many_JJ are_VBP
```

Figure 3.5. An Example Output of POS Tagger - 2

3.4. Proposed Models

After tagging the content of the articles with the Stanford POS Tagger, we parsed our tagged content and retrieve their individual sentiment scores from our knowledge-base, SentiWordNet. The result is a new string that shows positive, negative and objective scores of the words found in the SentiWordNet. Words that do not have any kind of noun, adjective, adverb or verb part-of-speech and the words that have those POS but somehow do not exist in the SentiWordNet are discarded and left out from the final result.

```
 4 With_0.0:0.0:0.0 international_0.69:0.0:0.31 condemnation_0.7:0.1:0.2 rising_0.0:0.0:0.0 over_0.0:0.0:0.0 the_0.0:0.0:0.0 death_0.8:0.03:0.17
   toll_1.0:0.0:0.0 in_0.0:0.0:0.0 Gaza_0.0:0.0:0.0 exceeding_0.0:0.0:0.0 650_0.0:0.0:0.0 in_0.0:0.0:0.0 the_0.0:0.0:0.0 warâ_0.0:0.0:0.0
   $_0.0:0.0:0.0 s_1.0:0.0:0.0 16th_1.0:0.0:0.0 day_0.95:0.05:0.0 ,_0.0:0.0:0.0 Israel_1.0:0.0:0.0 points_0.0:0.0:0.0 to_0.0:0.0:0.0
   its_0.0:0.0:0.0 adversariesâ_0.0:0.0:0.0 $_0.0:0.0:0.0 practice_0.85:0.13:0.03 of_0.0:0.0:0.0 embedding_0.0:0.0:0.0 forces_0.0:0.0:0.0
   throughout_0.0:0.0:0.0 the_0.0:0.0:0.0 crowded_0.0:0.0:0.0 ,_0.0:0.0:0.0 impoverished_0.0:0.0:0.0 coastal_1.0:0.0:0.0 enclave_1.0:0.0:0.0
   of_0.0:0.0:0.0 1.7_0.0:0.0:0.0 million_0.0:0.0:0.0 people_1.0:0.0:0.0 ._0.0:0.0:0.0
 5 â_0.0:0.0:0.0 $_0.0:0.0:0.0 Hamas_1.0:0.0:0.0 uses_0.0:0.0:0.0 schools_0.0:0.0:0.0 ,_0.0:0.0:0.0 residential_1.0:0.0:0.0
   buildings_0.0:0.0:0.0 ,_0.0:0.0:0.0 mosques_0.0:0.0:0.0 and_0.0:0.0:0.0 hospitals_0.0:0.0:0.0 to_0.0:0.0:0.0 fire_0.93:0.03:0.04
   rockets_0.0:0.0:0.0 at_0.0:0.0:0.0 Israeli_1.0:0.0:0.0 civilians_0.0:0.0:0.0 ,_0.0:0.0:0.0 â_0.0:0.0:0.0 $_0.0:0.0:0.0 Prime_0.84:0.16:0.0
   Minister_1.0:0.0:0.0 Benjamin_0.94:0.0:0.06 Netanyahu_0.0:0.0:0.0 told_0.0:0.0:0.0 his_0.0:0.0:0.0 Canadian_1.0:0.0:0.0
   counterpart_1.0:0.0:0.0 in_0.0:0.0:0.0 a_0.0:0.0:0.0 call_0.99:0.01:0.0 over_0.0:0.0:0.0 the_0.0:0.0:0.0 weekend_1.0:0.0:0.0 ,_0.0:0.0:0.0
   according_0.0:0.0:0.0 to_0.0:0.0:0.0 a_0.0:0.0:0.0 statement_0.91:0.05:0.04 from_0.0:0.0:0.0 Mr._1.0:0.0:0.0 Netanyahuâ_0.0:0.0:0.0
   $_0.0:0.0:0.0 s_1.0:0.0:0.0 office_1.0:0.0:0.0 ._0.0:0.0:0.0
 6 â_0.0:0.0:0.0 $_0.0:0.0:0.0 Hamas_1.0:0.0:0.0 uses_0.0:0.0:0.0 innocent_0.43:0.34:0.23 civilians_0.0:0.0:0.0 as_0.0:0.0:0.0 a_0.0:0.0:0.0
   human_1.0:0.0:0.0 shield_0.96:0.0:0.04 for_0.0:0.0:0.0 terrorist_0.0:0.0:0.0 activity_0.94:0.04:0.02 ._0.0:0.0:0.0
 7 $_0.0:0.0:0.0 Two_0.0:0.0:0.0 brothers_0.0:0.0:0.0 on_0.0:0.0:0.0 Wednesday_1.0:0.0:0.0 grieving_0.0:0.0:0.0 their_0.0:0.0:0.0
   father_1.0:0.0:0.0 ,_0.0:0.0:0.0 killed_0.0:0.0:0.0 in_0.0:0.0:0.0 shelling_0.0:0.0:0.0 in_0.0:0.0:0.0 Gaza_0.0:0.0:0.0 ._0.0:0.0:0.0
 8 Polls_0.0:0.0:0.0 find_0.98:0.02:0.01 Israel_1.0:0.0:0.0 is_0.0:0.0:0.0 losing_0.0:0.0:0.0 backing_1.0:0.0:0.0 in_0.0:0.0:0.0
   many_0.0:0.0:0.0 countries_0.0:0.0:0.0 ,_0.0:0.0:0.0 though_0.0:0.0:0.0 Americans_0.0:0.0:0.0 are_0.0:0.0:0.0
   supportive_0.88:0.13:0.0 ._0.0:0.0:0.0
 9 Nothing_0.5:0.25:0.25 is_0.0:0.0:0.0 ever_1.0:0.0:0.0 so_1.0:0.0:0.0 clear_0.54:0.4:0.06 in_0.0:0.0:0.0 the_0.0:0.0:0.0 complex_0.97:0.0:0.03
   and_0.0:0.0:0.0 often_0.79:0.21:0.0 brutal_0.5:0.03:0.47 calculus_1.0:0.0:0.0 of_0.0:0.0:0.0 urban_1.0:0.0:0.0
   warfare_1.0:0.0:0.0 ._0.0:0.0:0.0
10 There_0.0:0.0:0.0 is_0.0:0.0:0.0 no_0.0:0.0:0.0 evidence_1.0:0.0:0.0 that_0.0:0.0:0.0 Hamas_1.0:0.0:0.0 and_0.0:0.0:0.0 other_0.59:0.09:0.31
   militants_0.0:0.0:0.0 force_0.93:0.06:0.01 civilians_0.0:0.0:0.0 to_0.0:0.0:0.0 stay_0.95:0.01:0.03 in_0.0:0.0:0.0 areas_0.0:0.0:0.0
   that_0.0:0.0:0.0 are_0.0:0.0:0.0 under_0.0:0.0:0.0 attack_0.85:0.04:0.11 ._0.0:0.0:0.0
```

Figure 3.6. NYT Article That is Prepared for Naive Bayes Approach

In Figure 3.6, we can see an example output of an article (same NYT article from the previous section) after processing it with the information from our knowledge-base. Each word in each sentence is tagged with the corresponding objective ,positive and negative scores respectively, separated with semi-colons. It should be noted that there exists some words that do not have a non-zero score associated with them in any of the scoring categories: objective, negative and positive. These are the words that is either has a POS tag other than the ones mentioned in the Section 3.3 or they do not simply exist in the SentiWordNet Database. In this phase, they are not discarded yet to make the output more readable.

In the Figure 3.7, there is another example output of an article. This time the article is from Reuters and it is prepared for the use of our second method. The only notable change between the figures is the repeated words in the Figure 3.7. The reason for this repetition is, the first scores represents the probabilities we get from cumulative distribution function while the second one is the score we get from SentiWordNet (Details will be explained in the Section 3.4.2). It should be noted that the words that cannot be found in our knowledge-base do not have the second scores, thus the repetition.

```
   estimated_0.0:0.0:0.0 7,000_0.0:0.0:0.0 Palestinian_1.0:0.0:0.0 Palestinian_1.0:0.0:0.0 children_0.0:0.0:0.0 aged_0.15:0.94:0.0
   aged_0.73:0.28:0.0 12_0.0:0.0:0.0 to_0.0:0.0:0.0 17_0.0:0.0:0.0 ,_0.0:0.0:0.0 but_0.0:0.0:0.0 some_0.0:0.0:0.0 as_0.0:0.0:0.0
   young_0.13:0.88:0.86 young_0.7:0.15:0.15 as_0.0:0.0:0.0 nine_0.0:0.0:0.0 ,_0.0:0.0:0.0 had_0.0:0.0:0.0 been_0.0:0.0:0.0
   arrested_0.0:0.0:0.0 ,_0.0:0.0:0.0 interrogated_0.0:0.0:0.0 and_0.0:0.0:0.0 detained_0.0:0.0:0.0 ,_0.0:0.0:0.0 the_0.0:0.0:0.0
   U.N._0.0:0.0:0.0 report_1.0:0.0:0.0 report_1.0:0.0:0.0 said_0.0:0.0:0.0 ._0.0:0.0:0.0
16 Many_0.0:0.0:0.0 are_0.0:0.0:0.0 brought_0.0:0.0:0.0 in_0.0:0.0:0.0 leg_1.0:0.0:0.0 leg_1.0:0.0:0.0 chains_1.0:0.0:0.0 chains_1.0:0.0:0.0
   and_0.0:0.0:0.0 shackles_0.0:0.0:0.0 before_0.0:0.0:0.0 military_1.0:0.0:0.0 military_1.0:0.0:0.0 courts_0.0:0.0:0.0 ,_0.0:0.0:0.0
   while_0.0:0.0:0.0 youths_0.0:0.0:0.0 are_0.0:0.0:0.0 held_0.0:0.0:0.0 in_0.0:0.0:0.0 solitary_0.18:0.82:0.89 solitary_0.78:0.05:0.18
   confinement_1.0:0.0:0.0 confinement_1.0:0.0:0.0 ,_0.0:0.0:0.0 sometimes_0.0:0.0:0.0 for_0.0:0.0:0.0 months_0.0:0.0:0.0 ,_0.0:0.0:0.0
   the_0.0:0.0:0.0 report_1.0:0.0:0.0 report_1.0:0.0:0.0 said_0.0:0.0:0.0 ._0.0:0.0:0.0
17 It_0.0:0.0:0.0 voiced_0.0:0.0:0.0 deep_0.13:0.91:0.85 deep_0.68:0.2:0.12 concern_0.1:0.91:0.89 concern_0.6:0.2:0.2 at_0.0:0.0:0.0
   the_0.0:0.0:0.0 ``_0.0:0.0:0.0 continuous_0.18:0.0:0.9 continuous_0.75:0.0:0.25 use_0.51:0.84:0.0 use_0.95:0.05:0.0 of_0.0:0.0:0.0
   Palestinian_1.0:0.0:0.0 Palestinian_1.0:0.0:0.0 children_0.0:0.0:0.0 as_0.0:0.0:0.0 human_1.0:0.0:0.0 human_1.0:0.0:0.0 shields_0.0:0.0:0.0
   and_0.0:0.0:0.0 informants_0.0:0.0:0.0 ''_0.0:0.0:0.0 ,_0.0:0.0:0.0 saying_0.0:0.0:0.0 14_0.0:0.0:0.0 such_0.25:0.0:0.85 such_0.88:0.0:0.13
   cases_0.0:0.0:0.0 had_0.0:0.0:0.0 been_0.0:0.0:0.0 reported_0.0:0.0:0.0 between_0.0:0.0:0.0 January_0.0:0.0:0.0 2010_0.0:0.0:0.0
   and_0.0:0.0:0.0 March_0.97:0.0:0.04 March_0.98:0.0:0.02 2013_0.0:0.0:0.0 alone_0.18:0.86:0.85 alone_0.75:0.13:0.13 ._0.0:0.0:0.0
18 Israeli_1.0:0.0:0.0 Israeli_1.0:0.0:0.0 soldiers_0.0:0.0:0.0 had_0.0:0.0:0.0 used_0.0:0.0:0.0 Palestinian_1.0:0.0:0.0 Palestinian_1.0:0.0:0.0
   children_0.0:0.0:0.0 to_0.0:0.0:0.0 enter_1.0:0.0:0.0 enter_1.0:0.0:0.0 potentially_1.0:0.0:0.0 potentially_1.0:0.0:0.0
   dangerous_0.03:0.0:0.99 dangerous_0.31:0.0:0.69 buildings_0.0:0.0:0.0 before_0.0:0.0:0.0 them_0.0:0.0:0.0 and_0.0:0.0:0.0 to_0.0:0.0:0.0
   stand_0.26:0.01:0.85 stand_0.92:0.01:0.07 in_0.0:0.0:0.0 front_0.99:0.0:0.01 front_0.99:0.0:0.01 of_0.0:0.0:0.0 military_1.0:0.0:0.0
   military_1.0:0.0:0.0 vehicles_0.0:0.0:0.0 to_0.0:0.0:0.0 deter_1.0:0.0:0.0 deter_1.0:0.0:0.0 stone-throwing_0.0:0.0:0.0 ,_0.0:0.0:0.0
   it_0.0:0.0:0.0 said_0.0:0.0:0.0 ._0.0:0.0:0.0
19 ``_0.0:0.0:0.0 Almost_1.0:0.0:0.0 Almost_1.0:0.0:0.0 all_0.0:0.0:0.0 those_0.0:0.0:0.0 using_0.0:0.0:0.0 children_0.0:0.0:0.0 as_0.0:0.0:0.0
   human_1.0:0.0:0.0 human_1.0:0.0:0.0 shields_0.0:0.0:0.0 and_0.0:0.0:0.0 informants_0.0:0.0:0.0 have_0.29:0.11:0.74 have_0.93:0.02:0.05
   remained_0.0:0.0:0.0 unpunished_0.18:0.0:0.9 unpunished_0.75:0.0:0.25 and_0.0:0.0:0.0 the_0.0:0.0:0.0 soldiers_0.0:0.0:0.0
   convicted_0.0:0.0:0.0 for_0.0:0.0:0.0 having_0.0:0.0:0.0 forced_0.0:0.0:0.0 at_0.0:0.0:0.0 gunpoint_1.0:0.0:0.0 gunpoint_1.0:0.0:0.0
   a_0.0:0.0:0.0 nine-year-old_0.0:0.0:0.0 child_0.9:0.44:0.0 child_0.97:0.03:0.0 to_0.0:0.0:0.0 search_1.0:0.0:0.0 search_1.0:0.0:0.0
   bags_0.0:0.0:0.0 suspected_0.0:0.0:0.0 of_0.0:0.0:0.0 containing_0.0:0.0:0.0 explosives_0.0:0.0:0.0 only_0.24:0.84:0.85 only_0.86:0.05:0.09
   received_0.0:0.0:0.0 a_0.0:0.0:0.0 suspended_0.0:0.0:0.0 sentence_1.0:0.0:0.0 sentence_1.0:0.0:0.0 of_0.0:0.0:0.0 three_0.0:0.0:0.0
   months_0.0:0.0:0.0 and_0.0:0.0:0.0 were_0.0:0.0:0.0 demoted_0.0:0.0:0.0 ,_0.0:0.0:0.0 ''_0.0:0.0:0.0 it_0.0:0.0:0.0
   said_0.0:0.0:0.0 ._0.0:0.0:0.0
```

Figure 3.7. Reuters Article That is Prepared for the Relativistic Method

From this point, our work is separated into two different parts where we tried two different methods for analyzing the objectivity of the article.

3.4.1. Naive Bayes Approach

First proposed method uses Naive Bayes to calculate the sentiment on sentence level. Naive Bayes is a popular method for calculating the conditional probability of a concept or a category from the given features as shown in the formula below:

$$p(C_k|x_1, x_2, \ldots, x_n) \tag{3.1}$$

where C_k is the possible categories of the outcome, and $x_1, x_2, ..., x_n$ is the feature set we have. But the problem of this calculation is, we have to compute the probabilities for every feature which leads to an exponential complexity of 2^n. At this point, the most important assumption of Naive Bayes takes its place: Conditional Independence. If we assume that every feature (words and categories) is independent from each other, the complexity reduces greatly and the formula becomes as the famous formula of Naive Bayes:

$$p(C_k|x_1, x_2, \dots, x_n) = \frac{1}{Z} p(C_k) \prod_{i=1}^{n} p(x_i|C_k) \tag{3.2}$$

In our thesis, we have three possible categories of outcome: positive sentiment, negative sentiment and objective which means there are no sentiment. And our features are the words in the sentence. This situation does not violate the assumption of Naive Bayes, since the sentiment of each word in the sentence are indeed independent from each other. That's because we get sentiment of each of the words from SentiWordNet dictionary independently. In addition, we can discard the values $\frac{1}{Z}p(C_k)$, considering it is the same for each category and it will not affect the comparison and ratio of the scores of each category. So, now our calculation reduces to a more simple one where we just take the product of each word's sentiment in the sentence for each of the three categories. This operation was applied to every article and every sentence.

To demonstrate better how we applied Naive Bayes, here we have a sentence (without the words that cannot be found), 9th sentence from the Figure 3.6, part of the article that is given as an example in the previous section, tagged with the scores from SentiWordNet:

$Nothing\_0.5 : 0.25 : 0.25$ $ever\_1.0 : 0.0 : 0.0$ $so\_1.0 : 0.0 : 0.0$ $clear\_0.54 : 0.4 : 0.06$ $complex\_0.97 : 0.0 : 0.03$ $often\_0.79 : 0.21 : 0.0$ $brutal\_0.5 : 0.03 : 0.47$ $calculus\_1.0 : 0.0 : 0.0$ $urban\_1.0 : 0.0 : 0.0$ $warfare\_1.0 : 0.0 : 0.0$

Using the score next to words we have calculated the probabilities of each category as follows:

$$p(objective) = (0.5)(1.0)(1.0)(0.54)(0.97)(0.79)(0.5)(1.0)(1.0)(1.0) = 0.103 \tag{3.3}$$

$$p(positive) = (0.25)(0.0)(0.0)(0.4)(0.0)(0.21)(0.03)(0.0)(0.0)(0.0) = 0.000 \tag{3.4}$$

$$p(negative) = (0.25)(0.0)(0.0)(0.06)(0.03)(0.0)(0.47)(0.0)(0.0)(0.0) = 0.000 \tag{3.5}$$

However, the problem with this calculation is, as you can see from the positive and negative probability, if we have a zero as a probability of a word in some particular class (negative, positive or objective) like we have in the word "calculus", it automatically resets the posterior probability of that class to zero. Both positive score and negative score of the word “calculus” is zero, which leads the equations to a zero probability. Simply removing the words score from positive or negative calculation will not work here, since, if the number of products is different, the result will not and cannot be normalized properly. So, we cannot infer anything from the result. The reason for this inability to normalize is because the scores from SentiWordNet is dependent on each other; if you increase the negative score the positive scores decreases or vice versa (For further explanation refer to Section 4.1).

In SentiWordNet dictionary, there are a lot of words with a negative or positive score of zero, so that it poses a big problem. The solution of this zero probability values resides in the use of m-estimate of conditional probability. Instead of using the class probabilities that comes from the SentiWordNet directly, we used the m-estimate formula:

$$P(x_i|y_j) = \frac{n_c + mp}{n + m} \tag{3.6}$$

where n is the number of instances from class y_j, n_c is the number of training samples from class y_j that has the value x_i, m is the equivalent sample size and p is the prior probability of the class. So, for our method, we computed the m-estimate approach $P(word_i|class_i)$ for every word, where;

m, *the sample size*, is the total number of words in SentiWordNet
p, *prior probability*, is the average score of all the words in the SentiWordNet that belongs to $class_i$. This value is calculated for each class(negative, positive, objective), prior to m-estimate calculations.
n, *number of instance*, is the total number of words that belong to the $class_i$. In this thesis we heuristically assumed that the word belongs to a class if its score is higher than 0.5, i.e. if a word has 0.6 possibility of being positive, we counted this word as a positive word for calculating the number of instances that belong to the class “positive”.
$\boldsymbol{n_c}$, *the number of* $word_i$ *that belong to the* $class_i$. This value is the probability of a word being in a certain class, which is the value we get from the SentiWordNet. Even if we get zero here, the result will no longer resets to zero since the product $m * p$ next to the n_c will never be zero.

In the end, we get our new calculation without zero elements using m-estimate of conditional probability;

$$p(objective) \; = \; \frac{0.5 + (117659 * 0.9055)}{110644 + 117659} * \ldots * \frac{1.0 + (117659 * 0.9055)}{110644 + 117659} = 9.5326^{-4} \tag{3.7}$$

$$p(positive) = \frac{0.25 + (117659 * 0.0.0418)}{3572 + 117659} * \ldots * \frac{0.0 + (117659 * 0.0418)}{3572 + 117659} = 7.5149^{-21} \tag{3.8}$$

$$p(negative) = \frac{0.25 + (117659 * 0.0.0526)}{5683 + 117659} * \ldots * \frac{0.0 + (117659 * 0.0.0526)}{5683 + 117659} = 3.7189^{-20} \tag{3.9}$$

As we can see in the equations (3.8) and (3.9), zero scores are no longer problem for us and calculations end up as proper probabilities. Output of the article used in our calculations can be seen in the Figure 3.8. From left to right the columns represents the objectivity score, positive score, negative score and their ratios. Since the scores depend on the number of words used the sentence (The more words are used, the lower the score will be), we will use those ratios in the Chapter 4 to reach any conclusions.

sentence#	obj	pos	neg	ratio(pos/neg)	ratio(obj/pos)	ratio(obj/neg)
0	5.2453456227484215E-6	5.673285751995836E-31	3.273915748365984E-30	0.17328746944167334	9.245692623367566E24	1.6021626779389117E24
1	1.4447881547493594E-4	1.1296015067822708E-26	5.012440751816283E-25	0.022535957285339563	1.2790246348598758E22	2.882404453809919E20
2	0.060758433062717124	2.773458952190067E-11	1.848827601223322E-10	0.1500117669357024	2.19070965570769E9	3.28632226295814 8E8
3	5.463114213028912E-9	1.965772141131639E-57	6.3023994412917595E-53	3.11908529353501E-5	2.7791187486683748E48	8.668308417958946E43
4	6.104269912573225E-9	9.07472419314585E-65	5.9028795070233074E-61	1.5373385450861144E-4	6.726672659852061E55	1.034117316016751E52
5	0.004048836786761948	6.21822467380816E-19	9.346958512876932E-18	0.06652671738343077	6.511242354777064E15	4.33171579951278E14
6	1.0722560180721728E-4	1.3295256510122449E-21	3.2179421956222104E-20	0.041316020307045086	8.064951716093648E16	3.332117088774632E15
7	9.532602999034836E-4	7.514900276719182E-21	3.718971413715002E-20	0.20206932080750528	1.2684936124257569 6E17	2.5632364271153156E16
8	0.11882362552798001	4.601148037344507E-20	3.521632356893098E-19	0.13065384375908548	2.5824777649331 05E18	3.374106479212673 9E17
9	1.3847249566376063E-4	2.975224701063287E-27	5.099127330652002E-25	0.005834772321096087	4.654186139765285E22	2.715611646553152 2E20
10	1.417030164031195E-6	1.9450738129890145E-45	1.3348879594420991E-41	1.4571064179812778E-4	7.28522565348628 3E38	1.0615349056136714E35
11	4.86494351733729E-7	7.38120973905962E-48	9.054769505172355E-45	8.151736755798424E-4	6.590983984092956E40	5.372796640000929E37
12	3.731530310441876E-12	1.6051357390450398E-68	2.52280054300882E-63	6.3625154334661485E-6	2.3247443936809446E56	1.479122208365891 2E51
13	4.006240712519892E-6	5.936880228142012E-47	1.1517226787730613E-44	0.005154782776758911	6.74805715892583E40	3.4784768819415536E38
14	2.928125354252306E-8	6.652922751993396E-63	2.5289764037661917E-58	2.6306780648825973E-5	4.401261615994204E54	1.1578302391005688E50
15	7.174079128326865E-13	2.1493300072380713E-90	2.4156144253419285E-85	8.897653469401828E-6	3.3378211368973017E77	2.9698775818957032E72
16	7.01991811979758E-19	1.1471029320895672E-129	1.5859924008823103E-121	7.23271392379571E-9	6.119693292920165E110	4.4261990889062894E102
17	1.1091368359641806E-30	4.01652020192843E-201	2.0106314724904366E-187	1.9976411674056965E-14	2.7614372148101156E170	5.5163606615108134E156
18	2.687359481922964E-10	2.0163854796584416E-67	3.9069270698888145E-62	5.161052263296599E-6	1.3327607786474337E57	6.87844803307127 5E51
19	1.5970055997544273E-6	1.607537827507306E-35	5.28316213490995E-33	0.0030427569445295962	9.934482240027842E28	3.0228214826150657E26
20	6.0365045803213035E-9	1.1874947354084474E-44	2.530106773747012E-41	4.6934570024086505E-4	5.083394814584171E35	2.38586949885179E32
21	1.6661306895676898E-11	3.4718304842623916E-76	4.537179478987694E-72	7.651957566018534E-5	4.7989978114431695E64	3.67217276125789 5E60
22	2.560119811189736E-11	7.497883431590005E-76	9.09759647037251E-69	8.241609150293514E-8	3.4144566724036623E64	2.8140617354562765E57
23	8.513493291879319E-5	2.5286701203785143E-21	2.5297313496767005E-20	0.09995804972340949	3.366786843119316E16	3.3653744667264145E15
24	3.688790867092912E-7	1.4285940524118944E-50	4.184025867331633E-48	0.0034144006220568177	2.5821127148507504E43	8.81636725980722 1E40
25	3.645797557163144E-8	6.386387709341516E-49	2.981566664262742E-46	0.0021419570408702203	5.708700634993318E40	1.2227791519344235E38
26	2.7690751988539585E-5	3.8677990158157273E-29	5.128157637241336E-27	0.007542277927900025	7.159304781688493E23	5.399746643403823E21
27	1.3098824134228093E-5	7.395699139675302E-33	2.643544425960103E-31	0.02797645111255989	1.7711407517860682E27	4.9550232655805515E25
28	0.03798297617462973	2.1403600820104108E-14	1.2176199100079219E-12	0.017578228348750356	1.77460682872355E12	3.1194444064554276E10
29	2.0756944002955544E-4	9.705216236896198E-29	7.764367950319111E-27	0.012499686128987896	2.1387410126983192E24	2.673359136992271E22
30	2.820135414260972E-15	1.0689143789784024E-86	1.1282156250411875E-81	9.474380209362725E-6	2.638317408505881E71	2.4996422241165272E66
31	8.749848097598468E-4	1.4255008748125324E-16	2.547197330019824E-16	0.559635038091626 1	6.138086796157989E12	3.435088437977583 5E12
32	0.005619762464818291	1.8565224261173564E-19	4.858104005452116E-18	0.03821495842892274	3.0270372098715756E16	1.1567810113804452E15
33	3.4607376584678924E-15	1.3982298118089183E-115	6.331037913279218E-106	2.208531730438955E-10	2.475085017670068E100	5.466303797058407E90
34	1.3426638792697127E-10	1.0201463173494866E-61	1.023089088141711E-56	9.971236416981344E-6	1.3161483371896901E51	1.312362622993528E46
35	0.0366672206877870 2	1.479734261146517E-9	3.0808145954766712E-9	0.480306170750781 2	2.4779598371518947E7	1.1901794006566558E7
36	1.2689376828426611E-7	5.257059232946902E-56	3.244772526814546E-52	1.620162642990525 8E-4	2.413778553016882E48	3.910713840049679E44

Figure 3.8. An Example Output of Naive Bayes Method

3.4.2. Relativistic Method

In our second method, we made use of the statistical tools, including the Cumulative Distribution Function (CDF), to learn the nature of our data in knowledge-bases and to derive from that nature, an idea of what a negative or positive thought should be.

Implementation of such a method relies on a few ideas which are mostly inspired by the theory of relativity. Although the theory of relativity is a theory in physics, it affected most of the areas of research with its new philosophy, including both natural and social sciences. When we make any kind of calculation or observation (both quantitative or qualitative), we have to take something as a reference. It can be sun when we calculate planetary motions, earth when we calculate the speed of a car or a living organism such as a human when we try to observe animal behaviour. According to the theory of relativity, there are no universally true reference point, and if you change your reference point on your calculations, the result will be different everytime. So, theory of relativity states that there are no such thing as universal truth and everything is relative to some other entity that we take as a reference.

We have built this method on the idea of relativity by making two important analogies to it. First, it can be said that objectivity is actually not a metric that can be computed on its own, rather it is the lack of any positive or negative sentiment. Any kind of calculation on objectivity should be built relatively to the negative/positive scores. So, computing the objectivity scores without considering the negative or positive scores will not help us on achieving our goals. Now, we know that we should calculate objectivity relative to the negative/positive scores, but calculating negativity and positivity is another issue. Second, and the most important, analogy to relativity helps us particularly here. The notion of negative or positive sentiment comes from the fact that it is more negative/positive than the usual words used by the people in ordinary contexts in daily life. There is no such thing as the universal notion of negativity or positivity. So, if we want to determine the negativity of a word, we should find how different it is (how much more negative) than the average word used in English. In order to accomplish that, we had to find the nature of our data and find what are the properties of an average word.

Using R Studio, we have calculated and learned some of the properties of the data from our knowledge-base SentiWordNet. First step was to extract the score of every single word on SentiWordNet and parse them into the R Studio. Then, we created the histograms and density graphs for each category: positive, negative and objective.

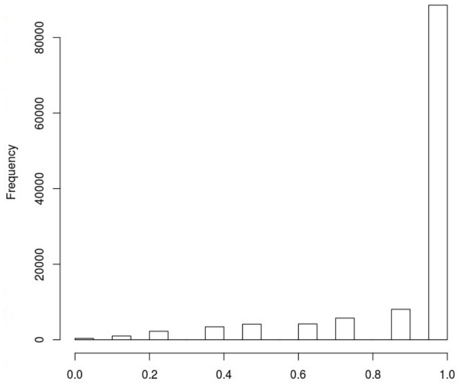

Figure 3.9. Histogram of Objectivity Scores

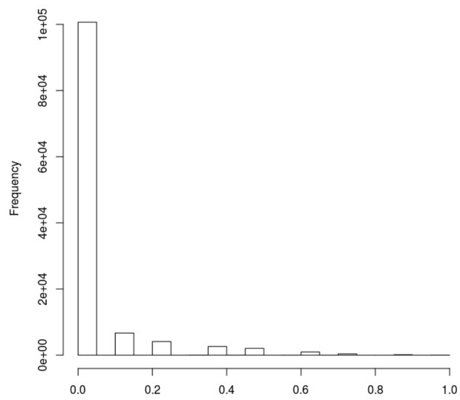

Figure 3.10. Histogram of Positivity Scores

Figures 3.9, 3.10 and 3.11 shows the histograms that belong to each of the scoring types. These histograms shows us why relativistic methods are necessary. The objectivity score of the most of the words are above 0.8, while there are only a small fraction of words that have such high scores for positivity or negativity. This unbalanced histogram suggests that without the use of a relativistic method any kind of calculation will tend to result in classifying the articles as objective more than 90 percent of the time (this is based on the fact that an average objectivity score of a word is 0.9055), even though it is positive or negative.

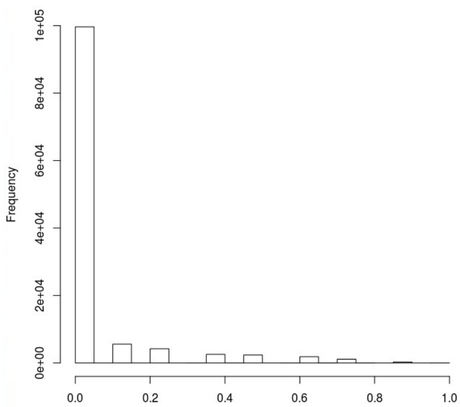

Figure 3.11. Histogram of Negativity Scores

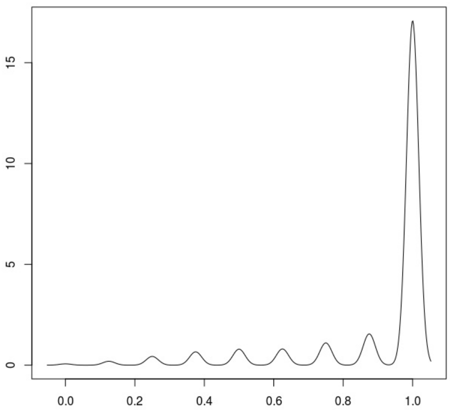

Figure 3.12. Density of Objectivity Scores

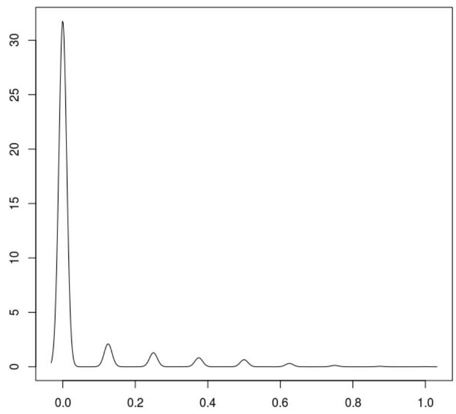

Figure 3.13. Density of Positivity Scores

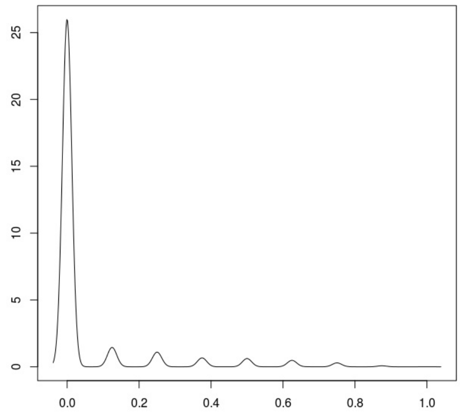

Figure 3.14. Density of Negativity Scores

We can also see the same problem from the density graphs, that we constructed, of each category in Figures 3.12, 3.13 and 3.14. From those density graphs, we have calculated and constructed the CDFs, which will be used as a reference for calculating the negative and positive scores of the words in the text.

These CDF graphs (Figures 3.15, 3.16 and 3.17) was used as a reference when we calculated the individual probabilities of words. For each word in an article, using the word's score, we calculated the probability of being positive/negative as the probability of seeing a word with such a positivity/negativity score in a text by finding the underlying area of the curve in CDF. Table 3.1 shows the sample of example outputs from the CDF graphs for each category.

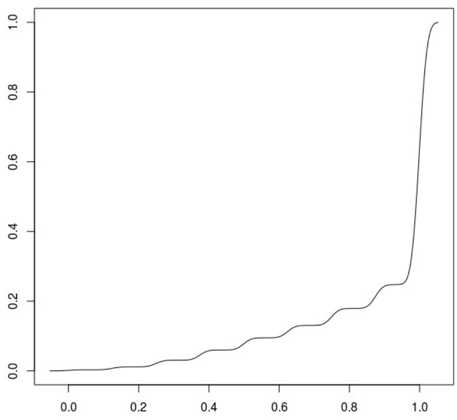

Figure 3.15. CDF of Objectivity Scores

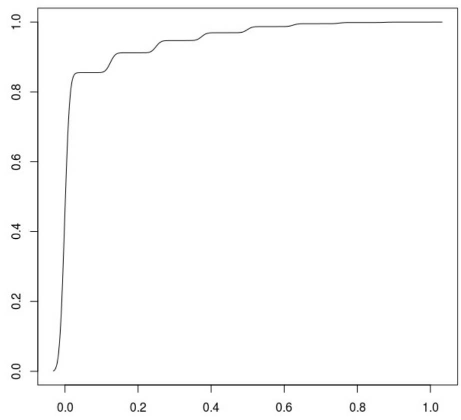

Figure 3.16. CDF of Positivity Scores

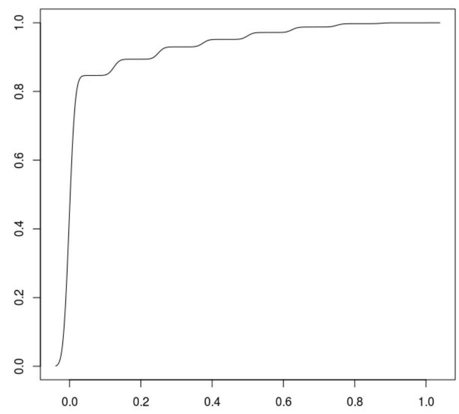

Figure 3.17. CDF of Negativity Scores

Here is an example of how we made the calculations using the 19th sentence from the Reuters article given in the Figure 3.7;

$Almost\_1.0 : 0.0 : 0.0$ $Almost\_1.0 : 0.0 : 0.0$ $human\_1.0 : 0.0 : 0.0$ $human\_1.0 : 0.0 : 0.0$ $have\_0.29 : 0.11 : 0.74$ $have\_0.93 : 0.02 : 0.05$ $unpunished\_0.18 : 0.0 : 0.9$ $unpunished\_0.75 : 0.0 : 0.25$ $gunpoint\_1.0 : 0.0 : 0.0$ $gunpoint\_1.0 : 0.0 : 0.0$ $child\_0.9 : 0.44 : 0.0$ $child\_0.97 : 0.03 : 0.0$ $search\_1.0 : 0.0 : 0.0$ $search\_1.0 : 0.0 : 0.0$ $only\_0.24 : 0.84 : 0.85$ $only\_0.86 : 0.05 : 0.09$ $sentence\_1.0 : 0.0 : 0.0$ $sentence\_1.0 : 0.0 : 0.0$

Table 3.1. Example of Scores from CDF

SentiWordNet Score	Objective prob. (%)	positive prob. (%)	negative prob. (%)
0	0	0	0
0.025	0.01	19.36	11.07
0.05	0.14	82.30	71.09
0.1	0.28	85.53	84.66
0.25	1.26	91.44	89.62
0.75	17.58	99.85	99.67
0.9	24.78	99.98	99.86
0.95	58.02	99.98	99.98
0.975	94.58	99.99	99.99
1	100	100	100

We simply take the average of each word's score given in the first instance of it;

$$p(objective) = \frac{1 + 1 + (0.29) + (0.18) + 1 + (0.9) + 1 + (0.24) + 1}{wordcount} = 0.734 \quad (3.10)$$

where the first two scores correspond to the objectivity score of the word "Almost" and "human" and the third score (0.29) corresponds to the word "have" and so on. Calculating the positivity and negativity probabilities are the same:

$$p(positive) = \frac{0 + 0 + (0.11) + 0 + 0 + (0.44) + 0 + (0.84) + 0}{wordcount} = 0.154 \quad (3.11)$$

$$p(negative) = \frac{0 + 0 + (0.74) + (0.9) + 0 + 0 + 0 + (0.85) + 0}{wordcount} = 0.276 \quad (3.12)$$

```
sentence#       PROBobj PROBpos neg

0       0.6790044235235045      0.0     0.3280264533111917
1       1.0     0.0     0.0
2       0.7271201944324799      0.270433052136337       0.163779771160792З
3       0.6628295037999378      0.32455738317031824     0.16945302044389823
4       0.5631386817495138      0.35701903520083433     0.3792400581517409
5       0.8159462377861313      0.22200369322920824     0.15267960308508777
6       0.6534698613700702      0.19869706105792964     0.30704518276976356
7       0.4756166700766702      0.5891452209874958      0.0
8       0.6683683739554799      0.32895901176906284     0.31056922421699756
9       1.0     0.0     0.0
10      0.7527194562273344      0.27369714767641407     0.23971070919031937
11      0.7094011682115122      0.2716177675325857      0.09817013498090574
12      0.9182035846996918      0.0943032362968984      0.007489093147506574
13      0.7348040620189923      0.07799879838757884     0.3147988222899124
14      0.5589885855698391      0.47792398216585574     0.18175320533458525
15      0.8631488408148801      0.13716858099940324     0.14868863544008218
16      0.4794390122023 0.39138465488627294     0.4862396638798343
17      0.8086114400231139      0.0015661411828512193   0.2052699276770885
18      0.732915400580467       0.1542531005134825      0.27607652079781153
19      0.786147318077721       0.08899661737725506     0.21094610806857866
20      0.8528143548188325      0.0     0.16024456804532267
21      0.8176987291759422      0.0     0.21159547103441256
22      0.9187262968349076      0.11419059992782465     0.07433988265874047
23      0.6825001972362398      0.36721423073638826     0.052913471976539395
```

Figure 3.18. An Example Output of the Relativistic Method

These final results give us the probability of the sentence belonging to those categories. As we can see from the results, although the initial scores from the SentiWordNet are adding up to 1 as total probability, these values do not. This is because the scores for each class, objective, positive and negative, are calculated from their respective CDF graphs individually. So, using this method gives us also the ability to solve the normalization problem mention in the Section 3.4.1. But, as we described above, objective scores on their own do not have any meaning. We will rather focus on the positive and negative scores and find a conclusion about objectivity from them in the next chapter. Figure 3.18 shows the output where all of the probabilities for all the sentences are shown.

CHAPTER 4

EXPERIMENTAL RESULTS

In this Chapter, you will find the test result we get from the methods explained in Chapter 3. In the following sections, we will use these results to prove that our methods are working and to prove our assumption that we made throughout the thesis. Our hypothesis and assumption are listed below:

- *Objectivity(hand-writtenarticles)* $<$ *Objectivity(editorials)* $<$ *Objectivity(NYT)* $<$ *Objectivity(Reuters)*

  As it was explained in the Section 3.1, it was our hypothesis that Reuters have the least sentiment in their news while editorials have the most (except for the articles we wrote for test purposes). So, the results should obey this nature above. Also, since we used the hand-written articles and editorials as ground-truth this inequality will prove that our methods are working properly to tag the sentiments.

- *Objectivity cannot be a metric that can be computed on its own.*

  We will try to show that, although the positive and negative sentiments are successfully found, the objectivity scores are not and will not be meaningful.

4.1. Naive Bayes Method

We applied Naive Bayes method to our sanity test data (hand-written texts and editorial that are manually tagged as positive or negative) to see whether it works or not. The result are shown in the Table 4.1.

Table 4.1. Results of Naive Bayes Method

	Positive Articles	Negative Articles
Sentences tagged as positive	19	4
Sentences tagged as negative	8	18

In positive articles we wrote, Naive Bayes method was able to classify 70 percent of the sentences as positive and in negative articles, we were able classify 80 percent as negative which was accurate enough to say that the method is working. Same thing can be said about the tests on editorials too. Our results were 70 percent in accordance with the manual tags and 66 percent in accordance with the relativistic method (refer to Section 4.2). On the other hand, the objectivity analysis on its own fails here as we have predicted before. This is caused by the multiplication of individual probabilities. Since words that do not carry any sentiment are very frequent in a sentence, objectivity score will be higher than any other sentiment's score no matter how emotional the text is. The reason for the high frequency of non-sentimental words is the presence of nouns such as "book, law, person, decision". Objective and sentimentally meaningless words being more frequent than the words carrying a sentiment makes the comparison between objectivity and, positivity or negativity impossible.

Naive Bayesian Classification is successful in determining the negativity and positivity of sentences, however, in this study we need to apply our theory to comparisons where the Naive Bayes approach fails.

The reason of this failure is the problem of normalization in Naive Bayes. Negativity and positivity scores are dependent on each other and furthermore objectivity score is dependent on both of them, so performing any kind of normalization on the scores is not possible. Inability to perform such a normalization reveals the first problem of the failure of Naive Bayes classifiers. The magnitude of the probability as a result from the Naive Bayes depends on the length of the sentence. Considering that we take the product of each word's individual probability, the smaller number of words there are in the sentence the smaller the probability of that sentence being positive. So, without any normalization the difference of the magnitude of individual probabilities of sentences does not mean anything. Even if we perform a length normalization to the results as mentioned in the Section 3.4.1, we would not be able to compare the results. The reason of this second problem is, even if we apply the length normalization, the result will not have a constant scale, each sentence will have a score in a different scale. In short, the probabilities will not add up to 1, and since they are dependent on each other, they cannot be made to fit in a scale of 0-1 properly.

These problems are caused by the same reason, non-existence of a reference point to compare the measurements. We solved this issue by introducing the relativistic method as explained in the Section 3.4.2 and further explained in detail in the next section.

4.2. Relativistic Method

First thing we did is to understand whether this method is working or not by applying it to the articles we wrote. As it is said before they are extremely opinionated texts, created for the sole purpose of testing. The result are shown in the Table 4.2.

Table 4.2. Results of Relativistic Method -1

	Positive Articles	Negative articles
Positivity Score	0,61781	0,37169
Negativity Score	0,17134	0,56098

The positively written texts have 61,78 percent chance of being positive and the negatively written texts have 56,10 percent chance of being negative. Although these numbers seem to be low, it should be noted that any text will have at least a small number of objective sentences and sentences of opposite sentiments. Also, considering the positive probability, which is 37,17 percent, of the negative texts, there is almost no chance of those articles being objective. Same thing applies to the positive texts too. There were only 7 sentences that are neither tagged as positive nor negative in all the documents. In the tagged editorials, we were able to tag 74 percent of the sentences correctly.

If we consider the objectivity analysis on its own, the results are not pleasing. The objective articles we wrote, have 54 percent chance of being objective which is not a good score considering the articles was purely objective. This result agrees with our assumptions that the objectivity cannot be computed on its own, rather it is the lack of any positive or negative sentiment. Furthermore, the objectivity tests on the editorials, NYT and Reuters articles are nearly equal with the percentages around 60-65. The reason for that is, the high average of objective words, like most of the nouns, that appears in each sentence. Since the frequency of purely objective words are very high, there is more chance of encountering them.This means that the longer the sentence is the more frequent objective words there is to increase the objectivity score to where we can see the ceiling effect.

So our sanity tests on ground-truth data (hand-written articles and editorials) is convincing. On average, our hand-written texts have almost 80-85 percent sentiment, positive and negative combined, and our editorials are tagged nearly as successful as them with 74 percent accuracy, which shows us that our relativistic method is working properly. It leads us to the next part of the thesis which includes the results of the articles from NYT, Reuters and editorials from NYT.

In order to show how our methods work, let's give example sentences from the editorials. Below sentence, which belongs to an article titled "The supreme court saves ObamaCare Again", has one of the highest positive score with 94 percent:

> *This is a remarkable success.*

This sentence, which belongs to an article titled "The Fight for Health Care isn't Over", on the other hand, is one of the most negative sentence in editorials with the score of 71 percent:

> *Yet the Republicans, gripped by an irrational hostility to helping poor, would rather hurt the uninsured and damage their state economies by refusing federal money.*

And this sentence from the same article is one of the most objective sentences with 8.8 percent chance of being sentimental (positive and negative combined):

> *The federal government will be paying 100 percent of the cost of expended benefits this year and next, gradually tapering down to 90 percent in the future.*

As can be seen from the examples, the relativistic method produces very decent scores on the sentence level. However, in some sentences problems may arise. Let's take the next sentence from he same article as an example:

> *But there are myriad ways the current Republican Congress, future Congresses or a future Republican president could **subvert important elements of the law** or **render it inoperative.***

Above sentence has both negative sentiment and positive sentiment. Our method scores 42 percent positive and 36 percent percent negative. Although these scores are acceptable because of the nature of the sentence, the result remains ambiguous, since it is both tagged as positive and negative. If we take a look at this sentence, we can see that, it can be considered as objective even though it has positive and negative thoughts in it. The fact that the author have used those thoughts in a way that every aspect is objectively defined, it can trick the system to produce such a result. However, it can be further argued that the existence of any positive or negative thoughts here means that the author is and

Table 4.3. Results of Relativistic Method - 2

	Editorials	NYT	Reuters
Mean of Positive Scores	0.32783	0.28363	0.22636
Median Positive Scores	0.32787	0.25194	0.23227
IQR of Positive Scores	0.06816	0.12195	0.12154
SD of Positive Scores	0.05323	0.09969	0.10252
Mean of Negative Scores	0.26938	0.20059	0.17295
Median of Negative Scores	0.26262	0.20885	0.17648
IQR of Negative Scores	0.04807	0.10634	0.12238
SD of Negative Scores	0.04395	0.08377	0.09523

will be explaining his/her ideas in some point. Since, in the scope of this thesis, we did not analyze the context of the article, we will leave this issue to future works on this subject.

Table 4.3 shows the statistical values of sentences (mean, median, interquantile range and standard deviation) we get from the entire corpus (Each sentence are taken into account independently without emphasizing which document they belong to). If we take a look at the positive scores, we will see that the score is gradually decreasing from 32.78 percent in editorials to 25.19 percent in NYT and finally 23.22 percent in Reuters, as expected. Same thing applies to negativity too. The probability of an average article to be negative is decreasing from 26.26 percent in editorials to 20.88 percent in NYT and finally to 17.64 percent in Reuters articles. There is almost 25 percent decrease in scores from editorial to NYT which is pretty significant enough to say that our method is working properly.

Table 4.4. Results of Analysis of Variance

	F	p-value
Positive scores	8765.38	2x10^{-16}
Negative Scores	2370.54	2x10^{-16}

An analysis of variance (refer to Table 4.4) shows us that F-score of negativity is 2370.54 and the F-score of positivity is 8765.38, and both of them lead to a p-value of $2x10^{-16}$. So, we can reject any hypothesis, stating that editorials, NYT articles and Reuters are equal in objectivity. Furthermore, we applied Scheffé test to see how those three sources relate to each other. By looking at the Table 4.5, we can safely say that all three sources are fairly different from each other with p-scores lower than 0.05.

Table 4.5. Results of Scheffé Tests

Comparison	F	p-value
Negative(Editorials vs. NYT)	7.99	3.38×10^{-4}
Negative(Editorials vs. Reuters)	15.65	1.59×10^{-7}
Negative(NYT vs. Reuters)	2355.56	0
Positive(Editorials vs. NYT)	2.82	5.98×10^{-2}
Positive(Editorials vs. Reuters)	14.85	3.55×10^{-7}
Positive(NYT vs. Reuters)	8751.78	0

For further proof, we tagged the sentences in articles in each source as either positive or negative in order to see the results are still acceptable or not. By doing this, we tried to extract the sentimental paragraphs written in the objective texts. Even there are extremely sentimental sentences in an article if we just take the average of all the sentences we might loose the sentimental parts, since the objective sentences decrease the average and eliminate the sentiment in that article. We used 0.33 as our threshold, and the reason for that is we have three classes to tag the articles when we do not have any information on it: 33 percent chance they are positive, 33 percent chance they are negative or if there are not any sentiments 33 percent chance they are objective. So, in order to tag a sentence, it has to have a score higher than 33 percent in any of the classes of outcome. The results are shown in the Table 4.6.

Table 4.6. Results of Relativistic Method - 3

	Editorials	NYT	Reuters
Sentence Count	632	2345221	8491252
Positive Sentence Count	301	880094	2314192
Negative Sentence Count	183	529823	1413685
Average Number of Sentences	21.067	28.004	10.568
Average Number of Positives	10.033	10.510	2.881
Average Number of Negatives	6.100	6.326	1.759

Results are in agreement with the previous data. The number of tagged sentences are decreasing from 76.5 percent in editorials to 60.1 percent in NYT and finally to 42.89 percent in Reuters. There are, on average, 4.63 sentence that are tagged in Reuters in each article which has 10.57 sentences on average. But, it should also be noted that a single sentence can both tagged as negative and positive as we have seen in the previous example sentence from the editorials. In general the total number of tagged sentences will be lower because of this reason.

A second outcome of this results is the difference in average number of sentences in NYT and Reuters. On average, New York Times have three times more sentences than Reuters. This is because, as we have said before, Reuters only deliver the information from the incidents that happen all over the world while New York Times authors use those information and add their own thoughts to it before they are published. This fact is also in agreement with our expectation that the Reuters are more objective than the NYT.

Now if we take a look at the question we asked previously in Chapter 1 about the two articles defending different parties on same topic, we can see which one is written in more objective manner. In those articles about the human shields in Israel-Palestine conflict we saw that the NYT were accusing Palestinians of using human-shields while Reuters were accusing Israelis. The results show that the positive and negative scores of Reuters article are 19.75 percent and 18.66 percent respectively, on the other hand NYT article have 28.1 percent positive score and 26.27 percent negative score. In conclusion we can say that both of them were pretty objective (they both have subjectivity level lower or equal then the editorials), but the Reuters article is more credible because of the fact that it involves less emotions (considering the fact that NYT article has a negativity score in the same level as editorials and Reuters article has a negativity score lower than the average of NYT articles), thus making it more believable.

CHAPTER 5

CONCLUSIONS

As the influence of internet grows the amount of news sources and news that becomes available to us are increasing. Thus, the important question, whether those sources are credible or not, is growing. In this study we aimed to find a solution to this problem by proposing two methods. In these methods, using the POS tagger and knowledge-base, we evaluated four different type of articles to test and prove the effectiveness.

We started by testing the methods on already-tagged articles to understand whether they are able to extract the sentiments or not, followed by a broad tests applied on the news articles from New York Times, Reuters and some editorials from New York Times.

We were able to prove our hypothesis that states the objectivity should be lowest in editorials and highest in Reuters articles, since the Reuters is a news agency and thus expected to be more objective. Also, we were able to overcome the problem of normalization in the sentiment analysis, that comes from the dependant nature of the scores, by introducing the relativistic approach.

In addition, as a by product of this thesis, we observed that the sentiments are mostly dictated by the use of adjectives in text. Although other type of POS' are affective too, the main contributors were found to be adjectives. Considering that the adjectives are used to describe everything that surrounds us, including our ideas on them, it is no surprise that they are found to be the cornerstones of sentiments. This subject will be argued further in the next section along with the other possible research directions and possible enhancements that can be made on this research.

5.1. Future Work

In this thesis, we proposed two methods which use syntax of the text to reach conclusions on subjectivity. Although the results are good and convincing, any future attempt at this subject should take the context in which the article and the words are written into consideration. It is expected to perform much better since the objectivity and sentiment of a word can change drastically depending on the context and the fact that whether the subject is highly controversial as Mejova et al. (2014) suggest or not. A research could further benefit from extracting the context by letting the researcher to use aspect-based sentiment analysis which works better than the other methods as proven by the Cambria et al. (2014). Furthermore, finding the context and the subject of the article (finance, sports, politics etc.) is important for any kind of analysis because it is acceptable to assume that in different area of news, emotions are expressed in different ways. Finally, with the context in hand, instead of taking average score of every instance of a word from SentiWordNet, like we did in the scope of this thesis, one could find the specific instance of the word and take that value as a score.

Examining the text without the search for context has a problem for our method too. In Naive Bayes approach we rightfully assume that the probabilities of the words in each category is independent from each other, but in some contexts sentiment score of a word might depend on the word used before (same thing applies for named entities). Although there is very small chance that this might happen, it still can lead to an improvement in the method's effectiveness of a small fraction.

Another issue to be addressed in any future work is the quotes in the text. Since the quotes are the words that does not belong to the author, they should be left out from the sentiment analysis. Giving a thought of another person should not be counted towards sentiments of the author. On the other hand, it can further be argued that if the author usually gives the quotes from the people with negative thoughts on the subject without any positive quotes, it can be counted as a bias towards negative opinion too. In conclusion, quotes should be evaluated separately from the text.

In this thesis probability of every word, no matter what POS they are, is treated as they have the same effect in text. In fact, some POS types can be more effective when it comes to sentiments. It can be argued that the adjectives carry more emotions than the verbs in general whereas the nouns have almost no sentiment. Even a simple, heuristic weighting between the POS should increase the efficiency of the method in the future

Finally, because of the reason we use SentiWordNet as our knowledge-base, we are bound to any inefficiencies coming from it. Any increase in the efficiency in SentiWordNet would directly lead to an increase in the efficiency of our methods.

REFERENCES

Baccianella, S., A. Esuli, and F. Sebastiani (2010, March). Sentiwordnet 3.0: An enhanced lexical reesource for sentiment analysis and opinion mining. In *Proceedings of the 2010 ACM Symposium on Applied Computing (SAC)*, pp. 1748–1754.

Balahur, A., R. Steinberger, M. Kabadjov, V. Zavarela, E. van der Goot, M. Halkia, B. Pouliquen, and J. Belyaeva (2010, May). Sentiment analysis in the news. In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC'2010)*, pp. 2216–2220.

Barnard, A. and J. Rudoren (2014). Israel says that hamas uses civilian shields, reviving debate. http://www.nytimes.com/2014/07/24/world/middleeast/israel-says-hamas-is-using-civilians-as-shields-in-gaza.html. Accessed: 2015-03-18.

Cambria, E., A. Livingstone, and A. Hussain (2011, February). The hourglass of emotions. In *Proceedings of the 2011 International Conference on Cognitive Behavioural Systems (COST'11)*, pp. 144–157.

Cambria, E., D. Olsher, and D. Rajagopal (2014, July). Senticnet 3: A common and common-sense knowledge base for cognitive-driven sentiment analysis. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, pp. 1515–1521.

"Credibility" (2013). *The American Heritage Dictionary of the English Language, 5th edition*. Houghton Mifflin Harcourt Publishing Company.

"Credibility" (2015). Merriam-webster.com. http://www.merriam-webster.com/dictionary/credibility. Accessed: 2015-02-12.

Gaziano, C. and K. McGrath (1986). Measuring the concept of credibility. *Journalism Quarterly 63*(3), 451–462.

Karamibekr, M. and A. A. Ghorbani (2012, December). Sentiment analysis of social issues. In *Proceedings of the 2012 ASE International Conference on Social Informatics*

(SocialInformatics 2012), pp. 215–221.

Lim, K. W. and W. Buntine (2014, November). Twitter opinion topic model: Extracting product opinions from tweets by leveraging hashtags and sentiment lexicon. In *Proceedings of the 23rd ACM International Conference on Information and Knowledge Management (CIKM)*, pp. 1319–1328.

Lin, W.-H., T. Wilson, J. Wiebe, and A. Hauptmann (2006, June). Which side are you on? identifying perspectives at the document and sentence levels. In *Proceedings of the 10th Conference on Computational Natural Language Learning (CoNLL-X '06)*, pp. 109–116.

Liu, B. (2012, May). *Sentiment Analysis and opinion Mining*. Morgan Claypool Publishers.

Marcus, M., G. Kim, M. A. Marcinkiewicz, R. MacIntyre, A. Bies, M. Ferguson, K. Katz, and B. Schasberger (1994). The penn treebank: Annotating predicate argument structure. In *Proceedings of the Workshop on Human Language Technology*, HLT '94, Stroudsburg, PA, USA, pp. 114–119. Association for Computational Linguistics.

Mejova, Y., A. X. Zhang, N. Diakopoulos, and C. Castillo (2014, October). Controversy and sentiment in online news. *Proceedings of Symposium on Computation + Journalism, poster track*.

Miller, G. A. (1995, November). Wordnet: A lexical database for english. *Communications of the ACM 38*(11), 39–41.

Nebehay, S. (2013). Palestinian children tortured, used as shields by israel: U.n. http://www.reuters.com/article/2013/06/20/us-palestinian-israel-children-idUSBRE95J0FR20130620. Accessed: 2015-03-18.

Pang, B. and L. Lee (2008, January). Opinion mining and sentimental analysis. *Foundations and Trends in Information Retrieval 2*(1-2), 1–135.

Pang, B., L. Lee, and S. Vaithyanathan (2002, July). Thumbs up? sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*, pp. 79–86.

Rose, T., M. Stevenson, and M. Whitehead (2002, May). The reuters corpus volume 1 - from yesterday's news to tomorrow's language resources. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC'2002)*, pp. 827–832.

Sandhaus, E. (2008). The new york times annotated corpus overview. https://catalog.ldc.upenn.edu/docs/LDC2008T19/new_york_times_annotated_corpus.pdf. Accessed: 2014-09-16.

Strapparava, C. and R. Mihalcea (2008, March). Learning to identify emotions in text. In *Proceedings of the 23rd Annual ACM Symposium on Applied Computing(APPLIED COMPUTING 2008)*, pp. 1556–1560.

Takala, P., P. Malo, A. Sinha, and O. Ahlgren (2014, May). Gold-standard for topic-specific sentiment analysis of economic texts. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'2014)*, pp. 2152–2157.

Toutanova, K., D. Klein, C. Manning, and Y. Singer (2003, May). Feature-rich part-of-speech tagging with a cyclic dependency network. In *proceedings of HLT-NAACL 2003*, pp. 252–259.

Yu, H. and V. Hatzivassiloglou (2003, July). Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing (EMNLP '03)*, pp. 129–136.

APPENDIX A

PENN TREEBANK TAG SET FOR WORDS

number	tag	description
1	CC	Coordinating Conjunction
2	CD	Cardinal Number
3	DT	Determiner
4	EX	Existantial “There”
5	FW	Foreign Word
6	IN	Preposition or Subordinating Conjunction
7	JJ	Adjective
8	JJR	Adjective, Comparative
9	JJS	Adjective, Superlative
10	LS	List Item Marker
11	MD	Modal
12	NN	Noun, Singular or Mass
13	NNS	Noun, Plural
14	NNP	Proper Noun, Singular
15	NNPS	Proper Noun, Plural
16	PDT	Predeterminer
17	POS	Possessive Ending
18	PRP	Personal Pronoun
19	PRP	Possessive Pronoun
20	RB	Adverb
21	RBR	Adverb, Comparative
22	RBS	Adverb, Superlative
23	RP	Particle
24	SYM	Symbol
25	TO	“to”
26	UH	Interjection
27	VB	Verb, Base Form
28	VBD	Verb, Past Tense
29	VBG	Verb, Gerund or Present Participle
30	VBN	Verb, Past Participle
31	VBP	Verb, Non-3rd Person Singular Present
32	VBZ	Verb, 3rd Person Singular Present
33	WDT	Wh-Determiner
34	WP	Wh-Pronoun
35	WP	Possessive Wh-Pronoun
36	WRB	Wh-Adverb

APPENDIX B

PENN TREEBANK TAG SET FOR PUNCTUATIONS

number	tag	description
37	#	Number Sign
38	$	Dollar Sign
39	“	Left Open Double Quote
40	”	Right Close Double Quote
41	,	Comma
42	.	Sentence Ending Punctuation(.!?)
43	:	Colon, Semi-Colon
44	-LCB-	Left Curly Brackets
45	-RCB-	Right Curly Brackets
46	-LRB-	Left Round Brackets
47	-RRB-	right Round Brackets
48	-LSB-	Left Square Brackets
49	-RSB-	Right Square Brackets