FEATURE BASED SENTIMENT ANALYSIS ON INFORMAL TURKISH TEXTS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BATUHAN KAMA

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
COMPUTER ENGINEERING

SEPTEMBER 2016

Approval of the thesis:

FEATURE BASED SENTIMENT ANALYSIS ON INFORMAL TURKISH TEXTS

submitted by **BATUHAN KAMA** in partial fulfillment of the requirements for the degree of **Master of Science in Computer Engineering Department, Middle East Technical University** by,

Prof. Dr. Gülbin Dural Ünver
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. Adnan Yazıcı
Head of Department, **Computer Engineering** _____

Assoc. Prof. Dr. Pınar Karagöz
Supervisor, **Computer Engineering Department** _____

Prof. Dr. İsmail Hakkı Toroslu
Co-supervisor, **Computer Engineering Department** _____

Examining Committee Members:

Prof. Dr. Ahmet Coşar
Computer Engineering Department, METU _____

Assoc. Prof. Dr. Pınar Karagöz
Computer Engineering Department, METU _____

Prof. Dr. İsmail Hakkı Toroslu
Computer Engineering Department, METU _____

Assoc. Prof. Dr. İsmail Sengör Altıngövde
Computer Engineering Department, METU _____

Assoc. Prof. Dr. Suat Özdemir
Computer Engineering Department, Gazi University _____

Date: Sept 01, 2016

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: BATUHAN KAMA

Signature :

ABSTRACT

FEATURE BASED SENTIMENT ANALYSIS ON INFORMAL TURKISH TEXTS

Kama, Batuhan
M.S., Department of Computer Engineering
Supervisor : Assoc. Prof. Dr. Pınar Karagöz
Co-Supervisor : Prof. Dr. İsmail Hakkı Toroslu

September 2016, 79 pages

Sentiment analysis (SA); in other words, opining mining, is the automatic extraction process of a author' s feeling on a specific topic. These feelings can be positive, negative or neutral.

However, most of the time authors do not carry the same opinion for all parts of a topic. While they have a positive attitude on some parts of a topic, they can criticise other parts. Therefore, feature or aspect based sentiment analysis (FBSA) a specialized version of sentiment analysis is used to analyse people' s attitude on each specific feature of a topic.

Nowadays, there are a lot of forums, blogs and review sites where customers and professional reviewers add their comments on different products. Since thousands of new entries are written to these sites each day, it is almost impossible to read all comments on a product and its features and learn powerful and weak sides of this product; and hence, the need for classifying these informal online

data automatically has increased and FBSA can be used to fulfill this need.

Up to date, most of the studies on both sentiment analysis and feature based sentiment analysis was on English. There are only several works on these topic on Turkish. In this thesis, a dataset created by collecting comments and reviews from forums is processed with existing feature based sentiment analysis methods for English. Moreover, new methods for Turkish feature based sentiment analysis is proposed.

Keywords: Sentiment Analysis, Feature Based Sentiment Analysis, Text Mining, Unsupervised Learning, Natural Language Processing, Turkish

ÖZ

KONUŞMA DİLİNDE YAZILMIŞ TÜRKÇE METİNLER ÜZERİNDE ÖZELLİK TABANLI DUYGU ANALİZİ

Kama, Batuhan
Yüksek Lisans, Bilgisayar Mühendisliği Bölümü
Tez Yöneticisi : Doç. Dr. Pınar Karagöz
Ortak Tez Yöneticisi : Prof. Dr. İsmail Hakkı Toroslu

Eylül 2016 , 79 sayfa

Duygu çözümleme, bir başka ifadeyle düşünce madenciliği, yazarın belirli bir konu üzerindeki düşüncelerini otomatik olarak çıkarma sürecidir. Bu düşünceler olumlu, olumsuz veya nötr olabilir.

Ancak çoğu zaman yazarlar bir konunun bütün parçaları hakkında aynı düşünceye sahip olmazlar. Konunun bir kısmı üzerinde olumlu bir düşünceye sahiplerken, diğer kısımlarını eleştirebilirler. Bu yüzden, kişilerin bir konunun her belirli özelliği hakkında tutumlarını analiz etmek için duygu çözümlemenin özelleşmiş bir türü olan özellik tabanlı duygu çözümleme kullanılır.

Günümüzde, müşterilerin ve profesyonel eleştirmenlerin değişik ürünleri hakkında yorumlarını ekleyebilecekleri birçok forum, blog ve eleştiri sitesi bulunur. Bu sitelere her gün binlerce yeni giriş yapıldığından bir ürünün güçlü ve güçsüz yanlarını öğrenmek için bütün yorumları okumak neredeyse imkansızdır, bu yüzden

bu çevrimiçi verilerin otomatik olarak sınıflandırılmasına olan ihtiyaç artmıştır ve özellik tabanlı duygu çözümleme bu ihtiyacı gidermek için kullanılabilir.

Bu güne kadar duygu çözümleme ve özellik tabanlı duygu çözümleme üzerine yapılmış olan çoğu çalışma İngilizce üzerinedir. Bu konular hakkında sadece birkaç tane Türkçe çalışma vardır. Bu tez kapsamında, forumlardaki yorum ve incelemelerin toplanmasıyla oluşturulan veri setine daha önceden İngilizce için geliştirilmiş işlemler uygulanmıştır. Ek olarak, Türkçe için yeni özellik tabanlı duygu çözümleme yöntemleri önerilmiştir.

Anahtar Kelimeler: Duygu Çözümleme, Özellik Tabanlı Duygu Çözümleme, Metin Madenciliği, Gözetimsiz Öğrenme, Doğal Dil İşleme, Türkçe

To my dearest wife.

ACKNOWLEDGMENTS

I would like to thank my supervisor Assoc. Prof. Dr. Pınar Karagöz for her guidance for the last two years.

I would like to thank my co-supervisor Prof. Dr. İsmail Hakkı Toroslu for his guidance for the last two years.

I would like to thank Murat Öztürk, for their help and ideas.

I would like to thank to my company MilSOFT Software Technologies for allowing me to continue my education and conduct my research while I am working there.

I would like to acknowledge and thank to Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TUBİTAK) for their 2210- Yurt İçi Yüksek Lisans Burs Programı which provided me financial support by their scholarship during my two tears long Master of Science education.

Finally, I would like to thank my family for their support.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

CRF	Conditional random field
EM	Expectation maximization
FBFE	Frequency based feature extraction
FBFE-SWS	Frequency based feature extraction with sentiment word support
FBSA	Feature based sentiment analysis
ME	Maximum entropy
NB	Naive bayes
NLP	Natural language processing
PMI	Pointwise mutual information
POS	Part-of-speech
SA	Sentiment analysis
SVM	Support vector machine
WSBFE	Web search based feature extraction

CHAPTER 1

INTRODUCTION

1.1 Motivation & Problem Definition

Obtaining the best possible solution is a part of human nature. People always want to get the best possible solution considering their resources. This situation can be also observed while they are shopping. They always want to buy the best products they can afford. To decide which product is *the best* for them, they usually investigate a lot of products.

For centuries people had conducted their research in a face-to-face way. In other words, they asked their friends, colleagues or other users about the product they were planing to buy and decided to whether they would buy or would not buy the product based on the others' answers. During the last decade this fashion has changed. Nowadays, before buying a product 81% of shoppers research online and 75% of them look over online reviews on the product they want to buy according to [52]. Another report written by Google [20] shows that websites on which product comments can be posted and read get very high ratings since the number of customers who research online - purchase offline is almost equal to the number of customers research and buy online. The other important point is that people read about different features of products. For example, research [52] shows that 66% of online researchers are interested in warranty information and 51% of them spend time on reading model's information.

Due to widespread use of the Internet all over the world and increase in the number of laptop computers, tablets and smart phones, terabytes of new data are

posted on the Internet by users each day. For example, the article in [21] shows that 500 million Tweets and 4.3 billion Facebook messages are shared each day. These high numbers are also valid for the websites where product comments are posted. Thousands of new positive, negative and neutral comments are added to these sites each day and therefore it is impossible to read all of these texts by just Googling about a product and reading page by page. Therefore, there is a need for tools which automatically find these entries and analyze and summarize them.

When these tools are run for a specific product, firstly, they have to crawl web and gather related reviews and comments about this product. Secondly, they must extract explicit and implicit meanings of each feature of product from the dataset. Finally, they should give a user readable summary of this data. In this thesis, we mainly worked on the second step which that tools should perform and developed techniques for feature based sentiment analysis on informal product comments in Turkish.

1.2 Sentiment Analysis

In [40], sentiment analysis is defined as a study area which tries to resolve people's thoughts, attitudes and feelings on different brands, individuals and groupings. The review in [50] summarizes that main application areas of opinion mining are usually based on economics and commerce. These areas are market and FOREX rate prediction, box office prediction, business analytic, recommender systems and marketing intelligence.

Although there are different levels of sentiment analysis, the most common levels are document level sentiment analysis, sentence level sentiment analysis and entity-aspect level sentiment analysis [40]. The studies [48] and [58] are based on document level sentiment analysis. In these studies, the sentiment orientation of whole document is calculated. On the other hand, although a document can be classified as positive or negative, each sentence in a document can carry different orientations; therefore, sentence level sentiment analysis is used for

finding orientation of each sentence in a document [37]. Despite the fact that document level and sentence level sentiment analysis gives an overall orientation, they do not explain what exactly people like or dislike [40]. Therefore, feature-based sentiment analysis is introduced.

Feature-based sentiment analysis inspects people' s attitude over each specific aspect of entities. The main steps of feature-based sentiment analysis are feature extraction and opinion analysis where feature extraction step is run to find different aspects of the topic mentioned in document and opinion analysis matches extracted sentiments to features to find overall sentiment over each feature [15] [61]. Moreover, if dataset is not available, web crawling is used to collect data [31] and data preprocessing is used when data are not ready to be analysed directly [15].

1.3 Contributions

Sentiment analysis for English texts has been a research area for more than a decade [48] [58] [37] [12] [51] [71]. In addition, there are several studies for Turkish released in the last five years [16] [62] [30]. On the other hand, although feature based sentiment analysis has been also popular for English, as far as I know, only a few studies has been made in Turkish [1] [60]. In this study, I developed methods to find this gap by applying methods developed for English to Turkish texts and proposing new techniques.

In detail, I aimed to create basics for feature based sentiment analysis in Turkish texts. To achieve my goal,

- a new dataset was created
- methods already developed for English are applied to this dataset
- new methods for Turkish by considering linguistic properties of Turkish are proposed.

1.4 Organization of the Thesis

The organization of this thesis is as follows:

Chapter 2- Literature Survey contains the summary of previous studies contacted on sentiment analysis and feature-based sentiment analysis.

Chapter 3- Background explains background information on the research fields and methods used in this thesis in detail.

Chapter 4- Methods is the part where the detailed information on the collected dataset and the methods used are explained.

Chapter 5- Results and Discussions reveals the results of the experiments and discusses these results.

Chapter 6- Conclusion and Feature Works gives a summary and concludes thesis. In addition, future work ideas to enhance studies conducted in this thesis are given in this chapter.

CHAPTER 2

LITERATURE SURVEY

Sentiment analysis; in other words, opinion mining is a study area focuses on extracting orientation of people' s thoughts, sentiments and attitudes towards goods, services, events and other topics [40]. Although there are several earlier publications on the basics of sentiment analysis such as extracting semantic orientation of adjectives [23] and subjectivity classifications [67], main research on opinion mining have been conducted since early 2000s.

In the scope of this chapter, firstly, sentiment analysis is introduced by explaining

- applications of sentiment analysis and
- opinion and opinion types.

Secondly, the studies on the following main sentiment analysis methods are summarized:

- document-level sentiment analysis
- sentence-level sentiment analysis
- aspect-level sentiment analysis

Finally, the literature on the segments of sentiment analysis which are helper to main methods is investigated by giving details of

- sentiment lexicon generation and
- cross domain sentiment analysis.

2.1 Applications of Sentiment Analysis

One of the main aims of sentiment analysis is facilitating users' information search on a specific topic because people usually consider others' opinions before making a decision. Individuals use other users' opinions and experiences before purchasing a product, take precaution on stock market depending on experts' opinions and are affected from people with same political view with them before an election. On the other hand, companies and organizations analyze their users' and customers' opinions to find their better and worse sides of them than their rivals.

Today, with the increase in the number of people using social media including micro-blog sites, forums and blogs, both individuals and companies have been conducting their research online. The information that the individuals access is not limited to their families' and friends' experiences anymore. Instead, they can read much more comments on each aspect of a specific topic and they can make a better decision. Moreover, businesses and organizations can conduct online surveys and they can reach higher number of users with lower workload.

Nevertheless, although online research eases both customers' and companies' reaching data, because of the high volume of data posted to internet every day, reaching and processing all the data on the internet almost impossible. Therefore, there is a need for applications which collect, process and analyse online data. Sentiment analysis is used in this applications.

Up to today, many studies have been conducted on applications of sentiment analysis. For instance, the work in [17] presented methods to predict movie box-office revenue by using box-office and micro-blog data. Sales performance were predicted with a sentiment analysis method in [73]. Another application is market and FOREX rate prediction. [33] focused on news to predict intra-day stock prices and Twitter feeds were analysed daily to match mood on Twitter to behaviour of stocks [10]. [57] and [64] proposed a application of sentiment analysis to predict results of elections by using Tweets. In [53], authors analyzed the online book reviews in terms of social influence. [45] was a study on detecting

the emotions in novels and fairy tails. An interesting study [13] tried to detect sentiments and feelings of texts on suicide notes. A sentiment analysis based application was developed for business intelligence which is used by managers of business to see opportunities, threats and competitors and take necessary actions in the study [34] by using micro-blog entries. There have been some applications of sentiment analysis on sports, too. For example, methods for American National Football League betting market were developed by analysing data from blog, Twitter and public news [25]. In addition to these researches, many articles which also define application of sentiment analysis on different topics can be found. It shows that applications of sentiment analysis have been used in almost all areas of life for many years.

2.2 What is Opinion?

Opinion, as stated in [44], is "*a belief, judgement, or way of thinking about something : what someone thinks about a particular thing*". Therefore, an opinion can be defined with two key components: a target g and sentiment on this target s, (g,s) [40]. The target g may also be named as topic. Moreover, sentiment s can take different values which express polarity explicitly, i.e, s can be *positive, negative, neutral* or *numeric value like 1 to 5 stars.*

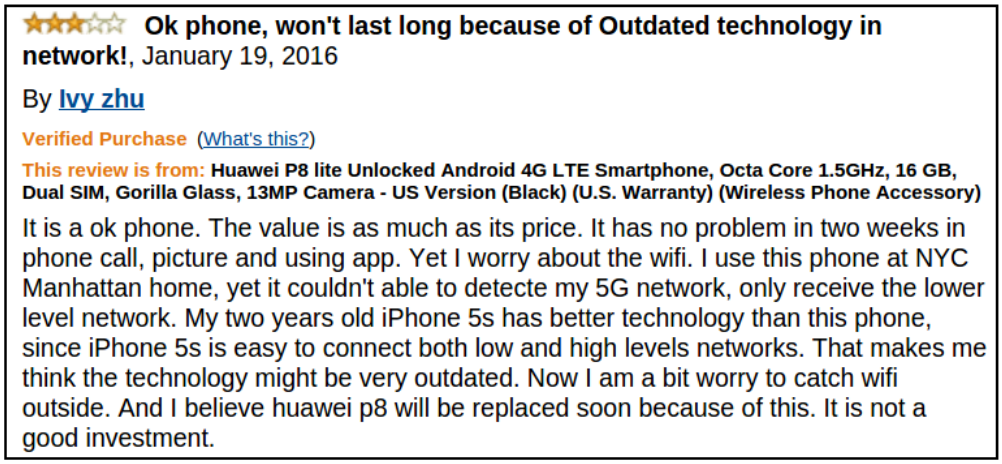

Figure 2.1: Review on Huawei P8 retrieved from Amazon.com

Figure 2.1 is a capture of a customer' s review on Huawei P8 mobile phone posted on Amazon.com. This review contains both positive and negative thoughts on

this product. However, overall sentiment over the target is neutral since user not only wrote *it is OK phone* but also gave 3/5-stars to target. Hence, the *(g, s)* 2-tuple for this review is (Huawei P8, Neutral).

Consumers' feelings over a topic can change in time. They may hate a product after they use it for a time since different problems may occur. On the other hand, users can get used to product and their attitude may change in a positive manner. Therefore, keeping opinion-holder; i.e., writer and time information is useful to detect changes on users' thoughts over time. Hence, opinion can be hold in a quadruple, *(g, s, h, t)* where h is used for opinion holder and t stored time information [40]. The opinion in Figure 2.1 can be defined as (Huawei P8, Neutral, Ivy zhu, January 19, 2016) by a quadruple.

Nevertheless, if a review or comment expresses different feelings for different aspects of topic, quadruple is not enough. In Figure 2.1, although writer expresses a neutral attitude over the mobile device, he also stated that phone calls do not create any problem, but with Wi-Fi quality is problematic. Hence, if only overall sentiment for a topic is considered, information loss occurs. To prevent this, definition of opinion is enhanced to quintuple, $(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$ where e_i stands for entity i, a_{ij} defines aspect j of entity i, h_k is opinion holder k, t_l states sentiment was written at time l and s_{ijkl} portraits sentiment written by k at time l for aspect j of entity i in [40]. For instance, the sentiment over Wi-Fi in Figure 2.1 can be formalized as (Huawei P8, Wi-Fi, Negative, Ivy zhu, January 19, 2016) and sentiment over phone calls can be written as (Huawei P8, Phone calls, Positive, Ivy zhu, January 19, 2016).

By using quintuple opinion formalization, a sentiment analysis over a document can be defined as a 5-steps process, which are entity extraction, feature extraction, opinion holder extraction, time extraction, aspect sentiment classification.

2.2.1 Regular versus Comparative Opinions

[38] defined regular opinion as opinion in the literature and proposes two types of opinion as direct and indirect. Direct opinion is defined as opinion whose target

is an explicit entity or aspect of entity. For example, *The screen resolution is more than enough.* On the other hand, indirect opinion was defined as opinion which affects an entity or aspect indirectly, e.g, *I have to spend time to charge my phone everyday* states that consumer has to recharge her/his phone every day and the negative thought over battery life was implicitly expressed.

Again, in [38] comparative opinion was explained as naming similar and different opinions over more than one entity or shared aspects of more than one entity. *The camera of Huawei P8 is better than iPhone 5S' s* is an example of comparative opinion.

2.2.2 Explicit versus Implicit Opinions

All of the subjective statements which states regular and comparative opinions are explicit opinions, e.g., Nutella tastes great. On the other hand, an objective statement gives an implicit opinion [74], usually a desirable or undesirable fact. For example, *The pictures taken by Nokia phones are more colorful than pictures taken by Samsung phones.* is a sentence with implicit opinion.

2.3 Document-Level Sentiment Analysis

Document-level sentiment analysis aims to classify the overall sentiment in a document as positive, negative or neutral. In other words, quintuple for document-level sentiment analysis process can be defined as *(e, GENERAL, s, h, t).* In [39], Liu assumed that if the sentiment in a document would be analysed, then this sentiment must be written by a single person on a specific entity. Most of the studies in the literature have been focused on either supervised learning or unsupervised learning methods. From now on, some studies on both methods are explained.

2.3.1 Supervised Learning Based Sentiment Analysis

One of the first known and popular studies on supervised learning was conducted by Pang, Lee and Vaithyanathan [48]. In their study, they used three different machine learning methods to classify documents: Naive Bayes (NB), maximum entropy (ME) and support vector machines (SVM). In their experiments, they created their dataset by collecting 700 positive and 700 negative movie reviews. They selected features of supervised learning methods in 8 different way and ran three supervised learning methods with each of these eight feature selection methods over 1400 reviews. In overall, SVM was the most successful method where NB and ME gives similar results. The highest performance in all 24 experiment was achieved by SVM with %82.9 when feature selection method was uni-grams (presence).

Liu stated that [40] the most important part for supervised sentiment classification was selecting effective features like in all other applications of supervised methods. He, also, summarized some of features could be used. First feature which is terms and their frequency described the single words (uni-grams) and multiple consecutive words (n-grams) with their frequencies. In addition, TF-IDF weighting schema could be used as feature, too. The other feature was part of speech. Since different part-of-speech tags could have different importance with respect to speech, this feature was applicable. For instance, in sentiment analysis, most of the studies have counted adjectives as the most important words since they explain sentiment. In addition, sentiment words and phrases were used as features highly. The sentiment words and word phrases are the words and words groups whose duty is to express orientation of sentiment in documents. They have been usually adjectives and adverbs. Nevertheless, some nouns could express sentiment, too. Rules of opinions were the features which were not sentiment words or phrases but gave opinions. The final feature explained was sentiment shifters which have been expressions changing orientation of a sentiment. They were important since one sentiment shifter word could change whole document' s sentiment orientation.

In [12] authors developed their own machine learning methods called as *score*

function. The study [51] have contained methods to enhance success of document classification by using emoticons. Again, Naive Bayes method was used in [32] to classify Chinese reviews. They collected 16000 reviews and trained with different feature representations. Nakagawa et al. used CRFs with hidden variables in dependency-tree based sentiment classification [46]. Bespalov et al. classified sentiments based on supervised latent n-gram analysis [6]. The study [70] proposed a multi-level model for document sentiment classification.

2.3.2 Unsupervised Learning Based Sentiment Analysis

In unsupervised learning based sentiment analysis, the first best known study was conducted by Turney in 2002 [58]. In this study, the average sentiment orientation of the phrases with adjectives and adverbs were used for classification of reviews. As a first step, phrases of two words where one of the words was an adjective or adverb and the second word which gave context were extracted. Rule-based approach for extraction process was used. Turney developed five different rules and word phrases obeying in one of these rules were extracted. An example rule was that RB, RBR, or RBS, JJ, not NN nor NNS which stated that when first word was one of adverb, comparative adverb or superlative adverb and second word was adjective and third word was different than singular or plural noun, the first two words should be extracted. Note that, rules could contain third words but only the first two words were used.

In the second step, he used PMI algorithm to find semantic orientation of these extracted phrases. The Pointwise Mutual Information (PMI) algorithm which is explained in Chapter 3.

In the final step, the average semantic orientations for each review was calculated. To calculate a review' s average sentiment orientation, each extracted phrase of a review was put into SO(phrase) formula and the average of calculated values was taken. Then, reviews with positive average semantic orientation value were marked as positive and vice versa. This method was run over 410 different reviews from 4 different domains and approximately 3 out of 4 review were classified correctly.

Another research [55] conducted by Taboda et al. In this study lexicon-based approach was preferred. The lexicon used to find direction of opinions in documents was consisted of sentiment words/phrases and their semantic scores. In addition, amplification and negation rules were considered to enhance results of operation.

2.4 Sentence-Level Sentiment Analysis

Sentence level sentiment classification focuses on extraction of orientation of sentiment in a sentence if exists. The sentence-level sentiment analysis can not be formalized as quintuple *(e, a, s, h,t)* since it can not find writer' s opinion over a entity or an aspect of entity because writer can express his/her feelings over a entity or an aspect in multiple sentence. While first two sentences have been about negative faces of a entity, the remaining sentence may turn thoughts on this entity by 180 degrees [40]. Therefore, sentence-level sentiment analysis is a intermediate state which helps document-level and aspect-based sentiment analysis procedures. Hence, resolving problems on this topic, helps us to enhance our methods on document-level and aspect-based sentiment analysis.

The sentence-level sentiment analysis can be solved in two steps: subjectivity classification and sentiment classification where subjectivity classification problem tries to reveal whether a sentence contains subjectivity and sentiment classification problem classifies subjective sentence as positive, negative or neutral.

2.4.1 Subjectivity Classification

The studies on subjectivity classification is older than the studies on sentiment analysis. Although there have been some unsupervised methods, most of studies focused on supervised methods and finding correct features.

Wiebe et al. expressed objective sentence as a sentence with facts and subjective sentence as a sentence with writers' opinions, beliefs and thoughts [67]. They used Naive Bayes for this problem. They defined different binary features with

respect to existence of the followings in a sentence: pronouns, adjectives, adverbs different than not, cardinal numbers and modals excluding will. Moreover, they considered whether the sentence was the first sentence of a paragraph or not.

An unsupervised method developed by Wiebe [66] have used a distributional similarity based word clustering method developed by Lin [36]. To refine features automatically learned polarity of adjectives [23] were added.

Another study [71] used sentence similarity and naive bayes for subjectivity classification . Sentence similarity was used based on the idea that a subjective sentence must be more similar to other subjective sentences than objective sentence. For Naive Bayes process, some of the features they chose were words, bi-grams, tri-grams, POS tags, counts of semantically oriented words and word groups, and average score of semantic orientation of the words.

The study of Wilson et al. [68] put expose that a sentence could be made of both subjective and objective phrase and finding subjectivity of phrases instead of subjectivity of sentence was useful for better sentiment analysis. They labeled subjectivity of phrases in four levels: neutral, low, medium and high where neutral phrases stood for objective phrases. In this study, supervised learning was used with a feature set containing words and phrases indicating subjectivity and syntactic clues extracted from dependency parse tree.

Barbosa and Feng developed a method on subjectivity classification of Tweets [5]. They used supervised learning with traditional features plus Twitter specific features such as re-tweets, presence of hashtags, links, punctuations, emoticons and upper cases. This study have been important since the data on Twitter have been noisy, hence, it was a good baseline for subjectivity classification on noisy sentences.

2.4.2 Sentiment Classification

Sentiment classification shows whether opinion in a subjective sentence is positive or negative. Both supervised and lexicon-based methods have been popular for this process.

In [71], log-likelihood ratio was used to find adjectives', adverbs', nouns' and verbs' orientation. In addition, high number of seed adjectives were used. In order to find out whether a sentiment express a positive or negative thoughts the average of log-likelihood ratio scores for each word in the sentence was used.

Hu and Liu proposed a lexicon-based approach which could be used both in sentiment classification and aspect-level sentiment classification [26]. They used a lexicon which consists of positive and negative words where the score of positive words are +1 and the score of negative word are -1. To calculate sentiment orientation of a sentence, the score of each word was assigned from lexicon and words out of the lexicon were assigned 0. Moreover, negation words such as but and however were considered and the scores of the words affected from these negation words were multiplied by -1. Finally, the scores of all words were summed and sentence was labelled as positive if the sum is greater than 0, and vice versa.

The study [17] used Expectation Maximization (EM), a semi-supervised method, to classify sentences as positive, negative and other (neutral or mixed sentiment). In experiments, a small set of labelled sentences were learned by EM, and a large set of unlabelled sentences were labelled.

Some studies showed that sentence-level and document-level sentiment classification could be worked jointly well. For example, in [42] authors created training dataset by labelling not only sentiment orientation of documents but also each sentences' orientation in the documents and they trained a hierarchical sequence learning model with these dataset. The results showed that accuracy for both levels of classification was increased.

Davidov et. al proposed a method for sentiment classification of Tweets. Their method was very similar to method of Barbosa and Feng mentioned in Subjectivity Classification part [5]. Again, a supervised learning method was chosen and the both traditional and Twitter-specific features were used.

2.4.3 Challenges

There two main unresolved challenges which should be handled during sentence-level sentiment analysis: conditional sentences and sarcastic sentences.

The conditional sentences describe implications and their consequences. Sometimes, in the conditional sentences, the sentiment words do not state the authors thought instead they describe hypothetical situation. For example, *If I can find a reliable phone, I will buy it.* sentence does not say the phone is reliable. Instead, it states writer is looking for a reliable phone, hence, this sentence cannot be labeled as positive. On the other hand, in conditional sentence *If I can sell my lousy car, I will buy a minivan.*, there is a negative sentiment on car. Therefore, detecting whether sentiment words in conditional sentences is an opinion or not is a challenging problem and to my knowledge, there is no study has been conducted on this topic yet.

The sarcastic sentences are sentences in which the writers write opposite of what they mean. For example, consider these two sentence : *What a great laptop! It does not let me even surf on the internet.*. The second sentence has explicit negative opinion on laptop. At the first look, first sentence seems positive; however, there is a sarcasm in this sentence; therefore, writer ; in fact, states laptop is not good. Hence, detecting sarcastic sentences are important to find sentiment orientation of sentences correctly. There have been some initial works on sarcastic sentences. Tsur et al. proposed a semi-supervised learning method [56]. They used a small set of labelled sentence and to enhance this set they used web search. In the study of Gonzalez-Ibanez et. al, they used SVM and logistic regression which were supervised-learning approaches to divide Tweets into group of two as sarcastic Tweets and non-sarcastic Tweets [19].

2.5 Aspect-Level Sentiment Classification

As mentioned before, even if a document has a negative attitude over a entity, it does not mean all of the aspects of this entity are bad or useless and vice versa. Therefore, aspect or feature based sentiment classification which investigates

reviews on aspects of entity rather than on entity, may be desirable in some cases. Recall that $(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$ quintuple is used to formalize aspect-based sentiment analysis.

There are two main steps of aspect-based sentiment analysis: aspect extraction and aspect sentiment classification. From now on, the researches on these two steps are summarized.

2.5.1 Aspect Extraction

The aim of this step is to discover aspects in a given text. This text may be a sentence or a document. One of the main key points used in this step have been that if there is a sentiment, then this sentiment must match a target. In this topic, this target is aspect of a entity or entity' s itself. The aspects can be both explicit or implicit. For example, *The camera of this phone rocks.* sentence defines positive attitude on phone's camera and camera is an explicit aspect. On the other hand, *This phone is so big to put in packet* sentence expresses negative feeling on phone' s size and since it is not given explicitly, size is an implicit aspect. Although, most of the studies focus on explicit aspect extraction, some studies have been conducted on implicit aspect extraction, too. In the remaining of this part, firstly, four main approaches to extract explicit aspects are introduced. Then, a summary of studies on implicit aspect extraction are given.

Finding Frequent Nouns and Noun Phrases

This method was firstly introduced in [26]. It was developed on the idea that, if a noun or noun group was used very frequently in a document about a specific topic, then the chance of its being an important aspect of the topic was very high because people usually wrote about the aspects of a topic mostly used. In this approach, firstly noun and noun groups were extracted by using POS tags and secondly frequency of each noun and noun group was calculated. Finally, the nouns and noun groups exceeds a experimental threshold were marked as

features of the topic.

Popescu and Etzioni in [49] enhanced this algorithm by using PMI operation. They focused on removing frequent but non-aspect noun and noun groups. They used the result of *PMI(a, d)*, where *a* refers to candidate aspect and *d* refers to domain. For example, assume that candidate aspect is *camera* and domain is *phone*. Then, *PMI(camera, phone) = hits(camera of phone) / hits(camera) hits(phone)* is calculated. If PMI score lied under a threshold value, this candidate aspect was removed since it means *a* and *d* did not co-occur together therefore *a* cannot be aspect of domain *d*.

There have been several studies to enhance this method. Scaffidi et al. used generic English corpus to compare occuruence of a noun in this corpus with occurrence of it in the review document [54]. Zhu et al. used Cvalue measure and a bootstrapping technique to find multi-word features [76]. The study of Long et al. used information distance in addition to frequency to find features [41].

2.5.1.1 Using Opinion and Target Relations

Firstly, Hu and Liu introduced this method to extract infrequent nouns [26]. The main idea under this method was that if a word was known to be sentiment word and this word was used to express opinion on an aspect of entity in a document, then same sentiment word could be used to express attitude on other aspects in the same document. Hu and Liu, extracted sentiment words which expressed opinion on frequent features as a first step. Then, they found *nearest* noun and noun groups to these sentiment words and added these noun and noun groups to aspects list.

In [77], a dependency parser was used to enhance Hu and Liu' s method. Using dependency parser instead of *nearest* function to match a sentiment word to noun it modifies, allowed researchers to extract infrequent aspects more correctly. Since dependency parser finds the relations between words, in [69] phrase dependency parser was used to correctly match not only nouns but also noun

groups to sentiment words.

2.5.1.2 Using Supervised Learning

Since information extraction is a generalized way of aspect extraction, supervised learning methods developed for information extraction process were used for aspect extraction. In [27], Conditional Random Fields (CRF) was used for aspect extraction. Features like POS tags, tokens, word distance were used to train CRF. In addition, reviews were chosen from different domains to train CRF better. Jin et al. chose lexicalized Hidden Markov Model (HMM) to extract features and opinions in a given text. Yu et al. used one-class SVM, a semi-supervised learning method, to extracted features [72]. In this method, only a set consist of positive samples was given to train SVM. In other words, only labelling features in training data was sufficient to train SVM.

2.5.1.3 Using Topic Models

Nowadays, topic models, an unsupervised method, are popular for extracting hidden topics from a huge set of text documents by using statistical methods. Topic modeling algorithms returns a group of word clusters where these clusters are used to find out topics [8]. Probabilistic Latent Semantic Analysis (pLSA) and Latent Dirichlet Allocation (LDA) are the two most common methods of topic modelling.

In sentiment analysis process, topics of topic modelling process have been expected to be the aspects. However, sometimes both aspects and sentiment words could be part of topics and they should be separated. Therefore, most of the researches on focused on joint model which extracted both aspects and sentiments.

A joint model based on pLSA was proposed by Mei et al. [43]. Their method extracted aspect model, positive sentiment model and negative sentiment model. Lin and He studied on a similar method by using extended LDA [35]. However, their method could not separate aspect words from sentiment words. A hybrid

model ,Maximum Entropy-LDA, was proposed by Zhao et. al [75]. The importance of this method was that it separated aspect words and opinion words on this aspects.

2.5.1.4 Implicit Aspect Extraction

As mentioned in *Aspect Extraction* section, aspects of a entity are given not only explicitly but also implicitly. However, implicit aspect extraction have been studied rarely.

An iterative clustering method which proposed a clustering method to connect sentiment words; in other words, implicit aspects to explicit aspects by Su et al. [43]. The mapping was found by clustering a group of explicit aspects to a group of sentiment words. Intra-set similarity and inter-set similarity were used to update pairwise similarities of the sets. A link was created between an aspect and sentiment word if they co-occurred in a sentence and this link became stronger when this two words co-occurred more often. As a final step, the aspect-sentiment word pair with links above a threshold were added to mapping.

Hai et al. proposed a two-step co-occurrence association rule mining method [22]. Firstly, association rules were created by labelling frequently co-occurring sentiment words as conditions and explicit aspects as consequences. Secondly, consequences were clustered and robust rules for conditions were generated. To find an explicit aspect referred by an implicit aspect (sentiment word), this implicit aspect was given to the association rule mining method after training and it was assigned to a cluster automatically.

2.5.2 Aspect Sentiment Classification

Aspect sentiment classification is the second step of the aspect-based sentiment analysis process. In this step, a mapping between extracted features and sentiment words is created. There have been two main approaches of aspect sentiment classification: lexicon-based approach and supervised learning approach.

The studies under lexicon-based approach have been mostly unsupervised. Sentiment lexicon, composite expressions, part-of-speech tags of sentence and rules have been commonly used in this approach. One of the most famous studies was proposed by Ding et al. [14]. Their method was a four-step approach. In the first step, sentiment words and phrases were marked and assigned their score from sentiment lexicon. In their study, they chose to assign +1 to positive sentiment words and -1 to negative sentiment words but different score assignment techniques existed in the literature. For example, different score for each word within in a range could be assigned to each sentiment word considering density of sentiment. As a second step, sentiment shifters were extracted and the score of the sentiment words they impacted were multiplied by -1. They were usually the negation words like not, never, none etc. After that, but-clauses were handled to detect score of context-dependent sentiment words. Context-dependent sentiment words were the ones which could have both positive and negative meaning depending on context. For instance, the sentiment word *high* states a positive condition in *performance of computer is high* phrase while it has a negative orientation in *temperature of the processor is very high* phrase. Therefore, they could not be assigned sentiment score directly and orientation of them could be found by using but-clauses. If the sentiment in the phrase before but-clause was negative, then context-dependent sentiment word was assigned positive score and vice versa. Finally, aggregate score for each sentiment word in a document was calculated. These four steps will be explained with an example over the sentence *The screen resolution of the phone is not enough; however, the photo quality is high.*. In the first step *enough* was assigned +1, but *high* was not assigned any score because it is context-dependent. *The screen resolution of the phone is not enough [+1]; however, the photo quality is high.* In the second step the negation word *not* before *enough* changed score of *enough. The screen resolution of the phone is not enough [-1]; however, the photo quality is high.*. Third step assigned a positive score to *high* since other part of the sentence is negative. *The screen resolution of the phone is not enough [-1]; however, the photo quality is high [+1].* Finally, closest aspect words to sentiment words were found and they were assigned corresponding score: screen resolution = -1 (negative) and photo quality = +1 (positive).

There have been several similar studies to mentioned above. For example, Wan applied similar techniques for Chinese [63] and Blair et al. proposed a lexicon-based supervised method [7].

There have been also some supervised studies. Firstly, supervised studies for sentence-level sentiment analysis have been valid for aspect-based sentiment analysis. In addition, a hierarchical classification approach based on sentiment ontology tree was used by Wei and Gulla [65]. Ganapathibhotla and Liu focused on supervised aspect-based sentiment analysis method on comparative sentences [18].

2.6 Sentiment Lexicon Generation

In this part, methods to create sentiment lexicons are mentioned. Sentiment lexicons consist of sentiment words, sentiment phrases and their polarity scores. Since a sentiment lexicon created in other studies have been used in this thesis, lexicon generation topic was not worked on. Therefore, only some basic methods are explained in this section.

One of the dictionary-based approaches was proposed by Hu and Liu [26]. Firstly, they created a small set with positive and negative sentiment words and their sentiment scores. Secondly, they iteratively improved these set by finding synonyms and antonyms of sentiment words currently in the set using WordNet,assigning scores to them and adding them to set. Until no new synonyms and antonyms were found, second step was run.

Another WordNet based approach was proposed by Kamps et al. [28]. WordNet has a distance function *dist(w1, w2)* which returns the length of shortest path between *w1* and *w2*. In this method, EVA value (described in Formula 2.1) was calculated for each word and the words with EVA value > 0 were assigned positive sentiment score while words with EVA value < 0 were treated

as negative.

$$EVA(w) = \frac{d(w,'bad') - d(w,'good')}{d('bad','good')} \tag{2.1}$$

Turney and Littman created sentiment lexicon with the help of PMI method [59]. They firstly created two sets namely *Pwords* and *Nwords*. *Pwords* defined a set consists of some positive sentiment words and *Nwords* defined a set consists of some negative sentiment words. Secondly, for each word SO-PMI score shown in Formula 2.2 was calculated and words with SO-PMI score is positive were labeled as positive and other words were labeled as negative.

$$SO - PMI(w) = \sum_{pword \in Pword} PMI(w, pword) - \sum_{nword \in Nword} PMI(w, nword) \tag{2.2}$$

All three methods mentioned above have been dictionary-based. However, some words can have different meanings with respect to domain. Therefore, generating sentiment lexicon by using domain corpus may be useful in some ways. In the following lines, some of the basic methods on corpus-based lexicon generation approaches are explained.

One of the first important studies on corpus-based sentiment lexicon generation was conducted by Hatzivassiloglou and McKeown [23]. They used the idea of sentiment consistency. In detail, they started with a set of seed adjectives and their sentiment scores. Then, they extracted sentiment words which were conjoined with adjectives from seed set with one of the connectives AND, OR, BUT, EITHER-OR or NEITHER-NOR. Finally, sentiment scores were assigned to extracted sentiment words by regarding connective which conjoins them to seed adjectives. In other words, two sentiment words conjoined with AND were assigned sentiment scores in the same orientation while two adjectives conjoined with BUT were given opposite sentiment scores. Assume the adjective *good*

was in seed set and it was known that it had a positive meaning. Then, in *The performance of my new computer is good and consistent.*, it is known that *good* is a positive sentiment word and it is conjoined to *consistent* with AND; therefore, *consistent* was assigned positive sentiment score, too.

The previous method enhanced by considering not only intra-sentence sentiment consistency mentioned in previous study but also considering inter-sentence sentiment consistency [29]. In other words, in addition to intra-sentence operations, they extracted the sentiment words from other sentences if this sentence was conjoined to another sentence which included a sentiment word from seed set. Finally, newly extracted inter-sentence sentiment words were assigned sentiment scores and added to lexicon.

In their study, Ding et al. used intra-sentence conjunction, pseudo intra-sentence conjunction and inter-sentence conjunction rules [14]. Moreover, they proposed that some sentiment words could have different sentiment orientations within the same domain when they explained sentiment on different entities and aspects. For example, in *My computer starts up in a short time.* and *My computer's battery life is short* sentences the sentence word *short* was used for both positive and negative attitude. To create sentiment lexicon by considering entity/aspect-sentiment word relations, they proposed a rule based approach with intra-sentence and inter-sentence methods.

2.7 Cross Domain Sentiment Analysis

If a word is used in different domains, it could have different meaning in each domain. Even, a word can have positive meaning in one domain while it has negative meaning in another domain. Therefore, to apply sentiment analysis methods on source domain to target domain, an adaptation process shall be applied. The method proposed in [9] used classified data from source domain and unclassified data from both source and target domains. The features co-occurred in both domains frequently were grouped into a set called pivot features. After that, a correlation matrix was created to calculate correlation between pivot

features and non-pivot features. Then, real-features set was chosen by using the correlation matrix. Finally, a new classifier was defined by using real-features and labelled data in source domain to classify data from both source and target domain. In [4], four different approaches were proposed to adopt a sentiment classifier to a new target domain. These approaches are as follows:

- a single classifier was chosen and it was trained with equal number of labelled data from n different domains
- again a single classifier was trained like in first approach but feature set was limited to target domain
- different classifier from different source domains were chosen and they were applied to target domain as a combination
- small number of classified data and larger number of unclassified data from target were combined to learn parameters and classified unclassified data. This approach only used target domain for training and classifying data

While SVM was chosen in the first three methods, expectation maximization was used in the fourth approach.

The article [47] proposed an approach to cross domain sentiment analysis which needed labeled source domain data and unlabeled target domain data. The source domain was adopted to target domain with spectral feature algorithm (SFA) which aligned domain-dependent sentiment words into a common cluster by using domain dependent words. SFA algorithm firstly created a bipartite graph between domain-dependent and domain-independent words. The weight of the link between two words were defined with number of times they co-occurred. Then, two domain-dependent words which had a lot of common links to domain-independent words were clustered together and two domain-dependent words were clustered if they had a lot common links to domain-depended words. Finally, these clusters were used to classify unlabeled data. In addition to these methods, joint topic modeling have also been used for cross domain sentiment extraction [24]. Firstly, topics covering both domains were extracted. Secondly, these topics were used as features in addition to original features in classification process while classifying unlabeled target domain data.

CHAPTER 3

BACKGROUND

3.1 Turkish Morphological Analysis

The previous researches have shown that almost all explicit features in a text have been nouns or nouns groups. Therefore, analyzing sentences morphologically and extracting POS tags are needed during feature extraction process. Morphological analysis refers to the mental system involved in word formation or to the branch of linguistics that deals with words, their internal structure, and how they are formed [3]. Morphological analysis of texts is a kind of text processing process and it is highly dependent on the language used in the text. In this context, since each language has its own rules, these rules have to be known to analyze sentences morphologically. Natural Language Processing (NLP) tools can be used for this purpose with high performance.

In this study, *Turkish possessive construction* is used as a part of the proposed aspect extraction method.

Turkish Possessive Construction

In Turkish, a *possessive construction* is a noun group where the second noun is possessed by the first noun. For instance, the nouns *telefon* (phone) and *ekran* (screen) generates the possessive construction *telefonun ekranı* (telefon-un ekran-ı), which refers to *phone's screen*. While producing a possessive construction, a suffix of ownership (genitive suffix) is added to the possessor, which is the first noun and third person suffix is added to the possessed object, which

Table 3.1: Turkish genitive and third person suffixes

Last Vowel	Genitive Suffix	Third Person Suffix
a, ı	-ın	-ı
e, i	-in	-i
o, ö	-un	-u
u, ü	-ün	-ü

is the second noun. The genitive suffix or the third person suffix which will be added to a word is chosen with respect to last vowel of the word as given in Table 3.1. In addition, if possessor or possessed object ends with a vowel, the character *n* is added to beginning of the genitive or the third person suffix.

Sentiment Enhancers

Sentiment enhancers are the words which usually take part before the sentiment words in both Turkish and English sentences and emphasize sentiment words. Some example sentiment enhancers are *very, pretty* and *quite*. For example, the sentiment word *good* has a positive meaning and if more positive meaning is demanded a sentiment enhancer can be used before *good*; e.g, *very good*.

Sentiment Shifters

Sentiment shifters are the words and phrases that change orientation of sentiment words they affect. Examples of sentiment shifters are *not, never* and *none*. In Turkish, usually, sentiment shifters exist within the two words after the sentiment word.

3.2 Pointwise Mutual Information

Pointwise mutual information (PMI) is a statistical measure to calculate the probability of two events being together [11]. In text processing domain, the PMI score of two words is calculated by Equation 3.1.

$$pmi(word_1; word_2) = \log_2 \frac{P(word_1, word_2)}{P(word_1)P(word_2)} \tag{3.1}$$

The problem with calculating PMI score for two words is to decide how to find scores of joint probability *P(word1,word2)* and probability *P(word)*. In [58], these two scores are calculated as in Equations 3.2 and 3.3.

$$P(word) \equiv hits(word) \tag{3.2}$$

$$P(word_1, word_2) \equiv hits(near(word_1, word_2)) \tag{3.3}$$

In these equations, *hits(X)* function refers to number of occurrence of a given phrase X on the Web and *near(Y,Z)* returns the word groups that start and end with words Y and Z, where at most 10 words exist between Y and Z according to [58].

This study uses PMI formula for texts with modifications to find co-occurrence rate of two words in the web for feature extraction process.

3.3 Tools

In this part the tools used for Turkish morphological analysis and web search are introduced. While Zemberek is chosen for Turkish morphological analysis, API support of Yandex is used for web search. Nevertheless, the proposed approaches do not directly depend on these tools.

Zemberek is the mostly used NLP tool for Turkish. It has NLP abilities such as morphological analysis, spellchecking, word suggestion for incorrect words, sentence extraction and spelling [2]. In my methodology, Zemberek is used for two purposes: to extract sentences from given texts, and to extract part of speech tags of words by analyzing the extracted sentences morphologically.

Yandex search engine API [1] is a service that allows users to send a query to Yandex search engine and get results in the XML format. In this thesis, search engine API is used while calculating hits(X) part of the PMI algorithm. While

[1] https://tech.yandex.com/xml/

several different parameters are returned in the result set from the search engine, only online frequency of the searched string is used. Since, near(Y,Z) function is not used in this study, it is used in Yandex search engine API. In addition, search engine API is also used to correct mistakes in the dataset as explain in Chapter 4 with details.

CHAPTER 4

METHODS

In this thesis, a two-step approach is followed. As seen in Figure 4.2 , firstly, data is collected and preprocessed in *Data Preparation* step. After data is preprocessed and corrected, *Aspect Based Sentiment Analysis* step, which is the main focus of this study, begins. At the beginning of this step, the sentences extracted in the *Data Preparation* part are passed to *Aspect Extraction* module and with the help of unsupervised algorithms, aspects in the dataset are extracted. Then, these extracted aspects and the sentences are given to *Aspect-Sentiment Classification* module and for all sentences in the dataset the orientation of thoughts over the extracted aspects are determined by using sentiment words in these sentences.

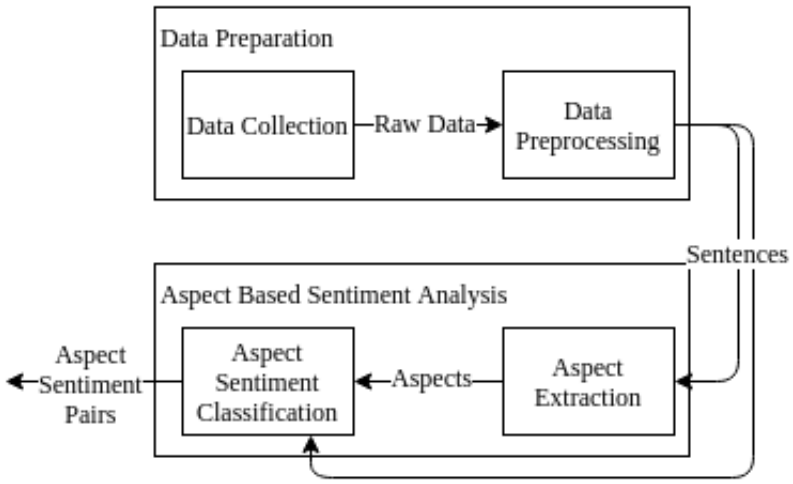

Figure 4.1: General steps

From now on, in Section 4.1, how data is prepared for further steps will explained by defining data collection and preprocessing techniques used. Section 4.2 will

be continued by giving details of *Aspect Extraction* algorithms developed in this study. At the end of this chapter, Section 4.3 will explain how *Sentiment-Aspect Classification* methods used in this study work.

4.1 Data Preparation

In this work, a dataset which consists of forum entries about mobile phones in Turkish is used for both development and experiment purposes. Since text mining in Turkish is not popular, a proper dataset for this work could not be found; therefore, a new dataset is created by collecting and preprocessing forum entries.

Donanimhaber [1] site is the biggest forum site in Turkish. It contains user reviews and comments on almost all gadgets used in Turkey as well as news about new technologies and devices. As seen in Table 4.1, more than 1.6 million users talks on approximately 900 main topics with the help of nearly 116 million entries with respect to key date May 2016. Since this work is supported by Huawei, the dataset is created by collecting entries written to Donanimhaber on Huawei mobile phones where these entries intensely state users' thoughts, comments and reviews on these devices.

Table 4.1: Donanimhaber statistics

Property	Count
Registered Users	1622708
Forums	905
Topics	8751574
Entries	115736613

It can be also stated that instead of the entries' in the dataset being about only Huawei mobile devices, the dataset does not have any labels or parts specific to this thesis and it leads that this dataset can be used easily in other studies, too. From now on, data collection and preprocessing approaches used and proposed in this work will be explained in details.

[1] http://www.donanimhaber.com

4.1.1 Data Collection

After choosing Donanimhaber for the reason that it is the most popular forum website in Turkish, all the pages under Huawei forum are crawled. Jsoup [2] which is a open-source crawling library is used to crawl all the entries requested. The reason why Jsoup is chosen is that it is a popular library which allows more support can be found on the Web than any other library. In addition, its capabilities are enough for what is needed from crawling library in this study.

After crawling operation is completed, the crawled pages which are in the format of raw HTML texts are cleaned up by deleting HTML tags and unnecessary information and extracting only the texts and other necessary informations of the entries. Finally, each of these entries are stored in different XML files with these three informations: author's name, entry and timestamp. What should be considered is that, forum entries can have quotes to others' entries which causes same entry appears more than once. Therefore, while creating XML files, these quoted parts are replaced with references to their original entries to prevent duplicate data.

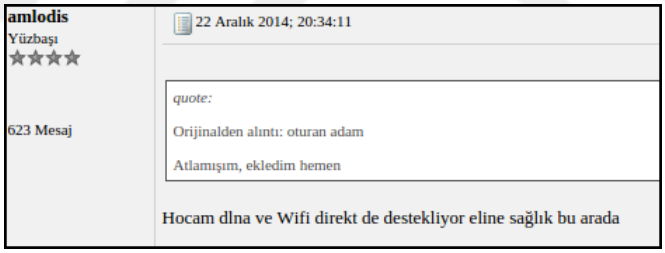

Figure 4.2: Example Donanimhaber entry

An example Donanimhaber entry can be seen in Figure 4.2. This entry is written by *amlodis* at *Dec 22, 2014*. Moreover, there is a quotation to *oturan_adam*' s entry which is replaced with reference to original entry in XML files.

After all XML files are created, to increase usability and portability, all entries in the XML files are imported into MySQL database. In this regard, fetching data and performing queries become much more easier. In addition, now, this dataset may be used in other studies easily by only importing dump of the database to

[2] http://jsoup.org/

Table 4.2: Dataset statistics

Property	Count
Topics	3458
Entries	179608
Authors	9305
Average entry per topic	51.94
Average entry per author	19.30
Total size of XML files	93.79 MB

the system. The detailed statistics of data collected at the end of data collection process can be seen in Table 4.2.

4.1.2 Data Preprocessing

Later on pages are crawled and their necessary parts are put into MySQL database by creating necessary MySQL tables to be used in this thesis. There are two main reasons of why this preprocessing step is needed: informal language used in the dataset and decreasing workload by applying some common operations for once.

Firstly, writing and grammar mistakes caused by informal language used are corrected. Since dataset is created from forum entries and forums contain authors from different educational and social levels, it can not be expected that all the entries written in formal language. Two main problems faced with are missing characters and Turkish characters whose English counterparts are used incorrectly.

Missing characters problem is a common problem in all languages. Users who write informally sometimes omits vowels in the texts they write. For instance, the Turkish word *tamam*, which corresponds to *OK* in English, usually written as *tmm* in informal websites. Since NLP tools which gives POS tags to words uses formal dictionaries and can not detect POS tags of words with missing characters, resolving missing character problem before sentiment analysis is important.

Table 4.3: Turkish specific characters

Turkish Character	English Counterpart
ç	c
ğ	g
ı	i
ö	o
ş	s
ü	u

The other problem is a Turkish related problem. Some characters used in Turkish are language specific and they have very similar English counterparts as seen in Table 4.3. For example, the Turkish word *kötü*, which corresponds to *bad* in English, sometimes used as *kotu* in the internet language. The reason of why English characters are chosen is not only keyboards does not support Turkish characters but also careless writing attitudes of authors. Again, this problem causes NLP tools run erroneously; therefore, it should be settled.

Because of these two problems, the collected data is preprocessed before starting sentiment analysis process. For this reason, a preprocessing method which has two legs is used.

In the first step, all entries put into MySQL database are fetched and separated into sentences with the help of NLP tool. It is a very straightforward operation. In detail, entries with more than one sentences are separated by using punctuations and special characters.

Secondly, Web search engine is used to resolve Turkish-English character problem and missing character problem. As known commonly, Web search engines have property of correcting misspelling search queries and proposing opportunity of searching with corrected query; in other words, *Did you mean?* feature. This property is used to correct sentences which contains words with Turkish-English character problem and missing characters. The sentences created in previous step are searched on the Web by using Yandex search engine API. Then, the sentences for which Yandex proposed a corrected sentence are replaced with these proposed ones.

Finally, a new MySQL table is created to store entries with sentences which are corrected during data preprocessing steps. New database created eases the operations in *Aspect Based Sentiment Analysis* and enhances results since it contains less erroneous data.

4.2 Aspect Extraction

In this study, a three-step model is proposed to extract aspects from an informal text dataset. At the beginning, to see whether the most common aspect extraction approaches can be used or not on Turkish informal texts, these approaches are implemented and applied to dataset. In other words, in the first two steps, state-of-the-art approaches *Frequency Based Aspect Extraction* and *Frequency Based Aspect Extraction with Sentiment Word Support* are applied to Turkish informal texts. Since the accuracy results of these approaches are not at acceptable level, a completely different method, that is *Web Search Based Aspect Extraction* method, is proposed to increase the accuracy of the aspect extraction process. The overall aspect extraction process is shown in Figure 4.3.

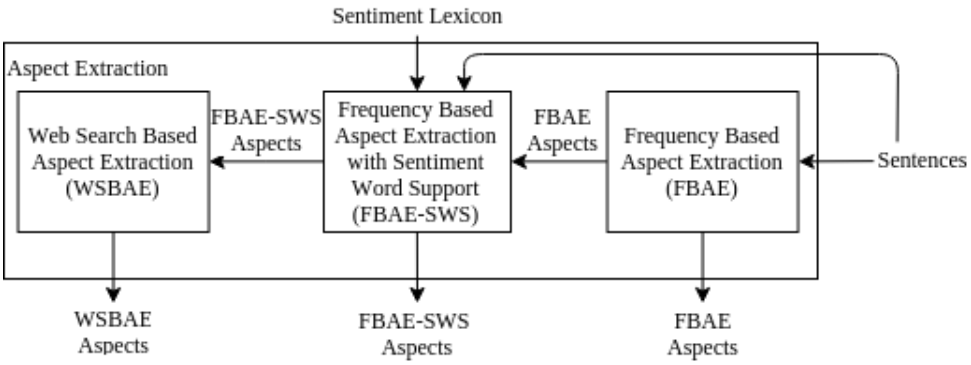

Figure 4.3: The aspect extraction process

4.2.1 Frequency Based Aspect Extraction (FBAE)

Frequency based aspect extraction is one of the most common aspect extraction methods. This approach is firstly introduced in [26]. Afterwards, it is used directly or with modifications in a lot of studies.

Algorithm 1 Frequency Based Aspect Extraction algorithm

Require: sl: list keeps sentences in the dataset

1: **procedure** FBAE
2: $noun\_freq\_list \leftarrow \{\}$ //list to keep nouns' occurrence in the dataset
//phase 1 - finding frequency of nouns in the dataset
3: **for** each sentence $s \in sl$ **do**
4: $word\_pos\_pairs \leftarrow$ give_POS_tags(s)
5: **for** each Word-POS pair $\{w, p\} \in word\_pos\_pairs$ **do**
6: **if** p is NOUN **then**
7: **if** $\exists$ $\{w, freq\}$ pair in $noun\_freq\_list$ **then**
8: update pair with $\{w, freq + 1\}$
9: **else**
10: add $\{w, 1\}$ pair to $noun\_freq\_list$
//phase 2 - extracting frequent nouns in the dataset
11: sort($noun\_freq\_list$)
12: $threshold \leftarrow$ get_experimental_threshold($noun\_freq\_list$)
13: $proposed\_aspects \leftarrow \{\}$
14: **for** each Word-Frequency pair $\{w, freq\} \in noun\_freq\_list$ **do**
15: **if** $freq \geq threshold$ **then**
16: add w to $proposed\_aspects$
return $proposed\_aspects$

As seen in Algorithm 1, there are two phases of this approach. In the starting phase, for sentences in the dataset which are extracted from entries, firstly, part-of-speech tags are found with the help of NLP tool. Secondly, nouns in these sentences are found from these POS tags and number of times each noun is seen are kept in a list. Second phase of the algorithm is the part where frequent nouns are extracted. To achieve that, the list which keeps frequency of nouns is sorted in descending order. Then, an experimental threshold is defined manually by investigating sorted list and deciding most effective cut-out point. Finally, nouns with frequency higher than or equal to this threshold are classified as aspects while others are classified as non-aspects.

4.2.2 Frequency Based Aspect Extraction with Sentiment Word Support (FBAE-SWS)

Algorithm 2 Frequency Based Aspect Extraction with Sentiment Word Support algorithm

Require: *sl*: list keeps sentences in the dataset, *fbae_aspects*: the FBAE aspects, *dict*: sentiment dictionary

1: **procedure** FBAE-SWS
2: *sent_word_freq_list* ← {} //keeps freq. of sentiment words in dataset
3: *freq_sent_words* ← {} //keeps frequent sentiment words in the dataset
4: *noun_freq_list* ← {} //keeps nouns' occurrence in the dataset
5: *freq_nouns* ← {} //keeps frequent nouns in the dataset
//phase 1 - finding sentiment words affects aspects frequently
6: **for** each sentence $s \in sl$ **do**
7: **for** each word $w \in s$ **do**
8: **if** $w \in fbae\_aspects$ **then**
9: *sent_word* ← find_nearest_sent_word(w, s, dict)
10: **if** $\exists$ $\{sent\_word, freq\}$ pair in *sent_word_freq_list* **then**
11: update pair with $\{sent\_word, freq + 1\}$
12: **else**
13: add $\{sent\_word, 1\}$ pair to *sent_word_freq_list*
14: *freq_sent_words* ← find_frequent_sent_words(*sent_word_freq_list*)
//phase 2 - finding nouns frequently affected from sentiment words
15: **for** each sentence $s \in sl$ **do**
16: **for** each word $w \in s$ **do**
17: **if** $w \in freq\_sent\_words$ **then**
18: *nearest_noun* ← find_nearest_noun(w, s)
19: **if** $\exists$ $\{nearest\_noun, freq\}$ pair in *noun_freq_list* **then**
20: update pair with $\{nearest\_noun, freq + 1\}$
21: **else**
22: add $\{nearest\_noun, 1\}$ pair to *noun_freq_list*
23: *freq_nouns* ← find_frequent_nouns(*noun_freq_list*)
return *fbae_aspects* ∪ *freq_nouns*

Frequency based aspect extraction with sentiment word support is the other common aspect extraction method used in this study. This approach, as FBAE, is firstly developed in [26]. In both of this study and [26], the main aim to work on this method is to enhance results of FBAE.

Again, two phase algorithm is followed for this approach. In the first phase of Algorithm 2, for each word in each sentence in the dataset it is checked that whether this word is an aspect proposed by FBAE or not. If so, the nearest sentiment word to this word in the same sentence is found, if exist. Then, frequency of such sentiment words are kept in a list. After all sentences are processed, the most frequent n sentiment words are passed to second phase.

Second phase which is very similar to first phase reveals nouns which are affected from sentiment words found out in the first phase. Accordingly, this time, for each word in each sentence in the dataset it is questioned that whether the word is a frequent sentiment word or not. If so, the nearest noun to this sentiment word in the same sentence is brought out, if exist. After, each occurrence of all these nouns are counted the union of the most common n nouns in addition to FBAE aspects are returned as aspects of this method.

Proceeding example reveals how second phase is applied to a sentence. Assume that *güzel* (well) and *kötü* (awful) are two of the common sentiment words of the dataset. To apply Phase 2 to sentence *GPS gayet güzel çalışıyor, ancak harita uygulaması kötü* (GPS works well, but map application is awful), algorithm goes over each word in the sentence. First frequent sentiment word detected is *güzel*. After it is detected, since nearest noun to this sentiment word is *GPS*, *GPS* is added to frequency list. The same process applies for *kötü* and *harita* (map), too.

This concludes explanation of how these two state-of-art methods can be applied to dataset. Before starting to explain the method developed in this thesis, an example is given in the following paragraphs for better understanding of FBAE and FBAE-SWS methods.

Table 4.4 gives an example on how FBAE and FBAE-SWS methods work. As-

sume that there is a dataset on mobile phones with limited number of nouns as shown in the table. In this table, *Is Aspect* column states whether corresponding noun or noun phrase is an aspect or not. *Data Freq* gives frequencies of nouns and noun phrases in the example dataset while *FBAE* column labels whether they are aspect (F) or not aspect (NF) with respect to FBAE method.The same applies for *Sent Based Freq* and *FBAE-SWS* columns, too.

Table 4.4: FBAE and FBAE-SWS example

Noun Phrase	Noun Phrase (English)	Is Aspect	Data Freq	FBAE (T=15)	Sent Based Freq	FBAE-SWS (T=25)
arkadaş	friend	No	13074	F	2321	F
ekran	screen	Yes	6743	F	1261	F
teşekkür	thank	No	4830	F	258	F
ihtimal	possibility	No	884	F	106	F
ekran ışık	screen light	Yes	70	F	15	F
ekran boyut	screen size	Yes	59	F	16	F
işlemci	processor	Yes	24	NF	20	F
ağaç	tree	No	9	NF	0	NF

To apply FBAE method on the dataset, firstly, all nouns and noun groups are extracted from the dataset with their frequencies and listed as sorted on their frequencies. In the example, six nouns and two noun groups given under *Noun* column are extracted and their frequencies shown in *Dataset Freq* column. With respect to FBAE, the nouns with frequencies higher than a threshold should be selected as aspects. In the example, the threshold is assumed as 15; therefore, as stated in *FBAE* column, first six nouns are labeled as aspects while last two nouns are classified as non-aspects. Then, these six nouns are given to FBAE-SWS method as input. In this method, firstly, common sentiment words are extracted by using these six nouns and sentiment directory. After that occurrence of each noun as a neighbor of frequent sentiment word is counted and written to *Sent Based Freq* column. Again, a threshold value is defined and nouns with occurrence counts above the threshold are labeled as aspects. For threshold value 15, as seen in the table, *işlemci* (processor) is classified as aspect in this step while it is not classified as aspect by FBAE method. After union of aspects processed in this step and FBAE is generated, classification results

shown in the *FBAE-SWS* column are gathered. In this example, 50% precision level and 75% recall value is achieved with FBAE method while FBAE-SWS achieves 57% precision and FBAE 100% recall value. Although recall values are high, precision values should be increased and this is achieved with the next method.

4.2.3 Web Search Based Aspect Extraction (WSBAE)

WSBAE is the main aspect extraction approach proposed in this thesis and given in Algorithm 3. In this method, search engine API and results of FBAE are used for aspect extraction. In this method, instead of using nouns' and noun groups' frequencies in the dataset, their frequencies on the Web are used.

During the first phase, firstly, for each noun or noun group labeled as aspect by FBAE-SWS method, a search query is generated to be used in search engine API. After these queries are searched on the Web, occurrence counts of these search results for all nouns and noun groups are stored in the list.

One of the key points of this algorithm is to generate search queries. They can be generated as stated in the Section 3.1. If a noun or noun group is an aspect of an item whose aspects are sought, then it is highly possible that this noun or noun group is used together with its item name on the Web many times. For example, for the noun *ekran* (screen) search query *telefonun ekranı* (phone' s screen) is created and it is searched on the Web as described in PMI method.

In the second phase, after all nouns and noun groups are searched on the Web, the list that keeps these nouns and noun groups with their occurrence counts on the Web is sorted in the descending order with respect to occurrence counts. Finally, an experimental threshold is defined and nouns and noun groups with frequencies above this threshold are returned as aspects. For instance, *telefonun ekranı* and *telefonun görüntüsü* (phone' s display) have high occurrence rates on the Web; therefore, *ekran* and *görüntü* are selected as aspects while *telefonun ürünü* (phone' s product) is an infrequent phrase on the Web and this is the reason why *ürün* is not selected as an aspect.

Algorithm 3 Web Search Based Aspect Extraction algorithm

Require: *item*: item in the dataset, *fbae_aspects*: aspects proposed by FBAE

```
1: procedure WSBAE
2:     aspect_occurrence ← {} //list to keep aspects' occurrence on the Web
   //phase 1 - finding occurrence count of nouns with the item on the Web
3:     for each aspect f ∈ fbae_aspects do
4:         if f does not contains item then
5:             search_query ← item + Genitive Suffix + " " + f + Third Person
   Suffix
6:         else
7:             search_query ← f
8:         occurrence_count ← get_occurrence_on_yandex(search_query)
9:         add {f, occurrence_count} to aspect_occurrence
   //phase 2 - extracting nouns and nouns groups frequent on the Web
10:    sort(aspect_occurrence)
11:    threshold ← get_experimental_threshold(aspect_occurrence)
12:    proposed_aspects ← {}, non_aspects ← {}
13:    for each {f, occurrence_count} ∈ aspect_occurrence do
14:        if occurrence_count ≥ threshold then
15:            add f to proposed_aspects
16:        else
17:            add f to non_aspects
   //phase 3 - extracting aspects that are infrequent on the Web
18:    for each f ∈ non_aspects do
19:        if f is a noun group then
20:            isAspect ← False
21:            for each word w ∈ f do
22:                if w ∈ proposed_aspects then
23:                    isAspect ← True
24:                    break
25:            if isAspect is True then
26:                add f to proposed_aspects
       return proposed_aspects
```

In the third phase, which is an enhancement phase, new aspects are extracted from noun groups below the threshold. For each such noun group, it is checked that whether at least one of its words belongs to aspects list created in Phase 2 or not. If there exists such a word, then this noun group is also added to the list. For example, *telefonun ekran kalitesi* (phone' s screen quality) is not a common noun phrase on the Web; however, since *telefonun ekranı* is a common noun phrase on the Web, *telefonun ekran kalitesi* is also selected as an aspect by using this phase.

In Table 4.5, an example applying WSBAE method to dataset given in the Table 4.4 is presented. First two columns are same as the Table 4.4. On the other hand, *Search Query* column gives noun and noun phrases which are searched on the Web with preceding *Telefonun* word. While *Search Results* column shows the online search counts for each search query, the last two columns give whether these nouns are aspects or not with respect to Phase 2 and Phase 3 of WSBAE algorithm.

Table 4.5: WSBAE example

Noun Phrase	Noun Phrase (English)	Is Asp	Search Query (Telefonun...)	Search Rslts	P2 (T=1000)	P3
arkadaş	friend	No	arkadaşı	3	NF	NF
ekran	screen	Yes	ekranı	28890	F	F
teşekkür	thank	No	teşekkürü	0	NF	NF
ihtimal	possibility	No	ihtimali	0	NF	NF
ekran ışık	screen light	Yes	ekran ışığı	354	NF	F
ekran boyut	screen size	Yes	ekran boyutu	1581	F	F
işlemci	processor	Yes	işlemcisi	3952	F	F
ağaç	tree	No	ağacı	N/A	N/A	N/A

All seven nouns proposed as aspects by FBAE-SWS are passed to WSBAE method. Then, in the first phase, search queries are created for these nouns. Secondly, all of these search queries are searched on the Web search engine and the occurrence count of each search query is stored as in the FBAE and FBAE-SWS approaches. In Phase 2, nouns with a higher Web occurrence counts than the threshold value are labeled as aspects. In the example, threshold is accepted as 1000 experimentally and *ekran, ekran boyutu* and *işlemci* are labeled

as aspects as stated in *WSBAE Phase 2* column. Finally, in Phase 3, each word of noun phrases classified as non-aspect in Phase 2 is checked whether it is an aspect or not. If at least one word of a noun phrase is an aspect, then this noun phrase is added to aspects list by Phase 3. In the example, *ekran ışık* is the only noun phrase that is a non-aspect with respect to Phase 2; nevertheless, it is classified as aspect by Phase 3 since Phase 2 classifies *ekran* as an aspect.

At the end of WSBAE process both precision and recall values are obtained as 100% which shows the results obtained by FBAE and FBAE-SWS methods are enhanced by WSBAE.

This concludes aspect extraction phase. In the remaining parts of this chapter aspect sentiment classification methods are explained.

4.3 Aspect Sentiment Classification

In this section, process of matching sentiment words to corresponding aspects and extracting overall score for each aspect is presented. To extract all aspect-sentiment couples, all comments in the dataset are given into comment classification process described below and at the end all sentiment scores for each different aspect is outputted.

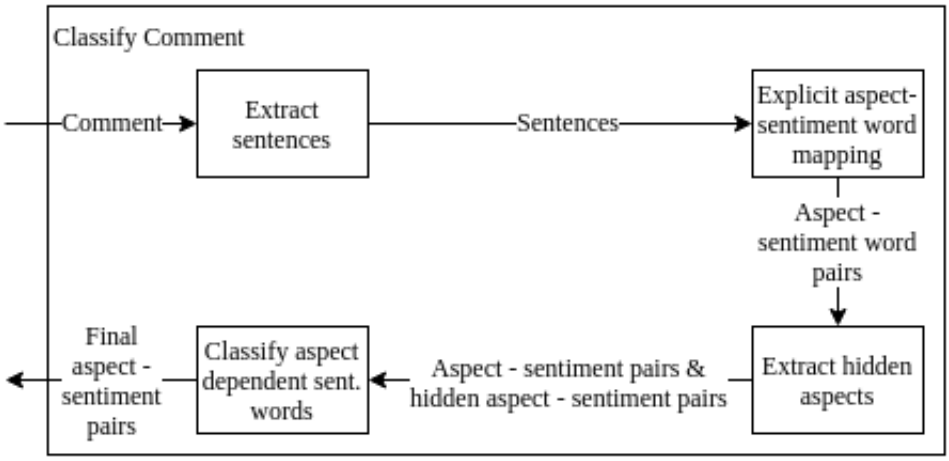

Figure 4.4: Comment classification steps

As seen in Figure 4.4, to classify a comment which is a group of sentences, firstly, this comment is split into sentences. While splitting comments into sentences,

sentences which end with a question mark or contain a Turkish question words are discarded with the help of a small preprocessing operation since question sentences do not specify sentiment, instead they investigate correctness of sentiments. After sentences are extracted, at the beginning, mappings between sentiment word groups and aspect groups are created for all sentences. Then, implicit sentiment words are extracted. Finally, sentiment scores of the sentiment words whose orientation depends on the aspects they classify are found out.

In the following subsections, the details of how aspects are mapped to sentiment words, how implicit sentiment words are extracted and how sentiment scores of aspect dependent sentiment words are extracted are mentioned in detail.

4.3.1 Explicit Aspect - Sentiment Mapping

Explicit Aspect - Sentiment Mapping process is the part where explicit aspects in a sentence and the sentiment words in the same sentence are revealed and connected. Explicit Aspect - Sentiment Mapping process consists of four steps as represented in Figure 4.5 and Algorithm 4: finding noun groups, finding sentiment word groups, matching noun groups to sentiment word groups and extracting scores for aspects.

4.3.1.1 Finding Noun Groups

This step takes a sentence as input and extracts all noun groups in this sentence. Noun group, in this work, is defined as group of nouns which are consecutive or just separated with comma or with words *ve* and *ile* (both of them correspond to *and* in English).

In order to find noun groups in a sentence, firstly, the position of the first noun in the sentence is detected with the help of NLP tool. Then, all consecutive nouns, commas and *ve* and *ile* words to this noun is grouped together and this new noun group is added to noun groups list. After that, the next noun in the sentence is searched by starting from the ending index of previously extracted

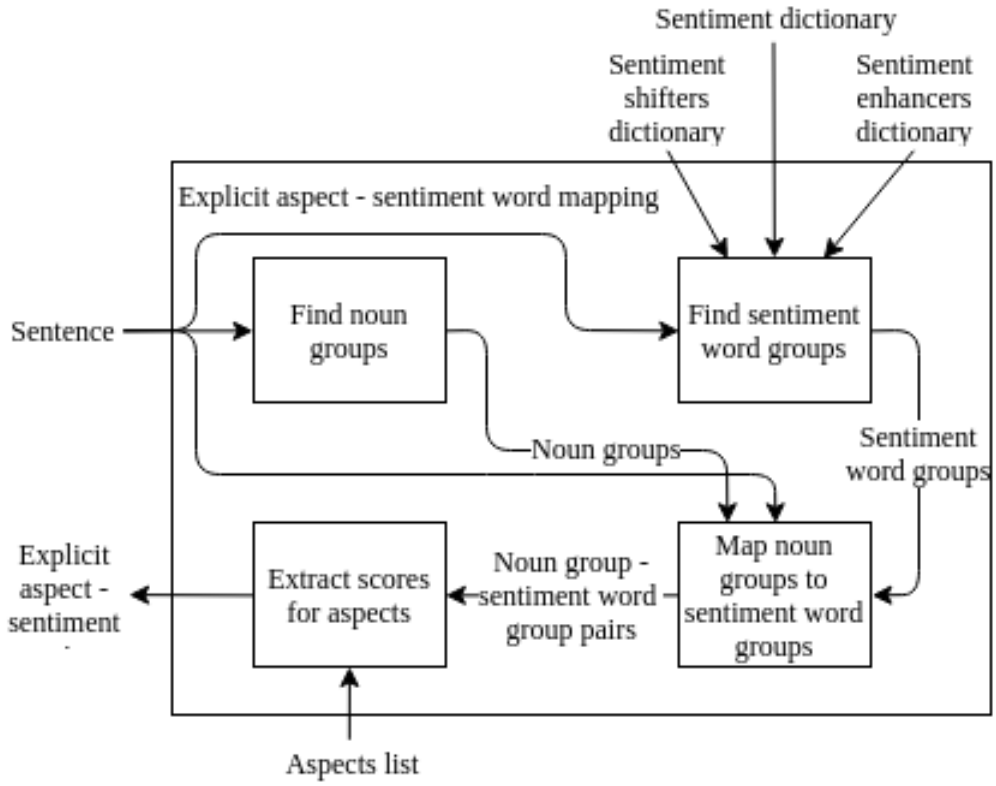

Figure 4.5: Explicit aspect - sentiment mapping steps

Algorithm 4 Explicit aspect - sentiment mapping algorithm

Require: *sentence*: sentence to classify, *aspects_list*: list of aspects, *sentiment_dictionary*: the list of sentiment word-score tuples, *enhancers_dictionary*: list of sentiment enhancers, *shifters_dictionary*: list of sentiment shifters

1: **procedure** EXPLICIT ASPECT - SENTIMENT MAPPING
2: *noun_groups_in_clause* ← Find noun groups(*sentence*)
3: *sentiment_word_groups_in_clause* ←
Find sentiment word groups(*sentence*)
4: *noun_group_sentiment_word_group_couples* ←
Map noun groups to sentiment word groups(
noun_groups_in_clause, sentiment_word_groups_in_clause)
5: *aspect_sentiment_score_couples* ← Extract aspect scores(
sentence, noun_group_sentiment_word_group_couples, aspects_list)
return *aspect_sentiment_score_couples*

noun group and, if a noun is found, new noun group is created by applying same grouping operations. This process continues until no new noun is found in the sentence. At the end, list of all extracted noun groups is returned as output of *Finding noun groups* step.

4.3.1.2 Finding Sentiment Word Groups

This is the second step of explicit aspect - sentiment mapping process where the sentiment word groups in the given sentence are extracted. The main part of a sentiment word group is very similar to a noun group. Consecutive sentiment words, commas and conjunctions forms the main part of sentiment word group. Moreover, sentiment score of each sentiment word in sentiment word group is also kept in sentiment word group. Finally, a sentiment word group can contain sentiment enhancer and sentiment shifter words, if they exist. Note that, sentiment enhancers and sentiment shifters are defined in detail in the Chapter 3.

A very similar process to *Finding noun groups* is applied to find main part of sentiment word group. This time, consecutive sentiment words, commas and conjunctions are extracted. To check whether a word is a sentiment word, sentiment words dictionary is used. After the main part of a sentiment word group is created sentiment, sentiment scores for each sentiment word is also added sentiment word group.

After sentiment words and their scores are extracted, sentiment enhancers and sentiment shifters are extracted and added to sentiment word group to complete *Finding sentiment word groups* step.

Extracting Sentiment Enhancer

This process is applied to find enhancer word of a sentiment word group, if exists. These step needs three inputs: the sentence, the sentiment word group, and sentiment enhancers dictionary. To check whether a sentiment enhancer which affects sentiment word group exists in the sentence, the word just before the sentiment word group is extracted and this word is searched in the sentiment

enhancers dictionary. If this word is found in this dictionary, then this word and enhancement coefficient is added to sentiment word group to use in the next steps. The enhancement coefficient is defined as 1.5 for all sentiment enhancers in the scope of this thesis. Nevertheless, this algorithm can be easily adopted to a new sentiment enhancers dictionary in which each different enhancer is assigned its own enhancer score. If no sentiment enhancer is found, then sentiment word group is returned with no change. After this step, *Extracting sentiment shifters* step is started.

Extracting Sentiment Shifter

The last step of *Finding sentiment word groups* extracts sentiment shifter, if exists. Again three inputs are given to this step: the sentence, the sentiment word group and the sentiment shifters dictionary. Since sentiment shifters usually take place within two words after the main part of sentiment word group in Turkish sentences, the two words which come after the main part of the sentiment word group are searched in the sentiment shifters dictionary. If any of these words is found in dictionary, then this word and the shifter coefficient which is -1 is stored in the sentiment word group. Again, this coefficient is defined same for all sentiment shifters and different shifter scores can be used for each different shifter if requested. If sentiment shifter is not located, then, as in the *Extracting sentiment enhancer* step, sentiment word group is returned without any change. This sub-step concludes *Finding sentiment word groups* step.

4.3.1.3 Mapping Noun Groups to Sentiment Word Groups

Mapping Noun Groups to Sentiment
Word Groupsss
This part is responsible from mapping extracted noun groups to extracted sentiment word groups in a sentence. Two main ideas are used to implement this part. Firstly, all elements of the group with smaller size must be mapped to different elements of the other group. For example, if less sentiment word groups exist in a sentence than noun word groups, then all sentiment word groups in this group should be paired with different noun groups. Secondly, the sum of

intra-pair distances of all noun group - sentiment word group pairs must be minimized.

To implement these ideas, firstly, sizes of two groups which are extracted by *Finding noun groups* and *Finding sentiment word groups* methods are calculated. Afterwards, all possible combinations with *size of the smaller group* elements are generated from elements of the larger group and the only possible combination with *size of the smaller group* elements is generated from elements of smaller group. For example, assume there exists 3 noun groups which are NG1, NG2 and NG3 and 2 sentiment word groups which are SWG1, SWG2 in a sentence. Since number of sentiment word groups is smaller, [NG1, NG2], [NG1, NG3], [NG2, NG3] combinations are created for noun groups and [SWG1, SWG2] combination is created for sentiment word groups.

After these combination are created, for each combination in the larger group, the sum of distances from its elements to corresponding elements of the combination of smaller group is calculated. For example, in the above example, firstly, sum of intra-pair distance between NG1-SWG1 pair and NG2-SWG2 pair, secondly, sum of intra-pair distance between NG1-SWG1 pair and NG3-SWG2 pair, finally, sum of intra-pair distance between NG2-SWG1 pair and NG3-SWG2 pair are calculated. Finally, the group of pairs with smaller total intra-pair distance is returned as mapping between noun groups and sentiment word groups.

If the size of two groups are equal, one of the groups is selected as smaller group randomly and same process is applied since only one combination can be generated for each group no matter which is selected smaller.

4.3.1.4 Extracting Scores for Aspects

In the last step of *Explicit Aspect - Sentiment Mapping* phase the aspects are extracted from noun groups and the corresponding sentiment scores are assigned to them. This phase takes the mappings created in the previous section and aspects list found in *Aspect Extraction* process as input.

For each noun group - sentiment word mapping, firstly, aspects in the noun

group are extracted. To extract aspects in a noun group, all nouns and noun groups in the noun group is searched in the aspects list and the nouns and noun groups found in this list are selected as aspects of noun group. Secondly, score of each sentiment word in sentiment word group is multiplied with sentiment shifter coefficient and sentiment enhancer coefficient of sentiment word group, if they exist. Finally, all possible pairings between aspects and sentiment words and their scores are generated and returned as results of Explicit Aspect - Sentiment mapping process.

This concludes *Explicit Aspect - Sentiment Mapping* part. In the following paragraphs how the whole *Explicit Aspect - Sentiment Mapping* process runs for a Turkish sentence is explained with an example. Assume that the sentence is *Ekran ve ses kalitesi pek iyi ve kaliteli değil, distribütör firma Hepsiburada. (Screen and sound quality is not very good, distributor company is Hepsiburada).* In addition, let *ekran* and *ses kalitesi* are the aspects of the domain, let *iyi* is a sentiment word with score +2, let *kaliteli* is a sentiment word with score +3, let *pek* is a sentiment enhancer and let *değil* is a sentiment shifter.

At the very first step of the *Explicit Aspect - Sentiment Mapping* process, firstly, noun groups are extracted. Extraction process starts for the first word of the sentence which is *ekran*. After *ekran* is found, consecutive nouns, commas and conjunctions to *ekran* are also added to noun group. Finally, first noun group *ekran ve ses kalitesi* is created and added to noun groups list. Then, process continues from the word *pek* and the second noun group which is *distribütör firma* is also added noun groups list.

The second step of the process, finds *iyi ve kaliteli* word phrase as the main part of sentiment word group with a similar method to first step. Then, the preceding word of this main part which is *pek* is searched in the enhancers dictionary and the following word of this main part *değil* is searched in the shifters dictionary and both of them are added sentiment word group as they are in corresponding dictionaries. In other words, first sentiment word group is created as *pek iyi ve kaliteli değil*. Starting from *firma* no new sentiment word is found until the end of sentence and this step finishes with one sentiment word group.

In the third step, since only one sentiment word group exists in the sentence, two combinations with one element is created from noun groups list: [*ekran ve ses kalitesi*] and [*distribütör firma*]. After that the distance from *ekran ve ses kalitesi* to *iyi ve kaliteli* and the distance from *distribütör firma* to *iyi ve kaliteli* is calculated. Since the distance from *ekran ve ses kalitesi* to *iyi ve kaliteli* is smaller, these pair is selected as the noun group - sentiment word group pair of the sentence.

At the end of *Explicit Aspect - Sentiment Mapping* process, firstly, aspects *ekran* and *ses kalitesi* are extracted from the noun group *ekran ve ses kalitesi*. Then, enhancer and shifter of sentiment word group is applied to score of sentiment words in this group. In other words, sentiment score of *iyi* is changed to -3 and *kaliteli* is changed to -4.5 for this sentiment word group. Finally, from these aspects and sentiment words, all possible aspect-sentiment score tuples are created. In other words, explicit aspect - sentiment mapping process returns the following two-tuples:*(ekran, -3)*, *(ses kalitesi, -3)*, *(ekran, -4.5)* and *(ses kalitesi, -4.5)*

4.3.2 Extracting Hidden Aspects

Some sentiment words not only state sentiment orientation but also implies a specific aspect. For example, the sentiment word *expensive* has a negative sentiment orientation and even it is used without an explicit aspect word, it is known that this word is used to state negative orientation on *price*. In some of the texts, if a sentiment word which implies an aspect is used, then aspect is not written explicitly to simplify the texts. Such aspects, implied by sentiment words are called as *implicit aspects*.

Extracting implicit aspects helps improvement of results achieved in this study in two ways: Firstly, more explicit results are listed after whole classification process is finished. In other words, presenting a sentiment word which implies an aspect together with this hidden aspect is more readable than just presenting the sentiment word. Secondly, the recall of classification process can be increased with the help of extraction of implicit aspects since the discarded sentiment

words with implicit aspect since there exists no explicit aspect which is mapped to them are not discarded any more.

Although, as stated before, in most of the sentences which contain a sentiment word with an implicit aspect do not have corresponding explicit aspects, in some sentences this situation is not followed. In other words, some sentences can have both sentiment words with implicit aspect and the explicit aspects which these sentiment words imply together. In this section, this property is used to extract implicit aspects.

In detail, to find whether a sentiment word implies and aspect or not, firstly, the number of times the sentiment word is coupled with a specific aspect is counted from the results set of *Explicit Aspect - Sentiment Mapping* section for each aspect the sentiment word affects. Secondly, the aspect-count pairs are sorted in the descending order with respect to count. Finally, the aspect with the highest count is classified as aspect implied by the sentiment word if the following three values are passed corresponding threshold values: the total number of times sentiment word is used, the ratio of the count of the aspect to total number of times sentiment word is used and the difference between counts of this aspect and the aspect with the second highest count.

This operations are applied for all sentiment words in the dataset to extract all implicit aspect. After all implicit aspects are extracted, all sentences containing sentiment words with implicit aspect are reprocessed by assuming extracted implicit aspect is the nearest noun group to a sentiment word with implicit aspect even it does not exist in the sentence.

4.3.3 Aspect Dependent Sentiment Word Classification

Although most of the sentiment words have polarity regardless of the aspects they affect, sentimental orientations of some of the sentiment words change with respect to aspects they are coupled with. One of the most well-known example to these words is *high*. For instance, in the sentence *The price of the phone is very high*, the sentiment word *high* is used to state level of the price of the phone;

therefore, it has a negative orientation. On the other hand, the same sentiment word is used in a positive manner in *It takes very high quality photos.*

This phase uses two different dictionaries: aspect dependent sentiment words dictionary and conjunctions dictionary. To find orientation of such a sentiment word on a specific aspect intra-sentiment word group and intra-sentence sentiment words and their scores are used. Firstly, for each occurrence of the aspect dependent sentiment word with the aspect, the following two operations are applied in order to aspect dependent sentiment word for extracting its orientation. If the sentiment score of this word is extracted during the first of these two operations, then second operation is not applied. On the other hand, if the sentiment score of this word cannot be extracted with these two methods, then no sentiment score is assigned to this word for its this occurrence. After all sentences containing aspect dependent sentiment word - aspect couple are processed as mentioned above, then average of all assigned score for this aspect dependent sentiment word - aspect couple is calculated and this score is assigned to aspect dependent sentiment word for its occurrences with the aspect. Finally, all the sentences in the dataset containing this aspect dependent sentiment word - aspect couple are reprocessed with a method similar to Explicit Aspect - Sentiment mapping method by using newly assigned score.

The above process is applied for all different aspect dependent sentiment word - aspect couples. From now on, the two approaches used to extract score of aspect dependent sentiment word over aspect are explained in detail.

4.3.3.1 Extracting Polarity of the Aspect Dependent Sentiment Words from Sentiment Words within the Same Sentiment Word Group

The first step of *Aspect Dependent Sentiment Word Classification* includes predicting score of an aspect dependent sentiment word from other sentiment words within the same sentiment word group, if exits. In other words, neighbor sentiment words determines the score of an aspect dependent sentiment word.

By assuming at least one non-aspect dependent sentiment word exist in the sentiment word group, firstly, a random sentiment word other than the aspect dependent sentiment word is selected from this group. Afterwards, it is checked that whether a negative conjunction exists between the randomly selected sentiment word and aspect dependent sentiment word. If no negative conjunction is found between them, than the sentiment score of randomly selected sentiment word is assigned to aspect dependent sentiment word. Otherwise, this sentiment score is multiplied by -1 before assigning to aspect dependent sentiment word.

For example, in the sentence *The design of the phone is thick but beautiful.*, *thick* is an aspect dependent sentiment word. Since the other sentiment word, *beautiful*, is a positive sentiment word and negative conjunction *but* exist between *beautiful* and *thick*, the negative of the sentiment score of *beautiful* is assigned to aspect dependent sentiment word *thick* for the aspect *design*.

4.3.3.2 Extracting Polarity of the Aspect Dependent Sentiment Words from Sentiment Words within the Same Sentence

If sentiment score cannot be predicted with the help of the first method, other sentiment word groups which are neighbor of the sentiment word group which contains the aspect dependent sentiment word are used to predict score. Only neighbor sentiment word groups are used since the effect of one sentiment word group on another sentiment words group in the same sentence decreases when the distance between them increases.

By assuming there exists two neighbor sentiment word groups of the sentiment word group which contains the aspect dependent sentiment word, firstly, one of these two neighbors is selected randomly and it is checked that whether there is a conjunction word between this sentiment word group and the sentiment word group with aspect dependent sentiment word. If a positive conjunction word is found between them, then the sentiment score of the nearest sentiment word of the selected sentiment word group to sentiment word group with aspect dependent sentiment word is extracted by considering sentiment shifter, if exists. Then, this score is directly assigned to aspect dependent sentiment word if there

is no sentiment shifter affecting aspect dependent sentiment word. Otherwise, it is multiplied by -1 before assigning. On the other hand, if a negative conjunction word exists between, same operations are applied until the end. At the end, the final version of the sentiment score is multiplied by -1 and then it is assigned to aspect dependent sentiment word.

If any conjunction word is not found between the randomly selected sentiment word group and sentiment word group with aspect dependent sentiment word, then other neighbor sentiment word group is selected and same process mentioned above is applied with newly selected sentiment word group. On the other hand, if and only if one neighbor sentiment word group exists, then the above process is applied with just this sentiment word group. Finally, if any neighbor sentiment word group is not found, then this step is passed directly.

For instance, the sentence *The design of the phone is not thin but screen quality is very good.* contains two sentiment word groups: *not thin* and *very good*. In addition, *thin* is an aspect dependent sentiment word. Since there exists a negative conjunction *but* between them, the negative of sentiment score of *beautiful* is assigned to sentiment word group *not thin*. Afterwards, due to the sentiment shifter *not*, this sentiment score is again multiplied by -1 and the same sentiment score with *beautiful* is given to the aspect dependent sentiment word *thin* for the aspect *design*.

This concludes, both aspect dependent sentiment word classification part and methods chapter. In the following lines the experiments and results are discussed in detail.

CHAPTER 5

RESULTS AND DISCUSSIONS

In this chapter, results of the aspect extraction and the aspect sentiment classification methods are discussed. Firstly, results for aspect extraction methods are mentioned in details. Then, aspect sentiment classification result are given in the following subsection.

5.1 Aspect Extraction

As mentioned before, three different aspect extraction methods are followed in this thesis. The third method used in this thesis is proposed since two very famous state-of-art methods are not fit to Turkish informal dataset. From now on, firstly, results obtained by using these two famous approaches are discussed and the reasons of why these algorithms are failed in this study are given. Secondly, outcomes of the experiments on WSBAE method are shown and the reason why this method reaches the success are explained. Before calculating results, all the noun and noun phrases in the dataset are extracted and they are labeled as aspect or non-aspect manually to compare extracted results from each method with correct labels.

5.1.1 Frequency Based Aspect Extraction (FBAE)

As seen in Table 5.1, for this baseline method, which is based on considering frequent nouns in the text as aspects, although very high recall is achieved, the

precision is very low which causes very low F-score. This is an expected result since nouns are expected to cover almost all of the aspects, however they include a high number of false positives.

Table 5.1: Experiment results

Method	Precision	Recall	F-score
FBAE	4.41	99.00	8.44
FBAE-SWS	4.55	99.00	8.70
WSBAE	59.24	84.90	69.79

On the other hand, in [26], the same method is used as a baseline and it is reported that 56% precision and 68% recall rates are obtained. There are several reasons for this performance difference. The first one is the content of the dataset. My dataset includes high number of sentences that do not refer to any aspect. As the second reason, the language used in the dataset is informal and hence morphological analysis fails to label some of the slang use of the words and word groups correctly. Moreover, since the approach is originally developed for English texts, linguistic features of Turkish may also affect the performance.

In addition, the points where FBAE is not sufficient can be seen in the example given in Table 4.4. Since the dataset used contains a lot of frequent nouns which are not aspects, there are a lot of mislabelled samples in the example. For instance, *ihtimal* is not a aspect of the item but it is a highly used word in Turkish, hence, it is frequent and selected as a aspect by FBAE. On the other hand, since a low threshold value is selected, almost all aspects are extracted. However, even this high recall value does not results with high F-score.

5.1.2 Frequency Based Aspect Extraction With Sentiment Word Support (FBAE-SWS)

In order to apply this method, which is presented in [26], I needed a sentiment corpus in Turkish. To this aim, I used the Turkish sentiment words collection of [62] by adding new sentiment words and removing some of the existing sentiment

words. In their study, a subset of English sentiment lexicon of SentiStrength [1] is translated to Turkish.

As seen in Table 5.1, in this method, very low precision and very high recall results are obtained as in that of FBAE method are achieved while much more better results are obtained in [26] with 59% precision and 80% recall. The reasons for these results is the informal language used in our dataset and linguistic features of Turkish language, as in the results of FBAE.

As can be seen in Table 4.4, similar results are obtained in FBAE-SWS. Since all nouns can have neighbor sentiment words even they are not a aspect of the item, low precision value is achieved. Again, much higher recall value is obtained but low F-score caused this method fail.

5.1.3 Web Search Based Aspect Extraction (WSBAE)

Experimental results show that (Table 5.1) the precision ratio is considerable improved through the proposed results. Hence, we can deduce that incorporating Web search results for the candidate terms can eliminate, especially slang use of nouns and noun groups that are not related with the item and hence are not aspects, effectively. On the other hand, this technique fails to recognize some of aspects that are specific to the item and that are not used frequently as search terms. Therefore, we observe some decrease in recall rate. However, on the overall, there is a huge increase in f-score.

As given in the experiment results, Table 4.5 shows WSBAE is much more efficient than other two methods. In the results of example given in Table 4.5, Phase 2 of WSBAE gives much better results than FBAE and FBAE-SWS. However, it misses a aspect which is a non-frequent noun phrase. After Phase 3 is applied, it is seen that this missing infrequent aspect is extracted without any loss on Phase 2' s results. Both high precision and recall values are obtained; therefore, it can be said that WSBAE approach is very efficient.

Table 5.2 and Figure 5.1 show the results of WSBAE algorithm under varying

[1] http://sentistrength.wlv.ac.uk

Table 5.2: Effect of threshold on accuracy

Threshold	Precision	Recall	F-score
10000	90.05	22.78	36.36
5000	84.13	30.20	44.44
2000	71.80	43.84	54.44
1000	67.13	57.88	62.16
500	62.99	74.83	68.40
300	59.24	84.90	69.79
250	57.27	86.62	68.95
100	49.51	92.98	64.61

threshold values. In these table and figure, precision, recall and F-score values are calculated for only intra-group of noun and noun phrases whose occurrence count on the web is equal to or more than 1 since even there is an aspect with this low frequency, it cannot be very important and frequent of the item. Increasing threshold value results with higher precision and lower recall values. The highest F-score value is achieved with threshold 300 and this value is used for the result reported in Table 5.1.

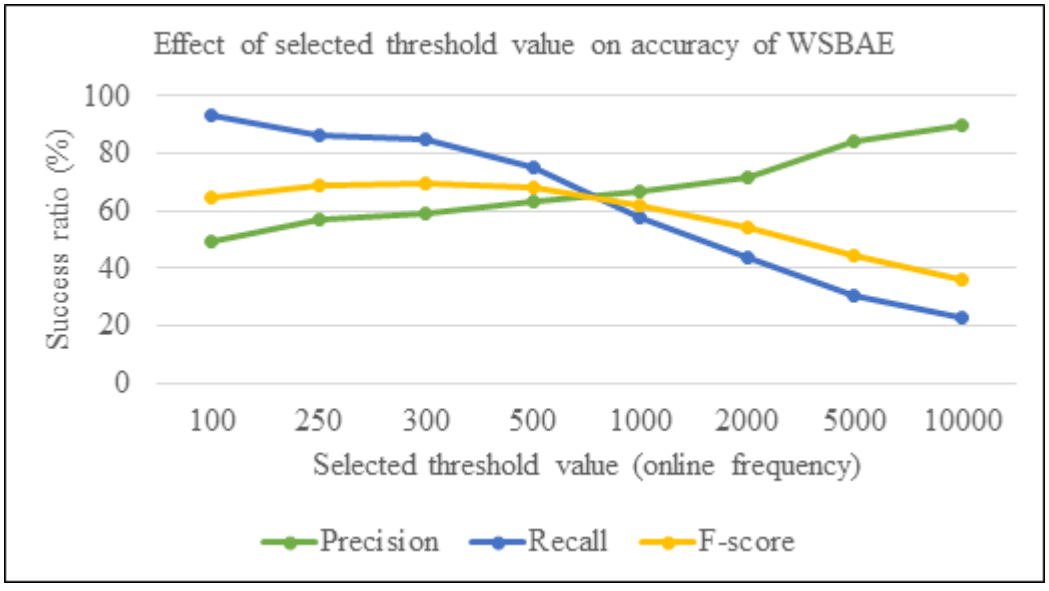

Figure 5.1: Effect of threshold on accuracy

5.2 Aspect Sentiment Classification

In this section the results of the three main steps of aspect sentiment classification process is discussed. As mentioned before, aspect sentiment classification process consists of three consecutive parts: explicit aspect-sentiment mapping, extracting hidden aspects, aspect dependent sentiment word classification. However before discussing the results of these parts, the success of the process on selecting sentences which contain sentiment - aspect pairs is revealed. Then, the achievement of the these three methods on the dataset defined before are discussed.

An important point on this part is that the complete dataset contains more than 500000 sentences. Although aspect sentiment classification process is applied to all of these sentences, limited sets of sentences are randomly selected and the results over these sentences are generalized to complete dataset since dataset is not labeled before and labeling all sentences in the dataset by hand is almost impossible. The number of randomly selected sentences to extract results are mentioned in the corresponding sections.

Success of Selection of Sentences with Sentiment - Aspect Pairs

In this section, it is discussed that whether the classified sentences with the three methods of aspect sentiment classification process are really sentences with sentiment-aspect pairs or not. In addition, the sentences which are not classified even they have sentiment-aspect pairs are also mentioned.

The reason why these results are presented is that in the following sections the success of three steps of aspect sentiment classification is investigated on both the complete dataset and on the sentences which are not directly discarded by aspect sentiment classification methods. The results on these sentences helps to visualize internal success rates of the steps of aspect sentiment classification.

To calculate results of these step, 1920 sentences are selected randomly and results are calculated from these sentences. 410 of these 1920 sentences contain at least one sentiment-aspect pair. However, 291 of 1920 sentences are processed

in aspect sentiment classification step while 28 of these 291 do not have any sentiment-aspect pair. On the other hand, 147 sentences are missed even they have sentiment-aspect pairs.

Table 5.3: Success of selection of sentences with sentiment-aspect pairs

Precision	Recall	F-score
90.37	64.15	75.04

Therefore, as mentioned in Table 5.3, almost 90% of the sentences are processed with methods of aspect sentiment classification process really contains aspect-sentiment pairs. On the other hand, approximately 64% of the sentences with aspect-sentiment pairs in the dataset are applied aspect-sentiment classification rules.

There are several reasons for why recall value is not as high as precision. Firstly, there are some aspect-based sentiment classification research areas which are out of scope of this thesis: co-reference resolution, comparative sentence, irony detection etc. Due to these areas, some aspect-sentiment pairs cannot be detected by methods of aspect sentiment classification process and it leads decrease in recall value. In addition, the informal language especially inverted sentences used in the dataset causes this problem.

From now on, the results of the three methods of aspect sentiment classification process is mentioned in detail. One important point for the following sections is that during calculation of these results complete aspects list about mobile phones is used instead of aspects returned by aspect extraction part since incomplete and erroneous aspects list returned by this section can cause huge deviations while calculating success of the aspect sentiment classification methods.

5.2.1 Explicit Aspect-Sentiment Mapping

As visualized in Table 5.4, to calculate the success of this method 1000 sentences which are processed by this method are selected randomly. After that, these 1000 sentences are classified manually to compare classification results obtained

Table 5.4: Explicit aspect-sentiment mapping results

Property	Count
Processed sentences	1000
Sentences without aspect-sentiment pair	92
Total number of aspect-sentiment pairs	1169
Sentences with aspect-sentiment pair	908
Total number of correctly classified aspect-sentiment pairs	863
Total number wrong aspect-sentiment pairing	223
Total number wrong sentiment score extraction	83

from aspect sentiment classification methods with correct classification of these sentences. Again, as seen in Table 5.4, 92 of these sentences do not contain any aspect-sentiment pair; in other words, they are processed unnecessarily and wrong aspect-sentiment pairs are created. The other 908 sentences contains 1169 aspect-sentiment pairs. 863 of these 1169 pairs are extracted correctly. On the other hand, 223 of these pairs are not extracted correctly because of wrong mappings between aspects and sentiment words. In addition, 83 pairs are classified incorrectly since sentiment score is not calculated correctly.

Table 5.5: Explicit aspect-sentiment mapping results (%)

Case	Precision	Recall	F-score
Only classified sentences consider	91.22	79.47	84.95
All sentences consider	91.22	50.98	65.41

Table 5.5 shows the precision, recall and F-score values for explicit aspect-sentiment extraction method. In the first line, the result when only the sentences classified by this method are considered. These results show 91% of all aspect-sentiment mappings returned by this method contains correct aspect-sentiment score mappings. In addition, nearly 80% of all aspect-sentiment pairs in these sentences are extracted and this results with 85% of F-score. On the other hand, second line of Table 5.5 shows results when not only 1000 sentences which this method processed are considered but also the discarded sentences even they have at least one aspect-sentiment pair are taken into consideration. Although precision value is not changed in this line of table, recall value decreases dramatically because of discarded sentences wrongly. This line can be

interpreted as 51% of all aspect-sentiment pairs in the dataset are classified by explicit aspect-sentiment classification method and 91% percentage success ratio is gathered from this classification.

The precision is equal to precision succeeded in [14] from which explicit aspect-sentiment extraction idea is adopted from. However, higher recall value is succeeded in [14]. The main reason for lower recall value obtained in this thesis is same as mentioned in *Success of Selection of Sentences with Sentiment - Aspect Pairs* part. In addition, mismatching of noun groups to sentiment words during explicit aspect-sentiment classification also causes this recall value.

5.2.2 Extracting Hidden Aspects

This section, firstly, mentions the results of hidden (implicit) aspect extraction method individually. Secondly, the change in results of explicit aspect-sentiment mapping part is revealed after extracted hidden aspects are applied to dataset and results of this part is combined with results of explicit aspect-sentiment mapping part.

In order to extract hidden aspects, complete dataset is used. However, to see effects of aspect sentiment classification by using this hidden aspects on the dataset the same 1000 sentences chosen randomly in the previous part is used.

Table 5.6: Extracting hidden aspects results

Precision	Recall	F-score
66.67	66.67	66.67

In the dataset, three frequent sentiment word - implicit aspect couples are used: pahalı-fiyat (expensive-price), uygun-fiyat (cheap-price) and ucuz-fiyat (cheap-price). On the other hand, extracting hidden aspects method is returned three sentiment word - implicit aspect couples: pahalı-fiyat (expensive-price), uygun-fiyat (cheap-price) and orjinal-rom (original-rom). Therefore, as shown in Table 5.6, both precision and recall values are measured as 66.67% in extraction part of this method.

The implicit aspect extraction method proposed in [22] uses co-occurrence of aspects and sentiment words to extract implicit aspect as my method. These precision, recall and F-score values of this study changes between 70-75%. The change between my method and method proposed in [22] is caused by the informal dataset used in my study.

The reason why the number of sentiment words with hidden aspects is very low is that although there are some words such as *ağır* (heavy) whose English counterpart is accepted as containing hidden aspect, these words are not returned by extracting hidden aspects method because these words are used with different meanings in Turkish. For example, *ağır* is used to not only mention about weight of a object but also the speed of an object in Turkish. Hence, such words cannot be accepted as sentiment words with hidden aspects. In addition, some sentiment words with implicit aspects are not used in the dataset or used with a very low frequency hence discarded.

Table 5.7: Aspect-sentiment mapping results (%) with hidden aspects

Case	Precision	Recall	F-score
Only classified sentences consider	90.36	80.89	85.36
All sentences consider	90.36	52.72	66.59

Table 5.7 presents the changes in results of explicit aspect-sentiment mapping step after sentences are reclassified by using extracted hidden aspects, also. The results shows that precision decreases slightly. The reason for this decrease is that, although sentiment words with hidden aspects are used to classify hidden aspect most of the time, they may used to classify other aspects, too. Matching all occurrence of the sentiment word with hidden aspect to corresponding aspect causes some wrong mappings and decreases precision.

On the other hand, the recall value increases more than precision value decreases. As mentioned before the occurrences of the sentiment word with hidden aspect is discarded if no explicit aspect can be mapped to this sentiment word in the explicit aspect-sentiment mapping part. This situation have a negative effect on recall. In this method, after hidden aspect word from a sentiment word is extracted, this hidden aspect is used to create sentiment - aspect mapping in the

each occurrence of this sentiment word without mapped to an explicit aspect. This prevents sentiment word being discarded. Therefore, recall value increases.

These results show that the F-score of this study approaches to F-score achieved in [14] which is the selected baseline study during the explicit aspect-sentiment classification step. It is achieved with the help of increasing recall value.

5.2.3 Aspect Dependent Sentiment Word Classification

This section discusses the results of aspect dependent sentiment word classification process and the results obtained when dataset is processed by considering these sentiment words. To extract this results six different aspect dependent sentiment words are used as shown in Table 5.8. Although there exists other aspect dependent sentiment words in Turkish, these six sentiment words are chosen since the dataset used in this thesis is about mobile phones and these six aspect dependent sentiment words are the most common sentiment words used to explain sentiments on mobile phones' aspects.

Table 5.8: Aspect dependent sentiment words

Turkish	English
büyük	big
küçük	small
kısa	short
uzun	long
ince	thin
kalın	thick

During classification of aspect dependent sentiment words, all the sentences; in other words, approximately 500000 sentences, in the dataset are used. Firstly, for each different aspect dependent sentiment words - aspect pair exist in the sentences, list of all predicted scores for this pair are extracted. In total 1123 scores are assigned to aspect dependent sentiment words - aspect pairs. After that, firstly, aspect dependent sentiment words - aspect pairs whose predicted score list contains less than 5 predicted scores is eliminated. Secondly, the pairs whose predicted score list does not contain scores in one orientation with at least

75% are also discarded. The average of remaining 639 scores for corresponding aspect dependent sentiment words - aspect pairs can be seen in Table 5.9.

Table 5.9: Aspect dependent sentiment word classification results

Aspect Dependent Sentiment Word - Aspect Pair			
Turkish	English	Score[-4,+4]	Frequency
kamera-büyük	camera-big	-2.5	52
ekran-küçük	screen-small	-2.4	93
ekran-büyük	screen-big	2.1	155
kılıf-kalın	case-thick	-2.8	23
malzeme kalitesi-uzun	material quality-long	-2.0	10
dil-küçük	language-small	-2.0	7
performans-uzun	performance-long	2.9	29
kapak-ince	cover-thin	2.5	35
şarj-uzun	battery-long	2.1	54
kutu-ince	box-thin	2.0	26
telefon-ince	phone-thin	2.3	50
telefon-kalın	phone-thick	-2.4	87
dizayn-ince	design-thin	2.3	18

Although, there exists two meaningless aspect dependent sentiment word - aspect pairs which are material quality-long and language-small, the other 11 aspect dependent sentiment word - aspect pairs are meaningful. Moreover, it can be seen that frequencies of these two meaningless pairs are much lower than frequencies of the other pairs. In addition, the sentimental orientation of predicted sentiment scores for these 11 meaningful pairs are correct.

In order to measure precision and recall, when the sentences containing aspect dependent sentiment word-aspect pairs are classified with newly predicted scores, again the same 1000 sentences are used. In other words, the results of Extracting Hidden Aspects is enhanced by classifying sentences which contain aspect dependent sentiment word-aspect pairs.

Table 5.10 displays the results after sentences which contains aspect dependent sentiment words are reclassified with predicted sentiment scores for aspect dependent sentiment word - aspect pairs predicted. It can be interpreted that precision does not change when aspect dependent sentiment words are used.

Table 5.10: Aspect-sentiment mapping results (%) with hidden aspects and aspect dependent sentiment words

Case	Precision	Recall	F-score
Only classified sentences consider	90.39	83.41	86.75
All sentences consider	90.39	53.70	67.37

The reason for this situation is that the meaningless aspect dependent sentiment word - aspect pairs leads mismatching between sentiment word groups and noun groups; therefore, precision can not be increased. On the other hand, recall value has increased because the discarded sentences and noun group - sentiment word group mappings in the explicit aspect-sentiment mapping step because of undefined scores of aspect dependent sentiment words are classified with the predicted scores in this step.

After this step, the F-score of the complete aspect sentiment classification process becomes closer to F-score achieved in [14] which is the selected baseline study during the explicit aspect-sentiment classification step. Again, increasing recall value decreases this gap.

Evaluation

In this section, results obtained during aspect sentiment classification are discussed all together. In Table 5.11, results after each step calculated by considering all sentences are shown.

Table 5.11: Aspect sentiment classification results (%) - all sentences considered

After Step	Precision	Recall	F-score
Explicit aspect - sentiment mapping	91.22	50.98	65.41
Hidden aspect extraction	90.36	52.72	66.59
Aspect dependent sentiment word classification	90.39	53.70	67.37

On the other hand, Table 5.12 demonstrates the results calculated by considering only classified sentences.

In both cases, the precision value obtained in the first step decreases lightly

Table 5.12: Aspect sentiment classification results (%) - only classified sentences considered

After Step	Precision	Recall	F-score
Explicit aspect - sentiment mapping	91.22	79.47	84.95
Hidden aspect extraction	90.36	80.89	85.36
Aspect dependent sentiment word classification	90.39	83.41	86.75

when the second step is applied; however, the third step does not change the precision level obtained after the second step. On the other hand, recall value and F-score value increase after each step in both tables. Since F-score increases with each step applied, it can be inferred that applying more steps gives better results.

Finally, when these two tables are compared it can be seen that recall and F-score values increase dramatically when only classified sentences are considered. In detail, although obtained recall and F-score values are far away from the values obtained in [14] when all sentences are considered, this gap is almost vanished when only classified sentences are considered.

CHAPTER 6

CONCLUSION AND FUTURE WORK

In this chapter, firstly, aspect extraction and aspect sentiment classification parts are concluded. After that, possible future developments which can be applied to both parts are explained.

6.1 Conclusion

Sentiment analysis is one of the most complex research topics and it is an active research area due to its impacts on both commercial and academical efforts. Aspect extraction is the first step of the aspect-based sentiment analysis and it is important to match correct sentiment words with correct features.

To solve aspect extraction problem in this thesis, I built a framework to extract features of a target item from informal texts by using an unsupervised learning based solution. I adapted the basic ideas from the literature and proposed a new approach that is based on constructing a search term by using the candidate features of the item and the Web search count.

For the experiments, I set the target item as a mobile phone model and constructed a dataset by collecting postings on a Turkish forum containing entries written about the item. At the end of experiments, %60 precision and %85 recall are achieved. Therefore, it can be revealed that the proposed approach improves the feature extraction performance in comparison to two basic approaches from the literature whose precision is %4.5 and recall is %99.

Second step of the aspect-based sentiment analysis is aspect-sentiment classification which connectes aspects to sentiment words which affect them. Again, unsupervised learning based methods are used and developed in this step. Firstly, the sentiment word groups and noun groups are extracted from sentences and they are matched to each other. After that, for each aspect in the noun groups the sentiment score of corresponding sentiment word group is calculated and mapping between aspects and sentiment scores is created. In this proceeding steps, to enhance precision and recall, implicit aspects are extracted from sentiment words referring them with the help of previously extracted aspect-sentiment word couples, finally, the sentiment scores of aspect dependent sentiment words are predicted by using sentiment scores of the other sentiment words.

The same dataset with the aspect extraction part is used during the experiments of this part. At the first step of the experiments; in other words, with the Explicit Aspect- Sentiment Mapping step, 91% precision and 51% recall values are achieved. After hidden aspects are extracted and the dataset is classified by using this hidden aspects in addition to explicit aspects, precision decreased slightly to 90% but recall value is increased to 53%. At the final step, scores for aspect dependent sentiment words are predicted for different aspect. When dataset is reclassified with this predicted scores in addition to other methods and parameters used until this step, precision value is stayed at 90%; however, recall values is increased to nearly 54%.

6.2 Future Work

The aspect extraction part can be improved further in several directions. One research direction is to build a hybrid model that incorporates both the term frequencies in the dataset and the search query result counts. Another point to work on is transferring the feature extraction knowledge among semantically similar target items in order to facilitate the process.

Moreover, although I developed and tested methods for Turkish, since I used the linguistic property that Turkish is an agglutinative language while creating Web

search queries, my aspect extraction approach may fit to other agglutinative languages well. Moreover, no words are used between the product word and feature word in Turkish; therefore, my approach can also easily be adopted to other languages having the same property. Hence, as a future work, aspect extraction method developed in this thesis may be applied to other languages having these properties.

Some improvements can be proposed for the aspect-sentiment classification part, too. The rules of Turkish can be analyzed more deeply and these rules can be applied to all phases of this part. For example, by analyzing the positions and effects of sentiment shifter words in Turkish sentences intensively, final sentiment scores can be calculated more precisely.

Furthermore, different methods on noun group - sentiment word group matching can be tried to see whether a better matching can be achieved or not. For example, sentence parse tree can be generated from a given sentence, then the distances in this tree can be used for matching.

In addition, new approaches can be proposed to analyze Turkish comparative sentences which possibly helps aspect-sentiment classification process end up with better aspect-sentiment score results. This is a very complicated and huge study area on which there exists some studies in the other languages but not in Turkish.

Finally, this work can be used commercially after all components are automatized and pipelined. In other words, the algorithms and methods proposed in this thesis can be used commercially by creating a system which automatically collects data, preprocesses them, finds features by using preprocessed data and, finally, extracts aspect-sentiment mappings. If such a system exists, all the comments and trending thoughts on a topic can be easily analyzed and summarized.

REFERENCES

[1] F. Akba, A. Uçan, E. A. Sezer, and H. Sever. Assessment of feature selection metrics for sentiment analyses: Turkish movie reviews. In *8th European Conference on Data Mining 2014*, 2014.

[2] A. A. Akın and M. D. Akın. Zemberek, an open source nlp framework for turkic languages. *Structure*, 10:1–5, 2007.

[3] M. Aronoff and K. Fudeman. Thinking about morphology and morphological analysis. In *What is morphology?*, pages 1–31. Blackwell Publishing, 2004.

[4] A. Aue and M. Gamon. Customizing sentiment classifiers to new domains: A case study. In *Proceedings of recent advances in natural language processing (RANLP)*, volume 1, pages 2–1, 2005.

[5] L. Barbosa and J. Feng. Robust sentiment detection on twitter from biased and noisy data. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pages 36–44. Association for Computational Linguistics, 2010.

[6] D. Bespalov, B. Bai, Y. Qi, and A. Shokoufandeh. Sentiment classification based on supervised latent n-gram analysis. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 375–382. ACM, 2011.

[7] S. Blair-Goldensohn, K. Hannan, R. McDonald, T. Neylon, G. A. Reis, and J. Reynar. Building a sentiment summarizer for local service reviews. In *WWW Workshop on NLP in the Information Explosion Era*, volume 14, pages 339–348, 2008.

[8] D. M. Blei. Probabilistic topic models. *Communications of the ACM*, 55(4):77–84, 2012.

[9] J. Blitzer, M. Dredze, F. Pereira, et al. Biographies, bollywood, boomboxes and blenders: Domain adaptation for sentiment classification. In *ACL*, volume 7, pages 440–447, 2007.

[10] J. Bollen, H. Mao, and X. Zeng. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8, 2011.

[11] K. W. Church and P. Hanks. Word association norms, mutual information, and lexicography. *Computational linguistics*, 16(1):22–29, 1990.

[12] K. Dave, S. Lawrence, and D. M. Pennock. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th international conference on World Wide Web*, pages 519–528. ACM, 2003.

[13] B. Desmet and V. Hoste. Emotion detection in suicide notes. *Expert Systems with Applications*, 40(16):6351–6358, 2013.

[14] X. Ding, B. Liu, and P. S. Yu. A holistic lexicon-based approach to opinion mining. In *Proceedings of the 2008 International Conference on Web Search and Data Mining*, pages 231–240. ACM, 2008.

[15] M. Eirinaki, S. Pisal, and J. Singh. Feature-based opinion mining and ranking. *Journal of Computer and System Sciences*, 78(4):1175–1184, 2012.

[16] U. Erogul. Sentiment analysis in turkish, 2009.

[17] M. Gamon, A. Aue, S. Corston-Oliver, and E. Ringger. Pulse: Mining customer opinions from free text. In *Advances in Intelligent Data Analysis VI*, pages 121–132. Springer, 2005.

[18] M. Ganapathibhotla and B. Liu. Mining opinions in comparative sentences. In *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*, pages 241–248. Association for Computational Linguistics, 2008.

[19] R. González-Ibánez, S. Muresan, and N. Wacholder. Identifying sarcasm in twitter: a closer look. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pages 581–586. Association for Computational Linguistics, 2011.

[20] Google. Influencing offline: The new digital frontier. https://ssl.gstatic.com/think/docs/influencing-offline-the-new-digital-frontier_research-studies.pdf, 2011. [Online; accessed December-2011].

[21] Gwava. How much data is created on the internet each day? https://www.gwava.com/blog/internet-data-created-daily, 2014. [Online; accessed 19-November-2014].

[22] Z. Hai, K. Chang, and J.-j. Kim. Implicit feature identification via co-occurrence association rule mining. In *Computational Linguistics and Intelligent Text Processing*, pages 393–404. Springer, 2011.

[23] V. Hatzivassiloglou and K. R. McKeown. Predicting the semantic orientation of adjectives. In *Proceedings of the 35th annual meeting of the association for computational linguistics and eighth conference of the european chapter of the association for computational linguistics*, pages 174–181. Association for Computational Linguistics, 1997.

[24] Y. He, C. Lin, and H. Alani. Automatically extracting polarity-bearing topics for cross-domain sentiment classification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 123–131. Association for Computational Linguistics, 2011.

[25] Y. Hong and S. Skiena. The wisdom of bookies? sentiment analysis vs. the nfl point spread. In *Proceedings of the international conference on Weblogs and Social media (icWSm-2010)*. Citeseer, 2010.

[26] M. Hu and B. Liu. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177. ACM, 2004.

[27] N. Jakob and I. Gurevych. Extracting opinion targets in a single-and cross-domain setting with conditional random fields. In *Proceedings of the 2010 conference on empirical methods in natural language processing*, pages 1035–1045. Association for Computational Linguistics, 2010.

[28] J. Kamps, M. Marx, R. J. Mokken, M. d. Rijke, et al. Using wordnet to measure semantic orientations of adjectives. In *Proceedings of LREC*, pages 1115–1118. European Language Resources Association (ELRA), 2004.

[29] H. Kanayama and T. Nasukawa. Fully automatic lexicon expansion for domain-oriented sentiment analysis. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 355–363. Association for Computational Linguistics, 2006.

[30] M. Kaya, G. Fidan, and I. H. Toroslu. Sentiment analysis of turkish political news. In *Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology-Volume 01*, pages 174–180. IEEE Computer Society, 2012.

[31] C. Leung and S. C. Chan. Sentiment analysis of product reviews., 2009.

[32] J. Li and M. Sun. Experimental study on sentiment classification of chinese review using machine learning techniques. In *Natural Language Processing and Knowledge Engineering, 2007. NLP-KE 2007. International Conference on*, pages 393–400. IEEE, 2007.

[33] X. Li, H. Xie, L. Chen, J. Wang, and X. Deng. News impact on stock price return via sentiment analysis. *Knowledge-Based Systems*, 69:14–23, 2014.

[34] Y.-M. Li and T.-Y. Li. Deriving market intelligence from microblogs. *Decision Support Systems*, 55(1):206–217, 2013.

[35] C. Lin and Y. He. Joint sentiment/topic model for sentiment analysis. In *Proceedings of the 18th ACM conference on Information and knowledge management*, pages 375–384. ACM, 2009.

[36] D. Lin. Automatic retrieval and clustering of similar words. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics-Volume 2*, pages 768–774. Association for Computational Linguistics, 1998.

[37] W.-H. Lin, T. Wilson, J. Wiebe, and A. Hauptmann. Which side are you on?: identifying perspectives at the document and sentence levels. In *Proceedings of the Tenth Conference on Computational Natural Language Learning*, pages 109–116. Association for Computational Linguistics, 2006.

[38] B. Liu. *Web data mining: exploring hyperlinks, contents, and usage data.* Springer Science & Business Media, 2007.

[39] B. Liu. Sentiment analysis and subjectivity. *Handbook of natural language processing*, 2:627–666, 2010.

[40] B. Liu. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1):1–167, 2012.

[41] C. Long, J. Zhang, and X. Zhut. A review selection approach for accurate feature rating estimation. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pages 766–774. Association for Computational Linguistics, 2010.

[42] R. McDonald, K. Hannan, T. Neylon, M. Wells, and J. Reynar. Structured models for fine-to-coarse sentiment analysis. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 432–439. Citeseer, Association for Computational Linguistics, 2007.

[43] Q. Mei, X. Ling, M. Wondra, H. Su, and C. Zhai. Topic sentiment mixture: modeling facets and opinions in weblogs. In *Proceedings of the 16th international conference on World Wide Web*, pages 171–180. ACM, 2007.

[44] Merriam-Webster-Dictionary. Opinion. http://www.merriam-webster.com/dictionary/opinion, 2016. [Online; accessed 26-July-2016].

[45] S. Mohammad. From once upon a time to happily ever after: Tracking emotions in novels and fairy tales. In *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, pages 105–114. Association for Computational Linguistics, 2011.

[46] T. Nakagawa, K. Inui, and S. Kurohashi. Dependency tree-based sentiment classification using crfs with hidden variables. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 786–794. Association for Computational Linguistics, 2010.

[47] S. J. Pan, X. Ni, J.-T. Sun, Q. Yang, and Z. Chen. Cross-domain sentiment classification via spectral feature alignment. In *Proceedings of the 19th international conference on World wide web*, pages 751–760. ACM, 2010.

[48] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, pages 79–86. Association for Computational Linguistics, 2002.

[49] A.-M. Popescu and O. Etzioni. Extracting product features and opinions from reviews. In *Natural language processing and text mining*, pages 9–28. Springer, 2007.

[50] K. Ravi and V. Ravi. A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowledge-Based Systems*, 89:14–46, 2015.

[51] J. Read. Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In *Proceedings of the ACL student research workshop*, pages 43–48. Association for Computational Linguistics, 2005.

[52] RetailingToday. Study: 81% research online before making big purchases. http://www.retailingtoday.com/article/study-81-research-online-making-big-purchases/, 2013. [Online; accessed 12-July-2013].

[53] P. Sakunkoo and N. Sakunkoo. Analysis of social influence in online book reviews. In *ICWSM*, 2009.

[54] C. Scaffidi, K. Bierhoff, E. Chang, M. Felker, H. Ng, and C. Jin. Red opal: product-feature scoring from reviews. In *Proceedings of the 8th ACM conference on Electronic commerce*, pages 182–191. ACM, 2007.

[55] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede. Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2):267–307, 2011.

[56] O. Tsur, D. Davidov, and A. Rappoport. Icwsm-a great catchy name: Semi-supervised recognition of sarcastic sentences in online product reviews. In *ICWSM*, 2010.

[57] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe. Predicting elections with twitter: What 140 characters reveal about political sentiment. *ICWSM*, 10:178–185, 2010.

[58] P. D. Turney. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 417–424. Association for Computational Linguistics, 2002.

[59] P. D. Turney and M. L. Littman. Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems (TOIS)*, 21(4):315–346, 2003.

[60] H. Türkmen, S. Ilhan Omurca, and E. Ekinci. An aspect based sentiment analysis on turkish hotel reviews. In *The Third International Symposium on Engineering, Artificial Intelligence and Applications, ISESIA*, 2015.

[61] G. Vinodhini and R. Chandrasekaran. Sentiment analysis and opinion mining: a survey. *International Journal*, 2(6), 2012.

[62] A. G. Vural, B. B. Cambazoglu, P. Senkul, and Z. O. Tokgoz. A framework for sentiment analysis in turkish: Application to polarity detection of movie reviews in turkish. In *Computer and Information Sciences III*, pages 437–445. Springer, 2013.

[63] X. Wan. Using bilingual knowledge and ensemble techniques for unsupervised chinese sentiment analysis. In *Proceedings of the conference on empirical methods in natural language processing*, pages 553–561. Association for Computational Linguistics, 2008.

[64] H. Wang, D. Can, A. Kazemzadeh, F. Bar, and S. Narayanan. A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In *Proceedings of the ACL 2012 System Demonstrations*, pages 115–120. Association for Computational Linguistics, 2012.

[65] W. Wei and J. A. Gulla. Sentiment learning on product reviews via sentiment ontology tree. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 404–413. Association for Computational Linguistics, 2010.

[66] J. Wiebe. Learning subjective adjectives from corpora. In *AAAI/IAAI*, pages 735–740, 2000.

[67] J. M. Wiebe, R. F. Bruce, and T. P. O'Hara. Development and use of a gold-standard data set for subjectivity classifications. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, pages 246–253. Association for Computational Linguistics, 1999.

[68] T. Wilson, J. Wiebe, and R. Hwa. Recognizing strong and weak opinion clauses. *Computational Intelligence*, 22(2):73–99, 2006.

[69] Y. Wu, Q. Zhang, X. Huang, and L. Wu. Phrase dependency parsing for opinion mining. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3-Volume 3*, pages 1533–1541. Association for Computational Linguistics, 2009.

[70] A. Yessenalina, Y. Yue, and C. Cardie. Multi-level structured models for document-level sentiment classification. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 1046–1056. Association for Computational Linguistics, 2010.

[71] H. Yu and V. Hatzivassiloglou. 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*, pages 129–136. Association for Computational Linguistics, 2003.

[72] J. Yu, Z.-J. Zha, M. Wang, and T.-S. Chua. Aspect ranking: identifying important product aspects from online consumer reviews. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 1496–1505. Association for Computational Linguistics, 2011.

[73] X. Yu, Y. Liu, X. Huang, and A. An. Mining online reviews for predicting sales performance: A case study in the movie domain. *Knowledge and Data Engineering, IEEE Transactions on*, 24(4):720–734, 2012.

[74] L. Zhang and B. Liu. Identifying noun product features that imply opinions. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pages 575–580. Association for Computational Linguistics, 2011.

[75] W. X. Zhao, J. Jiang, H. Yan, and X. Li. Jointly modeling aspects and opinions with a maxent-lda hybrid. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 56–65. Association for Computational Linguistics, 2010.

[76] J. Zhu, H. Wang, B. K. Tsou, and M. Zhu. Multi-aspect opinion polling from textual reviews. In *Proceedings of the 18th ACM conference on Information and knowledge management*, pages 1799–1802. ACM, 2009.

[77] L. Zhuang, F. Jing, and X.-Y. Zhu. Movie review mining and summarization. In *Proceedings of the 15th ACM international conference on Information and knowledge management*, pages 43–50. ACM, 2006.