NOVEL SYMBOLIC AND SUB-SYMBOLIC APPROACHES FOR TEXT BASED AND MULTIMODAL SENTIMENT ANALYSIS

KIA DASHTIPOUR

DEPARTMENT OF COMPUTING SCIENCE AND MATHEMATICS

Doctor of Philosophy

Division of Computing Science and Mathematics

University of Stirling

July 2019

DECLARATION

I hereby declare that this dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text and bibliography.

I also declare that this dissertation (or any significant part of my dissertation) is not substantially the same as any that I have submitted, or that is being concurrently submitted, for a degree or diploma or other qualification at the University of Stirling or similar institution.

I was admitted as a research student in [August 2019] and a candidate for the degree of Doctor of Philosophy in [August 2019]. This dissertation is a record of the work carried out at the University of Stirling between 2015 and 2019, under the supervision of Dr Jingpeng Li and Dr David Cairns.
I have read, and adhered to, the University's plaigarism policy, as detailed at: `http://www.plagiarism.stir.ac.uk/`

Stirling, July 2019

Kia Dashtipour

ABSTRACT

In the era of digital media, e-commerce and social networks, websites allow users to share opinions and feedback about products and services. Customers can make informed decisions by reading the experiences of other users. In addition, customer feedback can be used by the organizations to further improve the offered services. However, the quintillion bytes of data generated per day in different languages such as Persian consisting of user feedback cannot be manually read and analyzed by an individual or an organization, for gauging public opinion. Sentiment analysis is an automated process of computationally understanding and classifying subjective information in multi-disciplinary fields such as products, movies, news, public opinion etc.

In this thesis, we focus on developing novel methods for Persian text-based sentiment analysis. We exploit the developed text-based methods to improve multimodal polarity detection. Specifically, we develop a novel hybrid framework that integrates dependency-based rules and deep neural networks for detecting polarity in Persian natural language sentences. In addition, we develop a Persian multimodal sentiment analysis framework that integrates audio, visual and textual cues to computationally understand and harvest sentiments from videos posted on social media platforms such as YouTube and Facebook. Specifically, a first of its kind, multimodal Persian sentiment analysis dataset is developed, which is then used evaluate the proposed multimodal framework that exploits the hybrid dependency-based sentiment analysis framework and deep neural network based multimodal fusion. Extensive experimental results have proven the effectiveness of the proposed approaches as compared to state-of-the-art approaches including deep neural networks.

ACKNOWLEDGMENTS

First of all, I would like to thank to my supervisors Prof Amir Hussain, Dr Jingpeng Li and Dr David Carins, without whom this PhD would not have been possible. In addition, I cannot thank Prof Amir Hussain enough for appointing me on the prestigious University of Stirling.

I would also like to immensely thank Mandar Gogate also an outstanding former PhD student and collaborator of Prof Amir Hussain. I am deeply indebted to him for his continuous guidance and invaluable advice during my PhD research. I really appreciate his patience and kindness, and timely advice and encouragement, which helped me in every aspect of life. He also played a major role in my decision of pursuing research and graduate studies with Prof Amir and himself.

In advanced, I would like to thank to Dr Ahsan Adeel, who has been the key person behind whatever I have achieved till today. He always keeps supporting me with his valuable advice.

In addition, I would like thank to Prof Alexander Gelbukh, I cannot forget those sleepless nights we spent discussing research topics and papers.

Furthermore, I would like thank to Dr Amjad Ullah and Dr Jawad Ahmad for guiding me whenever I faced any problem during my research.

At the end, I would like to say big thanks to all my family and friends for their extensive support. An emotional thanks to everyone who has always been there to support, inspire and motivate me all the way.

LIST OF PUBLICATIONS

1. **Dashtipour, K.**, Poria, S., Hussain, A., Cambria, E., Hawalah, A.Y., Gelbukh, A. and Zhou, Q., 2016. Multilingual sentiment analysis: state of the art and independent comparison of techniques. Cognitive computation, 8(4), pp.757-771.
2. **Dashtipour, K.**, Hussain, A., Zhou, Q., Gelbukh, A., Hawalah, A.Y. and Cambria, E., 2016, November. PerSent: a freely available Persian sentiment lexicon. In International Conference on Brain Inspired Cognitive Systems (pp. 310-320). Springer, Cham.
3. **Dashtipour, K.**, Hussain, A. and Gelbukh, A., 2017, April. Adaptation of Sentiment Analysis Techniques to Persian Language. In International Conference on Computational Linguistics and Intelligent Text Processing (pp. 129-140). Springer, Cham.
4. **Dashtipour, K.**, Gogate, M., Adeel, A., Ieracitano, C., Larijani, H. and Hussain, A., 2018, July. Exploiting deep learning for Persian sentiment analysis. In International Conference on Brain Inspired Cognitive Systems (pp. 597-604). Springer, Cham.
5. **Dashtipour, K.**, Gogate, M., Cambria, E., Hussain, A, Jingpeng Li. 2019. Integrating Dependency-Based Rules and Deep Neural Networks for Persian Sentiment Analysis, Neurocomputing.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

ANN - Artificial Neural Network

SVM - Support Vector Machine

NB - Naive Bayes

AE - Auto Encoder

CNN - Convolutional Neural Network

PCA - Principal Component Analysis

LSTM - Long Short-Term Memory

BoC - Bag of Concepts

BoW - Bag of Words

LSTM - Long Short-Term Memory

AI - Artificial Intelligence

POS - Part Of Speech

RNN - Recursive Neural Network

ML - Machine Learning

SA - Sentiment Analysis

1

CHAPTER 1: INTRODUCTION AND MOTIVATION

Due to the Internet and technological evolution, the use of online forums has increased exponentially in the last few years. Additionally, with the advent of social media (Twitter, YouTube and Facebook etc), people are able to share their opinions frequently. Social media encourages people to engage in political discussions and enables them to share their thoughts on political issues. Online media provides a platform for sharing ideas and also encourages public to join group discussions. In addition, social media allow companies and organisations to get feedback regarding their products in the form of texts, images and videos [4].

In the past, people usually asked their family and friends about a product before making a purchase decision. However, these days, people generally share their opinions and comments about different products online as shown in Fig 1.1 One can see from Fig 1.1, through social media, people can comment on a product/service. Such opinions that are shared online can be used by the customers to make a final decision before they purchase a product. For example, an e-commerce website allows customer to share their experiences about the purchased products and services, thereby helping new customers decide whether or not to buy the products. Online public comments are generally unstructured. This unstructured data can be converted to a useful (structured) format that can be further used by companies for customer satisfaction and product improvement etc. Unstructured data can be converted through manual conversion process. However, the manual conversion will require a lot of effort. Therefore, the automatic extraction of useful information

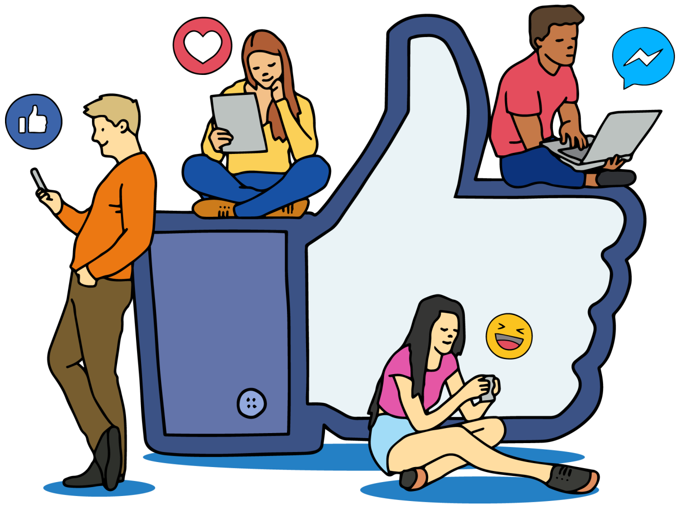

Figure 1.1: People share their opinions on a Product/Service in the form of an emoji

from people's opinions regarding different products and services is extremely useful for companies and organisations. For example, a person planning to buy a mobile phone can search for different websites to see the latest mobile phones and reviews about mobiles which can help them to make a better decision. The importance of people opinion can be illustrated via another example, if someone wants to go to the cinema, the person can look at online reviews before purchasing the ticket [5] [6].

From the aforementioned examples, the need and importance of automated systems for data understanding is very clear. The providers should be able to learn from online users' feedback. However, these comments and opinions can assist companies and organizations to continue to their product and services. Sentiment analysis (SA) has been used in different fields such as sport, movies, product and news [7].

SA can be used to automatically identify the polarity in the text and categorize the opinions into positive, neutral or negative. SA can be performed at the following two levels, i.e. the document level and sentence level. The document level is used to classify the sentiment expressed in the document, whereas the sentence level is used to identify the sentiments expressed only in a single sentence [8] [9]. SA can be applied in a variety of applications, however, some most important are outlined as: [10]:

E-commerce: Websites such as Amazon or Trip Advisor allow customers to write reviews about their services and products. Amazon website contains over seventy million reviews. Companies and organisations use SA to classify reviews and improve the quality of their products mainly based on customer opinions [6] [11] [12].

Movie Reviews: aims to find people's feeling towards the movie. SA helps film industries to obtain structured data from customers and improve their film industries [13] [14] [15].

Importance of SA in legislation: Government uses SA to understand people's opinions about their services. For example, with the question "What do people think about abortion legislation?", SA can be used to understand the overall public opinion about abortion legislation [16].

1.1 RESEARCH CHALLENGES IN TEXT-BASED SENTIMENT ANALYSIS

Text information is considered as an asset. It contain a lot of useful information which can be used for SA. However, the number of challenges involved in SA are outlined below:

Complex sentence structure: It is a difficult task for SA to understand the text structure. For example, the following sentence, من كَارگردَاني فيلم دوست ندَارم ولي بَازيهَا بسيَار خوب بود ("I do not like the movie direction, but I really like the acting”), contains positive and negative sentiment about the movie. Most of the current approaches are not able to detect the overall polarity of the sentence, therefore the system is unable to understand the different sentiment expressed in the sentence. Additionally, the presence of negation can flip the polarity from positive into negative [17].

Curse of dimensionality: Most of the current approach of SA use a bag of words and n-gram feature methodologies, extract term frequency of n-gram or words as features. Despite being simple, these approaches have achieved high performance in text classification. However, feature vectors generated by these methods are very sparse and high dimensional. These types of feature extraction require a large amounts of processing time and can cause over-fitting in the presence of regularisation.

Explicit sentiment: An opinion can be written in two different formats: (i) Explicit, (ii) Implicit. An explicit contains a word or expressions. For example, فیلم خوبي بود ("Movie is great"). the word خوبي ("great") is explicitly has positive opinion. However, people do not express their opinion explicitly. For example, من دوست دَارم امشب بیرون غذَا بخورم ولي رستورَان تعطیل است (“I want to eat something outside, but the restaurant was closed”). Although there is no sentiment available in the sentence, however, the overall polarity of the sentence is negative. Identifying the implicit sentiment is the most difficult challenge in the field of SA [18].

Noisy Data: The data from the web consists a lot of noise such as the presence of HTML tags, use of different scripting etc. However, these types of noise are not limited to aforementioned problems, it consists of ungrammatical text, slang, abbreviations words and phrases such as BRB ("Be Right Back") or Persian example, گنه is short term for گنَاه (sin). An automated SA system is unable to detect sentence polarity unless these challenges are addressed.

Multiple reviews: The online users can discuss various topics. For example, if they are talking about one product. Alternatively, they can discuss about other product. Additionally, they can discuss about different features of the product. شَارژ موبَایل اپل خیلي خوب بود ولي من شَارژ لپ تَاپ اپل دوست ندَارم, ("The charger of the Apple mobile is very good, but I do not like the charger for Apple laptop"). The sentence aspect is charger and the sentiment of the sentence for mobile is positive however, the second part of the sentence has a negative sentiment about the laptop.

Sarcasm detection: The sarcasm detection is a challenge for natural language processing. The presence of sarcasm in a sentence can change the overall polarity of the sentence, اَمروز همه چيز ارزون شده ("Everything is getting cheaper"). The identification of sarcasm in a sentence is a challenging task. These factors make the task of sarcasm detection very challenging [19].

There are lots of meaningful information is behind the text. For example, فيلم بدِ نبود ولِ اگر بَازيگرَان ديگري استفَاده كرده بودند مِتوَانست بهتر بَاشد, "It was not a bad movie, but if they are using different cast it could be a better movie". The natural language processing is used to understand the text. The use of keywords, punctuation and frequency of words are very useful to understand the text. However, the increase of web content in different languages make the English algorithm inefficient. In the order to overcome these challenges in this thesis we focus on Persian SA. Most of the previous studies have only focused on English SA and therefore most of the tools and resources are available in English. Hence the tools and research, in the most of the other languages such as Persian, are comparatively less developed. Persian SA has the following main challenges:

Lack of tools and resources: To the best of our knowledge, there is no valuable tools nor any online lexicon available for Persian language.

Utilizing many informal words: There are lots of informal words available in Persian language. Specifically, when online users share their opinion and comments online, they are using more of these informal words and phrases.

Lack of comprehensive approaches: There is no valuable comprehensive approach available in Persian language. Most of the current approaches utilize the available approaches of English and translate the Persian dataset into English. Motivation of this thesis is outlined as:

- In Persian language, there is a large amount of data that is not classified. Unfortunately, there is no valuable tools available to classify these data.
- Most of the current approaches translate the dataset into English and utilize English lexicon and Persian language needs the attention from the research community.
- There is no comprehensive approach available to detect polarity in Persian sentences.

1.2 Research challenges in multimodal sentiment analysis

The audio data can express the tone of the speaker and video can express the facial expression and movement. Video is important to identify the sentiment analysis. However, the multimodal is more complex task is compared with unimodal such as text sentiment analysis. There are number of challenges in multimodal sentiment analysis:

High dimensional feature: As compared to unimodal (text) sentiment analysis, the multimodal sentiment analysis consists of feature vector for all modalities such as video, audio and text.

Noisy data in Multimodal sentiment analysis: The noisy or incorrect data can effect the performance of the model. Therefore, a primary step is to identify this information and discard them. For example, some of the video may have poor quality or there are music (or unnecessary sounds) in the background, which can be discarded from the corpus.

Style of expression: The style of expression of opinions are different from person to person. For example, a person can express their opinion more visually, and another person may increase their voice while they are talking.

1.3 ORIGINAL CONTRIBUTIONS

In this thesis, there are two major contribution; (i) Persian text sentiment analysis and (ii) Persian multimodal sentiment analysis. The major contributions of this thesis can be summarized as follows:

1.3.1 *Persian Text-based Sentiment Analysis*

1. **Persian lexicon-based sentiment analysis**: In order to understand and develop SA in Persian language, there was a need for Lexicon in Persian Language. In this work, the Persian lexicon (PerSent) is proposed that consists of 1500 words along with their part-of-speech (POS) tag and their polarity.

2. **Dependency-based rules for Persian sentiment analysis**: A novel approach has been designed for extracting concepts from Persian sentences based on the dependency-based relation, which extract concepts using dependency parser. When compared with the state-of-the-art approaches, the accuracy of the proposed scheme is higher.
3. **Integrating Dependency-Based Rules and Deep Neural Networks for Persian Sentiment Analysis**: The novel hybrid approach proposed using deep learning classifiers such as Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) that enhanced the detection of polarity in Persian text sentiment analysis.
4. **Deep Learning Driven Sentiment Analysis**: The deep learning classifiers such as CNN, LSTM and Bidirectional LSTM were used for automated feature engineering and detecting polarity in Persian sentences. The experimental results shows impressive performance in sentiment analysis in comparison with manually engineering features such as POS tags (adjective, noun, verb and adverb) and ngrams.

1.4 PERSIAN MULTIMODAL SENTIMENT ANALYSIS

1. **Persian Multimodal Dataset**: In this thesis, the first Persian multimodal dataset is proposed which consists of 87 videos from YouTube. The dataset consists of product, movie, TV series, book and music reviews.
2. **Combining audio, visual and textual cues for multimodal sentiment analysis**: The trimodal using text, audio and video is developed to detect polarity in Persian sentences. The experimental results showed that the trimodal outperformed unimodal or bimodal.

This thesis is organised as follows.

- Chapter 1 introduces the thesis. Major contributions and motivation are outlined in Chapter 1.
- Chapter 2 discusses, the background related to SA.
- Chapter 3 reviews the recent literature on text and multimodal sentiment analysis
- Chapter 4 presents a lexicon-based method for extracting sentiment from Persian texts. The method uses a dictionary of tokens, called PerSent, annotated with their semantic orientation (i.e. polarity and strength). PerSent is applied to the polarity classification task, the process of categorizing a source text into positive or negative by capturing the text's overall opinion towards the main subject matter. Additionally, we describe the process of PerSent lexicon creation.
- Chapter 5 give details of the proposed novel context-aware approach using deep learning approaches on Persian movie and hotel reviews and the results has been compared with traditional classifiers such as SVM and MLP.
- Chapter 6 presents dependency-based rules for Persian sentiment analysis, that consider the dependency relations between keywords, the word order and, individual word polarities to address the issues with word co-occurrence frequencies based approaches. In addition, a novel hybrid framework for Persian sentiment analysis is proposed, that integrates dependency-based rules, and deep neural networks to address the limitation of unclassified sentences, associated with dependency-based rules.
- Chapter 7 focuses on multimodal sentiment analysis by integrating the text-based sentiment analysis approaches described in Chapters 3 to 5

with audio and visual features. We also discuss the development of deep learning driven multimodal fusion framework for cross modal sentiment analysis, incorporating audio, visual and textual cues.

- Chapter 8 concludes the thesis by summarizing the proposed symbolic and sub-symbolic approaches, explaining the key benefits and limitations of this research. We also propose a number of possible research directions to address the limitations of our current research.

2

CHAPTER 2: BACKGROUND

This chapter presents background information such as data pre-processing, feature selection, sentiment classification techniques, classification techniques, evaluation classification techniques, audio modality, visual modality, and then we discuss about the state-of-the-art of multilingual sentiment analysis and finally we discuss about multilingual multimodal sentiment analysis approaches.

2.1 DATA PRE-PROCESSING

This section discusses data pre-processing and feature selection. Data pre-processing is a key stage process in sentiment analysis for classification of text. Text can contain a lots of noise such as tags and advertisements. Data pre-processing is used to reduce noise and improve performance of the classification. The data pre-processing include different tasks such as tokenisation, normalisation, removing stop words, stemming and handling of negation [20].

2.1.1 *Tokenisation*

The process of breaking text into words and phrases is called tokenisation. The text is a linear sequence of symbols, after data pre-processing, the text is required to be segmented into linguistic units such as words, punctuation, numbers, alpha-numerics, etc. This process is called tokenization. In Persian, the words are often separated from each other by white spaces. The tokenisation is types pre-processing in sense, the identification of basic unit to be processed. The conventional to concentrate on analysis while taking unit is granted [21].

Tokenisation is the process of breaking the text into tokens consisting of words, phrases, symbols or other types of meaningful elements. For example, "I went to the cinema" will break into a list of tokens such as "I", "went", "to", "the", "cinema". Persian example, من به شَام دعوت شدم (I have been invited for dinner), it transform into من, به, دعوت, شدم [22].

2.1.2 *Normalisation*

Text normalisation is the process of converting text into canonical form. In this stage the tokens are transferred into normal form. For example, the sentence "The movie was perfeccccccccct" becomes "The movie was perfect". In the Persian example, شَام عَاَاَالي بود ("The dinner was greattttttttt") it becomes شَام عَالي بود, ("The dinner was great") [23].

2.1.3 *Stop-words*

The stop-words (such as "all", "the", "a", "and") are the most common words which can be removed easily without affecting the performance of the sentiment analysis system. Typical, stop-words in Persian are و, (and), بَا (with), به (to), etc. [24].

2.1.4 *Stemming*

Stemming is the process of reducing inflected words and changing the words to their root forms. For example, the word "stemming" will be changed into "stem" or the word "fishing" will be changed into "fish" (Savoy, 2006). In Persian, the word کشورها (countries) will be changed into کشور (country) or the word درختان (trees) will be changed into درخت (tree) [25]. There are mainly two types of error in stemming. (1) over stemming is when two words with different stems are stem in same root which is known as false positive (2) under stemming when the words that should be stemmed to the same root but it would not stemmed, this is also known as false negative [26]. In stemming, the translation of morphological form a words to stem by assuming the semantic relation between the words. There are two points are considered while using stemming technique:

- The words which do not have any meaning should be eliminated.
- Morphological forms of a word should have same base meaning and it should mapped to the same stem. These rules are good and sufficient in text mining or language processing applications. Stemming is usually considered as recall-enhancing device. For language such as Persian the power of stemming is less than those with more complicated morphology [27].

2.1.5 *Negation*

One of the most important tasks in sentiment analysis is the control of negation. For example, "I like the Rio movie" and "I do not like the Rio movie"; both sentences are very similar, but they are in fact the opposite of each other. A negation term (such as "not") can change the meaning of the sentence in the above examples. Negation can be handled directly or indirectly. Indirect negation can be used as a second feature and a feature vector can be used as the initial presentation, whereas direct negation is encoded. For instance, assigning the word "not" to negation words or phrases such as "no" or "don't"; For example, "I don't like classic movies" so the word "like" is changed to "not like". In the Persian language, the negation in the Persian sentences is located at the end of the sentence. For example, من فیلم دوست ندَارم ("I do not like this movie"). The word ندَارم means "is not" which is located in the end of the sentence [28].

2.2 FEATURE SELECTION

Feature selection can be divided into the following categories:

2.2.1 *N-gram*

An n-gram is a contiguous sequence of n items from a given sample of text. Usually, the stop-words are removed and then the unigram and bigrams are identified in the training dataset. For example, "I bought a new car", the unigrams are "I", "bought", "new", "car". In Persian example, فیلم خوبي بود (It was good movie), the unigrams will be فیلم خوبي and بود [29].

2.2.2 *Parts-of-Speech Tagging (POS)*

The parts of speech tagging is used in sentiment analysis to identify the type of words that can be used for disambiguation. In most languages, adjective can be useful to identify sentiment and helpful for feature selection during sentiment analysis [30]. For example, the phrase "Very nice dinner". The POS tag, “Very” is the adverb, “nice” is the adjective and “dinner” is the noun. For example, the Persian sentence من به سینمَا رفتم (I went to the cinema) will have the POS tag: من is the noun; به is the determiner; سینمَا is the noun; and رفتم is the verb.

2.2.3 *Syntactic Features*

This category uses only one feature from all of the available features such as N-grams, part of speech tag, phrase pattern and punctuation. In the phrase pattern, “n+aj” shows a noun with a positive adjective, and “n+dj” shows a noun with a negative adjective.

2.2.4 *Semantic Features*

This feature is used to show the relation between sign, symbols, words and phrases. This feature can be used to recognize the expression in the language.

2.2.5 *Link-based Features*

This feature is used to classify relation and the link between relations. It is used to identify the sentiment for documents or online views. The online view can be similar to each other or sometimes they contain similar sentiment. For example, when people are discussing different issues such as war, people have negative opinions about war [31].

2.2.6 *Term frequency (TF)*

Term frequency (TF) is the number of occurrences of an item (such as a word or n-gram) in a given document. It is often used in combination with inverse document frequency (logarithm of the inverse of the share of the documents in the collection that contain the given term) in the form of the TF-IDF feature. Mutual information (MI) is used to measure the dependence between two different variables. Mutual information is used in statistical language modelling [32] [33].

2.2.7 *Stylistic Feature*

This feature uses symbols to pass on messages to people. This symbol explains the situation of the person and uncovers the hidden meaning which can improve the performance of the sentiment. The stylistic feature is used more often compared to others [34].

2.2.8 *Word Embedding*

This technique is used to identify the contextual information in a low dimension vector. Each word is represented by small dimension, range from 300 to 1000. The vectors for each word is initialised and each and iterated in the whole corpus. The advantage of word embedding is low dimensions while keeping all the information of words.

2.3 Sentiment classification techniques

Sentiment classification can be divided into different approaches such as lexicon-based approach and machine learning approach.

2.3.1 *Lexicon-based approach*

A lexicon-based approach is used in most forms of sentiment analysis to express positive and negative opinion. The collection of terms and phrases is called a lexicon. A lexicon can divided into corpus-based and dictionary-based approaches as outlined below:

2.3.1.1 *Corpus-based approach*

A corpus is large body of natural language text used for accumulating statistics on natural language text. The corpus provides information about different words such as tags, part-of-speech and parse tree for each of the sentences [35].

2.3.1.2 *Dictionary-based approach*

Dictionary based approaches are used to extract positive, negative and neutral opinions from sentences. Different studies (for example, MPQA lexicon, WordNet and SentiWordNet) have employed dictionary-based approaches. The advantage of this approach is that once are created, they do not require any training data. However, they are designed as a dictionary of words based on WordNet and hence, they do not contain any technical terms or colloquial expressions even context-dependent expressions cannot be used as they have negative impact on accuracy [36].

SentiWordNet SentiWordNet is an important lexical resource for sentiment analysis which classifies tokens into three scores namely positive, negative and neutral. These three scores range within -1.0 to 1.0. The scores assigned to words indicate the term's positivity, negativity and neutrality [37, 1]. The Fig 2.1 presents the example of SentiWordNet.

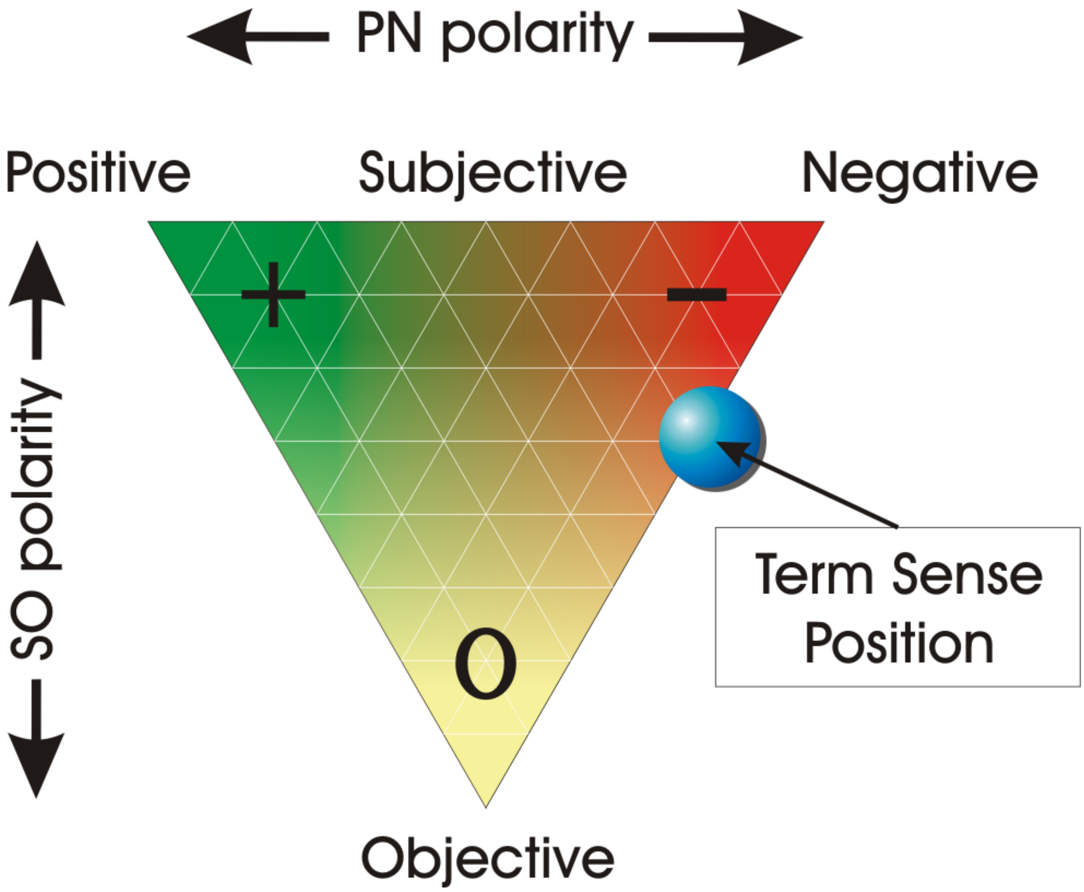

Figure 2.1: SentiWordNet [1]

SenticNet SenticNet is a new multi-disciplinary approach used to identify, interpret and process sentiment. The SenticNet is used for sentiment analysis and also to evaluate text based on common sense reasoning tools. These tools require large inputs and are not able to assess text with acceptable granularity [3]. The example of SenticNet can be seen in table 2.1

WordNet-affect WordNet-affect is a small lexical resource database which contains details about the emotions of words and phrases. Synsets denote the affective concepts. Affective concepts are the words which contain emotion.

Table 2.1: SenticNet [3]

Concept	Polarity
A lot	0.258
Abhorrent	-0.443
Able read	0.865
A little	0.032
Able run	0.775
Abandon	-0.566
Abroad	0.255

Some of the words express their emotions directly, such as sad, happy, etc. while some other words explain emotions implicitly like attitude, behaviour, physical states, etc. WordNet-affect uses six different emotions such as joy, anger, fear, sadness, disgust and surprise to classify words and phrases according to the emotion associated with them [38].

MPQA Multi-Perspective Question Answering (MPQA) is a subjective lexicon consisting of eight thousand terms were collected from different sources. The MPQA consists of different words with their parts of speech (POS) tags and also their polarity, such as positive, negative or neutral [39].

WordNet WordNet is a lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into synsets. They are linked by conceptual semantic and lexical relations. WordNet is a tool which is available for free to download and it's useful for computational linguistic and also natural language processing [40].

2.3.2 *Machine Learning Approaches for Sentiment Analysis*

English natural language processing has carried out significant research as compared to Persian language. Recently, Persian language has received more attention and several publication have been developed for text classification. However, developing machine learning approaches for Persian language requires more effort as compare to other languages, the sentence structure and removal of diacritics that represent the vowels. The aim of using machine learning is enable computers to learn without special programming. The machine learning has successfully applied to complex task such as natural language processing. In addition, the range of machine learning algorithm provide promising results. Machine learning consists of statistical techniques and algorithms that is able the generation of structures in the learning stage based on the training data in order to identify results for data in the testing stage. The learning stage contains optimizing a numerical measure to calculate the parameters which characterize a algorithm underlying model.

The machine learning approach is based on the various machine learning algorithm to identify the polarity of the text, audio and video. The machine learning used various features. Machine learning consists of different techniques such as supervised, semi-supervised and unsupervised:

2.3.2.1 *Supervised*

Supervised learning uses prior labelled data sets for prediction and classification tasks. Most of the research in sentiment analysis uses supervised learning techniques such as support vector machine, Naive Bayes, neutral network and decision trees [8].

2.3.2.2 *Semi-Supervised*

Semi-supervised learning is a new method which uses both labelled and unlabelled data to perform sentiment classification [8].

2.3.2.3 *Unsupervised*

An unsupervised learning method contains unlabelled data alone. This can be used for classifying texts at the word- or sentence level and also in feature-based opinion mining [41].

2.4 MACHINE LEARNING CLASSIFICATION ALGORITHMS

In this section we discuss about some important classifiers used in this research.

2.4.1 *Support Vector Machines (SVM)*

The support vector machine (SVM) is a algorithm which is used to assign label to the object. For example, the SVM is used to identify the polarity of the text by evaluating thousand sentences of the text. Alternatively, an SVM can learn to identify to recognise the handwritten digits by evaluating a large collection of scanned images of handwritten zeros, ones and so forth.

Support Vector Machines (SVM) is a supervised learning model that uses trained data for classification. The objective of the SVM is to find hyperplane a N-dimensional which classify the data points. The hyperplanes are decision boundaries which is used to classify the data into positive and negative. The data point can be falling to either side of the hyperplane which can be attributed to different classes. In addition, the dimension of the hyperplane is depends on the number of features. In our case, because we concentrated on two features (positive or negative), the hyperplane is a line. If the input features is more than three the hyperplane is two-dimensional plane [42].

The main idea behind the support vector machine is how to divide the space with decision boundary between data points. The w denote the vector per perpendicular to median of decision boundary, u is unknown vectors, b is constraint. The equation 2.1 and 2.2 shows the positive and negative equations respectively:

$$w.u + b > 0 \tag{2.1}$$

$$w.u + b < 0 \tag{2.2}$$

2.4.2 *Naive Bayes (NB)*

Naive Bayes (NB) classification is based on the Bayes Theorem. NB is categorised under supervised machine learning and it is used to classify of datasets. Additionally, can be used to predict data based on prior knowledge and assumption.

Naive Bayes is condition probability model, given problem isolation to be classified represent by vector x = c1, ..., c2 represent n-feature it assign to instance probabilities. The x denote the input train features and c is class:

$$P(ck|x) = \frac{P(ck)P(x|ck)}{P(x))} \tag{2.3}$$

The above equation can be written as:

$$Posterior = \frac{Posteriorxlikelihood}{evidence} \tag{2.4}$$

The Bayesian rule can be extended to get the following formula:

$$P(C|x1, x2, ..., xn) = \frac{P(x1, x2, .., xn)P(C)}{P(x1, x2, ..., xn)} \tag{2.5}$$

2.4.3 *Logistic Regression (LR)*

Logistic Regression (LR) is type of machine learning algorithms which is used for binary classification. The LR is used to find best fitting model to describe relationship between characteristic of interest and set of independent variables. In order to predict the values, the sigmoid function is used. The function maps between any real value into another value between 0 and 1. In machine learning, the sigmoid function is used to map predictions of probabilities.

$$h_\theta(x) = \frac{1}{1 + e^{-}\theta^T x}$$

h denotes the output between 0 and 1, and e is based on natural log.

2.4.4 *Convolutional Neural Network (CNN)*

The convolutional layers perform convolutional operation to input layers and send it to next layers. The fully connected layers can be used to learns features. In addition, classify the data is requires large number of neurons. Fully connected layer is used to connect a single neuron to every neuron for the next layer [43] [44] [45] [46] [47]. The pooling layer is consisting of local and global a pooling layer which is able to combine the output of one layer into a single neuron [48] [49]. There are set of weight between each layers, the layers map other layers using matrix multiplication. Another layers is available in neural network which is called convolutions layer. The operation called convolution:

$$s(t) = \int x(a)w(t-a)da \tag{2.6}$$

The convolutional operation is typically denoted with an asterisk:

$$s(t) = (x * w)(t) \tag{2.7}$$

If we assume x and w are defined only on integer t, we can define discrete convolution:

$$s(t) = (x * w)(t) = \sum_{\alpha=8}^{\alpha} x[a] \tag{2.8}$$

when (t - a) we flip the signal and then shift it, w and x are signals. The example of CNN classifier is show in Fig 2.2:

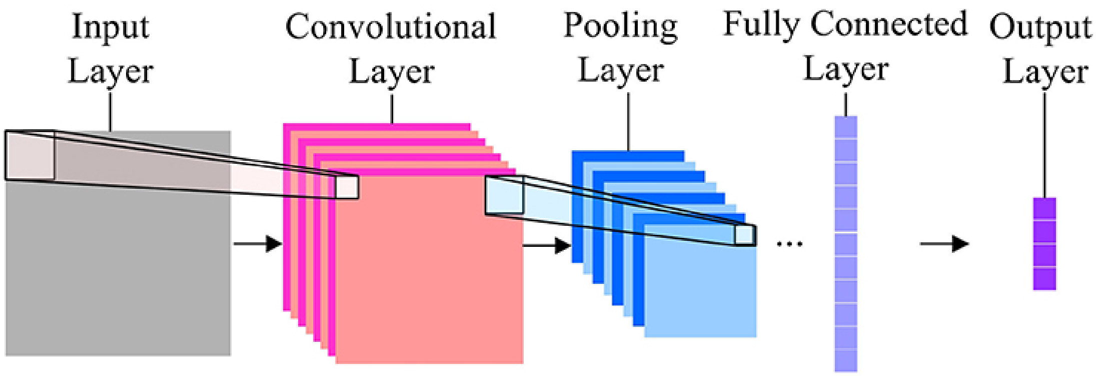

Figure 2.2: CNN Classifier[2]

2.4.5 *Long Short-Term Memory*

The LSTM consists of different gates input, forget and output, cell memory and activation vector. LSTM control the flow of information using input and output gates. The input gate is active to accept the signal in and output gate is used for signal out. The forgot gate is used to reset the cell's own state [50]. The LSTM is model which is able to calculate the hidden vector h = { $h_1, h_2, ...h_n$ } with sequence of input X = { $x_1, x_2, ...x_n$ } and provide the output vector Y = { $y_1, y_2, ...y_n$ } . The output layer is controlled by gates which is function to hidden layer and input as current time x_t , the forget gate f_t , the input gate i_t and output gate o_t [51]. These gates are used to decide the transition of the memory cell and current hidden state. The graph below displays the gates for LSTM. The LSTM function is defined as follows:

$$i_t = \sigma \ (W_i \ . \ \ [h_{t-1}, \ x_t] + \ b_i) \tag{2.9}$$

$$f_t = \sigma\ (W_f\ .\ \ [h_{t-1},\ x_t] +\ b_f) \tag{2.10}$$

$$f_t = \sigma\ (W_f\ .\ \ [h_{t-1},\ x_t] +\ b_f) \tag{2.11}$$

$$l_t = \tanh\ (W_l\ .\ \ [h_{t-1},\ x_t] +\ b_l \tag{2.12}$$

$$o_t = \sigma\ (W_o\ .\ \ [h_{t-1},\ x_t] +\ b_o) \tag{2.13}$$

$$c_t = f_t \odot c_{t-1} +\ i_t \odot l_t \tag{2.14}$$

$$h_t = o_t \odot \tanh\ (c_t) \tag{2.15}$$

σ is denotes the sigmoid function. Tanh is tangent function which provides output of [-1,1], $\odot$ is component wise multiplication. The old memory is controlled and discarded by f_t and i_t and they are used to control information which is stored in the new memory [52][50].

2.5 EVALUATING CLASSIFICATION TECHNIQUES

In order to evaluate the performance of classification techniques, the following four evaluation metrics are used: precision, recall, f-measure and accuracy. Precision is the fraction of correctly predicted positive observations to the total predicted positive observations. Recall is the fraction of correctly predicted positive observations to the all observations in actual class. F-measure is the weighted average of Precision and Recall, and Accuracy is simply the fraction of correctly predicted observation to the total observations.

$$Precision = \frac{TP}{TP + FP} \tag{2.16}$$

$$Recall = \frac{TP}{TP + FN} \tag{2.17}$$

$$F\_measure = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{2.18}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2.19}$$

where TP denotes true positive, TN is true negative, FP is false positive, and FN is false negative.

3

Chapter 3 Review of Multilingual Multimodal Sentiment Analysis

This chapter presents the state-of-the-art of multilingual sentiment analysis and finally we discuss about multilingual multimodal sentiment analysis approaches.

3.1 Multimodal Sentiment Analysis

In this section, the audio and visual modality will be discussed.

3.1.1 *Audio Modality*

Similar to text, the audio can express sentiment. The previous research showed that acoustic parameters cannot be changed through oral variations and depend on the personality traits. There are different research carried out to find the best feature for audio. However, the research showed that the pitch and energy are the most effective features. Some of the important audio features are described briefly below:

- Mel Frequency Cepstral Coefficients (MFCC) is short-term power spectrum for a sound which approximate the human auditory system is closely than any other available frequency band distribution.
- Spectral centroid is used to show the mass of the magnitude spectrum which can provides the brightness of a sound.

- Spectral flux is used to show how the power of spectrum of a signal is changing. The Euclidean distance is used to calculate the distance between two normalised spectra.
- Pause is the duration of the time speaker is silent in the audio.

3.1.2 *Visual Modality*

Facial expression gives primary cues to understand people's sentiment. The facial expression is first channel to present state of mind. Ekman et al. [53] mentioned it is possible to detect different emotions, such as sadness, anger, surprise, joy and disgust.

3.2 THE STATE-OF-THE-ART OF MULTILINGUAL SENTIMENT ANALYSIS

In the following subsections, the state-of-the-art approaches to sentiment analysis using English, other language and Persian language are discussed.

3.2.1 *English language*

Shi et al. [54] developed a supervised machine learning technique for sentiment analysis of online hotel reviews in English by using unigram features. They used term frequency and TF-IDF to identify the polarity of the document. The SVM classifier was chosen because it was reported to perform better than other classifiers [55], though Tong et al. [56] show that Naive Bayes and SVM are the most effective classifiers among machine learning techniques. The hotel-review corpus contained 4000 reviews; the reviews were pre-processed and tagged as positive and negative. Then, the obtained sentiment classification model was used to classify live information flow into positive and negative documents. The TF-IDF feature performed better than simple term frequency [54].

Another study [31] used supervised classification for identification of sentiment in documents. The method used sentences found in on web sites, in particular, blogs, forums and reviews [57]. After pre-processing, different features such as unigrams, stems, negation and discourse features were selected. The SVM, maximum entropy and Naïve Bayes classifiers was employed as machine learning algorithms. For linearly separable data, SVM achieves classification results with minimal error. The Naïve Bayes classifier is very simple to use for efficient classification with incremental learning [58]. A maximum Entropy classifier is efficient in extracting information that leads to good results [59]. English-language corpora were collected from blogs, reviews and forum sites such as livejournal.com or skyrock.com. The Maximum Entropy classifier gave 83% accuracy, which is better than other classifiers used in this study, namely, SVM and Naïve Bayes; however, other approaches [54] used SVM to evaluate datasets, and other machine learning techniques were reported to have accuracy lower than that of SVM.

Singh et al. [37] used the unsupervised semantic orientation with POS tagging on the Cornell movie review dataset; Feature extraction was done for all reviews. The semantic orientation was calculated for reviews, adjectives were extracted and the semantic orientation value was assigned to them. Aggregation was done for semantic orientation; for each positive term based on SentiWordNet +1 was added to the document score and for each negative term, −1. Thus, the semantic orientation of each review was the total semantic orientation values for the extracted terms. Then, a threshold of 5 on the absolute value of the score was used to classify a document as positive or negative basing on the aggregation score. The features were extracted, and then SentiWordNet was employed to check the scores for the selected features.

Two different datasets were used; one dataset contained one thousand positive and one thousand negative reviews, and another dataset contained seven hundred positive and seven hundred negative reviews.

Mizumoto et al. [60] introduced an unsupervised approach to identify sentiment polarity of stock market news. The polarity of the sentiment for stock market news was identified using a polarity dictionary that contained words and their polarities. In this method, for a small amount of words, polarity was determined manually. The polarity of new words was then identified automatically. The new dictionary method was built for unlabelled news. The dictionary contained a small number of words with their polarities such as positive and negative words. If a word was situated in one sentence with both positive and negative words, the co-occurrence of frequency for negative and positive polarity was calculated. The following equations were used:

$$rP(w) = (freq(w, wP))/(freq(w, wP) + freq(w, wN)) \tag{3.1}$$

$$rN(w) = (freq(w, wN))/(freq(w, wP) + freq(w, wN)) \tag{3.2}$$

The rP and rN defines as co-occurrence of positive and negative words, freq(w-1,w-2) denotes frequency of word 1 and word 2. wP and wN is positive and negative polarity dictionary. The bias of co-occurrence was measured; most of the words were occurring with positive and negative polarities; the rate of co-occurrence of positive and negative polarity of dictionary was used, and then the polarity of those words that were not added was estimated. Finally, the polarity of words was determined. Two different thresholds were introduced, namely, thresholdP and thresholdN. The thershold P value was used to add words to the positive polarity dictionary, and thresholdN was used to add words to the negative polarity dictionary. The threshold values varied from 0.5 to 1. Words with occurrence frequency lower than ten were excluded as not reliable. An online stock market news dataset was used for evaluation. It contained 62,478 news items. A polarity dictionary was built automatically with a semi-supervised technique. The method assigned 45% of correct polarity values for all news items.

Prabowo et al. [61] proposed an ensemble approach to detect polarity in documents. To optimize the performance, different classifiers such as SVM and Naïve Bayes are trained. Experimental results showed that the ensemble approach (89.23%) outperforms SVM (78.78%) and Naïve Bayes (74.23%). Thet et al. [62] proposed an approach to detect polarity in movie reviews. The SentiWordNet is used to assign polarity to extracted features (ngram features), and finally a movie reviews dataset (collected manually) is evaluated using SVM classifier achieving an accuracy of 67.54%. Poria et al. [63] proposed an rule-based approach to extract keywords from product reviews. The Amazon product reviews dataset is used to evaluate the performance of the approach, showing the SVM achieves a better accuracy (91.2%) as compared with the Naïve Bayes (89.25%). Chikersal et al. [64] proposed an approach based on a rule-based and SVM classifier to detect polarity in tweets. The n-gram and POS tag features are extracted from sentences and the SVM classifier is trained. This approach achieves an accuracy of 72.5%. Poria et al. [65] proposed a rule-based approach to extract keywords from sentences, in which SenticNet is used to assign polarity to the extracted keywords. An Extreme Learning Machine (ELM) classifier trained on Amazon product reviews dataset achieves an accuracy of 71.32%.

In addition, Ouyang et al.[66] proposed a framework to detect polarity in English movie reviews using a deep learning classifiers. The movie reviews were collected from rottentomatoes.com. The dataset consisted of five labels: positive, somewhat positive, neural, somewhat negative and negative. Experimental results showed that the RNN achieved an accuracy up to 78.34% as compared to SVM (62%). Chen et al. [67] proposed a novel framework to improve sentence-level sentiment analysis by employing BiLSTM. The comparative simulation results with benchmark datasets showed that their proposed framework improved the overall accuracy of sentence-level sentiment analysis. Appel et al. [68] proposed a hybrid approach to detect polarity in more than one million tweets collected manually. The approach contains three different classifiers: SVM, Naïve Bayes and maximum entropy techniques. The hybrid approach combined different lexicon such as SentiWordNet and SenticNet, thus further enhancing the performance of the approach. Experimental results showed that the hybrid approach achieves better accuracy (87.27%) as compared to other machine learning classifiers such as SVM and Naive Bayes. However, the proposed approach cannot be applied to different languages such as Persian.

On the other hand, Hassan et al. [69], proposed an approach to detect polarity using CNN and LSTM using pre-trained word vectors of IMDB movie reviews. The CNN classifier developed in this approach consisted of two convolutional layers and two pooling layers. The experimental results showed that the combined CNN and LSTM model achieved up to 88.3% accuracy and outperformed CNN and LSTM models. Similarly, Shen et al. [70] proposed a novel approach to combine CNN and bidirectional Long Short-Term Memory (BiLSTM) to detect polarity for movie reviews. The combined CNN-LSTM model outperformed CNN and LSTM classifiers by achieving accuracy up to 89.7% as compared to 83.9% and 78.4% respectively. Nguyen et al. [71] proposed a novel method to detect polarity in news articles using a deep learning classifiers. The authors used different news websites to collect more than one million news articles and fed the preprocessed embedding into CNN, LSTM and convolutional LSTM (CLSTM). The comparative simulation results showed that the CLSTM model outperformed CNN and LSTM classifiers by achieving accuracy up to 96.52% as compared to 92.3% and 91.19% accuracy achieved by CNN and LSTM respectively. Similarly, Liao et al. [72] proposed a framework to understand users' satisfaction of a product using deep learning classifiers. The CNN classifier developed in this framework contained one convolutional layer and one pooling layer. Their experimental results showed that the CNN outperformed SVM and Naive Bayes classifiers by achieving the accuracy up to 95% as compared to 70% and 62.25% accuracy achieved by SVM and Naive Bayes respectively. Manek et al. [73] proposed an SVM classifier based on selected features to detect polarity in large movie review dataset. The terms and their frequency, part of speech, negation and syntactic dependency are extracted. Experimental results showed that the proposed model outperforms the K-Nearest Neighbors (KNN) algorithm and the Naive Bayes model, with accuracy of 96.95 %, 92.68% and 88.54% respectively.

Recently, Zhao et al. [74] proposed a model to detect sentiment in product reviews using deep learning classifiers. The words are first converted into an embedding vector representation and then a deep learning classifier is used to classify sentiments. The experimental results showed the effectiveness of their proposed CNN (87%) as compared to LSTM (82%). Sohangir et al. [75] proposed a framework to detect sentiment in the financial market. The StockTwits dataset was used, and feature selection method such as chi-square was used to select features. Finally CNN classifiers were used to evaluate the performance of the approach. The proposed framework obtained an accuracy of 70.88%. Wagh et al. [76] proposed an approach to identify the sentiment in tweets and twitter API utilized to collect the tweets. Different classifiers such as Naive Bayes and Logistic Regression were used to evaluate the performance of the approach. Experimental results showed that logistic regression achieved better accuracy as compared to Naive Bayes. Pitsilis et al. [77], proposed a framework to discern hateful content in social media using LSTM classifier. The authors collected data from Twitter and fed extracted features into an LSTM classifier. The results demonstrated the effectiveness of LSTM as compared to SVM.

More recently, Rathore et al. [78] proposed an approach to collect movies reviews from a movie website and used sentiment analysis to detect polarity. Twitter movie reviews were used to evaluate the performance of the approach. Experimental results showed that Naïve Bayes achieved better accuracy as compared to Multilayer perceptron (MLP). However, the proposed approach does not contains any simulation results to identify the performance of the proposed approach on different application.

Table 3.1 summarizes some of the approaches used for English and their respective achieved accuracy.

Table 3.1: Summary of sentiment analysis approaches for English languages

Ref	**Purpose**	**Approach**	**Accuracy**
Sohangir et al. [75]	Detect sentiment in the financial market	CNN	70.88%
Wagh et al. [76]	Detect polarity in twitter	Naive Bayes Logistic Regression	68.3%
Chikersal et al. [64]	Detect polarity	SVM	72.5%
Poria et al. [65]	Extract keywords	ELM	71.32%
Ouyang et al.[66]	Detect polarity in movie reviews	RNN, SVM	78.34%
Nguyen et al. [71]	Detect polarity in news articles	CNN, LSTM CLSTM	96.52%

3.2.2 *Other languages*

In this section, we discuss the sentiment analysis approaches in Arabic, Chinese, Spanish, Korean, Swedish, Brazilian and Indonesian language.

3.2.3 *Arabic Language*

Farra et al. [79] proposed an approach to detect polarity in Arabic product reviews. A POS tagger is used to extract features from Arabic sentences, and SentiStrength is used to assign polarity to the extracted features. Experimental results showed that the SVM outperforms the decision tree. Al-Ayyoub et al. [80] proposed an unsupervised approach to sentiment analysis of Arabic tweets. This approach included two stages: The first stage was collecting and pre-processing the tweets. The pre-processing step included stop-word removal and stemming. The second stage was the development of a sentiment lexicon, with the sentiment scores in the range between zero and one hundred. Scores from zero to forty corresponded to negative sentiment, forty to sixty to neutral, and sixty to one hundred to positive. These values were combined with each other to calculate the sentiment value of the sentence. The disadvantage of this approach is that it is not able to handle different Arabic dialects.

Baniata et al. [81] proposed a sentiment analysis method to identify the polarity for Arabic text using a deep learning classifiers. The combination of CNN and LSTM classifier, trained on Arabic reviews, achieved a high accuracy of 86.43% in comparison with CNN (66.26%) and LSTM (65.34%). Dahou et al. [82] proposed an approach to detect polarity for Arabic text. The authors used a web crawler to build a corpus and trained word embedding to represent words in the corpus. The presented results showed that a CNN classifier outperformed other supervised learning algorithms achieving better accuracy as compared to SVM and Naïve Bayes. However, the main disadvantage of the proposed approach is evaluated on small dataset, hence, the deep learning approaches required large dataset.

Recently, Al-Saqqa et al. [83] proposed an ensemble machine learning classifier to classify Arabic text. This approach is based on the majority voting on different classifiers - Naive Bayes and SVM. The Arabic movie review dataset is used to evaluate the performance of the approach. Empirical results showed that their proposed ensemble classifier is limited to General Arabic and it unable to detect polarity for Arabic dialects such as Algerian dialects. Although, the traditional classifier including SVM and Naive Bayes are used to evaluate the performance of the approach. Therefore, Alayba et al. [84] proposed an approach to construct Word2Vec models from Arabic corpus obtained from ten newspapers in different Arab countries. The experimental results showed that the CNN outperforms the Naive Bayes, and linear support vector. However, the proposed approach is unable to detect negation in the sentences. Recently, Siddiqui et al. [85] proposed an approach to detect polarity in Arabic movie reviews collected manually. After pre-processing, rules are developed to extract keywords from sentences. Experimental results showed that the SVM outperforms the Naïve Bayes.

3.2.4 *Chinese Language*

Zagibalov and Carroll [86] used automated seed words for selection in the Chinese language. The approach did not require word segmentation. The lexical items was used to treat Chinese characters. In order to improve the classifier to find the seeds automatically, two assumptions were used: the first assumption was that the attitude was stated by using negation of word items with their opposite meaning; this assumption was used to find negative lexical items from positive seeds. The second assumption concerned the polarity of seeds that needed to be identified. To identify the polarity of a seed word, the lexicon was used to reach a gold standard for a positive lexical item. The sentiment classification and iterative technique was used in the unsupervised method. The method was used to find seeds automatically from raw text. To find positive seeds from the corpus, a special algorithm was developed. It operated over the sequence of characters that should be checked for containing negation or adverbs. This method did not use pre-segmentation or grammar analysis; the unit of processing is a lexical item. Input sequences of Chinese characters did not include punctuation marks and zone markers. A single zone was classified either as positive or negative and the corresponding scores were calculated [87].

In addition, Zhang et al.[88] proposed an approach to detect sentiment in Chinese articles. The approach determines sentiment based on dependency and aggregating sentences to predict the document sentiment, with the results show that the SVM outperformed decision tree and the Naïve Bayes. Xiao et al. [89] proposed an approach to detect polarity in Chinese text. The authors used 1D and 2D CNN classifiers and achieved highest 93.4% accuracy with 2D CNN. Furthermore, the experimental result showed that the character level approach outperformed word embedding in Chinese words. Shaung et al. [90] proposed an approach using CNN-LSTM to detect polarity in English and Chinese product reviews. Their proposed CNN-LSTM model outperformed individual CNN and LSTM classifiers and achieved 81.86% accuracy. However, the approah is not compared with current traditional classifiers such as SVM and Naive Bayes.

Day et al. [91] proposed a framework to explore the impact of deep learning for sentiment analysis on Google App mobile reviews in Chinese. CNN outperformed SVM and Naive Bayes classifiers. Recently, Zheng et al. [92]

proposed an approach to identify the polarity in Chinese reviews. The n-gram and POS features are extracted, and the experimental results showed the SVM outperforms the Naïve Bayes. The main disadvantage of the proposed approach is extracting simple features from text and there is no significant contribution in the approach.

3.2.5 *Spanish Language*

Martin et al. [93] proposed an approach to combine supervised and unsupervised learning to classify the polarity for Spanish movie reviews. After pre-processing, the SentiWordNet is used to assign polarity to the extracted features, and an ensemble of three machine learning classifiers is used to evaluate the performance of the approach. The experimental results indicated that the ensemble classifier outperforms the SVM, Naïve Bayes and decision tree. Valverde et al. [94] proposed a novel approach to detect polarity for Spanish product services using deep learning classifier. The product reviews were collected from e-commerce websites manually. After the pre-processing, the word embedding was used to obtain the vectors of the words, followed by the CNN classifier. It is to be noted that, the proposed approach is not compared with existing state-of-the-art approaches due to lack of proposed deep learning approaches in sentiment analysis in Persian.

Wehrmann et al. [95] proposed a method to detect polarity in four different languages. However, their proposed method used machine translation to translate English tweets into German, Portuguese and Spanish. The experimental results showed the effectiveness of their proposed model and achieved up to 76.2% accuracy with CNN as compared to LSTM that achieved 64.7% accuracy. Recently, Luque et al. [96] proposed a system to detect polarity in Spanish tweets. The Spanish tweets were collected and then converted into word embedding and then a MLP was applied to evaluate the performance of the approach. However, the proposed approach is not compared with existing state-of-art approaches.

3.2.6 *Korean, Swedish, Brazilian, Indonesian*

Avanco et al. [97] proposed a system to detect polarity in Brazilian Portuguese sentences. In order to evaluate the performance of the approach, a lexicon called SentiLex was developed. Furthermore, book reviews were used to evaluate the performance of the approach. The experimental results showed that the proposed approach using SVM outperformed Naive Bayes. In addition, the proposed approach is not able detect the polarity for different dialects of Brazilian. Le et al. [98] proposed an approach for sentiment analysis for the Indonesian language to detect the polarity in the sentences or documents. The authors used four thousand movie reviews labelled manually.

Ramadhani et al. [99] proposed an approach to detect polarity in Korean twitter sentiment analysis. There are different machine learning algorithms including CNN, SVM and Naive Bayes were used to evaluate the performance of the approach. The experimental results demonstrated that the CNN approach achieved better accuracy as compared to SVM and Naive Bayes. The limitation of the proposed approach is the only short text is preprocessed in only Korean and English.

Li et al. [100] proposed an approach to understand people's opinion about IKEA stores in Swedish. Swedish tweets were collected, and a Swedish corpus was used to identify the polarity of the tweets and SVM and Naive Bayes were used to evaluate the performance of the approach. The main disadvantage of the proposed approach is limited to the Swedish language and it cannot be

Table 3.2: Summary of sentiment analysis approaches for different languages

Ref	**Language**	**Purpose**	**Technique**	**Accuracy**
Sundström et al. [101]	Swedish	Detect polarity in multi domain	CNN	77%
Akhtar et al. [102]	Hindi	Detect polarity	CNN	65.96%
Zhang et al. [103]	Chinese	Detecting polarity in Chinese reviews	SVM	83.92%
Vural et al. [104]	Turkish	Detecting polarity in movie reviews	SVM	75.90%
Abdulla et al. [105]	Arabic	Developing lexicon and corpus	SVM, NB KNN	84.7%
Anta et al. [106]	Spanish	Detecting polarity in tweets	NB	58.52%

applied to different languages. Table 3.2 summarizes some of the approaches used for different languages and their achieved accuracy.

3.2.7 *Persian Language*

Persian uses 32 characters, that cover 28 Arabic characters. Its writing system includes special signs and diacritic marks that can be used in different forms or omitted from the word. Short vowels are not indicated in writing. There are letters with more than one Unicode encoding. Some words have more than one spelling variant. All this increases the number of homographs and synonyms, which presents problems in computational processing of Persian. The Fig 3.1 is example of Persian word [107, 108].

Figure 3.1: Example of Persian word (Farsi)

Saraee et al. [109] proposed a novel approach to detect polarity in Persian movie reviews using n-gram features. Their proposed approach consists of stemming and feature selection. The Naive Bayes classifier was used to evaluate the performance and achieved 82.26% accuracy. However, the main limitation of the proposed approach is not able to detect the polarity of Persian text and its limited to feature selection from the text.

Amiri et al. [110] developed a lexicon to detect polarity for multi-domain product and movie reviews for Persian. The authors collected sentences from peoples communication. The SVM classifier was used to evaluate the performance of the lexicon and achieved an accuracy of 69%. However, the developed lexicon is not able to detect abbreviation form of writing. Sabeti et al., [111] developed a lexicon for Persian. The lexicon achieved accuracy of 81.06% using K-nearest neighbors (KNN). The main limitation of the proposed lexicon is using machine learning classifiers and the deep learning classifiers were not used to improve the performance of the approach.

Alimardani et al. [112] proposed an approach to detect polarity in Persian hotel reviews. The developed lexicon contained more than one thousand words along with their polarity. The proposed approach collected hotel reviews and extracted part-of-speech tag features. The performance analysis revealed that logistic regression outperformed SVM classifier by achieving accuracy up to 85.9% as compared to 82.4% accuracy achieved by SVM. The main limitation of the proposed approach is limited to hotel reviews and the other domain such as product or politic reviews are not employed to evaluate the performance of the approach. Bagheri et al. [113] proposed a feature selection method based on mutual information to extract features from cellphone reviews, and the results showed that the Naive Bayes (69.35%) outperforms the SVM(64.23%). Despite that, the proposed approach is beneficial for small dataset.

Farhoodi et al. [114] proposed an approach to detect polarity in the Persian news, based on the Persian Hamashahri dataset containing Persian headline news from 1980 to 2000. The SVM and KNN are trained. However, the proposed approach is limited to politic domain. Similarly, Vaziripour et al. [115] proposed an approach to detect polarity in Persian tweets. More than one million tweets were collected on politics discussion, and the machine learning classifiers are trained on the data. However, the proposed approach is limited to political discussion and it does not extend to other domain such as product or movie reviews.

Basiri et al. [116] addressed the problem of lack of resources in Persian by developing a lexicon and an automatically labelled sentence-level corpus. A Naive Bayes classifier was trained on the collected corpus to determine the polarity in short sentences. However, the proposed approach is lack of comparison results with the state-of-the-art approaches. In addition, Ebrahimi et al. [117] proposed a method to detect polarity in Persian online reviews based on adjectives extracted from the sentence and translated SentiWordNet lexicon. However, the proposed method is limited to adjectives and it does not exploit nouns, adverbs and verbs that provide extra information on the underlying sentiment. Moreover, Razavi et al. [118] detect polarity based on extracted nouns and adjectives and a Persian lexicon. However, the proposed approach do not exploit the word order and hierarchical semantic dependency.

Aleahmad et al. [119] proposed a method to detect polarity based on n-gram features. Experimental findings showed that fourgram outperforms unigram, bigram and trigram features. On the other hand, Dashtipour et al. [120] proposed deep autoencoder based feature extraction for sentiment analysis.

Table 3.3: Summary of sentiment analysis approaches for Persian language

Ref	Purpose	Approach	Accuracy
Ebrahimi et al. [117]	Develop Lexicon	SVM	80%
Saraee et al. [121]	Detect polarity in Persian movie reviews	Naive Bayes	82.26%
Vaziripour et al. [115]	Detect polarity in Persian Tweets	SVM	70%
Basiri et al. [122]	Detect polarity in Persian product reviews	SVM	68%

The framework outperformed state-of-the-art CNN and MLP for detecting polarity in Persian text. Despite that, the proposed approach is evaluated and test on Persian movie reviews dataset. Table 3.3 summarizes some of the approaches used for different languages and achieved accuracy.

3.3 MULTILINGUAL MULTIMODAL SENTIMENT ANALYSIS

Morency [123] proposed an approach for multimodal sentiment analysis using the YouTube dataset which consists of 47 videos were used to evaluate the performance of the approach. The experimental results showed that the combination of audio, visual and text features can improve the performance of the approach. Text only achieved f-measure 0.43, visual achieved 0.439, audio only achieved 0.419. However, combination of text, visual and audio achieved f-measure 0.553. Poria et al. [47] presents a novel method to extract text features for short text and these method is used to extract feature for multimodal sentiment analysis. The dataset consists of 498 short videos and

Table 3.4: Comparison of Sentiment Analysis Approaches

Method	**Advantages**	**Disadvantages**
Shi et al. [54]	Very simple to implement	Ineffective features
Boiy [31]	Easily extend to other languages	Computationally Expensive
Singh et al. [37]	Useful for large datasets	Computationally expensive heavy PMI calculation
Mizumoto et al. [60]	Generate a dictionary for stock market SA	Only applicable to stock market sentiment analysis
Farra et al. [79]	Good precision	Need in translation affects precision
Wehrmann et al. [95]	Large dataset created, can be used for researchers	Needs further development
Day et al. [91]	Various feature selection	Requires more data
Zagibalov [86]	Extended to multilingual SA	Computationally expensive
Avanco et al. [97]	Extend to other languages	No resources available for multilingual SA
Ramadhani. [99]	Effective feature selection	Requires large dataset

finally, the audio, text and visual features are combined to train the model. The proposed method achieved accuracy of 88.60%.

Poria et al. [124] proposed a method for English multimodal sentiment analysis to identify sentiments used audio, text and video features and there are different techniques are used to combined audio, text and video features. The proposed method evaluated with YouTube dataset. The experimental results showed that the SVM achieved accuracy of 80%. The proposed approach is limited to a single speakers and it unable to identify the overall polarity of video with multiple speakers.

Rosas et al. [125] proposed an approach for Spanish multimodal sentiment analysis. Due to lack of resources, the Spanish videos were collected from YouTube and audio, video and text features are extracted from video. The SVM is used to evaluate the performance of the approach. The experimental results indicated that the combination of text, audio and video features (75%) are achieved better accuracy as compared text (64.94%) and audio (46.75%). The approach is identify the positive or negative polarity of the sentence and it unable to detect the polarity for sentences which is neutral.

Alqarafi et al. [126] proposed an approach to detect polarity in Arabic videos. There are 40 videos were collected from YouTube and manually transcribed. In order to evaluate the performance of the approach, there are text features such as ngram and visual features (smile, frown, head nod, and head shake) are extracted. The experimental results showed that the combination of visual and text features achieved accuracy of 58.42%. It is to be noted that, the approach did not consider the acoustic features and relies on text and visual features.

3.4 CONCLUSION

In this chapter an overview of state-of-the-art sentiment analysis methods were provided. We described data pre-processing, typical features and the main resources used for sentiment analysis. Then, we discussed different approaches applied by researchers to English, Persian and other languages.

Most of the aforementioned studies in this chapter exploit word co-occurrence frequencies and a lexicon to determine polarity of source text. These generally fail to exploit hierarchical semantic relations, and word order. In addition, current studies that use lexical rules to detect negation in various languages cannot be directly applied for Persian. Moreover, deep neural networks have shown state-of-the-art performance using a large supervised corpus for various natural language processing tasks including sentiment analysis. However, for scarce resource languages, we developed a lexicon for Persian consists of words along with their polarity and part-of-speech tag. In addition, we need innovative methods that jointly exploit deep learning models and dependency based rules approaches to go beyond current state-of-the-art performance. Therefore, we propose a novel hybrid framework for Persian sentiment analysis, that integrates dependency-based rules, and deep neural networks to improve the overall performance and robustness of polarity detection in real noisy data. However, none of these studies have explored Persian multimodal sentiment analysis. Therefore, we proposed a novel approach to detect polarity in text, audio and visual features. In next chapter, we discuss about proposed Persian lexicon.

4

CHAPTER 4 LEXICON-BASED SENTIMENT ANALYSIS

In recent years, people all around the world have been able to share their opinions on different topics over the Internet. This has resulted in a large amount of unstructured data being available online in different languages which is potentially very useful for companies and organisations as it could be used to improve their products and services [127]. Sentiment analysis (SA) is used to automatically classify the data into sentiment polarity. [128]. Sentiment polarity can be either binary (e.g. positive, neutral or negative) or can be multi-class such as strongly positive, positive, neutral, negative, and strongly negative. Most research has focused on the binary polarity classification, though identifying at least the neutral opinion in the sentence can be helpful [129].

In the recent years, SA has been a very active area of research. There have been numerous lexical resources and datasets compiled for the English language. However, very limited efforts has been made to the development of lexicons in other languages. Therefore, due to lack of available lexical resources, it is difficult for researchers to analyse text in other languages [130]. Particularly, there is no well-known dataset or lexicon available for Persian language [131]. In this chapter, the PerSent was presented, a Persian polarity lexicon for SA, which contains words and phrases along with their polarity and part-of-speech (POS) tag. In order to evaluate its quality and performance using POS-based features, the frequency of sentiment words, average polarity of words were used, and two machine learning algorithms:

Table 4.1: Examples from our Persian sentiment lexicon

Word	Translation	POS	Polarity	Score
خوب	good	adjective	positive	0.7
بد	bad	adjective	negative	-0.69
زشت	ugly	adjective	negative	-0.7
آزردن	annoying	verb	negative	-0.1689

support vector machines (SVM) and Logistic Regression (LR) used to evaluate the performance of the lexicon.

This remaining chapter is organized as follows. Section 4.1 presents PerSent, a Persian sentiment lexicon; Section 4.2 describes evaluation methodology, Section 4.3 provides the experimental results, Section 4.4 gives discussion of the results and Section 4.5 concludes this chapter.

4.1 PerSent: Persian Sentiment Lexicon

Many researchers highlighted that the main problem of multilingual SA is the lack of tools and resources. Classification of text into positive and negative is, however, a difficult task, because most of the reviews do not contain any subjective terms that would help to classify them as negative [132]. In order to overcome these challenges, we developed a Persian lexicon of 1500 Persian words along with their polarity and POS tag, which is called PerSent. Table 4.1 shows some examples of PerSent lexicon. Table 4.1 indicate the Persian words along with their translation, their POS tag, polarity (positive/negative) and their score.

Most of the previous research on identifying sentiment used adjectives to identify the polarity of sentences [133]. Some researchers used adverbs and

Table 4.2: POS Statistics

Part of Speech	Words
Adjective	1478
Adverbs	391
Nouns	780
Verbs	851

adjectives together to build a lexicon [134]; some used adjectives, adverbs, and verbs [135]. For our Persian sentiment lexicon, we used adjectives, adverbs, verbs, and nouns, because all these words and phrases are useful to determine the polarity of the sentence.

A lexicon can be developed in different ways, such as manually or by translating existing lexicons such as SentiWordNet [136] or General Inquirer [137]. The words and phrases used in the lexicon were taken from different resources such as a movie reviews' websites, blogs, and Facebook. The sources belong to the movies, news, mobile phones, and computers categories.

In order to assign polarity to the words and phrases, the TextBlob Python package was used to assign polarity to words, phrases, and sentences in English [138]. For this, we translated Persian words into English using Google Translate. We assigned a part of speech (POS) tag to each word or phrase using the HAZM python package. The degree of intensity was indicated: e.g., خوشحال (*happy*), بشاش (*cheerful*), and شاد (*delighted*) have different positive values. Table 4.2 show the distribution of the POS tags in the lexicon. As shown in Table 4.2 there are more adjectives than nouns, adverbs and verbs.

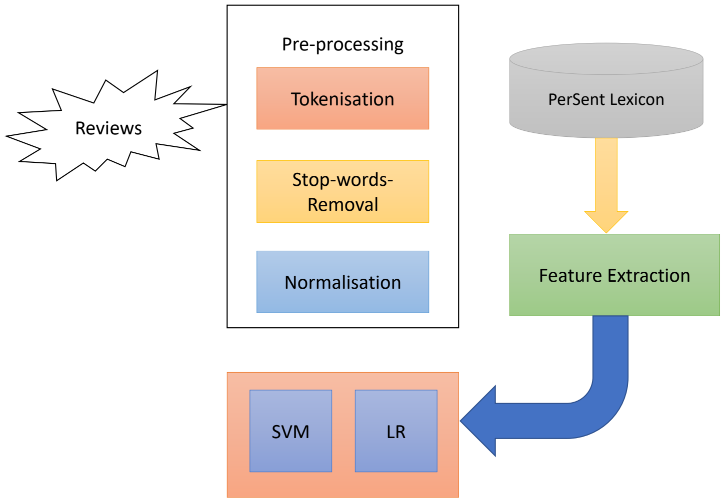

Figure 4.1: The Persian Framework

4.2 EVALUATION METHODOLOGY

In order to evaluate the performance of the lexicon, we used two classification algorithms (SVM and Logistic Regression). Fig 4.1 show the general framework we used to evaluate the performance of the lexicon.

Pre-processing: The pre-processing step consists of three parts: tokenisation, normalisation and stop-word removal.

Feature selection: The purpose of feature selection was to remove unnecessary features, which improved the performance of the classification. The features selected were word polarity and POS tag.

Frequency of sentiment words: The sentiment words identify the overall polarity for sentiment classification. Examples of positive words in Persian are زيبَا (beautiful) and عَالي (excellent), and of negative words are زشت (ugly) and بَد (bad). The features of presence of positive and of negative words (two different features) are binary, without considering the number of occurrences of a given word, while the other two features are integers and indicate the number of occurrences of positive and of negative words, correspondingly.

Table 4.3: Performance of different classifiers with all features

Classifier	Accuracy	Precision	Recall	F-measure
SVM	76.04	0.76	0.76	0.76
Logistic Regression	74.28	0.74	0.74	0.74

POS-based features: PerSent lexicon contains words along with their POS tag, such as adverb, verb, noun, or adjective. Most of the previous research used only adjectives and nouns to identify the polarity of sentences [29], but we consider eight different features: the frequencies of positive and of negative adjectives, adverbs, verbs, and nouns, correspondingly.

Word Polarity: PerSent lexicon contains polarity for words. As two different features, the overall polarity of negative and of positive words, correspondingly was used.

4.3 EXPERIMENTAL RESULTS

We applied approaches to SA using the lexicon to the Persian Product reviews, which contains 1500 positive and 1500 negative reviews [139]. The reviews collected from Persian product review website (www.digikala.com).

The SVM and Logistic Regression classifiers used for evaluation. The SVM achieved better performance than Logistic Regression as as shown in Table 4.3 In this experiment all the features were used.

Table 4.4 compared the frequency of unigram, bigram and trigram using SVM classifier. The experimental results reveal that the frequency of trigram achieved better performance as compared to unigram and bigram.

We compared the effectiveness of different features in order to determine their importance as shown in Table 4.5 Table 4.6 shows that the frequency of unigram, bigram and trigram. Exeprimental results show that the trigram

Table 4.4: Performance of the frequency features (SVM)

Feature	**Accuracy**	**Precision**	**Recall**	**F-measure**
Freq of Unigram	61.07	0.57	0.56	0.57
Freq of Bigram	60.53	0.58	0.59	0.58
Freq of Trigram	66.57	0.73	0.74	0.73

Table 4.5: Performance of the frequency features (LR)

Feature	**Accuracy**	**Precision**	**Recall**	**F-measure**
Freq of Unigram	58.96	0.58	0.58	0.58
Freq of Bigram	57.02	0.57	0.57	0.57
Freq of Trigram	64.45	0.69	0.69	0.69

Table 4.6: Performance of the POS features (SVM)

Feature	**Accuracy**	**Precision**	**Recall**	**F-measure**
Adjective	64.23	0.64	0.64	0.64
Verb	63.38	0.61	0.61	0.61
Noun	60.47	0.60	0.59	0.59
Adverb	58.24	0.69	0.68	0.69

outperforms unigram and bigram. The bigram achieved lowest accuracy as compared to other features.

The POS features, such as the frequency of positive and negative adjectives, adverbs, verbs, and nouns, correspondingly were compared; The Table 4.6 shows the SVM results with different POS tag. Experimental results show that the adjective achieved better accuracy as compared with other POS tag and adverb achieved the minimum accuracy.

Table 4.7: Performance of the POS features (LR)

Feature	Accuracy	Precision	Recall	F-measure
Adjective	61.42	0.61	0.61	0.61
Verb	60.56	0.61	0.61	0.61
Noun	60.09	0.60	0.59	0.60
Adverb	56.07	0.57	0.56	0.57

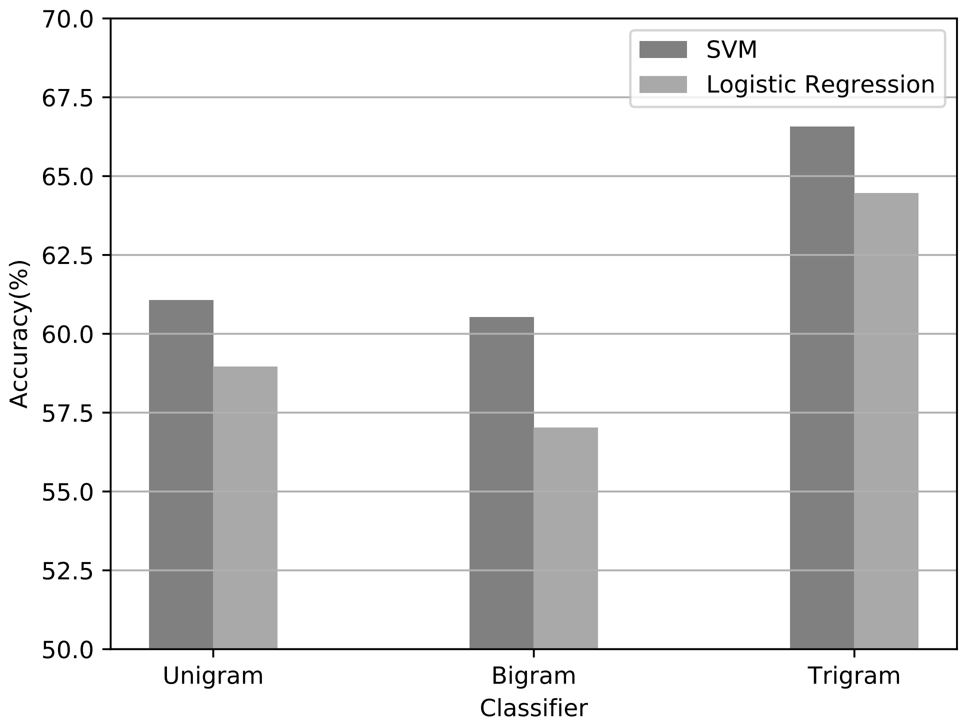

Figure 4.2: Comparison of Frequency SVM and Logistic Regression

The Table 4.7 shows the Logistic Regression results with different POS tag. The adjective outperforms the other POS tag. In addition, the adverb obtained lowest accuracy.

Fig 4.2 compare the frequency of unigram, bigram and trigram for SVM and Logistic Regression. The experimental results display that the SVM achieved better accuracy for frequency of unigram, bigram and trigram as compared to Logistic Regression.

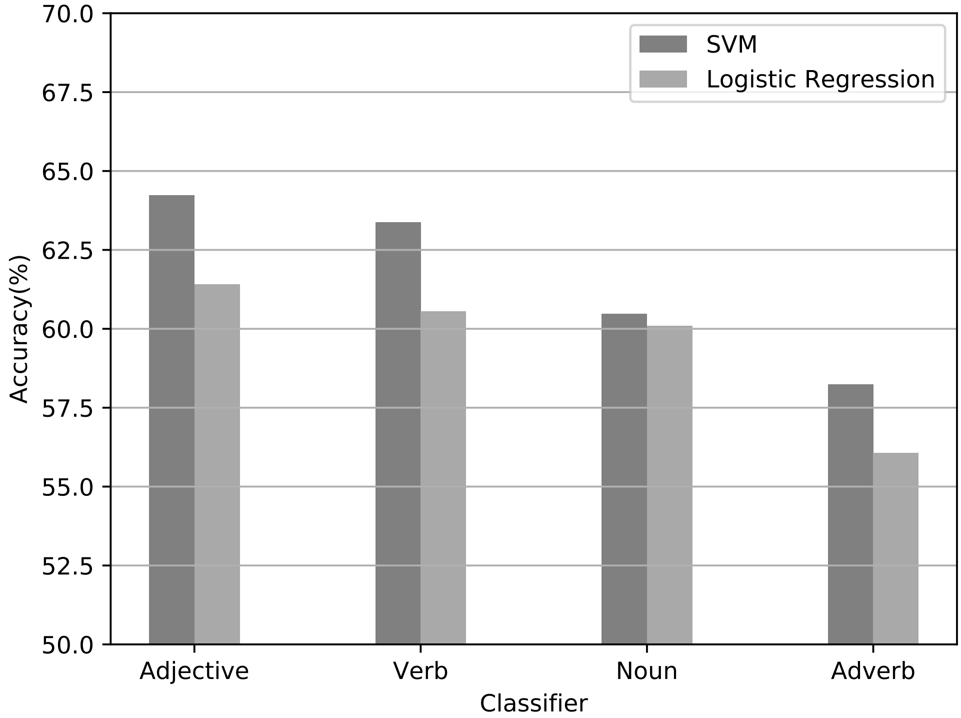

Figure 4.3: Comparison of POS SVM and Logistic Regression

Fig 4.3 compare the POS of unigram, bigram and trigram for SVM and Logistic Regression. The experimental results display that the SVM achieved better accuracy for adjective, adverb, noun and verb as compared to Logistic Regression.

4.4 DISCUSSION

We did not experiment with using purely supervised approaches, since the aim of our experiments was not to show how well one can classify Persian texts when one has a large enough corpus of manually labelled examples. Here we only to show how useful our lexicon is for classification of Persian texts in dictionary-based (distant-supervised) manner, without any manually labelled examples at all. A more complete distant-supervision approach could use for training both use manually labelled dataset and a much larger corpus automatically annotated with the help of our lexicon; however, currently we do not have at our disposal such a large corpus. In addition, our current experiments are sufficient to demonstrate the value of our PerSent lexicon.

In order to train the supervised classifier, the features were extracted from the reviews and the lexicon was used to assign polarity to the sentences. Then, we used this automatically labelled dataset to train a supervised classifier. We evaluated the trained classifier on our manually annotated dataset for which the polarity of the reviews was known (the main reason for using binary classification was its efficiency). We used five different classifiers: support vector machine (SVM), Logistic regression (LR).

The main problem of the lexicon is relatively small size: 3500 words are not enough for Persian because the language consists of many dialects and actively uses idiomatic expressions, and thus requires a larger lexicon, development of which would take time and effort [87].

Another problem is that our baseline model was not able to identify sarcasm. Thus, a much more sophisticated system needs to be developed to detect ironic and sarcastic sentences. In future, another tool will be developed to handle sarcasm in order to further improve the classification performance. In addition, the baseline model did not properly handle some sentences consisting of a mixture of Persian and English words.

Experimental results show adjectives gave better results in comparison with other POS tags, because generally the polarity in a sentence is often directly related to adjectives as compared with other words. For example, in این عکس زیبا است, which means "It is a beautiful picture", the adjective clearly indicates the sentiment. In addition, all features together gave better results than individual features separately, because in this way the algorithm had access to more information.

It is worth noting that the accuracy of the system crucially depends on the quality of the output of the part-of-speech tagger, which relies on grammatical correctness of the input sentence. Both datasets, however, contain ungrammatical sentences which penalize results. As shown in the results the adjective outperforms other POS tags. It has to be noted that, In Persian language adjective consists of more polarity as compared to other POS tag such as adverb, verb and noun.

4.5 CONCLUSION

We have developed a new lexicon for the Persian language, which can be used for Persian SA. The lexicon contains 3500 Persian words along with their polarity on a numeric scale from -1 to +1 and the POS tag. The majority of the values were assigned manually. The PerSent lexicon is freely available at: Footnote [1]

Experimental results show that the lexicon is a useful tool to determine the polarity of sentences in Persian. In the experiments, there are two classifiers were used, SVM and Logistic Regression, of which SVM achieved better accuracy.

1 http://www.gelbukh.com/resources/persent

5

CHAPTER 5 DEEP LEARNING DRIVEN SENTIMENT ANALYSIS

As discussed in the previous chapters, sentiment analysis (SA) is a method to automatically classify large amounts of text into positive or negative sentiments [3]. In this chapter, a novel corpus for Persian SA is developed and evaluated using both shallow and deep machine learning algorithms. For shallow learning, Logistic Regression (LR), Support Vector Machines (SVM), and Multilayer perceptron (MLP) classifiers are used. For deep learning, 1D Convolutional Neural Network (CNN), 2D-CNN, stacked Long Short-Term Memory (LSTM), and Bidirectional LSTM (BiLSTM) algorithms have been utilized.

In summary, this chapter reports two major contributions outlined below:

1. To the best of our knowledge, this is the first work on exploiting deep learning based automated feature engineering for Persian SA.

2. In addition, the fastText word embedding is used to obtain vector representation of Persian words. The contextual features are extracted using deep learning algorithms and their performance are compared with traditional state-of-the-art classifiers.

The rest of the chapter is organized as follows: Section 5.1 presents proposed novel approach for Persian SA. In section 5.2 experimental results are presented. Finally, section 5.3 concludes this chapter.

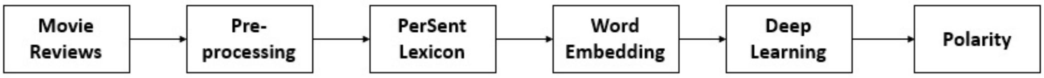

Figure 5.1: Proposed Framework

Positive	Negative
فیلم بینظیري بود بعد از چند سال واقعا بهترین فیلمي بود که دیدم ارزش چند بار دیدنو داره	ظاهراً کارگردان فقط قصد داشته با یک موضوع جنجالی و کنجکاو برانگیزجذب مخاطب کند بدون اینکه فیلمش پتانسیل لازم برای پاسخ به مخاطب را در این ژانر داشته باشد
پیشنهاد میکنم حتما ببینید بازیها خیلی طبیعی و خوب بود	بازی ها کاملاً متوسط است و شاهد سکانس بد هستیم
فیلم از همه لحاظ کارگردانی فیلم نامه بازی و بسیار خوب بود واقعا بدون رقیب بود	نا امید شدم خیلي خسته کننده بود طوري که درنهایت مردم فریاد زدند اخ جووون بالاخره تموم شد

Figure 5.2: Persian Sentences Examples

5.1 METHODOLOGY

In this section, the proposed approach for Persian movie reviews is discussed in detail. Fig 5.1 depicts the proposed framework and details are presented in subsequent sections.

5.1.1 *Data Pre-processing*

The novel dataset used in this work was collected manually from Persian movie websites: www.caffecinema.com and www.cinematicket.org . A subset of the dataset was used to train the neural network (60% training dataset) and rest of the data (40%) was used to test (30%) and validate (10%) the performance of the trained neural network. The dataset consists of two labels: positive and negative. The reviews were manually annotated by three native Persian speakers aged between 30 and 50 years old. After data collection, the dataset was pre-processed using tokenisation and normalisation techniques as discussed in section 2.1.

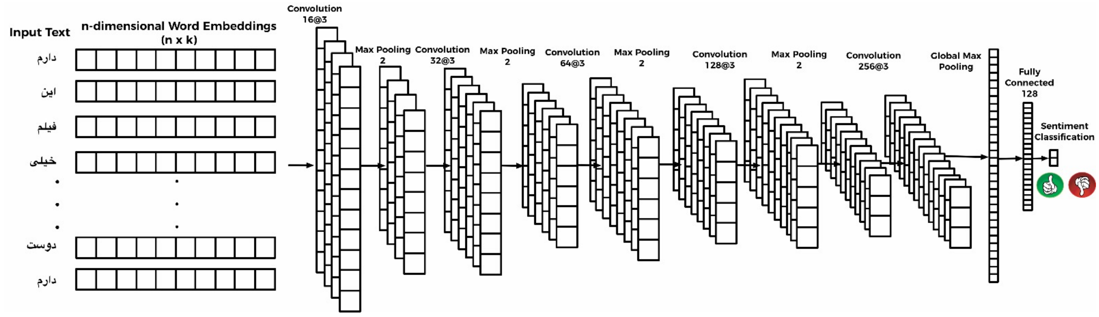

Figure 5.3: CNN Classifier

N-gram features are extracted from the text. The n-gram size one is called unigram, size two is bigram and size three is trigram [140]. The n-gram features are extracted from Persian movie reviews. Hazm reads the text and assigns part-of-speech tag (POS) (such as noun, verb, adjective, adverbs etc.) into words [141]. The Hazm tool is used to extract POS tag (Adjective, Adverb, Verb and Noun) from Persian movie reviews. After extracting n-grams and POS tag features, the PerSent lexicon is used to assign polarity. The lexicon contains 1500 Persian words along with their POS tag and polarity of the words [142].

Word embedding: In order to train the CNN classifier, the tokenised Persian words are converted into a three hundred dimensional vector using fastText pre-trained embedding [143]. Fig 5.4 shows the process of word embedding with Persian sentence example, من این فیلم خیلي دوست دَارم (I really like the movie).

5.1.2 *Classification*

5.1.2.1 *Convolutional Neural Network*

The developed CNN classifier is shown in Fig 5.3 The CNN classifier consists of input, output, and hidden layers. The hidden layers constitute of convolutional, pooling, fully connected and normalization layers [144] [145]. In experiments,

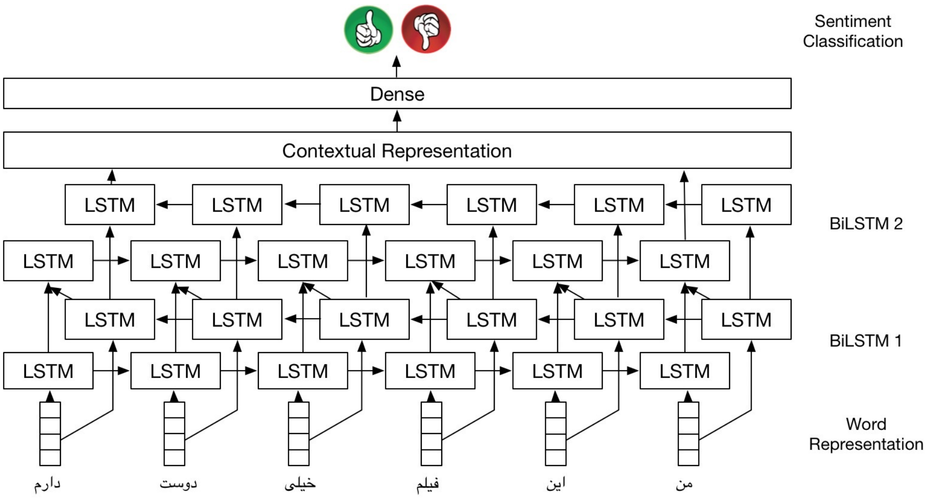

Figure 5.4: Long short-term memory

the best results were obtained using an 11 layered CNN architecture as shown in Fig 5.3 In the first convolutional layers, 16 feature maps with kernel size of 2 are used. In the second convolutional layer, 32 feature maps with a kernel size 2 are used. The layers are followed by a max pooling layer of size 2. In the fourth convolutional layer, 64 feature maps with kernel size 2 are used. The final convolutional layer is followed by a max pooling with window size 2. The feature extraction framework is followed by fully connected layers of size 5000, 500 and 2.

Long Short Term Memory (LSTM): LSTM was originally proposed in [146] by Hochreiter et al. The developed LSTM network, presented in Fig. 5.4, consists of an input layer, two stacked LSTM layers, and a fully connected layer. Specifically, the network consists of two bidirectional LSTM layers followed by two dropout layers, one dense layer and one activation function layer. The fastText word embedding are fed into stacked LSTM layers. The output of the second LSTM layer was then fed into the fully connected (dense) layer with two neurons. The architecture was trained using Adam optimizer with dropout probability 0.2.

MultiLayer perceptron (MLP): MLPs are made up of highly interconnected processing elements called neurons, processing information by their state response and learning from examples. A neuron in an MLP is connected to several inputs with different associated weights. The output of a neuron is the summation of all connected inputs, followed by a non-linear processing unit, called a transfer function. The main objective of MLP is to transform the inputs into meaningful outputs by learning the input-output relationship, and offer viable solutions to unseen problems (a generalization capability). Therefore, the capacity to learn from examples is one of the most desirable features of neural network models. The goal of training is to learn desired system behavior and adjust the network parameters (interconnections weights) to map (learn) the input-output relationship and minimize the cost function. The processed embedding vectors were fed into an MLP model with 1 hidden layer and 10 to 150 hidden neurons per layer.

Autoencoder: An autoencoder (AE) is a type of unsupervised learning algorithm, typically used for dimensionality reduction purposes. The AE standard configuration includes one input layer, one output layer and one hidden layer. It compresses the input data x into a lower dimension h through the encoding process:

$$h = g(xw + b) \tag{5.1}$$

where x, w, b are the input vector, weight matrix, the bias vector, respectively and g is the activation function. Then, it attempts to reconstruct the same set of input (x) from the compressed representation (h) through the decoding process:

$$\tilde{x} = g(hw^T + b) \tag{5.2}$$

5.2 EXPERIMENTAL RESULTS

In this section, we describe the experimental results followed by the obtained results and discussions. To evaluate the performance of the proposed approach, Persian movie reviews corpus is used. For data labelling, the PerSent lexicon is used to assign a polarity to sentences. The train and test sentences were converted into fastText word embedding vectors to train LSTM and CNN. In addition, the n-gram features (bigram, trigram) and POS features (noun, adjective, verb and adverb) are extracted from Persian movie reviews. The extracted features are converted into bag of words and principal component analysis (PCA) is used to reduce the dimensionality of data to two hundred dimensions. The extracted features are fed into a SVM and a LR classifier.

Furthermore, **Hotel reviews dataset** consists of 1800 positive and 1800 negative hotel reviews collected from the hotel booking website

(http://www.hellokish.com). The hotel reviews corpus is used to compare how our approach performs in a new domain compared to state-of-the-art approaches, including multilingual methods. The hotel reviews were collected, after data pre-processing, the extracted features are converted into TF-IDF and the machine learning is applied to evaluate the performance of the approach. The overall accuracy of their proposed approach is 87%. However, for their experiments, 5-fold cross-validation [112]. In our experiments, we used 60% for training, and the rest of the data was used for testing and validate the performance of the trained classifiers, 30% used for testing and 10% for the validation set.

The parameters of LR, SVM, LSTM, and CNN models are as follows: word embedding dimensions are 300, the number of epochs equal to 200, and batch size is 128. The ML classifiers are trained to classify the sentences into either positive or negative. In addition, it has been shown that the size of the filter has a positive impact on the final results and model received better performance when filter size is set to the smaller number such as 3.

Table 5.1 presents the results of SVM using various features. The results show that nouns outperformed other POS tags. Table 5.1 show the LR achieved

Table 5.1: SVM vs LR on Persian Movie reviews

Feature	Precision		Recall		F-measure		Accuracy (%)	
	SVM	LR	SVM	LR	SVM	LR	SVM	LR
Bigram	0.65	0.66	0.61	0.67	0.53	0.65	61.27	66.59
Trigram	0.81	0.88	0.75	0.88	0.71	0.88	74.53	87.96
Adjective	0.72	0.87	0.85	0.85	0.78	0.79	84.57	85.07
Adverb	0.76	0.81	0.87	0.86	0.81	0.82	86.95	85.88
Noun	0.79	0.91	0.89	0.89	0.84	0.85	88.93	89.37
Verb	0.78	0.78	0.88	0.88	0.83	0.83	88.4	88.4

better accuracy as compared with SVM. Experimental results show noun achieved better accuracy as compared to other features and bigram received the lowest accuracy as compared with other features. Table 5.2 presents the results of CNN and LSTM classifiers and comparison with MLP and AE. The experimental results demonstrate the effectiveness of the AE as compared to MLP. In addition, the stacked-BiLSTM achieved better accuracy, precision, recall and F-measure. However, the MLP achieved lower accuracy as compared with other classifiers. The Table 5.1 demonstrated that The logistic regression obtained better accuracy for bigram, trigram, adjective and noun as compared to SVM. However, the SVM obtained better results for adverb. The trigram and adjective achieved better results as compared to other features because they are more informative features.

Fig 5.5 presents the accuracy of both shallow and deep learning classifiers (Persian Movie reviews). It can be seen that the stacked bidirectional-LSTM outperformed all other machine learning algorithms.

Table 5.2: Deep Learning Classifiers Results on Persian Movie reviews

Classifier	Precision	Recall	F-measure	Accuracy (%)
MLP	0.78	0.78	0.78	78.49
MLP-AE	0.8	0.8	0.8	80.08
2D-CNN	0.82	0.82	0.82	82.47
1D-CNN	0.84	0.83	0.83	82.86
Stacked-LSTM	0.94	0.94	0.94	93.65
Stacked-BiLSTM	0.96	0.96	0.96	95.61

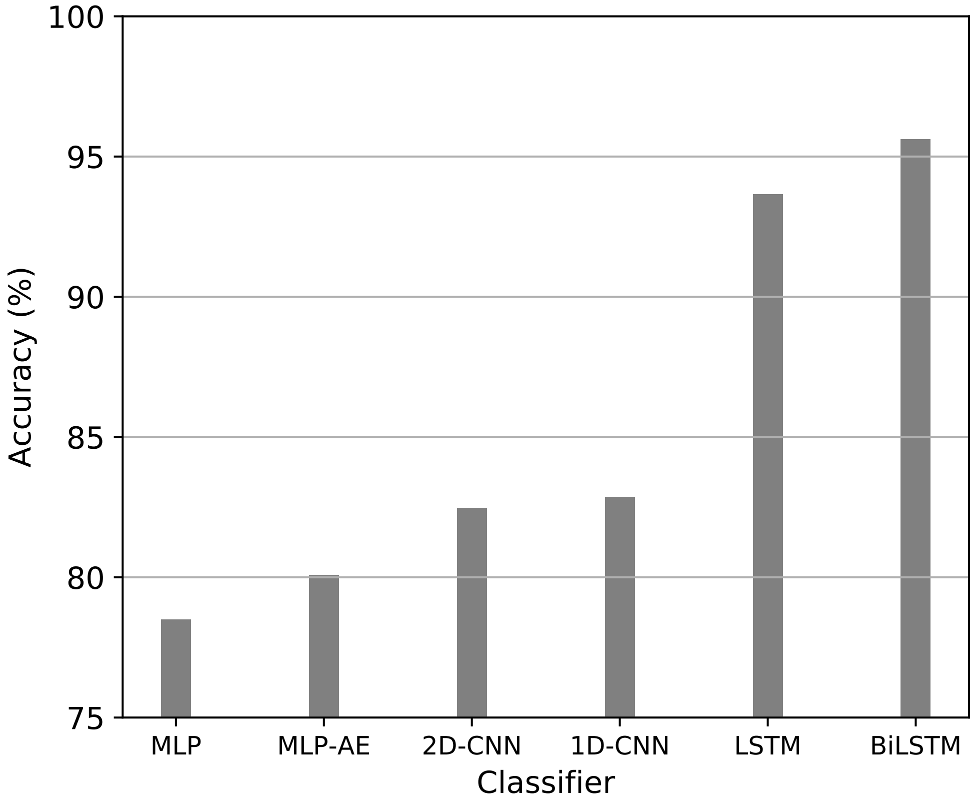

Figure 5.5: Deep Learning Classifiers Accuracy on Persian Movie reviews

Fig 5.6 shows the classification performance of SVM and LR using different feature types (Persian Movie reviews). It can be seen that the noun features help achieving better performance as compared to other features.

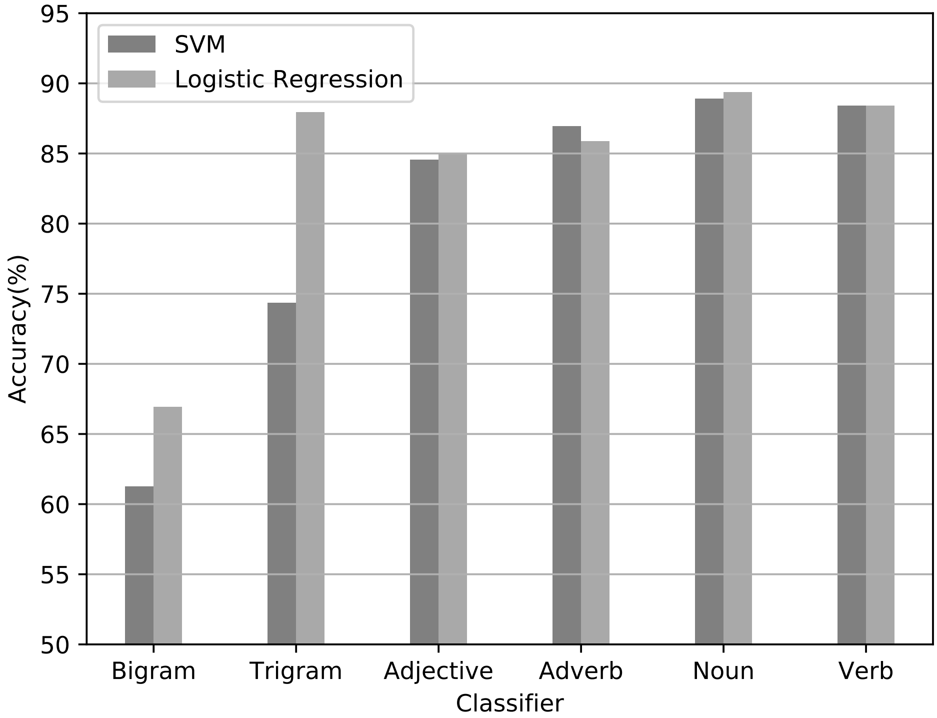

Figure 5.6: SVM vs LR Classifier Accuracy on Persian Movie reviews

Table 5.3: Comparison of 1D-CNN Layers

Layer	Precision	Recall	F-measure	Accuracy (%)	Time
2	0.72	0.73	0.72	73.55	2 m 31 s
3	0.74	0.74	0.74	74.26	3 m 12 s
4	0.81	0.79	0.79	79.01	3 m 44 s
5	0.84	0.83	0.83	82.86	4 m 22 s
6	0.78	0.76	0.76	78.42	5 m 24 s

Table 5.3 presents the comparison results for the proposed 1D-CNN with different layers. Experimental results show that the 5 layered 1D-CNN architecture generally outperforms other layered architecture. The time to train each model is shown in the last column of Table 5.3 Additionally, the recall and F-measure for 5-layer outperformed is compared with 2, 3, 4, and 6 layered model.

Table 5.4: Comparison of 2D-CNN Layers

Layer	Precision	Recall	F-measure	Accuracy (%)	Time
2	0.72	0.70	0.70	74.05	2 m 21 s
3	073	0.71	0.71	74.22	3 m 23 s
4	078	0.78	0.78	78.26	5 m 5 s
5	0.82	0.82	0.82	82.47	6 m 33 s
6	0.75	0.76	0.75	76.51	8 m 25 s

Table 5.5: Comparison of LSTM Layers

Layer	Precision	Recall	F-measure	Accuracy (%)	Time
2	0.94	0.94	0.94	93.56	6 m 49 s
3	0.84	0.83	0.83	84.26	7 m 28 s

Table 5.4 presents the comparison results for the proposed 2D-CNN with different layers. Experimental results show the 5-layer CNN generally outperforms as compared with other layers. The time to train each model is shown in the last column of Table 5.4. Additionally, the recall and F-measure for layer 5 outperformed as compared with other layers.

Table 5.5 presents the comparison results for different layers of LSTM. Experimental results show that the 2-layer LSTM achieved better results as compared with other layers. The time to train the model is shown in Table 5.5. Additionally, experimental results show that the 2-layer LSTM achieved better precision, recall, f-measure as compared with other layer.

Table 5.6 presents the comparison results for different layers of BiLSTM. Experimental results show that the 2-layer BiLSTM achieved better results as compared with 1-layer. Additionally, experimental results show the 2-layer BiLSTM achieved better precision, recall, f-measure as compared with 1-layer.

Table 5.6: Comparison of BiLSTM Layers

Layer	Precision	Recall	F-measure	Accuracy (%)	Time
2	0.96	0.96	0.96	95.61	7 m 23 s
3	0.84	0.86	0.84	85.05	9 m 24 s

Experimental results show the training time is increased as the number of layers is increasing.

The results demonstrate that the stacked-BiLSTM model outperformed all other methods. In addition, it is observed that CNN and LSTM classifiers can effectively detect polarity of the movie reviews. Furthermore, they also help in to detect contextual information as compared to traditional classifiers. It is shown that deep learning approaches are more optimal for sentimental analysis due to less over-fitting and better generalization.

Table 5.7 presents the results of SVM and LR using various features on hotel reviews. Experimental results indicate the LR achieved better accuracy as compared with SVM. The results show that the use of adjective outperformed other features. Additionally, the adverb feature achieved lower accuracy as compared with other features.

Furthermore, Table 5.8 presents the results of CNN and LSTM classifiers on the hotel reviews dataset. Experimental results show the 2D-CNN achieved better accuracy as compared with other classifiers. Additionally, the stacked-BiLSTM received lower performance as compared with other classifiers, however, it achieved better precision as compared with other classifiers.

Fig 5.7 presents the accuracy for deep learning classifiers (Hotel reviews). It can be seen that the 2D-CNN outperformed all other machine learning algorithm. However, the BiLSTM received lower accuracy as compared to other classifiers.

Table 5.7: SVM vs LR on Hotel reviews dataset

Feature	**Precision**		**Recall**		**F-measure**		**Accuracy (%)**	
	SVM	**LR**	**SVM**	**LR**	**SVM**	**LR**	**SVM**	**LR**
Bigram	0.71	0.72	0.71	0.72	0.71	0.72	71.25	72
Trigram	0.73	0.74	0.73	0.74	0.73	0.74	73.5	74
Adjective	0.75	0.76	0.78	0.79	0.76	0.77	76.24	77.06
Adverb	0.63	0.64	0.62	0.63	0.62	0.62	62.72	62.78
Noun	0.67	0.69	0.68	0.70	0.67	0.69	68.21	69.08
Verb	0.68	0.69	0.69	0.71	0.68	0.69	68.54	69.97

Table 5.8: Deep Learning Classifiers Results on Hotel reviews dataset

Classifier	**Precision**	**Recall**	**F-measure**	**Accuracy (%)**
Stacked-BiLSTM	0.53	0.91	0.67	74.49
1D-CNN	0.73	0.80	0.76	78.02
Stacked-LSTM	0.79	0.81	0.80	81.04
2D-CNN	0.89	0.89	0.89	89.76

Fig 5.8 depicts the classification performance of SVM and LR using different feature types (Persian Hotel reviews). It can be seen that the adjective features help to achieve better performance as compared to other features. In addition, adverb features received lower accuracy as compared to other features.

It is worth noting that the accuracy of the system crucially depends on the quality of the output of the architecture of the deep learning classifiers, however, the proposed architectures for CNN and LSTM is suitable for both Persian movie and product reviews dataset. It has to be noted that, the Persian language consists lots of sarcasm and idioms which cannot be handle by

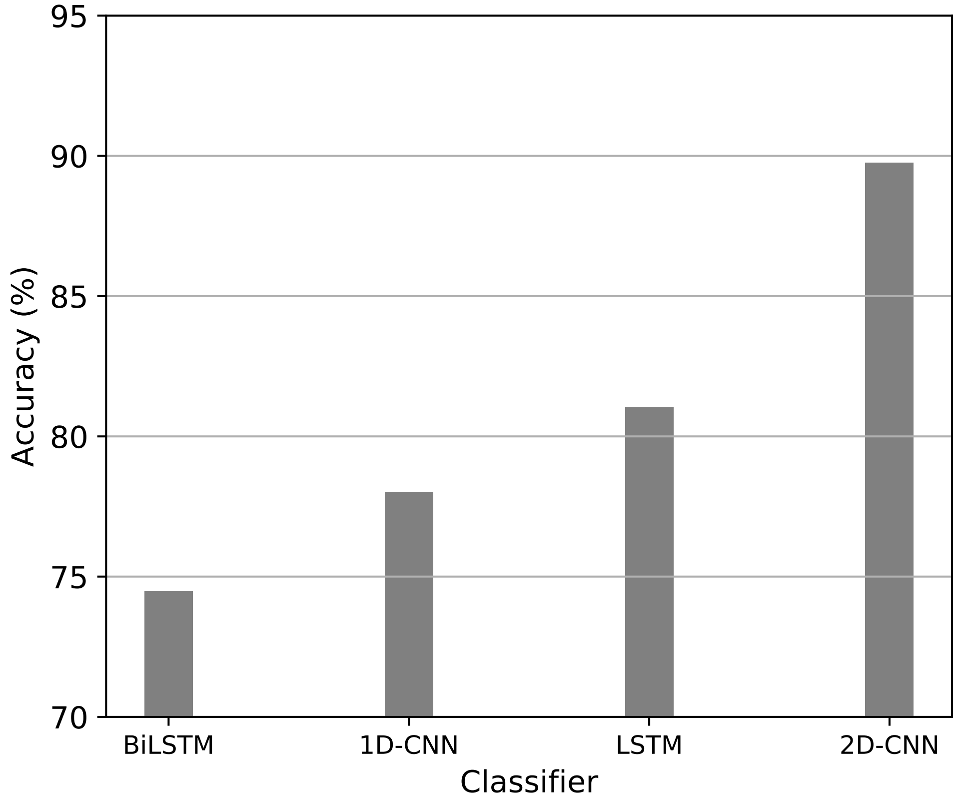

Figure 5.7: Deep Learning Classifiers Accuracy on Persian Hotel reviews

deep learning classifiers which consists of black box system and it cannot determines the polarity for this types of sentences.

In spite of the fact that the fast and accurate SVM classifier typically outperforms other classifiers in terms of accuracy, for large datasets it takes increasingly more time to train the model. A large dataset does not fit in memory when using SVM, which affects the efficiency of the classifier. While there are different techniques that help overcoming this issue, such as feature reduction, we evaluated some other algorithms for our task, which, however, did not perform as well as SVM in terms of accuracy.

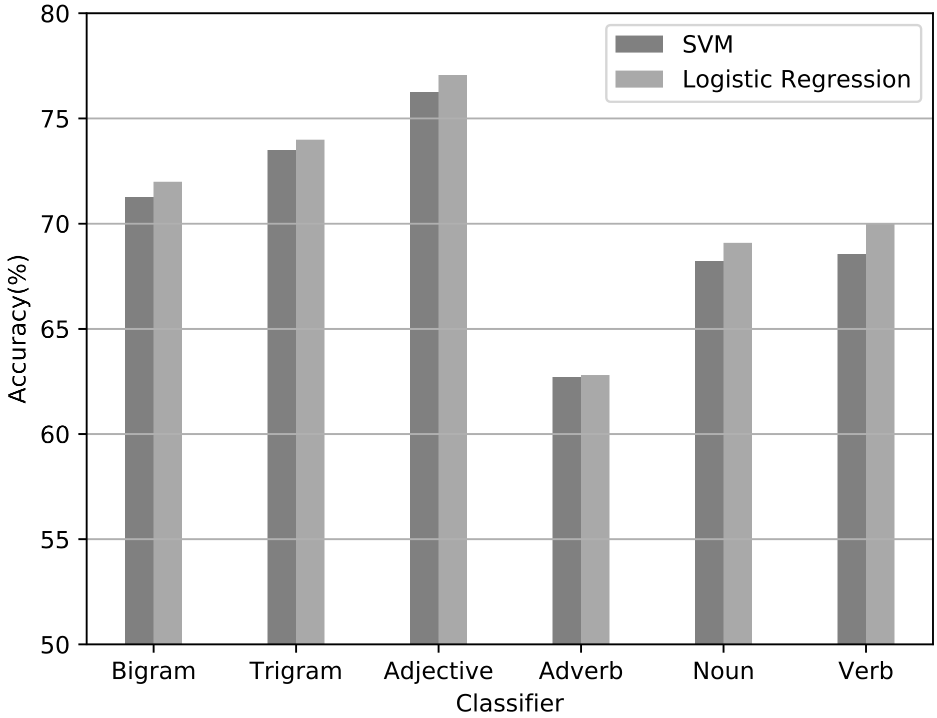

Figure 5.8: Features Accuracy on Persian Hotel reviews

5.3 CONCLUSION

In this chapter, a novel Persian movie reviews corpus is developed and validated using both shallow and deep learning algorithms. The experimental results demonstrate that deep learning outperformed state-of-the-art shallow machine learning approaches. Specifically, the Stacked-bidirectional-LSTM achieved the highest accuracy of up to 95.61%. The demo for the approach is available at: Footnote [1]. However, this chapter does not focus on concept extraction. The following chapters will focus on extracting concepts from Persian sentences using dependency parser.

1 https://cogbid.github.io/persian-sentiment-analysis/

6

CHAPTER 6 INTEGRATING DEPENDENCY-BASED RULES AND DEEP NEURAL NETWORKS FOR PERSIAN SENTIMENT ANALYSIS

Most of the current sentiment analysis approaches in Persian are based on word co-occurrence frequencies, that first extract frequency of the words containing polarity, and the overall polarity of source text is then determined. However, these approaches fail to consider the word order and the dependency relation between words, that play an essential role in determining the overall polarity of the sentence. For example, in "Certainly old, but I'd love to see more stories in that universe" though the reviewer is expressing a strong negative sentiment in the first part of sentence, the overall polarity is positive. Although, the words "old" and "love" have negative and positive polarities, the overall polarity not only depends on the word sentiment strength but also on the relative words position and dependency structure of the sentence [147, 144].

Moreover, current approaches of sentiment analysis fail to understand the real noisy text data consisting of sarcasm, idioms, informal words, phrases and spelling mistakes [3, 130]. For example, بي تفَاوت نيستم فقط ديگر كسي برَايم مهم نيست, a sarcastic sentence, cannot be accurately classified by the current approaches [148]. In addition, the scarce availability of natural language processing tools and resources such as lexicon, labelled corpus, parts-of-speech

(POS) tagger, etc. are a major bottleneck in the reliable implementation of sentiment analysis methods in Persian, a major language with 54 million speakers [149, 150, 151].

To address the aforementioned limitations, we propose a novel framework that integrates Persian linguistic grammar rules and deep learning for analyzing source text. This is shown to improve the overall performance and robustness of polarity detection for real noisy data.

Dependency grammar rules are based on linguistic patterns that allow sentiment to go from words to concepts based on dependency relations. As a result, dependency-based rules take into account hierarchical relations between keywords, the word order, and individual word polarities to accurately determine the underlying polarity.

We perform an extensive and comprehensive set of experiments using benchmark Persian product and hotel reviews corpora and compare the performance of our proposed dependency-grammar rule-based approach with state-of-the-art Support Vector Machine (SVM), Logistic Regression (LR), and DNN architectures including long short-term memory (LSTM) and Convolutional Neural Networks (CNN). Comparative simulation results reveal that dependency-based rules outperform state-of-the-art SVM, LR, and fastText classifiers by a margin of 10-15% and perform comparably with DNN classifiers. However, the dependency-based rules cannot classify 10-15% of the dataset due to non-availability of word polarity in the small sized Persian lexicon. Therefore, we propose a hybrid framework, that integrates our proposed dependency rule-based classifier and different DNN architectures, such as CNN and LSTM.

In summary, the chapter reports four major contributions outlined below:

1. Novel dependency-based rules for Persian sentiment analysis. These address the limitation of word co-occurrence frequency based approaches

by exploiting the hierarchical relations between keywords, the word order and, individual word polarities to accurately determine the underlying sentiment. To the best of our knowledge, this is a first Persian sentiment analysis framework that assigns polarity to sentences, without any feature engineering or training on large corpora.

2. A critical analysis of our proposed dependency-based rules with conventional Logistic Regression, Support Vector Machine as well as advanced DNN models, on two benchmark product and hotel reviews corpora, demonstrates the superior performance of our proposed model compared to conventional approaches and advanced DNN models.

3. An ablation study of our proposed dependency-based rules reveals the importance of individual rules in the context of the complete framework.

4. Addressed limitations of unclassified sentences with our proposed dependency-based rules approaches by proposing a hybrid framework that integrates Persian dependency-based rules and DNN models, including CNN and LSTM to further improve the performance.

The rest of the chapter is organized as follows: Section 6.1 presents our proposed dependency-based rules for Persian sentiment analysis. Section 6.2 describes our proposed hybrid framework to address the limitations of dependency-rule driven models, and architectural details for DNN models, including CNN and LSTM, used in our hybrid framework. Section 6.3 discusses comparative experimental results and ablation studies. Finally, section 6.4 concludes this work with limitations of our current approach and outlines future research directions.

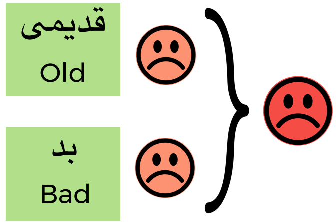

Figure 6.1: Traditional approaches extract words "old" and "bad" from a sentence and assign negative polarity into the sentence

6.1 PERSIAN DEPENDENCY-BASED RULES

This section describes our proposed dependency-based rules for Persian sentiment analysis. The proposed rules exploit hierarchical dependency relations to more accurately determine the underlying sentiment, compared to traditional word co-occurrence frequency based approaches.

For example, when using frequency of positive and negative words to classify a sentence like "The movie is very old but directing is not bad" (فیلم بسیَار قدیمي بود امَا کَارگردَاني بد نبود), the sentence will be classified as negative since the sentence consists of two negative words ("old" قَدیمي and "bad" بد) , as shown in Fig 6.1. However, the actual polarity of the sentence is positive due to the dependent words (i.e. "not" and "but"), which change the overall polarity of the sentence to positive. On the other hand, our proposed dependency-based rules take into account the syntactic relation between words, as shown in Fig 6.2 and 6.3. The exploitation of the dependent words in dependency-based rules establish a logical flow of sentiment, as shown in Fig 6.4, to determine the overall polarity. Specifically, the word "old" قدیمي following "very" بسیَار does not change the overall polarity, but the negative word "bad" بد following negation changes the overall polarity of the sentence into positive. Finally, the use of word "but" امَا changes the polarity of the second component of the sentence.

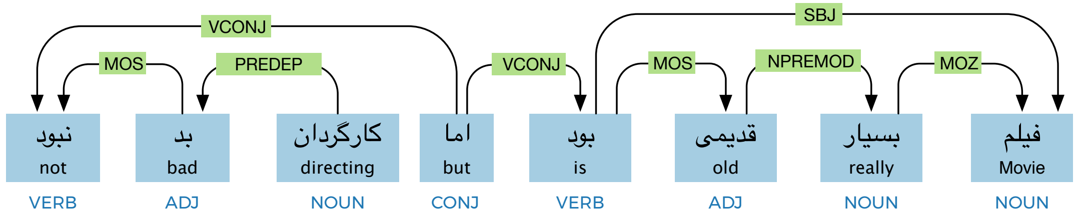

Figure 6.2: Dependency tree of a sentence (SBJ - Subject, MOZ - Extra Dependent, NPREMOD - Pre-Modifier of Noun, MOS - Mosnad, VCONJ - Conjunction of Verb, PREDEP - Pre-Dependent, ADJ - Adjective)

This section presents novel dependency-based rules for sentiment analysis that are adopted from [152] and [153].

6.1.1 *Polarity Inversion*

Trigger: When a sentence consists of at least one negation word such as نبود, نیست

Action: The negation in a sentence can change the polarity of the sentence. For example, if negation is used with a positive token, the overall polarity of the concept is negative and if negation is used with a negative token, the overall polarity of the concept is positive. For example, "I do not like the Samsung mobile" (من موبَایل سَامسونگ دوست ندَارم), the overall polarity of this sentence is negative.

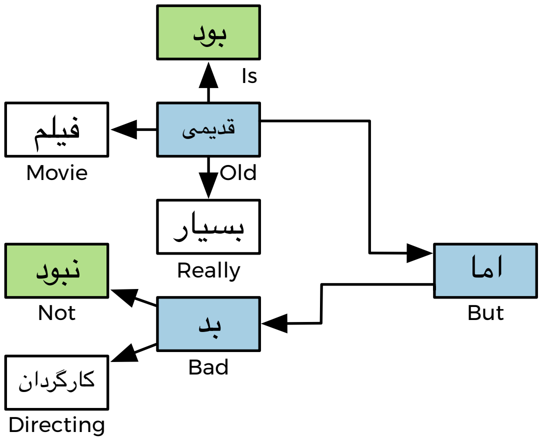

Figure 6.3: The dependency tree of a sentence resembles an electronic circuit

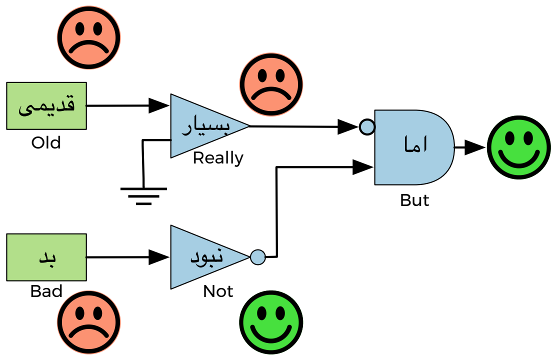

Figure 6.4: The final sentiment data flow of the signal in the circuit

6.1.2 *Complement Clause*

Trigger: The complement clause is introduced by “that” که in the sentence. Usually, the complement clause is connected by noun, verb and adjective in the sentence.

Action: The sentence is split into two parts based on the complement clause and the first part of the sentence is considered to identify the overall polarity of the sentence. For example, "I am happy that you did not buy old mobile" (من خوشحالم کِه این موبَایل قدیمي نخریدي), the sentiment expressed before the word "that" کِه is positive and the overall polarity of the sentence is positive.

6.1.3 *Adversative*

Trigger: The adversative can be a word, phrase or clause implying opposition or contrast in the sentence. An adversative word like "but" امَا, "although" اگر چِه, or "however" بَا اینکِه is used to connect two elements of opposite polarities.

Action: The sentences are split into two parts based on the conjunction such as "but" and the polarity of the second part is considered. For example, in the sentence "The iPhone is really good but it's very expensive" (موبَایل اپل خیلي خوب است ولي خیلي گرَان است), the first part before the word "but" is positive and the second part is negative. Hence, the overall polarity of the sentence is negative. Moreover, for "The Apple mobile is really expensive but it's very good" (موبَایل اپل خیلي گرَان است امَا خیلي خوب است), the first part of the sentence is negative and the second part is positive. Therefore, the overall polarity of the sentence is positive.

6.1.4 *Adverbial Clause*

Trigger: When a sentence contains an adverbial clause. The adverb clause is a group of words which count as an adverb in the sentence. The adverb clause must contain a subject and verb to trigger the rule.

Action: The role of “whereas” in a sentence is like the word “but”. If a sentence contains “whereas”, the sentence is split into two parts by identifying the subject in the sentence, since the overall sentiment is identified by the second part of the sentence. For example, “In the product description they said the mobile has good lens whereas the lens is not good” (در توضیحَات محصُول گفتِه بود کِه موبَایل لنز خوبي در صورتیکِه کِه لنزش لرزش دَارد). The polarity of the first part is positive and the polarity of the second part (after the whereas) is negative. Hence, the overall polarity of the sentence is negative.

6.1.5 *Adjective Clause*

Trigger: When a sentence contains an adjective clause. The rule is triggered when the sentence consists of pronouns such as “which” کِه ، این.

Action: The role of “which” کِه ، این is similar to the rule of “but” in the sentence. In this instance, the sentence is split into two parts and the polarity of the second part is considered. For example, “Read about things which are beautiful and good” (خوندن در مورد مسَایلي که این همه زیبَا و خوب هستند)

6.1.6 *Joint Noun and Adjective*

Trigger: When a sentence contains joint nouns and adjectives.

Action: When there is a relation between noun and adjective, both of these words are extracted from the sentence and the polarity of adjectives are considered. For example, "The mobile is bad" (این موبَایل بد بود). There is a subject relation between "mobile" موبَایل and "bad" بد.

6.1.7 *Preposition*

Trigger: When a sentence contains a preposition.

Action: The preposition "against" مخَالف is generally used in negative sentences. However, it can also be used in the positive sentences. Usually when an activity contains a negative sentiment and it follows a negative preposition modifier, the overall polarity of the sentence is changed to positive. For example, "I am against this request" (من مخَالف این درخوَاست هستم). On the other hand, if an activity is positive and it follows a negative preposition modifier, the overall polarity of the sentence is changed to negative. For example, "Hitler raised a great army and made war against other countries" (هیتلر ارتش بزرگي سَاخت و جنگي عليِه کشورهَاي مخَالف رَا شروع کرد).

6.1.8 *Additional rule*

Trigger: When a sentence contains the word "This" این in the middle of the sentence.

Action: The sentence is split into two parts based on the appearance of word (این) in the sentence, and the first part is considered. For example, "I had LG mobile, it was very bad, anyway, this mobile was for my brother" (من قبلَا موبَایل الجي دَاشتم خیلي بد بود به هر حَال این موبَایل مَال برَادرم بود), the sentiment expressed after the word "this" این does not contain any sentiment. However, the sentence contains negative words "I am not sure" which changes the polarity of the sentence into negative. The negative polarity is assigned to the sentence because in the first part, the reviewer is not sure about the quality of the mobile.

6.1.9 *Preposition Sub-rule*

The Persian sentences contain different prepositions that consist of polarity. If positive prepositions words "with" بَا, "happy" خوش and "enjoy" صفَاي are used before any adjective, they can change the polarity of the concept to positive. For example, من خوشحَال هستم "I am happy today", because the word خوش is appearing in the sentence, the polarity of the sentence is positive. However, when some preposition words such as "without" بي "anti" ضد, "Not", نَا, "poison" زهر and "No" لَا appear in the sentence, the polarity of the sentence is changed into negative. For example, in "Your answer is not wrong" (جوَاب شمَا نَادرست است) the polarity is negative, since the word "not" appears in the sentence. The word نَا is not detected by the negation rule as نَا is not a negation word and the rule cannot be triggered.

6.1.10 *Emoji Sub-rule*

Emojis were introduced as expressive components in the sentence. Emojis generally reinforce the polarity expressed in the sentence, except in sarcastic sentences. Emojis can be divided into positive :-), :), :-], :], :D or negative :(, :((, :-<, :<, >:(. Online reviews consist of different emojis. For example, من گوشي تَازه خريدم، گوشي خوبيه تو خريدش شك ندَاشته بَاشيد :), "I recently bought the mobile, do not hesitate to buy this mobile :)". Therefore, If a positive emoji is appearing in a sentence, the polarity of the sentence is positive and if a negative emoji is appearing in a sentence, the polarity of the sentence is negative.

Table 6.1 summarizes the proposed dependency-based rules for Persian.

Table 6.1: Overview of Dependency-Based Rules

Rules	**Behaviour**
Polarity Inversion	In this rule, if negation is used with a positive token, the polarity is negative and if negation is used with a negative token, the polarity is positive
Complement Clause	In this rule, the sentiment expressed before the word "that" (کِه) considered
Adverbial Clause	In this rule, the sentence is split into two parts and the polarity of the second part is used as the overall polarity of the sentence.
Adjective Clause	In this rule, if the word "which" (این کِه) is in the sentence, the sentence is split into two parts and the polarity of the second part is considered
Joint Noun and Adjective	In this rule, when there is a relation between noun and adjective, both noun and adjective are extracted. The lexicon is used to assign polarity to extracted words.
Adversative	In this rule, the word if a word like "but" (امّا), "although" (اگر), or "however" (بَا اینکِه) is used in the sentence. The sentences are split into two parts and the polarity of the second part is considered
Preposition	In this rule, if the word "against" (مخَالف) is in the sentence. It will change the polarity of the sentence into negative

Additional rule	In this rule, when the sentence contains the word "This" (این), the sentence is split into two parts and the first part of the sentence is considered the polarity of the sentence.
Preposition Sub-rule	In this sub-rule, if positive prepositions are used before any adjective, they can change the polarity of the sentence into positive and if negative prepositions are used before any adjective, they can change the polarity of the sentence into negative.
Emoji Sub-rule	In this sub-rule, if the positive emoji is appearing in the sentence, the polarity of the sentence is positive and if negative emoji is appearing in the sentence the polarity of the sentence is negative.

6.2 HYBRID FRAMEWORK

In this section, we discuss our proposed hybrid approach that integrates dependency-based rules with DNN classifiers including CNN and LSTM.

6.2.1 *Framework Overview*

Our proposed hybrid approach combines deep learning and dependency-based rules to address the problem of unclassified sentences in the aforementioned rule-based approach. The algorithm 1 depicts an overview of our proposed hybrid framework. First, sentences are preprocessed and the PerSent lexicon is used to assign polarity in extracted words. The word polarities and the sentence dependency tree is fed to a dependency-based rules classifier. If

Result: The polarity of the sentence

Tokenize and normalize sentences using hazm parser;

for *word in sentence* **do**

  if *word in lexicon* **then**

    assign opinion strengths to each extracted word;

  else

    assign zero polarity;

  end

end

apply dependency-based rules;

if *polarity assigned by dependency-based rules* **then**

  return polarity;

else

  apply DNN classifiers;

  return polarity;

end

Algorithm 1: Proposed hybrid framework

the rule-based classifier is unable to classify sentences either due to unavailability of subjective keywords in the lexicon, or if no rule was triggered, then the concatenated fastText embedding of the sentence is fed into a DNN classifier to determine the polarity of the sentence.

6.2.2 *Deep Neural Networks DNN architecture*

This section presents architectural details of DNN models used in our proposed hybrid framework, as well as standalone classifiers used for comparative performance evaluation. In recent years, DNN classifiers have achieved state-of-the-art performance as compared with other techniques due to their inherent

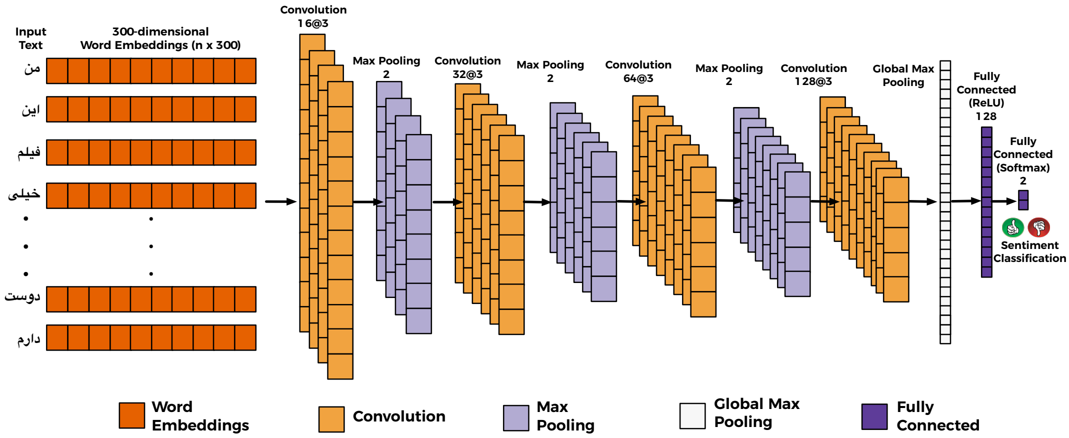

Figure 6.5: Convolutional Neural Network

ability to represent data at different levels of abstraction. The primary advantage of using deep learning classifiers such as CNN and LSTM is that they do not require any manual feature engineering [154]. However, word representations learned using a relatively small supervised corpus perform poorly as compared to unsupervisely trained fastText word embeddings. Therefore, we feed the concatenation of fastText embeddings, of size $(n \times k, 1)$ where n is the maximum number of words in a sentence and k is the embedding dimension), to the DNN architectures.

6.2.2.1 *Convolutional Neural Network (CNN)*

Our developed CNN architecture, is similar to the one used by Gogate et al. [155], consisting of input, hidden and output layers. The hidden layers consist of convolutional, max pooling, and fully connected layers. In our experiments, the best results are obtained with a 9-layered CNN architecture as illustrated in Fig 6.5 and Table 6.2

6.2.2.2 *Long Short-Term Memory (LSTM)*

Our developed LSTM architecture, is similar to the one used by Wang et al. [156], consisting of an input layer, two stacked LSTM layers and one output fully connected layer. Specifically, the LSTM part consists of two stacked

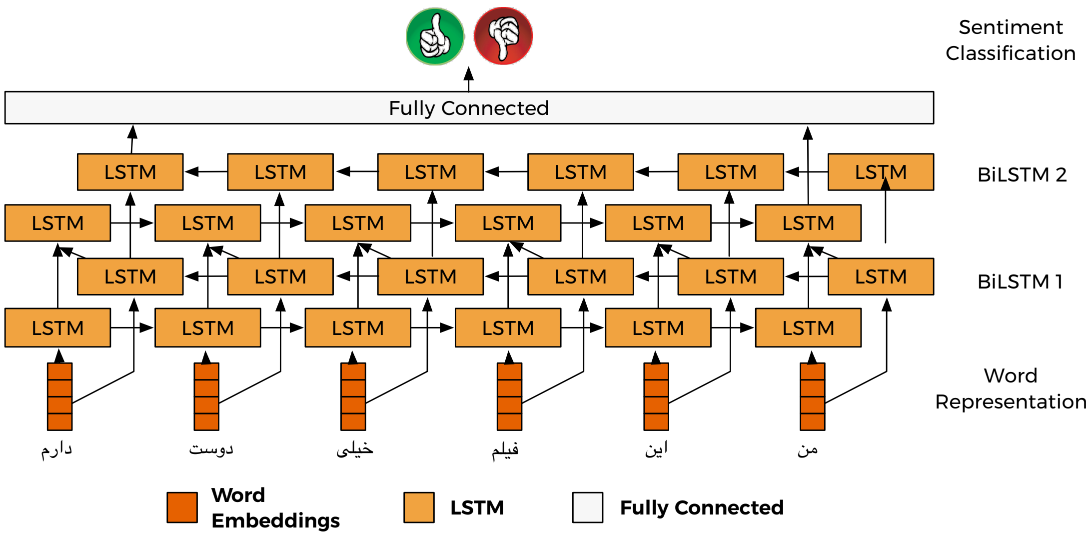

Figure 6.6: Long short-term memory

Table 6.2: CNN Architecture (Conv - convolutional layer, MaxPool - Maxpooling layer, GlobalMaxPool - Global Max Pooling layer, Fc - Fully connected layer, ReLU - Rectified Linear Unit Activation

Layer	1	2	3	4	5	6	7	8	9	10
Type	Conv	Max	Conv	Max	Conv	Max	Conv	Global	Fc	Fc
Filters	16	Pool	32	Pool	64		128	Max		
Kernal Size	3	2	3	2	3	2	3	Pool		
Neurons									128	2
Activation	ReLU		ReLU		ReLU		ReLU		ReLU	SoftMax

bidirectional LSTM layers with 128 and 64 cells, followed by a dropout layer with 0.2 probability, and one dense layer with two neurons and softmax activation. Our developed LSTM is illustrated in Fig 6.6.

6.3 PERFORMANCE EVALUATION

In this section we discuss the datasets, data preprocessing, experimental results and ablation studies.

6.3.1 *Dataset*

In this experiment, two benchmark Persian product and hotel reviews corpora are used to evaluate the performance.

Product reviews dataset [139]: The dataset consists of 1500 positive reviews and 1500 negative reviews collected from the product review website (www.digikala.com). There are two types of labels in the dataset (i.e. positive or negative) which are manually annotated.

Hotel reviews dataset [112]: The dataset consists of 1800 positive and 1800 negative hotel reviews collected from the hotel booking website

(http://www.hellokish.com). The hotel reviews corpus is used to compare how our approach performs in a new domain compared to state-of-the-art approaches, including multilingual methods.

6.3.2 *Data Preprocessing*

The corpus is preprocessed using tokenization and normalization techniques as discussed in section 2.1. In addition, PerSent lexicon [142] is used to assign polarity to the tokenized words. Zero polarity is assigned to the words that are not present in the lexicon. Finally, the Hazm Python package is used to identify the dependency tree for a sentence, the overall accuracy for hazm parser is 90.86% [157]. The dependency tree and assigned polarities are fed to our proposed dependency-based rules classifier.

6.3.3 *Experimental Setup*

Our proposed dependency-based rules do not require any training. However, to facilitate a fair comparison between dependency-based rules and other approaches we split the data into train (60%), validation (10%) and test set (30%), and evaluated the rule-based approach on the test set. For the rule-based approach alone, positive polarity is chosen when dependency-based rules are unable to classify the sentence mainly due to the unavailability of word polarity in the small sized Persian lexicon. In addition, SVM, Logistic Regression and fastText classifiers were used as a baseline to compare the performance of our proposed approach. The DNN architectures were trained and validated using Tensorflow library and NVIDIA Titan X GPU. The models were trained for 200 epochs using back propagation, with the Adam optimizer minimizing the categorical cross entropy loss function. Unclassified sentences from the rule-based approach were converted into 300 dimensional pretrained fastText word embedding and fed into deep learning classifiers as a part of the hybrid framework.

6.3.4 *Experimental Results*

To examine the effectiveness of our proposed approach with translation based approaches, we used the Google Translation API to translate our corpus into English. The translated English corpus was evaluated using state-of-the-art sentiment analysis classifiers in English [44]. In addition, we used n-gram based SVM and Logistic regression models to establish a baseline.

Comparative simulation results for product and hotel reviews, presented in Table 6.3 and Table 6.4 respectively, show that our proposed hybrid approach achieved better accuracy compared to CNN, LSTM, SVM and Logistic Regression classifiers. In addition, for the product reviews dataset, the hybrid CNN approach outperforms hybrid LSTM model. However, for the hotel reviews corpus, the hybrid LSTM approach outperforms hybrid CNN model. It is to be noted that, the dependency-based rules are unable to classify 5% (2% positive

Table 6.3: Summary of the results for Product Reviews

Classifier	Precision	Recall	F-measure	Accuracy
Kim et al.[44]	0.60	0.63	0.60	62.5
Dehkharghani et al.[158]	0.60	0.93	0.72	63.62
fastText Classifier[159]	0.67	0.67	0.67	70.01
SVM	0.75	0.75	0.75	74.8
Logistic Regression	0.75	0.75	0.75	75.2
Dependency-Based Rules	0.84	0.81	0.80	80.70
CNN	0.89	0.71	0.78	78.07
LSTM	0.90	0.74	0.81	79.77
Hybrid 1: CNN + Dependency-Based Rules	0.75	0.98	0.84	81.14
Hybrid 2: LSTM + Dependency-Based Rules	0.76	0.95	0.84	81.06

and 3% negative) of the product reviews test set and 16% (8% positive and 8% negative) of the hotel reviews test set.

Furthermore, we used the most widely used lexicon called SentiFarsNet [158] to compare our dependency-based rules approach with different lexicons. Experimental results show that the SentiFarsNet produced lower accuracy compared to PerSent lexicon. This can be attributed to the relatively smaller size of SentiFarsNet, that consists of 2500 words with polarity, whereas the PerSent lexicon comprises 3500 words.

Finally, dependency-based rules achieved better performance as compared to DNN classifiers on product reviews while DNN models outperformed dependency-based rules on product reviews. This is on account of the PerSent

Table 6.4: Summary of the results for Hotel Reviews

Classifier	Precision	Recall	F-measure	Accuracy
Kim et al.[44]	0.62	0.65	0.62	67.25
Dehkharghani et al.[158]	0.68	0.93	0.78	71.97
fastText Classifier[159]	0.79	0.79	0.79	80.34
SVM	0.70	0.70	0.70	70.72
Logistic Regression	0.70	0.70	0.70	70.75
Dependency-Based Rules	0.80	0.79	0.79	79.25
CNN	0.69	0.92	0.79	82.33
LSTM	0.77	0.90	0.83	85.03
Hybrid 1: CNN + Dependency-Based Rules	0.87	0.91	0.88	85.91
Hybrid 2: LSTM + Dependency-Based Rules	0.87	0.92	0.89	86.29

lexicon consisting of more words related to product reviews, compared to hotel reviews. For example گوشي بہتر "Better phone". However, in hotel reviews, the deep learning classifiers achieved better results as compared to dependency-based rules.

The main limitations with the dependency-based rules approach are as follows:

- Dependency-based rules are unable to detect word sense disambiguation as the multi-word expressions that can help in detecting such disambiguation are absent from the PerSent lexicon. For example, the کرم has a different meaning in Persian "generosity", "worms" and "cream". This types of words cannot be currently detected by the PerSent lexicon.

- The online reviews consist of informal words, idioms and sarcasm that cannot be detected by the PerSent lexicon. As a result the rule-based classifier cannot classify such sentences. In the future, we intend to include these words in PerSent lexicon.

- The dependency-based rules perform poorly on long sentences. For example,

  نگین مشکل ندَاره در خمن اگِه خوَاستِن این گوشیو بگیرین بدون گَارَانتیش بَا
  گَارَانتي هیچ فرقي نميكنه پول الكي به گَارَانتي ندين الَان يك مَاه گوشي ندَارم

  "The phone was great I was really satisfied, after three months the motherboard was burned, I feel sorry for LG, I had Sony long time back, I was very happy with it, in case if you want to buy this mobile, with guarantee without guarantee it does not make any difference."

- The noisy reviews data consists of numerous spelling mistakes that cannot be detected or auto corrected using a dependency-based rules approach. For example, هتل خیلي خوشگزشت, "I had good time in hotel".

  The word خوشگزشت "good time" has spelling mistakes which cannot be detected by the dependency-based rules approach. In order to auto

Table 6.5: Comparison of CNN Layers (Product Reviews)

Layer	Precision	Recall	F-measure	Accuracy	Time
2	0.85	0.67	0.75	72.17	6 m 21 s
3	0.89	0.67	0.77	73.27	8 m 01 s
4	0.91	0.73	0.81	79.37	12 m 42 s
5	0.87	0.74	0.80	78.77	14 m 13 s

Table 6.6: Comparison of LSTM Layers (Product Reviews)

Layer	Precision	Recall	F-measure	Accuracy	Time
2	0.91	0.75	0.81	79.95	4 m 49 s
3	0.91	0.71	0.80	77.77	7 m 08 s

correct spelling mistakes, we need to incorporate a contextual spell checker in the approach.

Table 6.5 presents the comparison results for the proposed CNN with different layers using product reviews. Experimental results show the 4-layer CNN generally outperforms as compared with other layers. The time to train each model is shown in the last column of Table 6.5. Additionally, the precision and F-measure outperformed as compared with other layers. However, the 5-layer recall outperformed as compared with other layers. Table 6.6 presents the comparison results for different layers of LSTM for product reviews. Experimental results show the 2-layer LSTM performs better as compared with other layer. The time to train each model is show in the table. As the number of layers are increased, the time to train each model is increased. Experimental results show that the 2-layer of LSTM achieved better recall, F-measure and accuracy as compared with one layer of LSTM.

Table 6.7: Comparison of CNN Layers (Hotel Reviews)

Layer	Precision	Recall	F-measure	Accuracy	Time
2	0.57	0.88	0.69	75.46	5 m 36 s
3	0.69	0.92	0.79	82.33	6 m 56 s
4	0.75	0.88	0.80	84.93	7 m 23 s
5	0.76	0.80	0.78	79.38	8 m 41 s

Table 6.8: Comparison of LSTM Layers (Hotel Reviews)

Layer	Precision	Recall	F-measure	Accuracy	Time
2	0.78	0.90	0.83	85.06	4 m 21 s
3	0.74	0.92	0.82	84.66	6 m 16 s

Table 6.7 presents the comparison results for the proposed CNN with different layers using hotel reviews. Experimental results show that the 4-layer CNN generally outperforms as compared with other layers. The time to train each model is shown in the last column of Table 6.7. However 3-layer achieved better recall as compared with other layers and 5-layer achieved better precision as compared with other layers. Table 6.8 presents the comparison results for different layers of LSTM for hotel reviews. Experimental results show the 2-layer LSTM performs better as compared with other layer. The time to train each model is show in the table 6.8. Experimental results indicated that the 2-layer of LSTM achieved better precision, F-measure as compared with one layer of LSTM. However, the 3-layer achieved better recall as compared with other 2-layer of LSTM.

We conduct an ablation study to better understand how each part of the system performs in isolation. Table 6.9 and Table 6.10 report the results of ablation studies on product reviews and hotel reviews corpora respectively.

Table 6.9: Ablation Study using Product Reviews Dataset

Rules	**Precision**	**Recall**	**F-measure**	**Accuracy**
Complement Clause	0.59	0.67	0.62	50.32
Adverbial Clause	0.75	0.56	0.64	51.25
Preposition	0.53	0.50	0.51	51.94
Adjective Clause	0.70	0.58	0.63	52.47
Additional rules	0.65	0.57	0.60	55.62
Polarity Inversion	0.65	0.71	0.67	65
Adversative	0.71	0.68	0.69	67.54
Joint Noun and Adjective	0.68	0.73	0.70	68.13

Table 6.10: Ablation Study using Hotel Reviews Dataset

Rule	**Precision**	**Recall**	**F-measure**	**Accuracy**
Complement Clause	0.51	063	0.56	53.17
Adjective Clause	0.55	0.67	0.59	54
Adverbial Clause	0.53	0.62	0.56	54.46
Additional rules	0.61	0.65	0.62	56.53
Preposition	0.60	0.65	0.62	56.55
Polarity Inversion	0.55	0.56	0.54	56.56
Adversative	0.63	0.71	0.66	63
Joint Noun and Adjective	0.71	0.74	0.72	68.99

Experimental results show that the joint-noun and adjective rules achieved better accuracy in both hotel and product reviews datasets as compared to other rules. In addition, the complement clause achieved the lowest performance compared to other rules.

Table 6.11: Selected examples of the Persian dependency-based rules approach

Persian	**English**	**Polarity**
وَاقعًا عَاليه وقتي خريدمش فهميدم ضد ضربه هست كه هيچ مشكلي بَا ويندوز ندَاره	It's really great. When I bought it, I realized it's a counter-hit, which means there's no problem with Windows	Positive
يكي از بهترين انتخَاب هَا بين هَارد هَاي اكسترنَال اين مدل از اي ديتًا هستش	One of the best choices between the external hard drives of this model is the ODATA	Positive
خيلي دَاغ ميكنه در حدي كه دستّون بسوزه	Its getting very hot, so it can damage your hand	Negative
هيچ چيز جديدي ندَاره	It does not have anything new	Negative

Table 6.11 presents some classified sentences by the hybrid approach from product review test set.

6.4 CONCLUSIONS

The quintillion bytes of data generated per day on e-commerce and social media websites, holds valuable information that can be exploited by both buyers to make informed decisions, and also by sellers to take into account past customers issues, in order to improve their products or services. However, the data, mainly consisting of user feedback, cannot be manually read and analyzed by an individual or an organization for gauging public opinion.

Sentiment analysis offers a solution to computationally understand and classify subjective information from user generated feedback. However, current approaches to Persian sentiment analysis are based on word co-occurrence frequencies, that fail to consider the words order and hierarchical relation between words, which are known to play an important role in determining the underlying sentiment. In this chapter, we first propose a novel framework based on Persian dependency-based rules, that consider the dependency relations between keywords, the word order and, individual word polarities to address the aforementioned issues. In addition, we propose a novel hybrid framework for Persian sentiment analysis, that integrates dependency-based rules, and deep neural networks to address the limitation of unclassified sentences, associated with dependency-based rules. A comparative evaluation using benchmark product and hotel reviews corpora demonstrates significant performance improvement of our proposed hybrid framework over state-of-the-art approaches based on SVM, Logistic Regression and advanced DNN models, including CNN and LSTM. In the next, chapter, we propose a novel approach for Persian multimodal sentiment analysis by utilizing text, audio and visual features.

7

CHAPTER 7 COMBINING AUDIO, VISUAL AND TEXTUAL CUES FOR MULTIMODAL SENTIMENT ANALYSIS

In the previous chapter, we explained different approaches for text-based sentiment analysis including dependency-based rules from Persian sentences. To date, most of the work has exploited text modality for sentiment analysis. However, the million hours of video recordings posted on social media platforms (such as YouTube and Facebook) every day also holds vital and unstructured information that can be exploited by organisations to gauge the public perception. On the other hand, the unlimited source of information, consisting of user reviews, cannot be manually viewed and analysed by an organization. Multimodal sentiment analysis offers a solution to computationally understand and harvest sentiments from the videos by developing a model that exploits audio, visual and textual cues. Accordingly, in this chapter, we present a novel Persian multimodal dataset consists of more than 900 utterances, that enables researchers to evaluate multimodal sentiment analysis approaches in the Persian language. In addition, we present a novel multimodal sentiment analysis framework, as shown in Fig. 7.1, that simultaneously exploits available acoustic, visual and textual cues to accurately determine the expressed sentiment. We used both decision-level (late) and feature-level (early) fusion methods to integrate affective cross-modal information. Experimental results demonstrated, textual and visual features significantly improve the performance of the multimodal sentiment analysis framework.

In summary, the paper reports two major contributions outlined below:

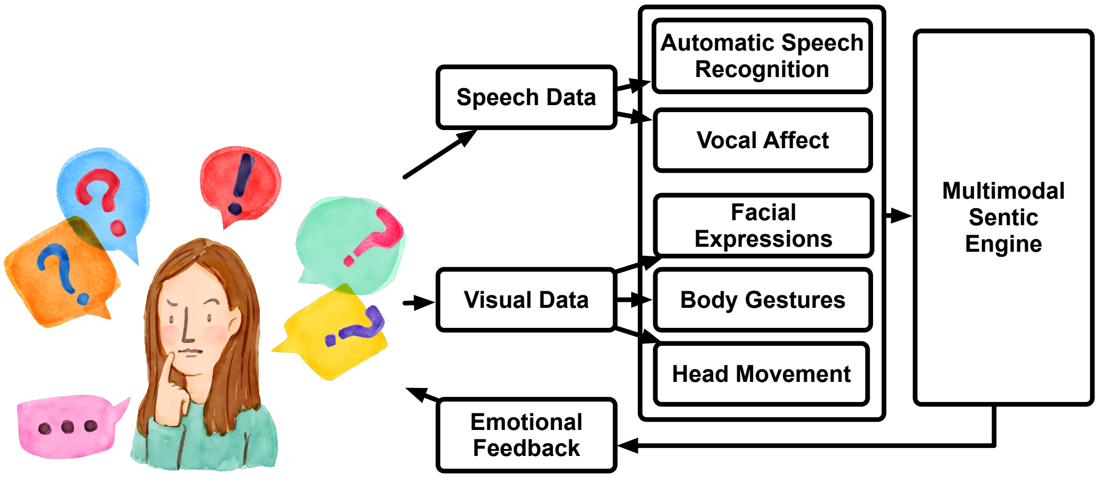

Figure 7.1: Multimodal Overview

1. Proposed a novel deep learning model based on Persian multimodal sentiment analysis. Specifically a stacked LSTM based data-driven model is proposed to approximate the textual features.

2. Proposed a first of kind Persian multimodal dataset collected from YouTube. The dataset consists of 91 videos.

The rest of the chapter is organized as follows: Section 7.1 presents the proposed novel approach for Persian multimodal sentiment analysis. Section 7.2 presents a novel Persian multimodal dataset. In section 7.3 experimental results are presented. Finally, section 7.4 concludes this chapter.

7.1 METHODOLOGY

In this section, the proposed multimodal sentiment analysis framework, as shown in Fig 7.2, is discussed. The audio, visual and textual features were first extracted and the extracted features, consisting of affective information, were fused to identify the overall polarity of the video.

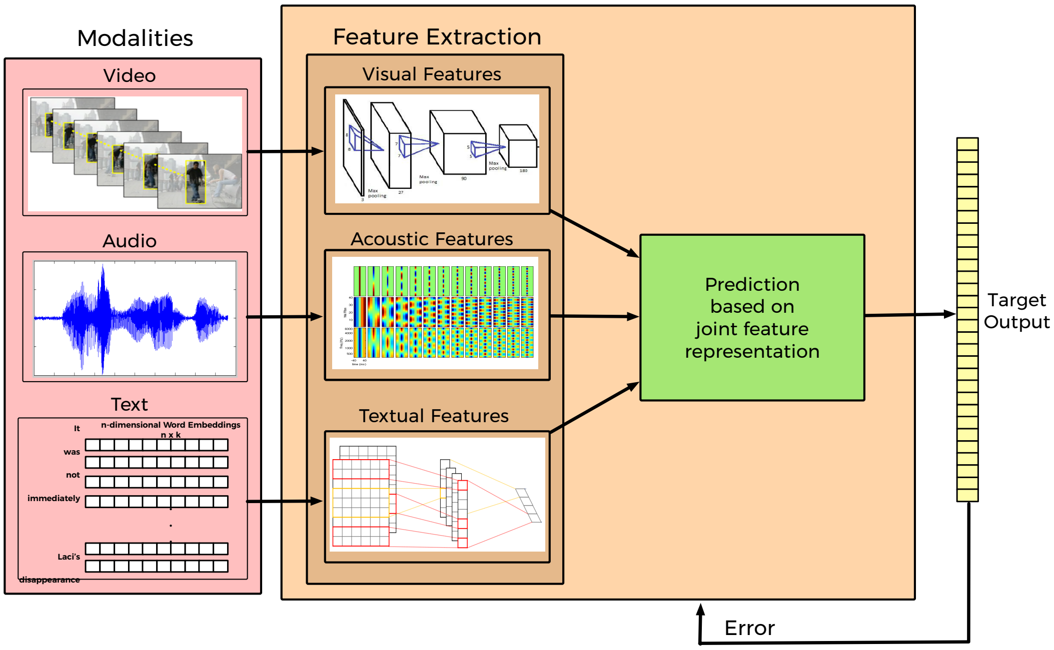

Figure 7.2: Overview of the Multimodal Sentiment Analysis Framework

7.1.1 *Unimodal Feature Extraction*

In this section, we discuss about the unimodal feature extraction techniques.

7.1.1.1 *Textual Feature Extraction: text-BiLSTM*

For extracting features from textual modality a stacked bidirectional LSTM (BiLSTM) model, as shown in Fig. 7.3 has been used. Each utterance is represented as a concatenation of 300 dimensional pretrained fastText word embeddings. Each utterance is either trimmed with a window of size 60 words or zero padded at the end to form a vector of dimension 60x300. The converted vectors were fed into stacked BiLSTM model. The model consists of two bidirectional LSTM layers with 128 cells each. The output of the last bidirectional LSTM is concatenated and fed to a fully connected layer with 128 neurons (ReLU activation) and 2 neurons (Softmax activation) respectively.

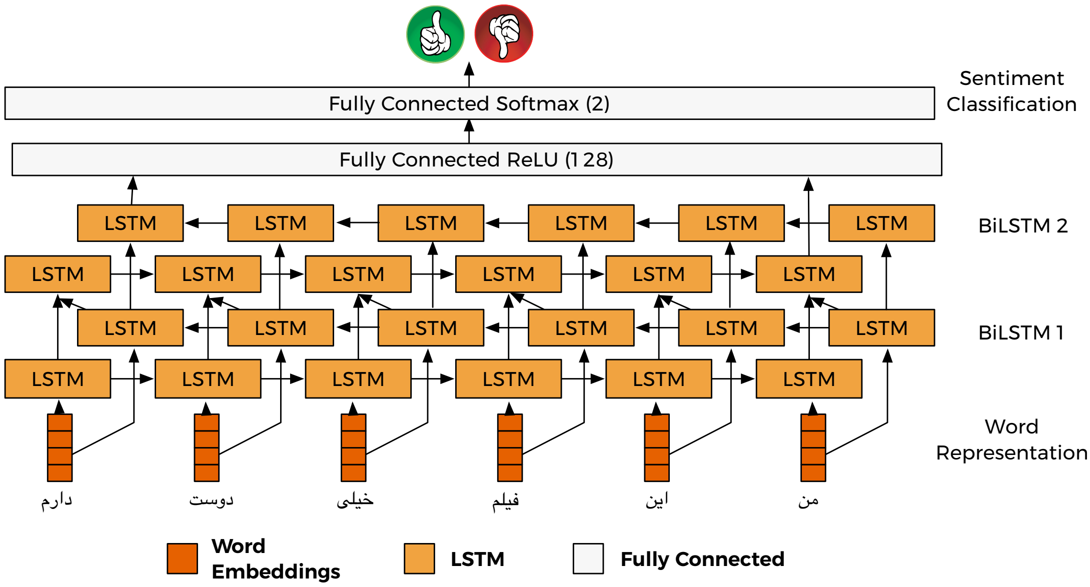

Figure 7.3: Bidirectional LSTM architecture: Textual-cues based Sentiment Analysis

The network learn the levels of abstract representations and implicit semantic information, that spans over the entire utterance.

7.1.2 *Audio Feature Extraction: openSMILE*

The audio features are automatically extracted from the speech of each utterance using a widely used OpenSMILE software. The features are extracted at a frequency of 40 samples per second. The extracted features consist of the following acoustic sub-features:

- Prosody feature: This feature consists of intensity, loudness and pitch that describes the speech signal in terms of amplitude and frequency.
- Energy features: The energy feature depicts the human loudness perception.
- Voice probabilities: The voice probabilities provides an estimate of percentage of voiced and unvoiced energy in the audio.

Table 7.1: MLP Architecture

Layer	1	2	3	4
Type	Relu	ReLU	ReLU	ReLU
Neurons	1024	512	128	1

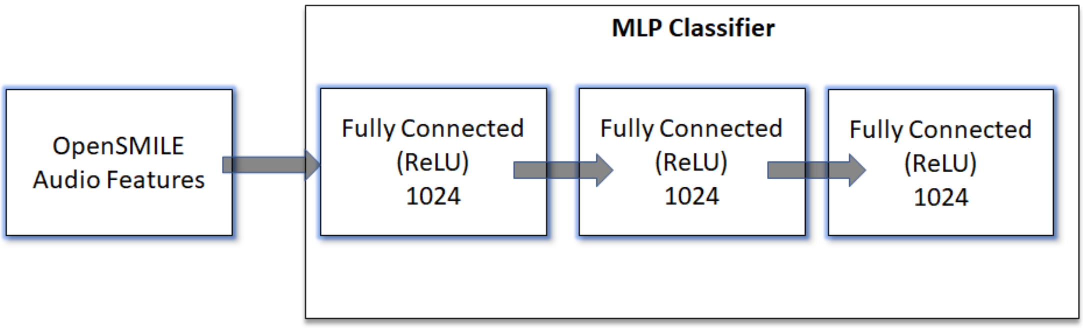

Figure 7.4: Acoustic cues based Sentiment Analysis

- Spectral features: The spectral features are based on the characteristic of human ear, which uses a nonlinear frequency unit to simulate the auditory system

- Cepstral features: The cepstral features emphasize the changes in the spectrum features measured by frequencies, the 12-model Mel-frequency cepstral coefficients used which is calculated based on the Fourier transform of a speech fame.

The overall audio features for a single utterance consists of 6373 features. Speaker normalization is performed using z-standardization. The voice intensity is a threshold to identify the samples with and without speech. The features are averaged over all the frames in an utterance, to obtain a feature vector for each utterance. The audio feature extraction framework is shown in Fig 7.4. The MLP architecture, as depicted in Table 7.1, is used to exploit the extracted audio feature for determining the opinion strength based on acoustic cues. The architecture of used MLP is shown in Fig.

7.1.3 *Visual Feature Extraction: 3D-CNN*

Human expression play a significant role in identifying the emotion expressed in day-to-day conversations. The facial expressions helps in decoding the expressed affect by providing visual cues. Therefore, the visual features are important in multimodal sentiment analysis. The Facial Action Coding System (FACS) is a system used for measuring and describing facial behaviours. According to [160], facial behaviour can be categorized into 64 action units. The Computer Expression Recognition Toolbox (CERT) is employed to automatically extract the following visual features:

Smile and head pose estimates: The smile feature depicts the probability of a person smiling given an image. Head pose detection consists of three dimensional head orientation yaw, pitch and roll. These dimensions provide information about the face position while expressing positive or negative opinions.

Facial action units: The facial action units estimate the thirty related muscle movements related to eyes, nose, eyebrows and chin. This feature provides information about facial behaviours which can be exploited to find differences between positive and negative opinions.

The visual features are extracted from videos using a 3D Convolutional Neural Network (CNN). The 3D-CNN exploits both spatial and temporal patterns to accurately find the spatio-temporal association between a subjective and an objective utterance. In our experiments, the best results are obtained with a 9-layered 3D-CNN architecture as illustrated in Fig 7.5. The architecture details are presented in Table 7.2

Table 7.2: 3D-CNN Architecture: Visual cues based Sentiment Analysis)

Layer	**Type**	**Feature Map**	**Kernel**
1	Convolutional3D	16	2 x 2 x 2
2	Convolutional3D	32	2 x 2 x 2
3	Max pooling3D		1 x 2 x 2
4	Convolutional3D	64	2 x 2 x 2
5	Max Pooling3D		2 x 2 x 2
6	convolution3D	64	2 x 2 x 2
7	Max pooling3D		1 x 2 x 2
8	Fully connected	5000	
9	Fully connected	500	
10	Fully connected	2	

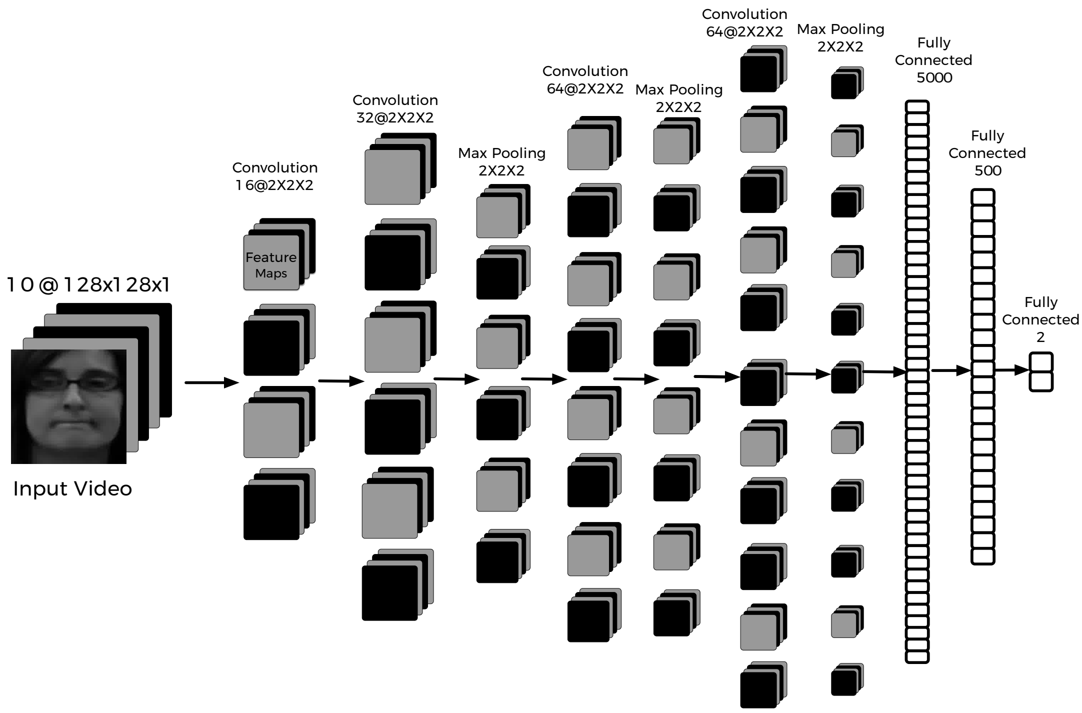

Figure 7.5: 3D-CNN based Visual Feature Extraction

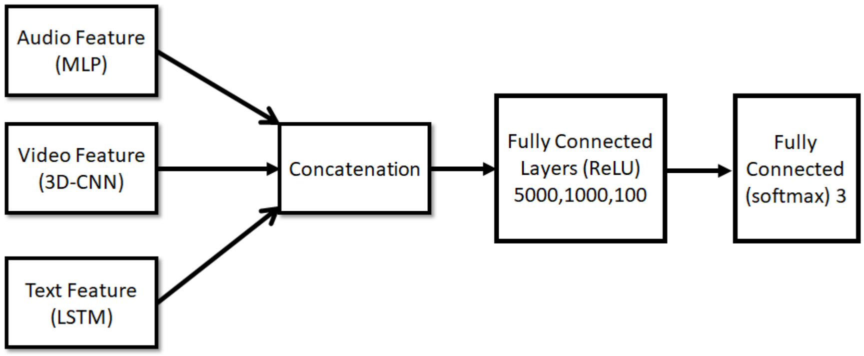

Figure 7.6: Early (Feature level) Fusion

7.1.4 *Multimodal Fusion*

In this section, we discuss about the multimodal fusion techniques. The multimodal fusion is process of integrating features collected from different modalities for sentiment analysis task. The multimodal fusion technique can be divided into early or feature-level fusion and late or decision-level fusion. The multimodal systems often outperform unimodal because the correlation and discrepancies between modalities often help in achieving superior [161].

7.1.4.1 *Feature-level (Early) Fusion*

In the early fusion, first, the features are extracted from the input modalities using either deep neural networks or state-of-the-art feature extraction. The input features are concatenated and fed into a classifier. For example, the audio features are extracted using OpenSMILE software, visual features are extracted using 3D-CNN and text features are extracted using BiLSTM . The features are fused and fed into a MLP classifier. The main advantage of feature-level fusion is the cross-correlation between multiple modalities at an early stage helps in achieving better performance and the main disadvantage of early fusion is that the modalities must be tightly time synchronized, because the incorrect time synchronization could lead to an improper functioning system as the model will be unable to learn any cross-modal correlation [162].

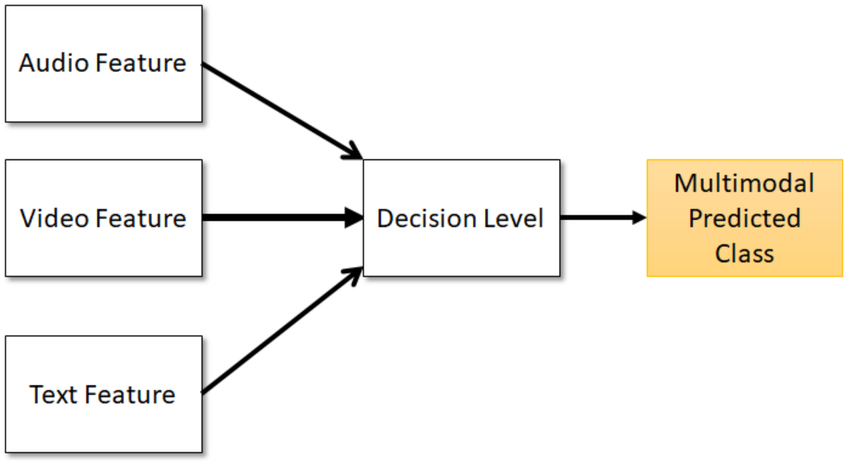

Figure 7.7: Late (Decision-Level) Fusion

7.1.5 *Decision-Level (Late) Fusion*

In the late or decision-level fusion, unimodal classifiers are used to identify the prediction for each modality. The local predictions are concatenated and is further classified to achieve the final decision. The advantage of late fusion is that, the sampling rate at which the predictions are generated is same, therefore the extracted predictions can be easily concatenated without up-sampling or down-sampling. However, the disadvantage is that the model cannot learn the cross-modal correlation[163]. The Fig display the late fusion feature-level.

7.2 PERSIAN MULTIMODAL DATASET

In this section, we introduce the novel Persian multimodal dataset:

7.2.1 *Dataset Collection*

The YouTube website was used to collect Persian videos with focus on product, movie and music reviews. videos were found using the following keywords

Figure 7.8: Example snapshots of videos from our new Persian multimodal dataset

such as: نقد فيلم ("Movie critize"), محصولَات ايرَاني ("Persian product"), نقد موزيك ("Critize music").

There are total of 91 videos were selected that respected the following guidelines:

- The video must contain only one speaker
- The speaker should look directly in the camera
- The speaker face should be clearly visible
- There should not be any background noise or music in the recording
- The video should be recorded with high quality microphone and camera

While in the collected videos, the speaker has similar distance from camera, the background and lighting is variable between different videos. The length of the video is between 1-5 minutes. In addition, each video consists of 10–50 utterances. The dataset includes 10 male and 6 female speakers between 20 to 60 years old. Sample snapshots of our Persian multimodal dataset are shown in Fig 7.8.

7.2.2 *Segmentation and Transcription*

All the collected videos are transcribed manually with utterance start and end times. The transcription process consists of two stages. First, an expert transcriber manually transcribed all the videos. The transcripted text was reviewed by two other native speakers. Second, the transcriptions were divided into utterances using pause details such as أُه هم (hm, oh, and etc.).

Each video was segmented into average ten utterances, resulting in a final dataset of 945 utterances with 754 subjective utterances and 191 objective utterances. Each of the utterance is linked to audio and video stream as well to manual transcription. The utterances have an average duration of 6 seconds.

Three expert annotators categorized segmented utterances into subjective (utterances with expressed emotion) and objective (utterances with no polarity for e.g. fact, figures etc.) category. The subjective segmentation is important to achieve fine-grained sentiment analysis. The utterance is categorized into subjective category, if the sentence is carrying an opinion, belief, thoughts, feeling or emotions. The three rules used to identify the subjectivity in the sentence are outlined below:

- Explicitly criticising an entity. For example: كه يك فيلم كمدي طنز خوب فرَاهم ميكنه لهجه هَاي خوب بري مخَاطب, "The comedy movie which is really good and the viewers have good time when they are watching the movie".

Table 7.3: Persian Multimodal dataset statistics

Total number of positive segmented	373
Total number of negative segmented	381
Total number of subjective	754
Total number of objective	191
Total number of unique word in the dataset	4065
Total number of speakers	16

- Referencing an opinion expressed by a third person. For example, امّا رَاضي نيستند ديدن فيلم از منتقدين, "The movie critic are not satisfying with watching this film".
- Implicitly expressing a subjective opinion. که این فیلم بَا من پیشنهَاد نمیکنم خَانوَاده نگاه کنید, "I am not recommended to watch this movie with family".

Detailed statistics of the dataset can be found in Table 7.3.

7.2.3 *Sentiment Polarity*

The utterances were annotated by three native Persian speakers between 30 to 50 years old. The annotators had three choices, positive (+1), neutral (0), negative (-1). The polarity assignment included all three modalities (visual, audio and text). Table 7.4 shows the example of utterances obtained from one of the videos in the multimodal dataset, along with their translation and polarity. It can be observed that, a single video consist of both positive and negative utterances.

Table 7.4: Example utterances from Persian Multimdal Sentiment Analysis Dataset

Persian Sentence	**English Translation**	**Polarity**
بعد چند سَال شَادمهر يك كَار خوب منتشر كرد به اسم بَا تو عشقم	After few years, Shadmehr release good work called "with you my love	1
بَازيگر دوست دَاشتني رضَا عطآرَان مَا رو به ديدن فيلم ترغيب ميكنه	The good acting of Reza Attaran help to people watch the movie	1
ارشَاد قصد مجوز به فيلم ندَاره	The ministry of culture did not give the permission to the movie	-1
تو دورَان پهلوي يك مدت به خَاطر ترَانه سيَاسي زندَان بود	During Pahlavi Dynasty he was jailed for few years	-1

Furthermore, the manual gesture including smile, frown, head node and head shake were annotated manually to study the relation between words and gestures. The annotation was simply carried out by marking the utterances into these gestures. Expert coders manually annotated each utterance with gesture information. The average agreement of the gestures was 89.23%.

7.3 EXPERIMENTAL RESULTS

In this section, the experimental results for Persian multimodal dataset are discussed in details. In this study, we focus on identifying affective strengths as compared to finding if an utterance is subjective or objective. Therefore, we removed all objective utterances for training the multimodal sentiment analysis framework.

Table 7.5 summarizes the results for dependency-based rules based sentiment analysis, as proposed in section 6.1, for the transcribed utterances. As discussed in the aforementioned section, the deep learning based classifier is used to classify the unclassified sentences (the sentences which cannot be categorized by dependency-based rules). The sentence is converted into a 300-dimensional vector using fastText and the concatenation of word embedding is fed into deep learning classifiers. Alternatively, the sentences are converted into Bag-of-words and fed into logistic regression and SVM. The dataset is divided into 60% train set, 10% validation set and 30% testing set. The CNN, LSTM, SVM and LR are trained on the train set, tuned on validation set and evaluated on the test set. Experimental results show that the hybrid 2: LSTM + Dependency-based achieved better accuracy as compared with other approaches including DNN based classifiers.

Table 7.6 presents the comparative simulation results for validation set using proposed CNN with different number of hidden layers using the annotated utterance transcriptions for multimodal dataset. Experimental results show that the 4-layered CNN outperforms other CNN architectures.

Table 7.7 presents the comparison results for proposed LSTM with different layers using transcribed text from multimodal dataset. Experimental results show that the 2-layered LSTM achieved better accuracy, precision and F-measures as compared with other architectures.

The results of text-based, audio-based and video-based sentiment analysis is summarized in Table 7.8, Table 7.9 and Table 7.10 respectively. Experimental results show that the text-based classifiers achieved better accuracy as compared to audio-based and video-based sentiment analysis models.

Table 7.5: Comparison of results (Text Modality)

Classifier	**Precision**	**Recall**	**F-measure**	**Accuracy**
Kim et al. [44]	0.61	0.63	0.61	61.24
Dehkharghani et al. [158]	0.78	0.82	0.79	70.12
fastText Classifier [159]	0.67	0.67	0.67	70.01
SVM	0.65	0.65	0.65	65.01
Logistic Regression	0.64	0.64	0.64	64.23
Dependency-based Approach	0.83	0.93	0.87	75.94
CNN	0.91	0.63	0.75	68.53
LSTM	0.92	0.83	0.87	86.14
Hybrid 1: CNN + Dependency-Based	0.78	0.77	0.77	76.16
Hybrid 2: LSTM + Dependency-Based	0.86	0.95	0.85	88.01

Table 7.6: Comparison of CNN Layers (Text)

Layer	**Precision**	**Recall**	**F-measure**	**Accuracy**	**Time**
2	0.59	0.52	0.54	52.05	4 m 5 s
3	0.94	0.59	0.73	64.04	4 m 21 s
4	0.91	0.63	0.75	68.53	6 m 13 s
5	0.62	0.65	0.64	63.67	7 m 12 s

Table 7.9 shows the results for audio-based sentiment analysis. Experimental results demonstrated that the positive received better recall as compared to negative features. However, the negative features achieved better F-measure as compared positive features.

Table 7.10 display the results for video-based sentiment analysis. As discussed, the 3D-CNN is used to extract features from videos. Experimental

Table 7.7: Comparison of LSTM Layers (Text)

Layer	Precision	Recall	F-measure	Accuracy	Time
1	0.85	0.81	0.83	82.39	2 m 22 s
2	0.92	0.83	0.87	86.14	2 m 54 s
3	0.89	0.83	0.86	85	3 m 46 s

Table 7.8: Prediction Results: Text-Based Sentiment Analysis

	Precision	Recall	F-measure	Accuracy
Positive	0.92	0.83	0.87	
Negative	0.88	0.94	0.91	
Average	0.90	0.88	0.89	89.24

Table 7.9: Prediction Results: Audio-Based Sentiment Analysis

	Precision	Recall	F-measure	Accuracy
Positive	0.78	0.84	0.81	
Negative	0.78	0.82	0.84	
Average	0.78	0.83	0.82	82.79

Table 7.10: Prediction Results: Video-Based Sentiment Analysis

	Precision	Recall	F-measure	Accuracy
Positive	0.76	0.85	0.80	
Negative	0.87	0.79	0.83	
Average	0.81	0.82	0.81	81.72

results show that the negative features achieved better precision and F-measure as compared to positive. However, the positive received better recall as compared to negative.

Table 7.11: Prediction Results: Late Fusion

Modality		**Precision**	**Recall**	**F-measure**	**Accuracy**
A+V	Positive	0.78	0.79	0.79	
	Negative	0.84	0.83	0.83	
	Average	0.79	0.78	0.78	81.18
V+T	Positive	0.89	0.89	0.89	
	Negative	0.91	0.91	0.91	
	Average	0.88	0.88	0.88	90.32
A+T	Positive	0.84	0.93	0.88	
	Negative	0.94	0.87	0.90	
	Average	0.92	0.84	0.88	89.24
A+V+T	Positive	0.88	0.90	0.89	
	Negative	0.92	0.90	0.92	
	Average	0.90	0.87	0.89	90.32

Table 7.11 summarized the results for late or decision-level fusion. Experimental results show that the A + V + T and V + T modalities achieved better accuracy as compared to other modality combinations. However, the A + T modality achieved better precision as compared with other modalities. In addition, the V + T modality achieved better recall as compared with other modalities and A + V + T achieved better F-measure as compared with other modalities. In addition, the experimental results show that A + V modality achieved the least accuracy as compared to other modality combinations.

Table 7.12 summarised the results for early or feature-level fusion. Experimental results show that the A + V + T and V + T modalities achieved better accuracy as compared with other modalities. However, the A + T modality achieved better precision as compared with modalities. In addition, the V + T

Table 7.12: Prediction Results: Early Fusion

Modality		**Precision**	**Recall**	**F-measure**	**Accuracy**
A+V	Positive	0.82	0.79	0.81	
	Negative	0.84	0.87	0.85	
	Average	0.79	0.82	0.80	83.33
V+T	Positive	0.90	0.89	0.89	
	Negative	0.92	0.92	0.92	
	Average	088	0.90	0.89	90.86
A+T	Positive	0.91	0.88	0.89	
	Negative	0.91	0.93	0.92	
	Average	0.87	0.91	0.89	90.86
A+V+T	Positive	0.85	0.98	0.91	
	Negative	0.98	0.87	0.92	
	Average	0.97	0.84	0.90	91.39

modality achieved better recall as compared with other modalities. Moreover, the experimental results show that the A + V achieved the least accuracy as compared with other modalities.

Fig 7.9 presents the accuracy of unimodal sentiment analysis models for text, audio and video. It can be seen that, the text achieved better accuracy as compared to audio and video modalities.

In our experiments, all the possible fusion combination including A + V, A + T, T + V, A + V + T were considered. Fig 7.10 presents the accuracy of early and late multimodal fusion. Experimental results show that the early fusion outperformed late fusion. In addition, the A + V + T modality in early fusion achieved the highest accuracy as compared to late fusion and other early fusion combinations.

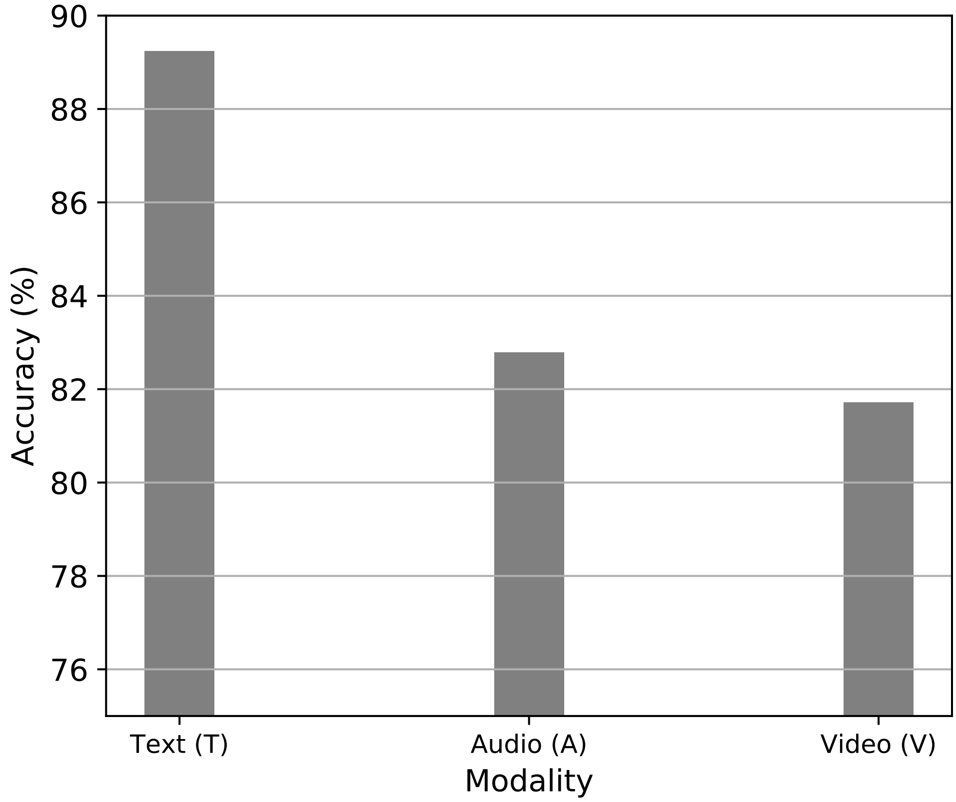

Figure 7.9: Comparison of proposed Unimodal

The main limitations with the proposed Persian multimodal sentiment analysis framework are outlined below:

- The Persian multimodal sentiment analysis framework cannot detect subjective/objective utterances. For example, بریم قسمت‌هاي از این فیلم ببّینیم و برگردیم, "Let's go and see some part of the movie and come back". The proposed approach is unable to detect objective utterances.

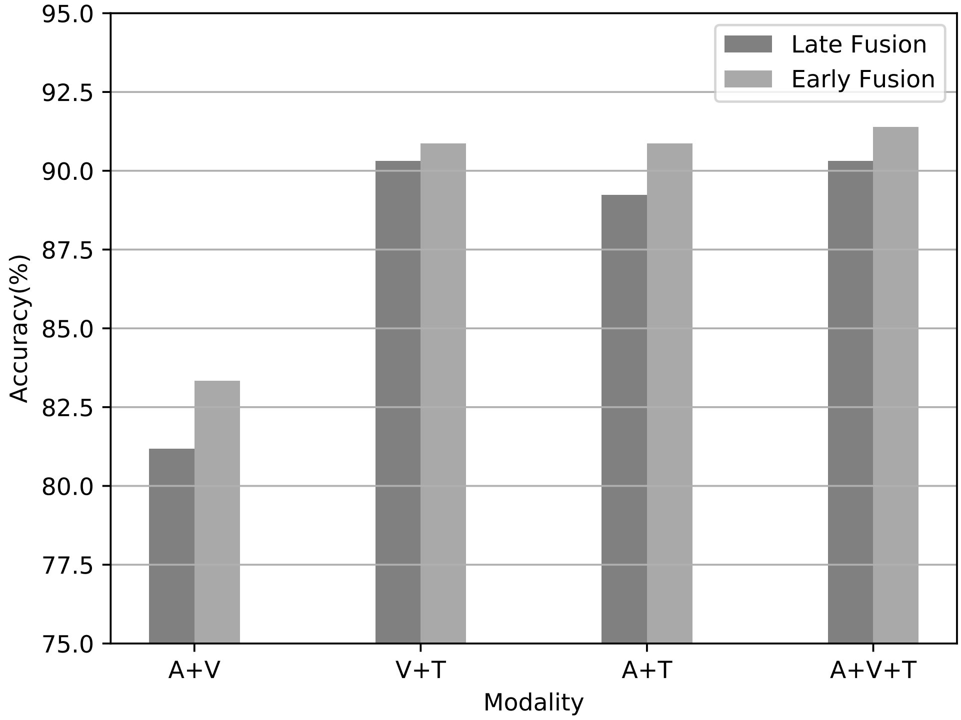

Figure 7.10: Feature-level vs Decision-level Multimodal Fusion approaches: Comparison of prediction accuracy, precision, recall and f1-score

- The speakers in the videos are not speaking formally i.e. most of the videos consists of informal words to express their opinion which is difficult for the proposed approach to detect. For example, يَاسر بختيَاري بَا موزيك رپش حرف دل جوَانهَا رو ميزنه و خيلي تَاثير گذر هست "Yasr Bakhtiari with his rap music saying what young people wants and his music is really effective".

- The current model is restricted to a single speaker and does not work in multi-speaker scenarios. For example, when the video consists of two speakers it cannot detect the polarity of the sentences which they are expressing.

- The proposed approach is unable to detect the polarity of the sentence, when background noise or music is present in the video.

In text-based sentiment analysis, the only source of information consists of n-gram, word-order, dependency relation and part-of-speech features which sometime prove inadequate to identify the overall polarity of the sentence. Instead, the video consists of multiple modalities including text, audio and visual features. In this chapter, we presented a novel multimodal dataset for Persian consisting of utterances along with their sentiment polarity extracted from YouTube videos. In addition, a novel mutlimodal sentiment analysis framework for combining audio, visual and textual features is proposed. Experimental results show that the use of multiple modalities especially visual and textual can significantly improve the performance of sentiment analysis model. Finally, it can be seen from the experimental results that the fusion of audio, visual and textual features outperformed all other unimodal classifiers including text-based, audio-based and video-based sentiment analysis models.

8

CHAPTER 8: CONCLUSION AND FUTURE WORK

In this research thesis, we proposed a novel framework based on dependency-based rules and deep learning to detect polarity in the Persian sentences. Specifically, we developed dependency-based rules for Persian to extract concepts from text based on the dependency parser. The concepts were extracted based on the linguistic patterns, which were extracted from dependency trees of the sentence. Afterwards, the PerSent lexicon was used to assign polarity into concepts and identify the overall polarity of the sentence. To evaluate the performance, deep learning classifiers such as CNN and LSTM were used. Comparative simulation results revealed that the proposed dependency-based rules outperformed state-of-the-art SVM, LR, and fastText based classifiers (by a margin of 10-15%) and performed comparably to deep learning classifiers (including CNN and LSTM).

The proposed framework was then evaluated with multimodal inputs incorporating audio, video, and text. The empirical results showed that the proposed approach can effectively improve polarity detection when text, audio and video cues were fused. Particularly, the experimental results revealed that the fusion achieved accuracy up to 91.39% as compared to unimodal approaches.

The proposed SA framework was then used to validate a novel Persian lexicon (labelled corpus), which was developed to address the limited available Persian corpora. The developed PerSent lexicon consists of 1500 words along with their polarity and part-of-speech tag. The lexicon was evaluated with Persian product reviews dataset. The experimental results demonstrated the effectiveness of the proposed lexicon for the Persian language. The obtained results demonstrated that the SVM achieved accuracy of 75.23% on Persian product reviews corpus.

Furthermore, a context-aware deep learning approach for Persian sentiment analysis was developed. The proposed approach automated feature engineering and classified the Persian movie and hotel reviews into positive and negative sentiments. The deep learning classifiers such as CNN, LSTM (stacked and Bidirectional) were applied. The obtained results were compared with manual feature engineering and with SVM and Logistic Regression, where the stacked-bidirectional-LSTM achieved the highest accuracy up to 95.61% as compared to 89.37% and 88.93% achieved by SVM and LR, respectively.

8.1 SOCIAL IMPACTS

The primary goal of this research was to develop novel techniques for Persian sentiment analysis. We believe that this research is likely to have an impact on both industry and consumers.

Impact on Industry: It is important for organisations to understand public perception towards their products and services. Most of the current approaches focus on English sentiment analysis. However, with the growth of the Internet, people from all around the world express their opinions in different languages. Therefore, it is vital for companies to develop a system to detect polarity in a different language, especially in Persian. The proposed Persian approach can help the industry to understand customers behaviours towards their product and services.

Impact on consumer: Currently, whenever, consumers intend to buy a product or service, they always search customer reviews over the Internet. However, reading customer reviews manually is not an easy task. The proposed Persian sentiment analysis approach helps customers to make a decision before purchasing the product.

8.2 ADVANTAGES

- The proposed approach on extracting concepts in the context of Persian sentiment analysis goes beyond the Persian text by focusing on the hierarchical semantic relation between two concepts. Identify the semantic dependencies in Persian allow us to assign polarity into the sentence.
- The role of dependency-based rules to identify the overall polarity of the sentences are ignored by most of the current word co-occurrence frequency based approaches in sentiment analysis. Developing and using grammar rules based on dependency parser jointly exploit the word order, hierarchical dependency relation and word polarities. The proposed approach shows that integrating dependency-based rules and deep learning classifiers achieves superior performance as compated to state-of-the-art approaches including deep neural networks.
- Mutlimodal sentiment analysis: Most of the works in this field has been carried out in English text. In order to enhance the text modality, we collected first of its kind Persian multimodal dataset consisting of over 900 utterances. In addition, we proposed an deep learning driven fusion approach for combing the audio, visual and textual cues.

8.3 LIMITATIONS

The proposed approaches has the following limitations:

- The dependency-based rule techniques presented in this thesis only works in the case of sentence level sentiment analysis and cannot be used for document level. For example, it can be used to identify the polarity of a tweet but not a long paragraph.
- The dependency-based rules developed in this thesis, such as grammar rules are language-dependent, i.e. Persian specific. Developing grammar rules for other languages is not addressed in this thesis.
- Some of the proposed techniques that use deep learning classifiers are computationally expensive.
- The dependency-based rules can be applied to Persian sentences with large number of grammatical errors. Since, the dependency-based rules are based on grammatical sentences and it required to parse the sentences before trigger the rules.
- The proposed approaches are not detecting different dialects for Persian language.
- The proposed Persian multimodal approach was tested and evaluated on a small dataset.
- The current multimodal sentiment analysis works in a special case of single speaker talking. It is unable to detect polarity when multiple speakers are present in the video.

8.4 FUTURE WORK

The following aspects of this research work are planned to be carried out in future:

Multilingual Sentiment Analysis: The current work is limited to be only Persian language. In future, we will be extending the applicability to other languages. Hence, we aim to extend our research towards multilingual sentiment analysis by developing multilingual dependency-based rules. In order to extend our work to multilingual sentiment analysis, we plan to build a corpus for various languages such as Urdu, Turkish, Kurdish, Arabic and French and train deep learning classifiers to evaluate the performance of the approach.

Multilingual Lexicon: We plan to extend our lexicon to a wider variety of tasks such as sarcasm, idiom, dialects etc. In addition, we plan to develop a lexicon for multilingual sentiment analysis. Furthermore, in order to build a multilingual lexicon, we are required to collect words and phrases online and assign polarity into the words that are a part of our ongoing/future work.

Detecting idiom and sarcasm: Detecting sarcasm is very important in sentiment analysis because idioms and sarcasm can change the overall polarity of the sentence. Most of the current approaches for detecting sarcasm have used English and Spanish corpora. However, there is not any research on Persian. In future, we intend to develop an approach to detect sarcasm in Persian language by developing linguistic rules that exploit the semantic relationship between words.

Multilingual Dependency-based rules: We plan to develop dependency-based rules to extract multilingual concepts from a mixture of Persian and English sentences. Figure 8.1 displays the proposed approach to detect a mixture of English and Persian sentences. The dependency parser used to extract concepts from sentences and SenticNet and PerSent lexicons are used to assign polarity into extracted concepts and finally, machine learning classifier is used to evaluate the performance of the approach.

Dependency-based rules using Graph Neutral Network: The proposed approach on dependency-based rules are developed manually to extract con-

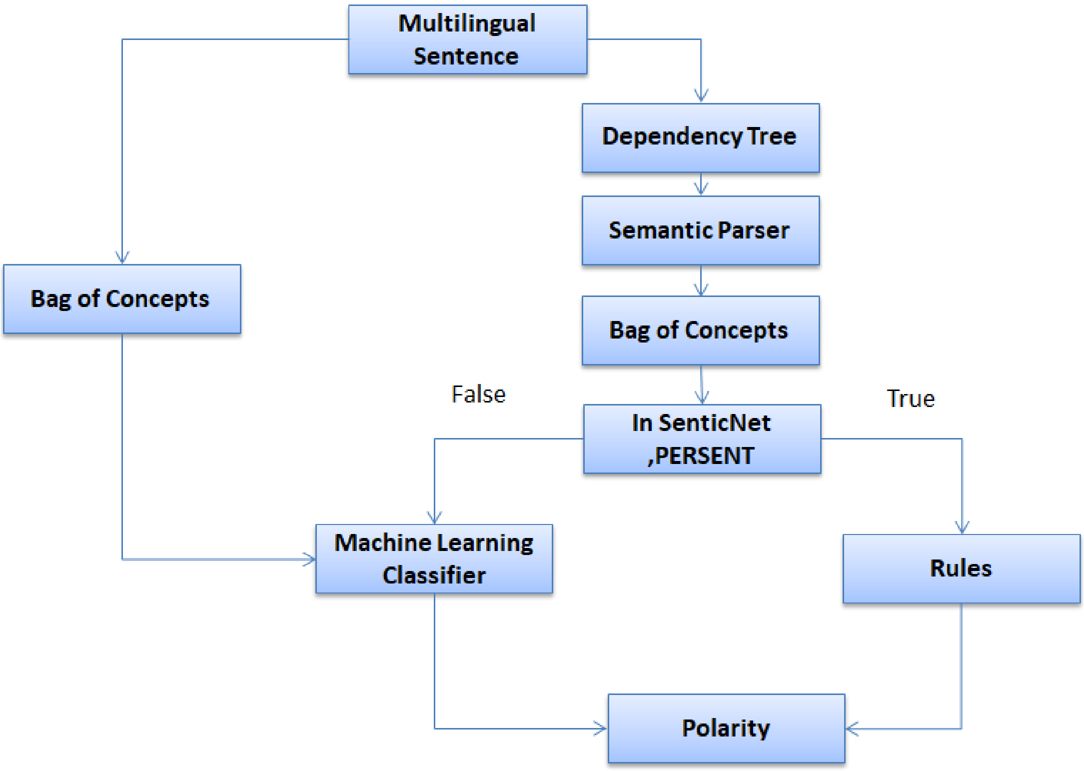

Figure 8.1: Mixture of English and Persian Dependency-based rules

cepts and assign polarity. As ongoing work, we plan to exploit Graph Neural Networks to automatically learn dependency-based rules from multilingual corpus.

Detecting Writing Styles: The proposed sentiment analysis approach is limited to the modern Persian language. However, there are different writing styles such as Persian/Afghani (Dari) accent or Persian Kurdish and Persian Azeri accent, that can be exploited to contextually determine the polarity of the sentence.

Exploiting Transliterated Persian Texts: The proposed sentiment analysis approach is limited to detect Persian sentences and it is unable to detect transliterated Persian sentence. The transliterated sentence are those which are written using English characters but they contains Persian pronunciations. For example, "film khob bod" (The movie was good).

Document-Level sentiment Analysis: The proposed sentiment analysis approach is limited to detecting polarity in Persian sentences. However, there are cases where we are required to detect polarity at a document level. As future work, we plan to detect polarity in document level.

Multilingual Multimodal Sentiment Analysis: There is not any approach available to identify the polarity for multilingual multimodal sentiment analysis. As future work, we plan to propose an approach to detect polarity in multilingual multimodal sentiment analysis.

BIBLIOGRAPHY

[1] A. Esuli and F. Sebastiani, "Sentiwordnet: A publicly available lexical resource for opinion mining." in *LREC*, vol. 6. Citeseer, 2006, pp. 417–422.

[2] P. N. Kieu, H. S. Tran, T. H. Le, T. Le, and T. T. Nguyen, "Applying multi-cnns model for detecting abnormal problem on chest x-ray images," in *2018 10th International Conference on Knowledge and Systems Engineering (KSE)*. IEEE, 2018, pp. 300–305.

[3] E. Cambria, S. Poria, D. Hazarika, and K. Kwok, "SenticNet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings," in *AAAI*, 2018, pp. 1795–1802.

[4] Z. Li, Y. Fan, B. Jiang, T. Lei, and W. Liu, "A survey on sentiment analysis and opinion mining for social multimedia," *Multimedia Tools and Applications*, vol. 78, no. 6, pp. 6939–6967, 2019.

[5] P. Chiranjeevi, D. T. Santosh, and B. Vishnuvardhan, "Survey on sentiment analysis methods for reputation evaluation," in *Cognitive Informatics and Soft Computing*. Springer, 2019, pp. 53–66.

[6] R. S. Jagdale, V. S. Shirsat, and S. N. Deshmukh, "Sentiment analysis on product reviews using machine learning techniques," in *Cognitive Informatics and Soft Computing*. Springer, 2019, pp. 639–647.

[7] J. Serrano-Guerrero, J. A. Olivas, F. P. Romero, and E. Herrera-Viedma, "Sentiment analysis: A review and comparative analysis of web services," *Information Sciences*, vol. 311, pp. 18–38, 2015.

[8] B. Liu and L. Zhang, "A survey of opinion mining and sentiment analysis," in *Mining text data*. Springer, 2012, pp. 415–463.

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

[10] M. S. Vohra and J. Teraiya, "Applications and challenges for sentiment analysis: A survey," *International journal of engineering research and technology*, vol. 2, no. 2, pp. 1–6, 2013.

[11] C. Zhang, Y.-X. Tian, Z.-P. Fan, Y. Liu, and L.-W. Fan, "Product sales forecasting using macroeconomic indicators and online reviews: A method combining prospect theory and sentiment analysis," *Soft Computing*, pp. 1–14, 2019.

[12] L. Gou, M. Hu, Y. Li, and H. Yang, "Mining product aspects from opinion text," May 7 2019, uS Patent App. 10/282,467.

[13] T. Gadekallu, A. Soni, D. Sarkar, and L. Kuruva, "Application of sentiment analysis in movie reviews," in *Sentiment Analysis and Knowledge Discovery in Contemporary Business*. IGI Global, 2019, pp. 77–90.

[14] S. J. S. Chowdary, S. W. Farooq, S. A. Siddique, and S. Dinesh, "Sentiment analysis of movie reviews using microsoft text analytics and google cloud natural language api," 2019.

[15] M. Anandarajan, C. Hill, and T. Nolan, "Sentiment analysis of movie reviews using r," in *Practical Text Analytics*. Springer, 2019, pp. 193–220.

[16] M. Karamibekr and A. A. Ghorbani, "Sentiment analysis of social issues," in *2012 International Conference on Social Informatics*. IEEE, 2012, pp. 215–221.

[17] M. T. Khan, M. Durrani, A. Ali, I. Inayat, S. Khalid, and K. H. Khan, "Sentiment analysis and the complex natural language," *Complex Adaptive Systems Modeling*, vol. 4, no. 1, p. 2, 2016.

[18] A. Montejo-Raez, M. C. Díaz-Galiano, F. Martinez-Santiago, and L. Ureña-López, "Crowd explicit sentiment analysis," *Knowledge-Based Systems*, vol. 69, pp. 134–139, 2014.

[19] W. O. Beeman, "Affectivity in persian language use," *Culture, Medicine and Psychiatry*, vol. 12, no. 1, pp. 9–30, 1988.

[20] E. Haddi, X. Liu, and Y. Shi, "The role of text pre-processing in sentiment analysis," *Procedia Computer Science*, vol. 17, pp. 26–32, 2013.

[21] C. Trim, "The art of tokenization," *IBM Developer Works*, 2013.

[22] A. B. Carus, "Method and apparatus for improved tokenization of natural language text," Mar. 30 1999, uS Patent 5,890,103.

[23] R. Sproat, A. W. Black, S. Chen, S. Kumar, M. Ostendorf, and C. Richards, "Normalization of non-standard words," *Computer speech & language*, vol. 15, no. 3, pp. 287–333, 2001.

[24] J. Martineau, T. Finin *et al.*, "Delta tfidf: An improved feature space for sentiment analysis." *Icwsm*, vol. 9, p. 106, 2009.

[25] J. B. Lovins, "Development of a stemming algorithm," *Mech. Translat. & Comp. Linguistics*, vol. 11, no. 1-2, pp. 22–31, 1968.

[26] S. Kannan and V. Gurusamy, "Preprocessing techniques for text mining," *International Journal of Computer Science & Communication Networks*, vol. 5, no. 1, pp. 7–16, 2014.

[27] S. Vijayarani, M. J. Ilamathi, and M. Nithya, "Preprocessing techniques for text mining-an overview," *International Journal of Computer Science & Communication Networks*, vol. 5, no. 1, pp. 7–16, 2015.

[28] B. Pang, L. Lee *et al.*, "Opinion mining and sentiment analysis," *Foundations and Trends® in Information Retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008.

[29] E. Kouloumpis, T. Wilson, and J. D. Moore, "Twitter sentiment analysis: The good the bad and the omg!" *Icwsm*, vol. 11, no. 538-541, p. 164, 2011.

[30] S. Ghosh, S. Ghosh, and D. Das, "Part-of-speech tagging of code-mixed social media text," in *Proceedings of the Second Workshop on Computational Approaches to Code Switching*, 2016, pp. 90–97.

[31] E. Boiy and M.-F. Moens, "A machine learning approach to sentiment analysis in multilingual web texts," *Information retrieval*, vol. 12, no. 5, pp. 526–558, 2009.

[32] V. Narayanan, I. Arora, and A. Bhatia, "Fast and accurate sentiment classification using an enhanced naive bayes model," in *International Conference on Intelligent Data Engineering and Automated Learning*. Springer, 2013, pp. 194–201.

[33] Y. Yang and J. O. Pedersen, "A comparative study on feature selection in text categorization," in *Icml*, vol. 97, no. 412-420, 1997, p. 35.

[34] R. S. Rahate and M. Emmanuel, "Feature selection for sentiment analysis by using svm," *International Journal of Computer Applications*, vol. 84, no. 5, pp. 24–32, 2013.

[35] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Engineering Journal*, vol. 5, no. 4, pp. 1093–1113, 2014.

[36] A. Moreo, M. Romero, J. Castro, and J. M. Zurita, "Lexicon-based comments-oriented news sentiment analyzer system," *Expert Systems with Applications*, vol. 39, no. 10, pp. 9166–9180, 2012.

[37] V. Singh, R. Piryani, A. Uddin, and P. Waila, "Sentiment analysis of movie reviews and blog posts," in *Advance Computing Conference (IACC), 2013 IEEE 3rd International*. IEEE, 2013, pp. 893–898.

[38] V. Bobicev, V. Maxim, T. Prodan, N. Burciu, and V. Angheluş, "Emotions in words: Developing a multilingual wordnet-affect," in *International Conference on Intelligent Text Processing and Computational Linguistics*. Springer, 2010, pp. 375–384.

[39] J. Wiebe and R. Mihalcea, "Word sense and subjectivity," in *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2006, pp. 1065–1072.

[40] G. A. Miller, "Wordnet: A lexical database for english," *Communications of the ACM*, vol. 38, no. 11, pp. 39–41, 1995.

[41] Z. Zhai, H. Xu, and P. Jia, "An empirical study of unsupervised sentiment classification of chinese reviews," *Tsinghua Science & Technology*, vol. 15, no. 6, pp. 702–708, 2010.

[42] E. Osuna, R. Freund, and F. Girosi, "An improved training algorithm for support vector machines," in *Neural networks for signal processing VII. Proceedings of the 1997 IEEE signal processing society workshop*. IEEE, 1997, pp. 276–285.

[43] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences," *arXiv preprint arXiv:1404.2188*, 2014.

[44] Y. Kim, "Convolutional neural networks for sentence classification," *In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1746–1751, Doha, Qatar, Association for Computational Linguistics*, 2014.

[45] Y. Lu, K. Sakamoto, H. Shibuki, and T. Mori, "Are deep learning methods better for twitter sentiment analysis?" in *Proceedings of The 23rd Annual Meeting of Natural Language Processing (Japan)*, 2017, pp. 787–790.

[46] N. Ketkar, *Deep Learning with Python*. Springer, 2017.

[47] S. Poria, E. Cambria, and A. Gelbukh, "Deep convolutional neural network textual features and multiple kernel learning for utterance-level multimodal sentiment analysis," in *Proceedings of the 2015 conference on empirical methods in natural language processing*, 2015, pp. 2539–2544.

[48] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "System and method for parallelizing convolutional neural networks," Feb. 7 2017, uS Patent 9,563,840.

[49] J. Wang, L.-C. Yu, K. R. Lai, and X. Zhang, "Dimensional sentiment analysis using a regional cnn-lstm model," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, vol. 2, 2016, pp. 225–230.

[50] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[51] D. Tang, B. Qin, and T. Liu, "Deep learning for sentiment analysis: Successful approaches and future challenges," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 5, no. 6, pp. 292–303, 2015.

[52] S. Vosoughi, P. Vijayaraghavan, and D. Roy, "tweet2vec: Learning tweet embeddings using character-level cnn-lstm encoder-decoder," in *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. ACM, 2016, pp. 1041–1044.

[53] P. Ekman, "Universal facial expressions of emotions," *California mental health research digest*, vol. 8, no. 4, pp. 151–158, 1970.

[54] H.-X. Shi and X.-J. Li, "A sentiment analysis model for hotel reviews based on supervised learning," in *Machine Learning and Cybernetics (ICMLC), 2011 International Conference on*, vol. 3. IEEE, 2011, pp. 950–954.

[55] 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*. Association for Computational Linguistics, 2002, pp. 79–86.

[56] S. Tong and D. Koller, "Support vector machine active learning with applications to text classification," *Journal of machine learning research*, vol. 2, no. Nov, pp. 45–66, 2001.

[57] J. M. Ponte and W. B. Croft, "A language modeling approach to information retrieval," in *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 1998, pp. 275–281.

[58] H. Schütze, C. D. Manning, and P. Raghavan, *Introduction to Information Retrieval*. Cambridge University Press, 2008, vol. 39.

[59] A. L. Berger, V. J. D. Pietra, and S. A. D. Pietra, "A maximum entropy approach to natural language processing," *Computational linguistics*, vol. 22, no. 1, pp. 39–71, 1996.

[60] K. Mizumoto, H. Yanagimoto, and M. Yoshioka, "Sentiment analysis of stock market news with semi-supervised learning," in *Computer and Information Science (ICIS), 2012 IEEE/ACIS 11th International Conference on*. IEEE, 2012, pp. 325–328.

[61] R. Prabowo and M. Thelwall, "Sentiment analysis: A combined approach," *Journal of Informetrics*, vol. 3, no. 2, pp. 143–157, 2009.

[62] T. T. Thet, J.-C. Na, and C. S. Khoo, "Aspect-based sentiment analysis of movie reviews on discussion boards," *Journal of information science*, vol. 36, no. 6, pp. 823–848, 2010.

[63] S. Poria, E. Cambria, L.-W. Ku, C. Gui, and A. Gelbukh, "A rule-based approach to aspect extraction from product reviews," in *Proceedings of the second workshop on natural language processing for social media (SocialNLP)*, 2014, pp. 28–37.

[64] P. Chikersal, S. Poria, and E. Cambria, "Sentu: Sentiment analysis of tweets by combining a rule-based classifier with supervised learning," in *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, 2015, pp. 647–651.

[65] S. Poria, E. Cambria, A. Gelbukh, F. Bisio, and A. Hussain, "Sentiment data flow analysis by means of dynamic linguistic patterns," *IEEE Computational Intelligence Magazine*, vol. 10, no. 4, pp. 26–36, 2015.

[66] X. Ouyang, P. Zhou, C. H. Li, and L. Liu, "Sentiment analysis using convolutional neural network," in *Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM), 2015 IEEE International Conference on*. IEEE, 2015, pp. 2359–2364.

[67] H. Du, X. Xu, X. Cheng, D. Wu, Y. Liu, and Z. Yu, "Aspect-specific sentimental word embedding for sentiment analysis of online reviews," in *Proceedings of the 25th International Conference Companion on World Wide Web*. International World Wide Web Conferences Steering Committee, 2016, pp. 29–30.

[68] O. Appel, F. Chiclana, J. Carter, and H. Fujita, "A hybrid approach to the sentiment analysis problem at the sentence level," *Knowledge-Based Systems*, vol. 108, pp. 110–124, 2016.

[69] A. Hassan and A. Mahmood, "Deep learning approach for sentiment analysis of short texts," in *Control, Automation and Robotics (ICCAR), 2017 3rd International Conference on*. IEEE, 2017, pp. 705–710.

[70] Q. Shen, Z. Wang, and Y. Sun, "Sentiment analysis of movie reviews based on cnn-blstm," in *International Conference on Intelligence Science*. Springer, 2017, pp. 164–171.

[71] D. Nguyen, K. Vo, D. Pham, M. Nguyen, and T. Quan, "A deep architecture for sentiment analysis of news articles," in *International Conference on Computer Science, Applied Mathematics and Applications*. Springer, 2017, pp. 129–140.

[72] S. Liao, J. Wang, R. Yu, K. Sato, and Z. Cheng, "Cnn for situations understanding based on sentiment analysis of twitter data," *Procedia Computer Science*, vol. 111, pp. 376–381, 2017.

[73] A. S. Manek, P. D. Shenoy, M. C. Mohan, and K. Venugopal, "Aspect term extraction for sentiment analysis in large movie reviews using gini index feature selection method and svm classifier," *World wide web*, vol. 20, no. 2, pp. 135–154, 2017.

[74] W. Zhao, Z. Guan, L. Chen, X. He, D. Cai, B. Wang, and Q. Wang, "Weakly-supervised deep embedding for product review sentiment analysis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 1, pp. 185–197, 2018.

[75] S. Sohangir, D. Wang, A. Pomeranets, and T. M. Khoshgoftaar, "Big data: Deep learning for financial sentiment analysis," *Journal of Big Data*, vol. 5, no. 1, p. 3, 2018.

[76] B. Wagh, J. Shinde, and P. Kale, "A twitter sentiment analysis using nltk and machine learning techniques," *International Journal of Emerging Research in Management and Technology*, vol. 6, no. 12, pp. 37–44, 2018.

[77] G. K. Pitsilis, H. Ramampiaro, and H. Langseth, "Detecting offensive language in tweets using deep learning," *arXiv preprint arXiv:1801.04433*, 2018.

[78] A. S. Rathore, S. Arjaria, S. Khandelwal, S. Thorat, and V. Kulkarni, "Movie rating system using sentiment analysis," in *Soft Computing: Theories and Applications*. Springer, 2019, pp. 85–98.

[79] N. Farra, E. Challita, R. A. Assi, and H. Hajj, "Sentence-level and document-level sentiment mining for arabic texts," in *Data Mining Workshops (ICDMW), 2010 IEEE International Conference on*. IEEE, 2010, pp. 1114–1119.

[80] M. Al-Ayyoub, S. B. Essa, and I. Alsmadi, "Lexicon-based sentiment analysis of arabic tweets," *International Journal of Social Network Mining*, vol. 2, no. 2, pp. 101–114, 2015.

[81] L. H. Baniata and S.-B. Park, "Sentence representation network for arabic sentiment analysis," *Korean Technical Journal*, pp. 470–472, 2016.

[82] A. Dahou, S. Xiong, J. Zhou, M. H. Haddoud, and P. Duan, "Word embeddings and convolutional neural network for arabic sentiment classification," in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, 2016, pp. 2418–2427.

[83] S. Al-Saqqa, N. Obeid, and A. Awajan, "Sentiment analysis for arabic text using ensemble learning," in *2018 IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA)*. IEEE, 2018, pp. 1–7.

[84] A. M. Alayba, V. Palade, M. England, and R. Iqbal, "Improving sentiment analysis in arabic using word representation," *arXiv preprint arXiv:1803.00124*, 2018.

[85] S. Siddiqui, A. A. Monem, and K. Shaalan, "Evaluation and enrichment of arabic sentiment analysis," in *Intelligent Natural Language Processing: Trends and Applications*. Springer, 2018, pp. 17–34.

[86] T. Zagibalov and J. Carroll, "Automatic seed word selection for unsupervised sentiment classification of chinese text," in *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*. Association for Computational Linguistics, 2008, pp. 1073–1080.

[87] Y. He and D. Zhou, "Self-training from labeled features for sentiment analysis," *Information Processing & Management*, vol. 47, no. 4, pp. 606–616, 2011.

[88] C. Zhang, D. Zeng, J. Li, F.-Y. Wang, and W. Zuo, "Sentiment analysis of chinese documents: From sentence to document level," *Journal of the Association for Information Science and Technology*, vol. 60, no. 12, pp. 2474–2487, 2009.

[89] K. Xiao, Z. Zhang, and J. Wu, "Chinese text sentiment analysis based on improved convolutional neural networks," in *Software Engineering and*

Service Science (ICSESS), 2016 7th IEEE International Conference on. IEEE, 2016, pp. 922–926.

[90] K. Shuang, X. Ren, J. Chen, X. Shan, and P. Xu, "Combining word order and cnn-lstm for sentence sentiment classification," in *Proceedings of the 2017 International Conference on Software and e-Business*. ACM, 2017, pp. 17–21.

[91] M.-Y. Day and Y.-D. Lin, "Deep learning for sentiment analysis on google play consumer review," in *Information Reuse and Integration (IRI), 2017 IEEE International Conference on*. IEEE, 2017, pp. 382–388.

[92] L. Zheng, H. Wang, and S. Gao, "Sentimental feature selection for sentiment analysis of chinese online reviews," *International journal of machine learning and cybernetics*, vol. 9, no. 1, pp. 75–84, 2018.

[93] M.-T. MartíN-Valdivia, E. MartíNez-CáMara, J.-M. Perea-Ortega, and L. A. UreñA-LóPez, "Sentiment polarity detection in spanish reviews combining supervised and unsupervised approaches," *Expert Systems with Applications*, vol. 40, no. 10, pp. 3934–3942, 2013.

[94] M. A. Paredes-Valverde, R. Colomo-Palacios, M. d. P. Salas-Zárate, and R. Valencia-García, "Sentiment analysis in spanish for improvement of products and services: A deep learning approach," *Scientific Programming*, vol. 2017, 2017.

[95] J. Wehrmann, W. Becker, H. E. Cagnini, and R. C. Barros, "A character-based convolutional neural network for language-agnostic twitter sentiment analysis," in *Neural Networks (IJCNN), 2017 International Joint Conference on*. IEEE, 2017, pp. 2384–2391.

[96] F. M. Luque and J. M. Pérez, "Atalaya at tass 2018: Sentiment analysis with tweet embeddings and data augmentation," in *Proceedings of TASS*

2018: Workshop on Semantic Analysis at SEPLN (TASS 2018), vol. 2172, 2018.

[97] L. V. Avanço and M. d. G. V. Nunes, “Lexicon-based sentiment analysis for reviews of products in brazilian portuguese,” in *Intelligent Systems (BRACIS), 2014 Brazilian Conference on*. IEEE, 2014, pp. 277–281.

[98] T. A. Le, D. Moeljadi, Y. Miura, and T. Ohkuma, “Sentiment analysis for low resource languages: A study on informal indonesian tweets,” in *Proceedings of the 12th Workshop on Asian Language Resources (ALR12)*, 2016, pp. 123–131.

[99] A. M. Ramadhani, H. R. Choi, and S. B. Lim, “Korean and english sentiment analysis using the deep learning,” , vol. 23, no. 3, pp. 59–71, 2018.

[100] Y. Li and H. Fleyeh, “Twitter sentiment analysis of new ikea stores using machine learning,” in *2018 International Conference on Computer and Applications (ICCA)*. IEEE, 2018, pp. 4–11.

[101] J. Sundström, “Sentiment analysis of swedish reviews and transfer learning using convolutional neural networks,” 2018.

[102] M. S. Akhtar, A. Kumar, A. Ekbal, and P. Bhattacharyya, “A hybrid deep learning architecture for sentiment analysis,” in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, 2016, pp. 482–493.

[103] W. Zhang, H. Xu, and W. Wan, “Weakness finder: Find product weakness from chinese reviews by using aspects based sentiment analysis,” *Expert Systems with Applications*, vol. 39, no. 11, pp. 10 283–10 291, 2012.

[104] 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*. Springer, 2013, pp. 437–445.

[105] N. A. Abdulla, N. A. Ahmed, M. A. Shehab, and M. Al-Ayyoub, "Arabic sentiment analysis: Lexicon-based and corpus-based," in *Applied Electrical Engineering and Computing Technologies (AEECT), 2013 IEEE Jordan Conference on*. IEEE, 2013, pp. 1–6.

[106] A. F. Anta, L. N. Chiroque, P. Morere, and A. Santos, "Sentiment analysis and topic detection of spanish tweets: A comparative study of nlp techniques," *Procesamiento del lenguaje natural*, vol. 50, pp. 45–52, 2013.

[107] S. Karimi, "Aspects of persian syntax, specificity, and the theory of grammar," Ph.D. dissertation, dissertation, University of Washington, 1989.

[108] M. Seraji, B. Megyesi, and J. Nivre, "A basic language resource kit for persian," in *Eight International Conference on Language Resources and Evaluation (LREC 2012), 23-25 May 2012, Istanbul, Turkey*. European Language Resources Association, 2012, pp. 2245–2252.

[109] M. Saraee and A. Bagheri, "Feature selection methods in persian sentiment analysis," in *International Conference on Application of Natural Language to Information Systems*. Springer, 2013, pp. 303–308.

[110] F. Amiri, S. Scerri, and M. Khodashahi, "Lexicon-based sentiment analysis for persian text," in *Proceedings of the International Conference Recent Advances in Natural Language Processing*, 2015, pp. 9–16.

[111] B. Sabeti, P. Hosseini, G. Ghassem-Sani, and S. A. Mirroshandel, "Lexipers: An ontology based sentiment lexicon for persian," in *GCAI*, 2016, pp. 329–339.

[112] S. Alimardani and A. AGHAIE, "Opinion mining in persian language using supervised algorithms," *JOURNAL OF INFORMATION SYSTEMS AND TELECOMMUNICATION (JIST)*, 2015.

[113] A. Bagheri, M. Saraee, and F. de Jong, "Sentiment classification in persian: Introducing a mutual information-based method for feature selection," in *Electrical Engineering (ICEE), 2013 21st Iranian Conference on*. IEEE, 2013, pp. 1–6.

[114] M. Farhoodi and A. Yari, "Applying machine learning algorithms for automatic persian text classification," in *Advanced Information Management and Service (IMS), 2010 6th International Conference on*. IEEE, 2010, pp. 318–323.

[115] E. Vaziripour, C. G. Giraud-Carrier, and D. Zappala, "Analyzing the political sentiment of tweets in farsi," in *ICWSM*, 2016, pp. 699–702.

[116] M. E. Basiri and A. Kabiri, "Sentence-level sentiment analysis in persian," in *Pattern Recognition and Image Analysis (IPRIA), 2017 3rd International Conference on*. IEEE, 2017, pp. 84–89.

[117] F. Ebrahimi Rashed and N. Abdolvand, "A supervised method for constructing sentiment lexicon in persian language," *Journal of Computer & Robotics*, vol. 10, no. 1, pp. 11–19, 2017.

[118] S. A. Razavi and M. Asadpour, "Word embedding-based approach to aspect detection for aspect-based summarization of persian customer reviews," in *Proceedings of the 1st International Conference on Internet of Things and Machine Learning*. ACM, 2017, p. 33.

[119] A. Aleahmad, P. Hakimian, F. Mahdikhani, and F. Oroumchian, "N-gram and local context analysis for persian text retrieval," in *Signal Processing*

and Its Applications, 2007. ISSPA 2007. 9th International Symposium on. IEEE, 2007, pp. 1–4.

[120] K. Dashtipour, M. Gogate, A. Adeel, C. Ieracitano, H. Larijani, and A. Hussain, "Exploiting deep learning for persian sentiment analysis," in *International Conference on Brain Inspired Cognitive Systems*. Springer, 2018, pp. 597–604.

[121] A. Bagheri and M. Saraee, "Persian sentiment analyzer: A framework based on a novel feature selection method," *arXiv preprint arXiv:1412.8079*, 2014.

[122] M. E. Basiri, A. R. Naghsh-Nilchi, and N. Ghassem-Aghaee, "A framework for sentiment analysis in persian," *Open Trans Inf Process*, vol. 1, no. 3, pp. 1–14, 2014.

[123] L.-P. Morency, R. Mihalcea, and P. Doshi, "Towards multimodal sentiment analysis: Harvesting opinions from the web," in *Proceedings of the 13th international conference on multimodal interfaces*. ACM, 2011, pp. 169–176.

[124] S. Poria, E. Cambria, N. Howard, G.-B. Huang, and A. Hussain, "Fusing audio, visual and textual clues for sentiment analysis from multimodal content," *Neurocomputing*, vol. 174, pp. 50–59, 2016.

[125] V. P. Rosas, R. Mihalcea, and L.-P. Morency, "Multimodal sentiment analysis of spanish online videos," *IEEE Intelligent Systems*, vol. 28, no. 3, pp. 38–45, 2013.

[126] A. S. Alqarafi, A. Adeel, M. Gogate, K. Dashitpour, A. Hussain, and T. Durrani, "Toward's arabic multi-modal sentiment analysis," in *International Conference in Communications, Signal Processing, and Systems*. Springer, 2017, pp. 2378–2386.

[127] S. Poria, E. Cambria, G. Winterstein, and G.-B. Huang, "Sentic patterns: Dependency-based rules for concept-level sentiment analysis," *Knowledge-Based Systems*, vol. 69, pp. 45–63, 2014.

[128] 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*. Association for Computational Linguistics, 2002, pp. 417–424.

[129] H. Tang, S. Tan, and X. Cheng, "A survey on sentiment detection of reviews," *Expert Systems with Applications*, vol. 36, no. 7, pp. 10 760–10 773, 2009.

[130] K. Dashtipour, S. Poria, A. Hussain, E. Cambria, A. Y. Hawalah, A. Gelbukh, and Q. Zhou, "Multilingual sentiment analysis: State of the art and independent comparison of techniques," *Cognitive computation*, vol. 8, no. 4, pp. 757–771, 2016.

[131] A. Neviarouskaya, H. Prendinger, and M. Ishizuka, "Sentiful: A lexicon for sentiment analysis," *IEEE Transactions on Affective Computing*, vol. 2, no. 1, pp. 22–36, 2011.

[132] M. Al-Ayyoub, A. A. Khamaiseh, Y. Jararweh, and M. N. Al-Kabi, "A comprehensive survey of arabic sentiment analysis," *Information Processing & Management*, vol. 56, no. 2, pp. 320–342, 2019.

[133] 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*. ACM, 2004, pp. 168–177.

[134] F. Benamara, C. Cesarano, A. Picariello, D. R. Recupero, and V. S. Subrahmanian, "Sentiment analysis: Adjectives and adverbs are better than adjectives alone." in *ICWSM*. Citeseer, 2007.

[135] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-based methods for sentiment analysis," *Computational linguistics*, vol. 37, no. 2, pp. 267–307, 2011.

[136] A. Esuli and F. Sebastiani, "Sentiwordnet: A high-coverage lexical resource for opinion mining," *Evaluation*, vol. 17, pp. 1–26, 2007.

[137] P. J. Stone, D. C. Dunphy, and M. S. Smith, "The general inquirer: A computer approach to content analysis." *MIT press*, 1966.

[138] Y. Yang, "Application of latent dirichlet allocation in online content generation," Ph.D. dissertation, UCLA, 2015.

[139] M. E. Basiri and A. Kabiri, "Words are important: Improving sentiment analysis in the persian language by lexicon refining," *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, vol. 17, no. 4, p. 26, 2018.

[140] A. Celikyilmaz, D. Hakkani-Tür, and J. Feng, "Probabilistic model-based sentiment analysis of twitter messages," in *Spoken Language Technology Workshop (SLT), 2010 IEEE*. IEEE, 2010, pp. 79–84.

[141] A. Voutilainen, "Part-of-speech tagging," *The Oxford Handbook of Computational Linguistics*, pp. 219–232, 2003.

[142] K. Dashtipour, A. Hussain, Q. Zhou, A. Gelbukh, A. Y. Hawalah, and E. Cambria, "Persent: A freely available persian sentiment lexicon," in *International Conference on Brain Inspired Cognitive Systems*. Springer, 2016, pp. 310–320.

[143] A. Joulin, E. Grave, P. Bojanowski, M. Douze, H. Jégou, and T. Mikolov, "Fasttext. zip: Compressing text classification models," *arXiv preprint arXiv:1612.03651*, 2016.

[144] S. Poria, I. Chaturvedi, E. Cambria, and A. Hussain, "Convolutional mkl based multimodal emotion recognition and sentiment analysis," in *2016 IEEE 16th international conference on data mining (ICDM)*. IEEE, 2016, pp. 439–448.

[145] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015.

[146] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with lstm," *IET*, 1999.

[147] C. W. Leung, "Sentiment analysis of product reviews," in *Encyclopedia of Data Warehousing and Mining, Second Edition*. IGI Global, 2009, pp. 1794–1799.

[148] M. Nodoushan, "Conversational strategies in farsi complaints: The case of iranian complainees," *International Journal of Language Studies*, vol. 2, no. 2, pp. 187–214, 2008.

[149] M. Ghayoomi, "Word clustering for persian statistical parsing," in *Advances in Natural Language Processing*. Springer, 2012, pp. 126–137.

[150] K. Dashtipour, M. Gogate, A. Adeel, A. Algarafi, N. Howard, and A. Hussain, "Persian named entity recognition," in *Cognitive Informatics & Cognitive Computing (ICCI* CC), 2017 IEEE 16th International Conference on*. IEEE, 2017, pp. 79–83.

[151] K. Dashtipour, M. Gogate, A. Adeel, A. Hussain, A. Alqarafi, and T. Durrani, "A comparative study of persian sentiment analysis based on different feature combinations," in *International Conference in Communications, Signal Processing, and Systems*. Springer, 2017, pp. 2288–2294.

[152] G. L. Windfuhr, *Persian Grammar: History and State of Its Study*. Walter de Gruyter, 2011, vol. 12.

[153] L. P. Elwell-Sutton, *Elementary Persian Grammar*. Cambridge University Press, 1972, vol. 206.

[154] C. dos Santos and M. Gatti, "Deep convolutional neural networks for sentiment analysis of short texts," in *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, 2014, pp. 69–78.

[155] M. Gogate, A. Adeel, and A. Hussain, "A novel brain-inspired compression-based optimised multimodal fusion for emotion recognition," in *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2017, pp. 1–7.

[156] Y. Wang, M. Huang, L. Zhao *et al.*, "Attention-based lstm for aspect-level sentiment classification," in *Proceedings of the 2016 conference on empirical methods in natural language processing*, 2016, pp. 606–615.

[157] M. Khallash, A. Hadian, and B. Minaei-Bidgoli, "An empirical study on the effect of morphological and lexical features in persian dependency parsing," in *Proceedings of the Fourth Workshop on Statistical Parsing of Morphologically-Rich Languages*, 2013, pp. 97–107.

[158] R. Dehkharghani, B. Yanikoglu, D. Tapucu, and Y. Saygin, "Adaptation and use of subjectivity lexicons for domain dependent sentiment classification," in *2012 IEEE 12th International Conference on Data Mining Workshops*. IEEE, 2012, pp. 669–673.

[159] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of tricks for efficient text classification," in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*. Association for Computational Linguistics, April 2017, pp. 427–431.

[160] M. A. Turk and A. P. Pentland, “Face recognition using eigenfaces,” in *Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 1991, pp. 586–591.

[161] S. K. D’mello and J. Kory, “A review and meta-analysis of multimodal affect detection systems,” *ACM Computing Surveys (CSUR)*, vol. 47, no. 3, p. 43, 2015.

[162] C. G. Snoek, M. Worring, and A. W. Smeulders, “Early versus late fusion in semantic video analysis,” in *Proceedings of the 13th annual ACM international conference on Multimedia*. ACM, 2005, pp. 399–402.

[163] P. Natarajan, S. Wu, S. Vitaladevuni, X. Zhuang, S. Tsakalidis, U. Park, R. Prasad, and P. Natarajan, “Multimodal feature fusion for robust event detection in web videos,” in *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2012, pp. 1298–1305.