北海道大学
HOKKAIDO UNIVERSITY

Language Media Laboratory
Research Group of Information Media Science and Technology
Division of Media and Network Technologies
Graduate School of Information Science and Technology
Hokkaido University

Da Li

A Study on Deep Learning-based Humor Detecting Methods for Sentiment Analysis of Social Media

A doctoral dissertation
supervised by
Prof. Kenji Araki

Sapporo 2020

A Study on Deep Learning-based Humor Detecting Methods for Sentiment Analysis of Social Media
（ソーシャルメディアにおける感情分析のための深層学習を用いたユーモア検出手法に関する研究）

学位論文の要旨

情報通信の重要なファクタの一つであるスマートフォンは近年急速なスピードで普及しており，それに伴いソーシャルメディア(SNS) は私たちの生活に不可欠なものとなり，その中には感情情報が多くふくまれている．「Weibo」とは，中国語圏最大のソーシャルメディアであり，中国本土のみならず，日本やアメリカなど中国語話者がいる地域で幅広く利用されている．現在，自然言語処理の分野においてソーシャルメディアの感情分析は研究テーマとして注目を集めている．しかし，従来手法ではテキストのみを考慮し，絵文字などはノイズとして除外される場合が大半であった．さらに,ユーモアという感情表現が人の生活において多様な役割を果たすことが明らかとなってきている．近年，コンピュータがユーモアを理解することが可能となり，その有用性が認知されてきている．先行研究によると，SNS のユーザは絵文字を使用することで，ユーモアの感情を表す傾向が見られる．そのために，ソーシャルメディアの感情分析に着目し，絵文字を使用したユーモアの投稿を中心にし，本研究を行うことは有効である．

絵文字は，感情のグラフィカル表現としてソーシャルメディアで広く使用され，重要な役割を果たしている．絵文字と同様に，インターネットスラングは，日常のオンラインコミュニケーションで使用される非公式の言語であり，感情表現として多く用いられている．スラングや絵文字がソーシャルメディアの感情分析に与える影響を十分に理解することが重要であると考えられる．したがって，本研究は絵文字の感情極性アノテーションを行い，SNS によく使用されているスラングとフェイシャル絵文字を分析し，それぞれの

レキシコンを構築した．さらに，絵文字を考慮したソーシャルメディアにおける感情分析のための深層学習を用いたユーモア検出手法を提案したことが本研究の新規性である．

評価実験により，以下に述べる 3 点の絵文字極性，絵文字レキシコンとスラングレキシコンの有効性及び提案手法の有効性が確認された．1）絵文字極性の有効性を検証するために，絵文字の極性を考慮した感情分析モデル再帰型ニューラルネットワーク Long Short-Term Memory(EPLSTM)を提案した．実験結果により，従来手法と比べ，提案手法が感情極性を予測する性能が大幅に改善することが確認された．2）感情分析の精度を向上するために，構築した絵文字とスラングのレキシコンを形態素解析ツールに応用し，感情分析の機械学習モデルを提案した．サポートベクターマシン(SVM)，ロジスティック回帰(LR)，単純ベイズ分類器(NB)などの機械学習モデルを利用し，感情分析において，構築したスラングレキシコンと絵文字レキシコンが有効であることを検証した．3）絵文字辞書とスラング辞書を用い，SNS データを深層学習モデル Attention-based Bi-directional LSTM に学習させ，細粒度ユーモア検出システム「HEMOS」(Humor-EMOji-Slang)を提案した．実験の結果，従来手法である深層学習モデルLSTMやtextCNNと比べて，提案手法を用いることで，F 値が 6.86～20.59 ポイント上回ったことが確認された．有意差検定により，p 値は 0.05 以下になっていることから，有意差が認められることが確認された．

本研究により，絵文字に関する分析の有効性及び絵文字とスラングを考慮したソーシャルメディアにおける感情分析のための深層学習を用いたユーモア検出手法の有効性が確認された．

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Nowadays, social media has become the essential part of our lives. With the tremendous popularity of web 2.0 applications, people are free to express opinions on social networks such as Twitter, Facebook or Weibo[1] (the biggest Chinese social media network that was launched in 2009). These new communication channels generate massive amounts of data daily, which become an insightful information source for various research purposes, often using Natural Language Processing (Kavanaugh et al., 2012). One of its popular and essential tasks is sentiment analysis, which mainly focuses on automatic sentiment predicting from online user-generated content but also draws interest from other fields as psychology, cognitive linguistics, or political science. Sentiment analysis can be useful in election result forecasting, stock prediction, opinion mining, business analytics, and

[1] https://www.weibo.com

other data-driven tasks (Li, Rzepka, and Araki, 2018). Sentiment analysis has been widely used in real-world applications that analyze the online user-generated data, such as opinion mining, product reviews analysis, and business-related activity analysis (Zhao et al., 2018). Sentiment analysis in English has achieved noteworthy success in the recent two decades. On the other hand, classifying sentiment in other languages, for example Chinese, remains at an early stage (Wang et al., 2013). The current sentiment analysis mainly focuses on text-based user-generated content. However, various new forms of semiotic information on social networks such as emoji, images, and memes, allow users to express themselves more creatively. Emojis and slangs have become a noticeable part of online informal expression on social networks.

Humor is one of the personal aspects which defines us as human beings and social entities. It is a very complex, as well as ubiquitous concept, which we could simply define by the presence of amusing effects, such as laughter or well-being sensations, a set of phenomena which play a relevant role in our lives (Reyes, Rosso, and Buscaldi, 2012). Before we had spoken or written language, humans used laughter to express our enjoyment or accession in certain situations (Mathews, 2016). In today's society, people tend to write

funny jokes, play word games, use emojis or memes to comment on current affairs by "white sarcasm"[2], and relieve life stress by self mocking, often in very creative ways.

Emojis are ideograms and smileys used in electronic messages and web pages. Originating from Japanese mobile phones in 1997, this kind of pictograms became increasingly popular worldwide in the 2010s after being added to several mobile operating systems. In 2015, Oxford Dictionaries named the "Face with Tears of Joy" emoji 😂 the "Word of the Year". In my opinion ignoring emojis in sentiment research is unjustifiable because they convey a piece of significant emotional information and play an important role in expressing emotions and opinions in social media (Novak et al., 2015; Guibon, Ochs, and Bellot, 2016; Li et al., 2019a).

Internet slang is similarly ubiquitous on the Internet. The emergence of new social contexts like micro-blogs, discussion groups, community-driven question-and-answer websites, and social networks has enabled slang and non-standard expressions to abound on the Web. Despite this, slang has been traditionally viewed as a form of non-standard language and a form of language that is not the focus of linguistic analysis. In consequence, it has mostly been neglected (Kulkarni and Wang, 2017).

[2]Similar to "white lie", I define this special kind of sarcasm with good intentions as "white sarcasm".

Furthermore, I also noticed that the conventional sentiment analysis generally considers only three dimensions (positive, negative, and neutral) for text classification, while the usage of some pictograms in sentiment analysis is hard to be simply classified as either positive or negative. Especially, many emojis and slangs seem to be used purely for laughter, self-mockery, or jocosity, which communicates an implied irony that is intensively used in Chinese conversation on social networks (Li et al., 2019a). Thus, I believe that humor, as one of the most common expressions in natural language, can also be regarded as a useful dimension for sentiment analysis with an emoji dataset.

1.2 Contributions of the Thesis

My main contributions are as follows:

- First, I focus on the emojis used on Weibo in order to establish if pictograms improve sentiment analysis by recognizing humorous entries which are difficult to polarize. Because the emojis probably play an equal or sometimes even more important role in expressing emotion than textual features, I analyzed the characteristics of emojis, and report on their evaluation while dividing them into three categories: positive, negative and humorous.

- Second, I also noticed that among the resources of Chinese social media sentiment analysis, the labelled Weibo data sets containing emojis are extremely rare which makes considering them in machine learning approaches difficult. To resolve this problem, I applied several neural network model using emoji polarity to improve sentiment analysis on smaller annotated data sets.

- Third, I collected 576 frequent Chinese Internet slang expressions as a Chinese slang lexicon. I converted the 109 Weibo emojis into textual features creating Chinese emoji lexicon. To test the influence of slang and emojis on sentiment analysis task, I utilized both lexicons with several machine learning-based classifiers, namely k-Nearest Neighbors, Decision Tree, Random Forest, Logistic Regression, Naïve Bayes and Support Vector Machine for detecting humorous expressions on Chinese social media.

- Fourth, I empirically confirmed inherent humor characteristic to Chinese culture visible on Weibo and divided Weibo posts into four categories: positive, negative, optimistic humorous, and pessimistic humorous. I applied the slang lexicon and emoji lexicon to an attention-based bi-directional long short-term memory recurrent neural network (AttBiLSTM) and proposed HEMOS, a fine-grained humor detecting method for sentiment analysis of social media.

- Finally, to the best of my knowledge, my fine-grained humor detecting method for sentiment analysis is the only project currently in development which considered slang and emoji in Chinese sentiment analysis.

1.3 Structure of the thesis

This thesis consists of nine chapters apart from Introduction.

Chapter 2 describes related research in the field of sentiment analysis, analysis of slang and emoji. This chapter is meant to discuss research related to the ultimate goal of this study.

Chapter 3 introduces theories of humor, and shows several examples of humorous expression in Chinese culture.

Chapter 4 details my work on Internet slang, and shows the detail of Chinese slang lexicon.

Chapter 5 presents my work on emoji, includes the emoji polarity and emoji lexicon.

Chapter 6 and Chapter 7 presents my experiments for testing the influence of slang and emoji lexicons on humor detection task, and verifying the validity of emoji polarity on sentiment analysis task.

Chapter 8 presents the HEMOS system, a fine-grained humor detecting method for sentiment analysis of social media.

Chapter 9 summarizes my findings and presents future directions for the study.

Chapter 2

Related Research

2.1 Sentiment Analysis

Traditionally, sentiment analysis is a binary approach that classifies the text emotions into positive, negative and neutral. Peng and colleagues (Peng, Cambria, and Hussain, 2017) have conducted a comprehensive literature review on Chinese sentiment analysis. According to their findings, the methodologies for Chinese sentiment analysis can be generally categorized with supervised machine learning approach and unsupervised knowledge-based approaches. For this study, I mainly focus on supervised machine learning approaches.

2.1.1 Rule-based methods

(Zhang et al., 2009) proposed a rule-based approach with two phases: a) the sentiment of each sentence is first decided based on word dependency to aggregate the sentences sentiments and then b) the sentiment of each document is calculated. (Zagibalov and Carroll, 2008) presented a method

that does not require any annotated corpus training data and only requires information on commonly occurring negations and adverbials. Li et al. stated that polarities and strengths judgment of sentiment words comply with a Gaussian distribution, and thus proposed a Normal distribution-based sentiment computation method which allows quantitative analysis of semantic fuzziness of sentiment words in Chinese language (Li et al., 2014). Zhuo et al. presented a novel approach based on the fuzzy semantic model by using an emotion degree lexicon and a fuzzy semantic model (Zhuo, Wu, and Luo, 2014). Their model includes text preprocessing, syntactic analysis, and emotion word processing. However, optimal results of Zhuo's model were achieved only when the task was clearly defined.

2.1.2 Machine learning-based methods

There are generally three separate stages for machine learning based sentiment analysis tasks, which are text segmentation, feature extraction, and sentiment classification. Text segmentation divides a text into meaningful tokens. Feature extraction retrieves both sentiment features and raw segmented words features and represents them in a bag of words (BoW). Finally, the dataset is fed into a machine learning model for assigning the sentiment score to the given

text. The commonly used algorithms are Naïve Bayes, maximum entropy, SVM, neural networks and others (Vinodhini and Chandrasekaran, 2012; Khan et al., 2010; Dhande and Patnaik, 2014).

One representative example of such studies is an empirical work performed by Tan and Zhang, who categorized sentiment in Chinese documents (Tan and Zhang, 2008). Four features, namely mutual information, information gain, chi-square, and document frequency, were tested separately on five different algorithms, including centroid classifier, k-Nearest Neighbor, Winnow classifier, Naïve Bayes (NB) and Support Vector Machine (SVM). Among these algorithms, the information gain and SVM features were found to yield the best performance under topic-dependent classifiers. Chen et al. proposed a novel sentiment classification method that incorporated the existing Chinese sentiment lexicon and convolutional neural network (Chen et al., 2015). Their results showed that the proposed approach outperforms the convolutional neural network (CNN) model with only word embedding features (Kim, 2014).

These Chinese sentiment classification approaches, although they have achieved satisfying outcomes, only consider a pure text-based dataset from a polarity-base perspective. In addition to the classical binary label classification problem (positive, negative, or neutral), Liu et al. proposed a multi-label sentiment analysis prototype for micro-blogs (Liu and Chen, 2015). They also compared the performance of eleven

state-of-the-art classification methods (BR, CC, CLR, HOMER, RAkEL, ECC, MLkNN, RF-PCT, BRkNN, BRkNN-a and BRkNN-b) on two microblog datasets. Although Liu and colleagues consider multi-label classification, none of the papers consider humorous as a dimension for sentiment classification.

2.2 Internet Slang

Recently, for sentiment analysis in the English language, more researchers have realized the value of emoji and slang dataset. Wu et al. constructed an English slang dictionary (named SlangSD) for sentiment analysis tasks and proved its ease of use (Wu, Morstatter, and Liu, 2016). In the research of (Soliman et al., 2014), the authors constructed a Slang Sentimental Words and Idioms Lexicon (SSWIL) of opinion words. They also proposed a Gaussian kernel SVM classifier for Arabic slang language to classify Arabic news comments on Facebook[1]. The proposed classifier achieved precision of 88.63% and recall of 78%. (Manuel, Indukuri, and Krishna, 2010) presented an approach for finding the sentiment score of newly found slang sentiment words in blogs, reviews and forum texts on Internet. A simple mechanism for calculating sentiment score of documents using slang words with the help of Delta Term Frequency and Weighted Inverse Document Frequency technique is also presented in their paper.

[1]https://www.facebook.com/

2.3 Emoji

In 2017, Felbo and collegues (Felbo et al., 2017) proposed a powerful system utilizing emojis in their Twitter sentiment analysis model called DeepMoji. During this study, 1,246 million tweets containing one of 64 common emojis were trained by a Bi-directional Long Short-Term Memory (Bi-LSTM) model to interpret the sentiment within the online tweets. DeepMoji also works well for sarcasm detection task with a verified accuracy rate of 82.4%. Their system even outperforms human detectors who managed to acquire a 76.1% accuracy rate. Although sarcasm and irony tend to convey negative sentiments in general, since these two factors may reverse the overall sentiment score of given text, I found that in Chinese social media (Weibo in my example), in addition to the represented positive and negative sentiments, users tend to express an implicit humor that escapes the traditional bi-polarity.

Chen et al. proposed a novel scheme for Twitter sentiment analysis with extra attention to emojis (Chen et al., 2018). They first learned bi-polarity emoji embeddings under positive and negative sentimental tweets individually, then trained a sentiment classifier by attending on these bi-polarity emoji embeddings with an attention-based Long Short-Term Memory network (LSTM). Their experiments have shown that the bipolarity embedding was effective for extracting sentiment-aware embeddings of emojis. When it

comes to Chinese social media Weibo, Zhao and collegues built a system called MoodLens (Zhao et al., 2012), which is the first system for sentiment analysis of Chinese posts in Weibo. In MoodLens, 95 emojis are mapped into four categories of sentiments (angry, disgusting, joyful, and sad), which serve as class labels of the entries. They collected over 3.5 million labeled posts as the corpus and trained a fast Naïve Bayes classifier, with an empirical precision of 64.3%. However, the precision in their method was still relatively low, and emotion of many emojis has been changed dramatically. More specific analysis of Weibo emojis is discussed in Chapter 5.

More recently, in my previous research (Li et al., 2018), I analyzed the usage of the emojis with a facial expression used on Weibo. I applied the emojis polarity in an Long Short-Term Memory recurrent neural network (LSTM) and classified Weibo posts into three categories: positive, negative, and humorous. In (Li et al., 2019b), I proposed a attention-based GRU network model using emoji polarity to improve sentiment analysis on smaller annotated data sets. My experimental results shown that the proposed method can significantly improve the performance of sentiment polarity prediction. Both of these researches assign a hyperparameter to the probability of the deep learning model's softmax output, and apply the labelled emojis from the work of (Li, Rzepka, and Araki, 2018) as emoji polarity. Then we

assign a hyper-parameter to the emoji polarities to calculate the final probability output of sentiment classification. However, I found that many Weibo entries carry humorous aspects like self-mockery and contain some optimistic or pessimistic emotional load. It is quite easy to make a wrong prediction by simply classifying them into a "humorous" category without polarizing it. Therefore, to tackle this problem, I propose a fine-grained humorous classification method to improve the results of bi-polarity sentiment analysis.

Chapter 3

Humor in Chinese Culture

3.1 Theories of Humor

Humor is a tendency of experiences to provoke laughter and provide amusement, and it also can be defined as a reliable elicitor of exhilaration (Ruch, 1993). However, many theories exist about what humor is and what social function it serves.

In the long course of history, regions with different cultural backgrounds have had different interpretations of humor. In ancient Greece, Plato first suggested a view in the "Philebus" that the nature of humor is an ignorance in the weak, who are thus unable to retaliate when mocked. Later, in his famous "Poetics", Aristotle proposed that an ugliness that does not disgust is fundamental aspect of humor (Jones, 2005). In ancient Sanskrit drama titled "Natya Shastra", Bharata Muni defined humor as one of the nine emotional reactions, which can be invoked in the audience by the simulations of emotions that the actors perform (Sharma, 2011). Humor expressions were first documented around

2,500 BC in China when the first Chinese poetry and literary books appeared Yue, 2010. In ancient China, humor was considered traditionally as subversive or unseemly in the Confucian culture, which ritual and propriety. It was usually perceived as irony and sarcasm. On the other hand, Taoism placed a high value on humor. Both core ideas and expressions of Taoism reflect the wisdom of humor (Yue, 2014). At present, Chinese people tend to express implicit humor on social media, and some posts can be seen as pessimistic irony, while other entries can be considered as a pure joke or optimistic self-deprecation. More specific analysis of these phenomena is discussed in the next section.

Nowadays, humor became a ubiquitous human phenomenon that occurs in all types of social interaction. It is also a form of communication (often creative) that bridges the gap between various languages, cultures, ages, and demographics (Furnham, 1984). Most of us laugh at something funny many times during a typical day (Martin, 2006). Laughter releases endorphins, relaxes the body, and helps to relieve stress. Although it is a form of play, humor serves several cognitive, emotional, and social functions (Martin and Ford, 2018). However, humor is an emotion that is difficult to define. Linguists, psychologists, and anthropologists have taken humor to be an all-encompassing category, covering any event that amuses, or is felt to be amusing (Attardo, 2010). Different people will not necessarily find the same things equally funny. Many things which strike one

group as funny may bore another group, and some jokes are private or individual, often restricted in their funniness to just one or very few individuals. As Raskin (Raskin, 2012) states, "This universality of humor is further reinforced by the fact that surprisingly many jokes or situations will strike surprisingly many, if not all people, as funny. Therefore, I am dealing with a universal human sentiment, humor. Responding to humor is a part of human behavior, ability, or competence, other parts of which comprise such important social and psychological manifestations of human as language, morality, logic, faith, etc." Therefore, I hypothesize that the recognition of humorous emotion is of great significance for sentiment analysis.

3.2 Humor in Chinese Culture

One interesting thing I found from analyzing Weibo is that people tend to express implicit humor that escapes the traditional bi-polarity of positive and negative. Moreover, emojis and slang seem to play an essential role in enhancing such effect. Figure 3.1 shows an example of a Weibo microblog that includes emojis and Internet slang. In the second line of the post, 累觉不爱 (*lei jue bu ai*[1]) is an abbreviation of 很累，感觉自己不会再爱了 (*hen lei, gan jue zi ji bu hui zai ai le* which means "too tired for romance"). As illustrated in Figure 3.1, the forms of such expressions are often shortened to

[1]In this paper I use italic to indicate romanization of Chinese language (*pinyin*).

a "phrase" that consists of four Chinese characters. These expressions are similar to the usage in English, such as "lol" (laugh out loud) "idk" (I don't know), "ASAP" (as soon as possible), but for Chinese expression, it carries more creative, complex and ambiguous meaning.

FIGURE 3.1: Example of Weibo post with Internet slang and emojis. The entry says "After jogging, I'm starving. Someone sent me a picture of skewers. I'm too tired for romance".

In general, this particular form of expression can be seen from Chinese *chengyu*, which is a four-character phrase whose

meaning was inherited from the older generations containing moral concepts, pearls of wisdoms or previous experiences. Nowadays, *chengyu* still plays a vital role in Chinese conversations and education. However, the young generation tends to adopt this form and give it new meanings that carry a sense of implicit humor, self-mocking or amusement inside the digital context. Typically, Internet slang expressions are humorous, satirical, or ironic, which is also a crucial aspect of what makes them appealing and widely accepted. Such informal articulations are popular and extensively used in Chinese social media platforms and more examples can be found: *lei jue bu ai*, *ren jian bu chai* (life is so hard that some lies are better not exposed), *xi da pu ben* (news so exhilarating that everyone is celebrating and spreading it around the world).

Emojis are generally used to enhance or emphasize the sentiment of certain content. However, another interesting phenomenon I observed is that the meaning of certain emojis may change with time. For instance, 🫣 was initially designed for expressing a "bye-bye" gesture. However, it seems that more people gradually started to use this emoji for a funny way to express an artificial smile of refuse or self-mockery, and this usage is currently becoming trendy inc Chinese social media. In the research of (Li et al., 2018), the authors confirm this effect, underlining that this emoji tends to carry more humor rather than negative polarity. For example, in the following post: "After jogging,

I'm starving. Someone sent me a picture of skewers. I'm too tired for romance 🥲🥲🥲 ". In this context, the post is trying to express a humorous nuance of a pessimistic attitude. In such situations, emojis and slang seem to play a pragmatic role in denoting humor rather than merely exhibiting positive or negative moods.

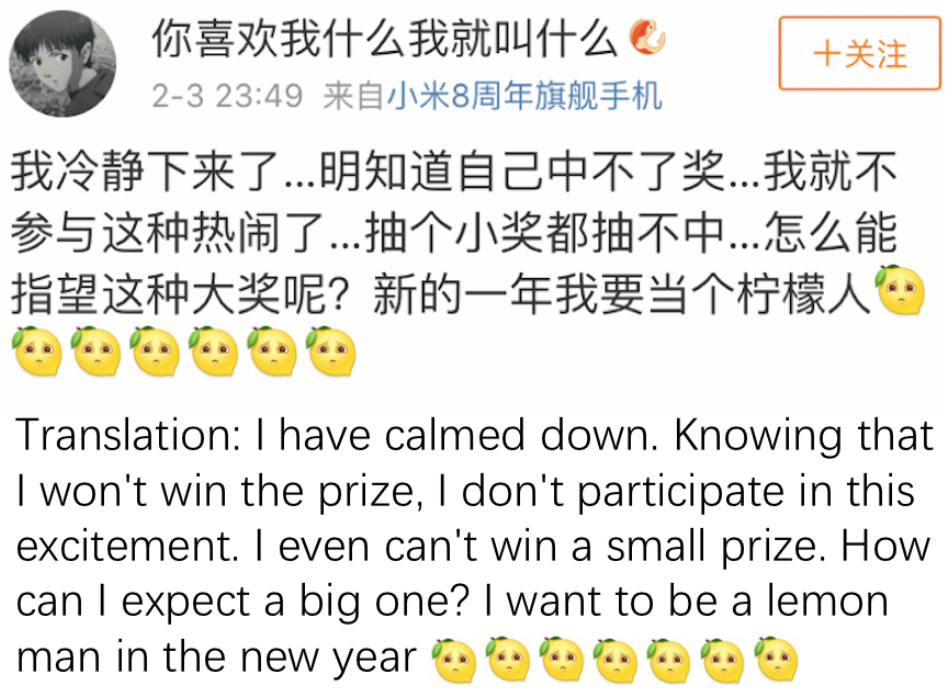

FIGURE 3.2: Example of Weibo post with "lemon man" emojis.

Figure 3.2 shows anther example of a Weibo microblog posted with emojis and slang. In the third line of the post, 柠檬人 (*ning meng ren*) is a new word that appeared in early 2019 on Chinese social media and means "lemon man". Accordingly, to address this new popular phrase, 🥲 was added to the pictogram list by social media companies in January 2019. This lemon with a sad face is also called "lemon man"

which expresses the same emotion as slang *ning meng ren* – "sour grapes" or "jealous of someone's success". This entry seems to convey a humorous nuance of a pessimistic attitude. Emojis seem to play an important role in expressing this kind of emotions. And I also noticed that Internet slang and emoji has been constantly changing. There is a high possibility that this phenomenon can cause a significant difficulty in sentiment recognition task.

Chapter 4

Analysis of Internet Slang

4.1 Internet Slang in Social Media

Slang, as a trendy form of informal language expression are widely used on social media for posting or commenting (Jones and Schieffelin, 2009). It appears in people's everyday life, which has changed the way how people communicate with each other's from a great extent. However, due to the unstructured nature and tricky meaning under different contexts, it causes great difficulties for machine to extract the sentiment polarity directly.

4.2 Lexicon of Chinese Online Slang

Because Internet slang is not easy to process automatically, it can cause significant difficulty in the polarity recognition task. For improving the performance of Chinese social media sentiment analysis, I created a Chinese Internet slang lexicon (examples shown in Table 4.1).

576 Chinese slang phrases that are frequently used on the Internet were identified and stored in my Chinese Internet

TABLE 4.1: Examples of my Chinese Internet Slang Lexicon.

Type	Examples (Origin)	English Translation
Numbers	233 (哈哈哈)	"laughter"
Latin alphabet abbreviations	TMD (他妈的)	"Damn"
Chinese contractions	人艰不拆 (人生已经如此的艰难有些事情就不要拆穿)	"Life is so hard that some lies are better not exposed."
Neologisms	屌丝	"Loser"
Phrases with altered or extended meanings	壕 (土豪)	"Vulgar tycoon"
Puns and wordplay	河蟹 (和谐)	"Harmony"
Slang derived from foreign language	欧尼酱 (お兄ちゃん)	"Brother"

Slang Lexicon. They originate from various sources, including the Internet New Words Ranking List, Baidu Baike[1], Wikipedia[2] and social media systems such as Baidu Tieba[3] and Weibo[4] (the time span of the processed data is between 2010 and 2019). After analysis, I divided the entries into seven following categories:

- Numbers: such as 233 ("laughter/lol": Chinese use 233 to express "can't stop laughing" because 233 is an emotional sign in a Chinese BBS site[5] and the sign is the character number 233 in the list of all the emojis); 213 ("a person who is very stupid"); 520/521 ("I love you").

[1] https://baike.baidu.com
[2] https://en.wikipedia.org
[3] https://tieba.baidu.com
[4] https://www.weibo.com
[5] https://www.mop.com

- Latin alphabet abbreviations: Chinese users commonly use a QWERTY keyboard with pinyin enabled. Upper case letters are quick to type, and no transformation to ideograms follows. (lower case letters are automatically converted into Chinese characters). Latin alphabet abbreviations (rather than Chinese characters) are also sometimes used to evade censorship. Such as SB ("dumb cunt"); YY ("fantasizing/sexual thoughts"); TT ("condom").

- Chinese contractions: e.g. *ren jian bu chai* ("life is so hard that some lies are better not exposed": This comes from the lyrics of a song entitled "*Shuo Huang*" ("Lies"), by Taiwanese singer Yoga Lin. This slang reflects that some people, especially young people in China, are disappointed by reality); *lei jue bu ai* ("too tired for romance": this slang phrase is a literal abbreviation of the Chinese phrase "too tired to fall in love anymore". It originated from an article on the Douban[6] website, a Chinese social networking service website allowing registered users to record information and create content related to movies, books, music, recent events and activities in Chinese cities. The article was posted by a 13-year-old boy who grumbled about his single status and expressed his weariness and frustration towards

[6] https://www.douban.com

romantic love. The article went viral on the Chinese Internet, and the phrase started to be subsequently used as a sarcastic way to convey depression when encountering misfortunes or setbacks in life); *gao da shang* ("high-end, impressive, and high-class": a popular meme used to describe objects, people, behavior, or ideas that became popular in late 2013).

- Neologisms: *diao si* ("loser": The word *diao si* is used to describe young males who were born into a low-income family and are unable to improve their financial status. People usually use this phrase in an ironic and self-deprecating way); *ye shi zui le* ("nothing to say": it is a way to gently express your frustration with someone or something that is entirely unreasonable and unacceptable); *dan shen gou* ("single dog": a term which single people in China use to poke fun at themselves for being single).

- Phrases with altered or extended meanings: *hao* or *tu hao* ("vulgar tycoon": This word refers to irritating online game players who buy large amounts of game weapons to be gloried by others. Starting from late 2013, the meaning has changed and now is widely used to describe nouveau-riche people in China who are wealthy but less cultured.); *bei tai* ("spare tire": A girlfriend or boyfriend kept as a "backup", "plan B" , just in case of breaking up with the current partner).

- Puns and wordplay: 河蟹 (“river crab”: pun on 和谐, another Chinese characters pronounced *he xie*, meaning “harmony”).
- Slang derived from foreign language: 工口 (The word *gong kou* comes from the Japanese katakana *ero*, which translated from English “erotic" into the abbreviation of the katakana エロチック, meaning “sensual”).

Chapter 5

Analysis of Emoji

5.1 Emoji in Social Media

Facial expressions are a powerful tool in social communication (Batty and Taylor, 2003). Basic facial expressions of emotion are universal, Ekman and Friesen (Ekman and Friesen, 1971) reported that six (anger, happiness, fear, surprise, disgust and sadness) are readily recognized in different cultures. Although there are many unknown factors in continually changing moods of human beings, expressing feelings is one of the most basic functions of our communication, also by using emojis which have become a global phenomenon. However, due to different cultural backgrounds and different contexts, misunderstandings commonly occur when communicating with emojis (Rzepka, Okumura, and Ptaszynski, 2017).

Moreover, according to my previous observations, the meaning of certain emojis might change over time, and it is often difficult to interpret them as positive or negative. It seems that some emojis are used just for fun, self-mockery,

or jocosity, which expresses an implicit humor characteristic in Chinese culture. For instance, 🐶 originated from the Japanese Shiba meme in 2013[1], and has now become one of the most widely used emoji on Weibo. Weibo also developed a cat 😸 version of 🐶 emoji. It is a clue for users to indicate that they know the real thought behind the words. The 🐕 is also known as “doge saves life", which is a humorous/informal way to express disagreement and avoid being attacked at the same time. Some users include this emoji in their messages just for being cute/funny without specific meaning. 🤤 is a reference to the phrase *chi gua qun zhong*, which refers to onlookers who are watching a situation just for fun. By using 🤤 in a post or comment, users are usually mocking themselves by showing they have no interest in the given topic and do not want to join the discussion. 🧼 is a bar of soap and its existence is probably inspired by an iconic joke about bending over for a bar of soap in a men’s shower room. In most situations at Weibo, the passed message is “be careful of the situation," but sometimes it seems not to carry the original meaning and to make the content entertaining. Another example of changing the original meaning is 🙂, which, instead of being just a cheerful smile, started to carry a more passive than a positive attitude. By using this emoji in Chinese social networks, users could be expressing sarcasm, a passive-aggressive attitude, or even contempt. 🙂 can sometimes reverse the sentiment of the

[1] https://en.wikipedia.org/wiki/Doge_(meme)

sentence.

FIGURE 5.1: 109 Weibo emojis which can be transformed into Chinese character tags.

The examples given above show that pictograms seem to play an important role not only in expressing emotions but also in conveying humorous content. There is also a high possibility that this phenomenon can cause significant difficulty in sentiment detecting task. Therefore, I decided to build a lexicon of emojis before adding them to my system to facilitate classifying emotions in Weibo.

5.2 Lexicon of Chinese Social Media Emojis

Because Weibo emojis have no corresponding Unicode[2] assigned, they are formatted to Chinese letters within square brackets when scrapped. For example, 🙂 is transformed into [微笑] ("smile"). This conversion provided us with the

[2] https://unicode.org/emoji/charts-13.0/full-emoji-list.html

possibility of building a Chinese emoji lexicon based on these tags. Consequently, 109 Weibo emojis (see Figure 5.1) were selected and converted into textual features for building the lexicon. More examples are shown in Table 5.1.

5.3 Emoji Polarity

In the real-life (offline) dialogue between human beings, besides tone changes, I usually express emotions with body language. In social networks, this can partially be achieved by using emojis (Aldunate and González-Ibáñez, 2017). I have noticed previously mentioned humorous emotion in Weibo microblog entries containing emojis which are often difficult to interpret as positive or negative. It seems that some emojis are used just for fun, self-mockery or jocosity which expresses an implicit humor characteristic in Chinese culture. Emojis seem to play an important role in expressing this kind of emotion. There is a high possibility that this phenomenon can cause a significant difficulty in sentiment detecting task, therefore I decided to analyze Weibo emojis before adding them to my system for classifying emotions in Chinese microblogs.

TABLE 5.1: Examples of Chinese Emoji Lexicon.

Emoji	Textual Feature	Emotion/Implication
🙂	[微笑]	"smile"
😊	[可爱]	"lovely"
🤗	[太开心]	"too happy"
👏	[鼓掌]	"applause"
😁	[嘻嘻]	"hee hee"
🤤	[吃瓜]	"watermelon-eating"
😜	[挤眼]	"wink"
🤪	[馋嘴]	"greedy"
😑	[黑线]	"speechless/awkward"
😅	[汗]	"sweat"
🤧	[挖鼻]	"nosepick"
😤	[哼]	"snort"
😡	[怒]	"anger"
😠	[委屈]	"upset/fell wronged"
🥺	[可怜]	"pathetic"
😞	[失望]	"disappointment"
😟	[悲伤]	"sad"
😭	[泪]	"weep"
😳	[害羞]	"shy"
🤮	[污]	"filthy"
🥰	[爱你]	"love face"
😚	[亲亲]	"kissy face"
😍	[色]	"leer"
🤤	[舔屏]	"lick screen"
😱	[憧憬]	"longing"
🤡	[二哈]	"dog leash"
🤷	[摊手]	"smugshrug"

5.3.1 Annotation

For the analysis I chose 67 most common emojis out of 144 emojis used on Weibo (see Figure 5.2) which represent facial expression, because this type is usually used in sentiment analysis in other languages. I asked 12 Chinese native speakers to label these 67 emojis by applying one of three following categories: positive, negative and humorous. The evaluators were 6 males and 6 females in their 20's, each of them being a Weibo user for more than five years.

FIGURE 5.2: Most common emojis of Weibo.

The annotation results showed that there are 44 emojis with obvious positive or negative emotion, and the remaining 23 emojis are considered more as humorous than positive or negative. The results are summarized in Table 5.2.

My hypothesis is that the emojis which are considered to be humorous in more than 50% of cases (see Table 5.2) usually convey tongue-in-cheek type of message. Moreover, the sentiment polarity of emojis such as 🙂, 🤔, 😲 and other cases show rather clearly their ambiguity, and they should be treated with special care when I add emoji recognition

TABLE 5.2: Twenty-three emojis conveying humor typical for Chinese culture.

Emoji	Humorous (%)	Negative (%)	Positive (%)
🙂	41.7	25.0	33.3
😂	58.3	0.0	41.7
🤭	66.7	33.3	0.0
🤣	91.7	8.3	0.0
😤	58.3	0.0	41.7
🐶	83.3	0.0	16.7
🐱	58.3	25.0	16.7
🐼	66.7	8.3	25.0
🤔	66.7	8.3	25.0
🤨	41.7	33.3	25.0
😴	75.0	25.0	0.0
😵	58.3	41.7	0.0
🤫	50.0	25.0	25.0
😳	66.7	33.3	0.0
🤯	50.0	50.0	0.0
🤗	58.3	41.7	0.0
🤧	50.0	50.0	0.0
🥱	50.0	50.0	0.0
🤰	50.0	33.3	16.7
😪	41.7	41.7	16.7
🤤	75.0	8.3	16.7
🙅	58.3	33.3	8.3
🤦	75.0	0.0	25.0

capability to my system in the future. 🙂 was originally a smile emoji, but it seems that gradually Weibo users started using this emoji for expressing artificial smile, ridicule or self-mockery. On the other hand, when dealing with such emojis on Weibo, we should be careful with jumping to conclusion and straightforwardly label it as a humorous. Instead we need to use a significant number of labelled examples containing these emojis to train a machine learning model.

5.3.2 Applications

The application of sentiment analysis using emoji polarities is described in next chapter. In addition to sentiment analysis, emoji polarities also can be applied to spam detection.

According to my current research, I noticed that a lot of spam in Weibo contains many emojis, which are abundant. Nowadays, various spam messages often obfuscate the current data-related tasks standing on their way to achieving better accuracy. Furthermore, some users are being mislead by malicious links or message causing financial or reputation losses. To mitigate the damage caused by spam, various classification algorithms have been applied to distinguish spam from non-spams (Dhaka and Mehrotra, 2019; Tang, Ding, and Zhou, 2019). On the other hand, majority of works in this domain treat emojis as noise, and focus on detecting spammers rather than spam messages. Most papers

focus on English spam detection on social media networks such as Twitter not Weibo. Therefore, we should focus on emojis used in spam content instead of detecting spamming behaviors, and investigate the characteristics of unsolicited or undesirable posts on Weibo.

My hypothesis is that by analyzing the characteristics of spam comments and the sentiment polarities of emojis used in Weibo posts, it will help enhance the accuracy of social media spam detection in the future. To verify this hypothesis, I collected first 500 Weibo comments containing emojis from the "Trending Topics" category. All comments were posted in February, 2020. Then I asked three Chinese native speakers to annotate them in two categories: "spam" or "not spam". After one annotator labeled polarities of all posts, two other native speakers confirmed the correctness of his annotations. Whenever there was a disagreement, all evaluators decided the final polarity through discussion. The results show that the data set was divided into 367 spam comments and 133 general, non-spam comments. The analysis of Weibo emoji usages and the characteristics of Weibo spam comments, are summarized bellow:

- Some spammers are real users, not bots. Often spammers compromised accounts of existing users and send spam messages as these users. We should focus on spam messages rather than spamming accounts in the future, since a spam content can be sent from an actual

user's account.

- Some spam comments are relevant to the topic of a post, which highly increases the difficulty for detection.

- Majority of the spam comments describe erotic content and advertising, due to the strict censorship in the Chinese online communities (it is illegal to make sex-related material open to public).

- Another interesting phenomenon I observed is that a great number of spam comments are semantically ill-structured. Spammers replace the sensitive words with other words/symbols (with similar meaning or pronunciation) or separate sensitive words with emojis to avoid being detected (an example shown in Figure 5.3). It can be problematic to detect such examples of deliberate content concealing.

Translation: "username: charming/tender [awesome] school belle [Clapping Hands] [awesome] @username"

FIGURE 5.3: A example of erotic message which avoids automatic detection by inserting a pictogram inside the text.

- Since many spam posts contain emojis, it is necessary to analyze how the Weibo emojis are used in spam comments. I selected several representative emojis and listed

TABLE 5.3: Occurrence of emojis used in 500 randomly chosen Weibo comments on trending topics.

Emoji	Spam	Non-spam
❤️	863	19
🌹	247	2
🈳	150	0
👈	87	0
😊	60	1
🤔	6	41
🐶	0	83

how many times they were used in my data set. The results (see Table 5.3) show that spam messages rarely use humorous emoji which were annotated in my previous research. This phenomenon suggests a probable use for applying humorous sentiment polarities of emojis to automatically classify spam and non-spam comments in the future. Furthermore, I noticed that spammers tend to use emojis without facial expression, such as ❤️ and 🌹. I also observed that pointing fingers (👈) are almost always used to point to links or accounts inviting users to click.

I believe that by applying sentiment polarities of emojis and neural network, a high-quality spam detection method can be constructed in the future.

Chapter 6

Experiment of Emoji Polarity for Sentiment Analysis

6.1 EPLSTM Approach

In order to verify the validity of my analysis in emoji polarity, I performed series of experiments described below.

I propose a emoji polarity-aware Long Short-Term Memory (LSTM) model for sentiment classification on Weibo. I use LSTM (Hochreiter and Schmidhuber, 1997) as the framework because it can overcome the problem of gradient vanishing compared with recurrent neural network (RNN). An overview of the proposed method for emoji prediction is shown in Figure 6.1. In the first step, I extract the textual features from a Weibo post containing only one type of emoji. Then, the textual features are sent into the LSTM model to learn an output representation. Second, I use a softmax classifier to obtain the predicted results of textual features and output their probability. Next, I give a hyperparameter for each text and emoji, calculate the summation of both features with the hyperparameter. Finally, I can obtain the

sentiment probability of a Weibo post which considers the effect of emojis.

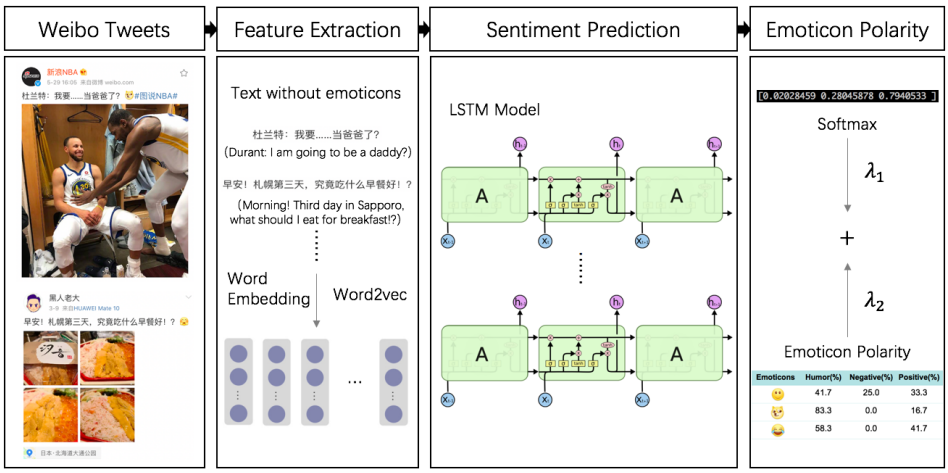

FIGURE 6.1: Overview of the proposed EPLSTM method for sentiment analysis.

6.1.1 Model structure

An LSTM network computes a mapping from an input sequence $x = (x_1, ..., x_T)$ to an output sequence $y = (y_1, ..., y_T)$ by calculating the network unit activations using the following equations iteratively from $t = 1$ to T. The equations of the LSTM cell are as follows:

$$i_t = \sigma(W_{ix}x_t + W_{im}m_{t-1} + W_{ic}c_{t-1} + b_1) \tag{6.1}$$

$$f_t = \sigma(W_{fx}x_t + W_{fm}m_{t-1} + W_{fc}c_{t-1} + b_f) \tag{6.2}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g(W_{cx}x_t + W_{cm}m_{t-1} + b_c) \tag{6.3}$$

$$o_t = \sigma(W_{ox}x_t + W_{om}m_{t-1} + W_{oc}c_t + b_o) \tag{6.4}$$

$$m_t = o_t \odot h(c_t) \tag{6.5}$$

$$y_t = \phi(W_{ym}m_t + b_y) \tag{6.6}$$

where the W terms denote weight matrices (e.g. W_{ix} is the matrix of weights from the input gate to the input), W_{ic}, W_{fc}, W_{oc} are diagonal weight matrices for peephole connections, the b terms denote bias vectors (b_i is the input gate bias vector), σ is the logistic sigmoid function, and i, f, o and c are respectively the input gate, forget gate, output gate and cell activation vectors, all of which are the same size as the cell output activation vector m, $\odot$ is the element-wise product of the vectors, g and h are the cell input and cell output activation functions, generally and also in this paper *tanh*, and ϕ is the network output activation function, softmax in my method.

The outputs of softmax layer $S(z_i)$ are the probability of each category. The softmax function is defined as following (Merity et al., 2016):

$$S(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{i} e^{z_j}} \tag{6.7}$$

where the input of softmax layer z_i is defined as:

$$z_i = w^i x + b_i \tag{6.8}$$

where w is the weight and b is bias, both of them are calculated by training the model.

6.1.2 Emoji polarity

In order to predict sentiment category of Weibo posts considering the influence of emojis for Chinese social media sentiment analysis, I give the probability of the LSTM model's softmax output $S(z_i)$ a hyperparameter λ_1. At the same time, I take the labelled emojis as polarity P_{e}, and assign a hyperparameter λ_2. P becomes the final probability output of the classification:

$$P = \lambda_1 S(z_i) + \lambda_2 P_e \tag{6.9}$$

where the summation of λ_1 and λ_2 is equal to 1.

6.2 Experiments

6.2.1 Preprocessing

Initializing word vectors with those obtained from an unsupervised neural language model is a popular method to improve performance in the absence of a large supervised training set. Considering that there are many neologisms

and Internet slang expressions on Weibo, for my experiment I collected a large dataset (7.31 million posts) from Weibo API from May 2015 to July 2017 to be used for word embedding. First, I deleted the emojis, images, and videos treating them as noise. Next, I used Python Chinese word segmentation module Jieba[1] to segment the sentences of the microblogs, and applied the segmentation results into the Word2vec model (Mikolov et al., 2013) for training word vectors. The vectors have dimensionality of 100 and were trained using the continuous skip-gram architecture.

After that, I collected 3,000 Weibo posts containing the 🙂, 😹 , 😂, 🤷 emojis, ensuring each entry has only one emoji of given type (cases with more emojis of the same type were allowed). To use these posts as my training data, I asked three Chinese native speakers to annotate them into three categories: "positive", "negative", and "humorous". After one annotator labelled polarities of all posts, two other native speakers confirmed correctness of his annotations. Whenever there was a disagreement, all decided the final polarity through discussion.

6.2.2 LSTM model

I deleted all emojis from training data. After training a word embedding model, I passed its output into the LSTM model to train and save the model. I trained the model with

[1] https://github.com/fxsjy/jieba

TABLE 6.1: Precision comparison results.

	Humorous	Negative	Positive
Emoji Only Polarities	60.8%	0.0%	0.0%
LSTM	58.7%	50.0%	26.9%
EP-LSTM	63.0%	75.0%	62.5%

10 epochs and the performance achieved the highest value when the dropout rate was 0.5 and the batch size was 32. The validity of the model was examined by holdout method (95%/5%, training/validation) and it achieved 78.0% accuracy.

6.2.3 Testing

I collected 120 Weibo entries with the four emojis mentioned above, deleted the emojis and processed sentiment classification with the trained LSTM model using the softmax layer which outputs probability of each of three categories. Then I used the proposed method (described in Section 4) to calculate probability of each category and confirmed the precision. Because I assumed that in emotion expression emojis might play a greater role than text, in my experiment, I set the hyper-parameters λ_1 and λ_2 to 0.4 and 0.6 respectively. As the baseline methods, I compared the results of sentiment classification by emoji polarity and of LSTM network only. The results are shown in Table 6.1.

The results showed that my proposed method is more effective than a) LSTM network only categorization and b)

considering just emoji polarities. Limited to small annotated data, the sentiment classification precision predicted by the LSTM model is relatively low, but my proposed approach has improved the performance showing that low-cost, small-scale data labelling is able to outperform widely used state-of-the-art methods.

6.3 Considerations

昨天美国堪萨斯州出现了奇怪的闪电，是从下往上的，是因为发誓的人太多了吗？🐶

Yesterday, there was a strange thunder in Kansas, USA. It was from the ground up to the sky. Is it because there were so many people swearing (they have not lied)? 🐶

FIGURE 6.2: Example of correct prediction.

In my proposed EPLSTM approach, I paid more attention to the weight of the emojis in microblogs, therefore compared to LSTM, the proposed method obtained better performance in some unclear posts. For example (see Figure 6.2), this is a microblog of correct prediction with my EP-LSTM method, which is labeled as “humorous", but it is predicted by LSTM as “negative". There is an idiom in Chinese culture: “I swear if I tell half a lie, I will be killed by the thunder and God". So the microblog expressed a humorous view of the natural phenomenon, and I found that my

proposed method can predict the implicit humorous meaning to correct category with considering the polarity of the emojis.

FIGURE 6.3: Example of unconventional Chinese character choice in a Weibo post with images. Entry says "A notch on the screen.... speechless.", however one of characters of the word "notch" is replaced with another homonymous one changing the meaning to person's name.

I also analyzed the wrongly predicted examples of results. I found that the short text or text with images brings an impact on the results. Weibo microblogs contain a lot of wrongly chosen characters and new words which are difficult to detect, especially for a short entry containing only a few words. I think that adding image processing or adding

a lexicon of neologisms may improve the results in the future.

An example of short text with unconventional character choice and images of Weibo microblog is shown in Figure 6.3.

I also found that some Weibo entries posted with a humorous expression like self-mockery also contain optimistic or pessimistic emotional load. It is quite easy to make a wrong prediction by simply classifying them into "humorous" category. Therefore, A fine-grained humorous classification should be considered in the future.

An example of Weibo entry predicted as "humorous" and annotated as "negative" is shown in Figure 6.4 below.

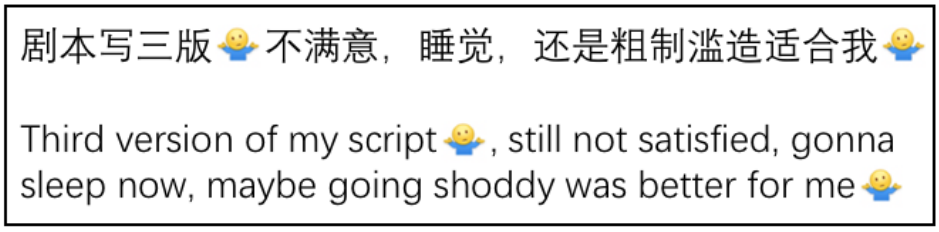

FIGURE 6.4: Example of wrong classification into "humorous" category.

Chapter 7

Experiment of Emoji Lexicon and Slang Lexicon for Humor Detection

7.1 Machine Learning Approaches

Inspired by above mentioned works on Internet slang and emojis, in order to test the influence of them, I utilized both lexicons with several machine learning approaches, k-Nearest Neighbors (k-NN), Decision Tree (DT), Random Forest (RF), Logistic Regression (LR), Naïve Bayes (NB) and Support Vector Machine (SVM) for detecting humorous expressions on social media.

In the first step, I add the Chinese slang lexicon and Chinese emoji lexicon to segmentation tool for matching new words and emojis. Then I use the updated tool to segment the sentences of large data set. Second, I apply the segmentation results into the word embedding tool for training word vectors. Next, I apply the word embedding model

which considered Internet slang and emojis to train a machine learning model with training data. Finally, I input testing data into machine learning model, and I can obtain the sentiment probability of a Weibo post which considers the effect of emojis and Internet slang.

7.2 Experiments

In order to verify the validity of my proposed method, I performed series of experiments described below.

7.2.1 Preprocessing

Initializing word vectors with those obtained from an unsupervised neural language model is a popular method to improve performance in the absence of a large supervised training set. For my experiment I collected a large dataset (7.6 million posts) from Weibo API from May 2015 to July 2017 to be used for calculating word embeddings. First, I deleted the images, and videos treating them as noise. Second, I applied Chinese Internet slang lexicon and Chinese emoji lexicon into the dictionary of Python Chinese word segmentation module Jieba. Next, I used Jieba to segment the sentences of the microblogs, and applied the segmentation results into the word2vec model (Mikolov et al., 2013) for training word vectors. The vectors have dimensionality of 300 and were trained using the continuous skip-gram model.

Next, I collected 3,000 Weibo posts containing the emojis. To use these posts as my training data, I asked three Chinese native speakers to annotate them into two categories: "humorous", and "non-humorous". After one annotator labelled polarities of all posts, two other native speakers confirmed correctness of his annotations. Whenever there was a disagreement, all decided the final polarity through discussion.

7.2.2 Applied Classifiers

Logistic Regression

Logistic regression model is confirmed to be used in many tasks such as document classification (Yu, Huang, and Lin, 2011). In Logistic regression model, I generally correct overfitting with regularization. Regularization adds a penalty term on model to reduce the freedom of the model. Hence, the model will be less likely to fit the noise of the training data and will improve the generalization abilities of the model. I train the model with L2 penalty regularization called Ridge regression in my experiments.

Support Vector Machine

Support vector machine (Cortes and Vapnik, 1995) is a supervised learning model with associated learning algorithms that analyzes data used for classification. An SVM model is a representation of the examples as points in space, mapped

so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall. In addition, it uses kernel trick, implicitly mapping their inputs into high-dimensional feature spaces. In my experiments, I used the radial basis function kernel.

Naïve Bayes

Naïve Bayes classifier is based on applying Bayes theorem with strong independence assumptions between the features. Naïve Bayes has been studied extensively since the 1950s. It was introduced under a different name into the text retrieval community in the early 1960s, and remains a baseline method for text categorization (Rish, 2001), the problem of judging documents as belonging to one category or the other with word frequencies as the features. With appropriate pre-processing, it is competitive in text classification task with more advanced methods including support vector machines. In my experiments, I set the parameter of alpha to 0.01.

k-Nearest Neighbors

In pattern recognition, the k-Nearest Neighbors algorithm is a non-parametric method used for classification and regression. In both cases, the input consists of the k closest training examples in the feature space (Weinberger and Saul, 2009). The output depends on whether k-NN is used for classification or regression. The number of neighbors is set to 5 in my experiments.

Random Forest

Random forests is an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees (Ho, 1995). Random decision forests correct for decision trees habit of overfitting to their training set.

Decision Tree

Decision tree is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. Decision tree is commonly used in operations research, specifically in decision analysis to help identify a strategy most likely to reach a goal, but is also a popular tool in machine learning (Sharma and Kaur, 2013).

TABLE 7.1: Comparison results of machine learning without emojis and Internet slang lexicon.

	Precision	Recall	F1-score
DT	56.21%	55.56%	55.88%
RF	60.26%	54.97%	57.49%
k-NN	63.48%	66.08%	64.76%
NB	59.17%	83.04%	69.10%
LR	60.16%	86.55%	70.98%
SVM	57.00%	100.00%	**72.61%**

7.2.3 Performance Test

Using trained word2vec model, I passed word vectors of training data into the machine learning models to train the model. I collected and annotated 300 Weibo entries with emojis as a test set, deleted images, and videos. Then I used the above mentioned methods to calculate scores of the precision, recall and f1-score. I compared the results of humorous detecting by machine learning only, machine learning considering Internet slang only, and machine learning approaches considering emojis only. The results are shown in Table 7.1, Table 7.2 and Table 7.3, respectively. The Table 7.4 introduces results of the experiment where both Internet slang and emojis were used, and Table 7.5 shows the results of F1-score with above methods.

The results show that considering Internet slang and emojis. Limited to small annotated data, the precision of the humor / non-humor classification was relatively low, but by considering Internet slang and emojis, the F1-score of each classifier outperformed previous method by 1.39% (LR), 2.13%

TABLE 7.2: Comparison results of machine learning with Internet slang lexicon only.

	Precision	Recall	F1-score
RF	59.76%	57.31%	58.51%
k-NN	60.10%	69.59%	64.50%
DT	66.46%	63.74%	65.07%
NB	59.92%	83.04%	69.61%
LR	60.16%	88.30%	71.56%
SVM	57.00%	100.0%	**72.61%**

TABLE 7.3: Comparison results of machine learning with Chinese emojis lexicon only.

	Precision	Recall	F1-score
RF	61.59%	54.39%	57.76%
DT	63.37%	63.74%	63.56%
k-NN	60.70%	71.35%	65.59%
NB	59.92%	83.04%	69.61%
LR	60.08%	88.89%	71.70%
SVM	58.44%	100.00%	**73.77%**

TABLE 7.4: Comparison results of machine learning with both emojis and slang.

	Precision	Recall	F1-score
RF	61.29%	54.75%	58.33%
DT	62.50%	57.26%	59.77%
k-NN	61.37%	70.11%	65.45%
NB	63.60%	82.96%	72.00%
LR	62.30%	86.31%	72.37%
SVM	59.67%	100.00%	**74.74%**

TABLE 7.5: Comparison results of F1-score between feature sets.

	Baseline	Slang	Emojis	Both
RF	57.49%	58.51%	57.76%	58.33%
DT	55.88%	65.07%	63.56%	59.77%
k-NN	64.76%	64.50%	65.59%	65.45%
NB	69.10%	69.61%	69.61%	72.00%
LR	70.98%	71.56%	71.70%	72.37%
SVM	**72.61%**	**72.61%**	**73.77%**	**74.74%**

(SVM), 2.90% (NB), 0.69% (k-NN), 0.84% (RF) and 3.89% (DT). My proposed approach has improved the performance showing that low-cost, small-scale data labeling is able to outperform widely used state-of-the-art when emoji and slang information is added to the learning process.

7.3 Considerations

Post: 理想生活，一颗赛艇 🙂
Pinyin: *Li xiang sheng huo, yi ke sai ting* 🙂
Segmentation: 理想/生活/，/一颗赛艇/[微笑]
Translation: Ideal life, exciting. 🙂

FIGURE 7.1: Example of correct classification of humorous post.

In my proposed approach, I paid more attention to the emojis and Internet slang in microblogs and investigated how adding these features separately and together influences the previously proposed method for recognizing humorous posts which are problematic when it comes to semantic analysis. Figure 7.1) shown an example of a microblog which was correctly classified by my proposed method as “humorous" while the baseline recognized it incorrectly as non-humorous. This post contains word 一颗赛艇 (*yi ke sai ting* which is a homophone of English word “exciting”). The baseline does not know this expression and the parser

divides it as 一颗/赛艇 (*yi ke / sai ting* which means "a rowing boat"). When this expression is accompanied by 🙂 emoji, they both improve the performance of classification and predict the implicit humorous meaning.

Post: 大麦网真会玩，2.14给我发个5.21号演唱会的票务信息 🤷

Pinyin: *Da mai wang zhen hui wan, 2.14 gei wo fa ge 5.21 hao yan chang hui de piao wu xin xi* 🤷

Segmentation: 大麦/网真会/玩/，/2.14/给我发/个/5.21/号/演唱会/的/票务/信息/[摊手]

Translation: Damaiwang really knows how to live it up, they sent me a concert ticket information of May 21st on February 14 🤷

FIGURE 7.2: Example of wrong classification into "non-humorous" category.

Error analysis showed that some posts were wrongly predicted due to proper nouns missing in the parser's dictionary which brought clearly negative impact on the results. In Figure 7.2 I show an example of such misclassification into "non-humorous" category annotated as "humorous" by annotators. Name of a ticketing website *Da mai wang* was parsed incorrectly, and one shifted character caused

mis-recognition of humorous word. Weibo microblogs contain numerous ideograms deliberately altered from their everyday meaning, what makes them difficult to parse and match. I think that adding new named entities into the parser's dictionary may significantly improve the results in the future. I observed that when emotions are expressed online, emojis might play a greater role than it is usually considered, therefore I will experiment with weight of the emojis in the future.

Chapter 8

Humor Detection

8.1 HEMOS System

Inspired by the works on Internet slang and emojis mentioned above, I utilized both lexicons with attention-based bi-directional long short-term memory recurrent neural network for sentiment analysis of Chinese social media to build my **HEMOS** (**H**umor-**EMO**ji-**S**lang-based) system. In the real-life (offline) dialogue between human beings, besides tone changes, we usually express emotions with body language. In social networks, this nonverbal type of communication can partially be mimicked by using emojis (Aldunate and González-Ibáñez, 2017). Hence, as a working hypothesis, I assume that all microblogs with emojis convey non-neutral emotions. First, I add the Chinese slang lexicon and Chinese emoji lexicon to a segmentation tool for matching new words and pictograms. Then I use the updated tool to segment the sentences of a large data set. Second, I apply the segmentation output into the word embedding tool for acquiring word vectors. Next, I apply the word embedding

model, which considers Internet slang and emojis, to train an attention-based bi-directional long short-term memory recurrent neural network model (AttBiLSTM) with training data to learn an output representation. In the last step, I input testing data into the AttBiLSTM model, and I use a softmax classifier to obtain the predicted results and output their probability.

To solve the problems of humor detection encountered in the previous research (Li et al., 2018), I applied AttBiLSTM model to achieve a fine-grained classification method for classifying Weibo divided into four categories: positive, negative, optimistic humorous and pessimistic humorous. The "optimistic humorous" category includes jokes, self-mockery and jocosity and etc., while the "pessimistic humorous" category contains sarcasm and irony. I believe that the proposed fine-grained classification method can more effectively detect humorous expressions and improve the results of "positive/negative" bi-polarity sentiment classification.

In my experiments, I focus on the humorous posts using Internet slang and emojis on Weibo in order to verify:

1) if both slang and emojis lexicons improve sentiment analysis results by recognizing humorous entries which are difficult to polarize;

2) if adding new "optimistic humorous type" and "pessimistic humorous type" categories improve the result of bi-polarity sentiment prediction.

8.1.1 Attention-based Bi-Directional Long Short-Term Memory Recurrent Neural Network

The architecture of attention-based bi-directional long short-term memory recurrent neural network model (AttBiLSTM) is shown in Figure 8.1; my proposed model mainly consists of word encoder, attention layer and softmax layer. The details are presented below.

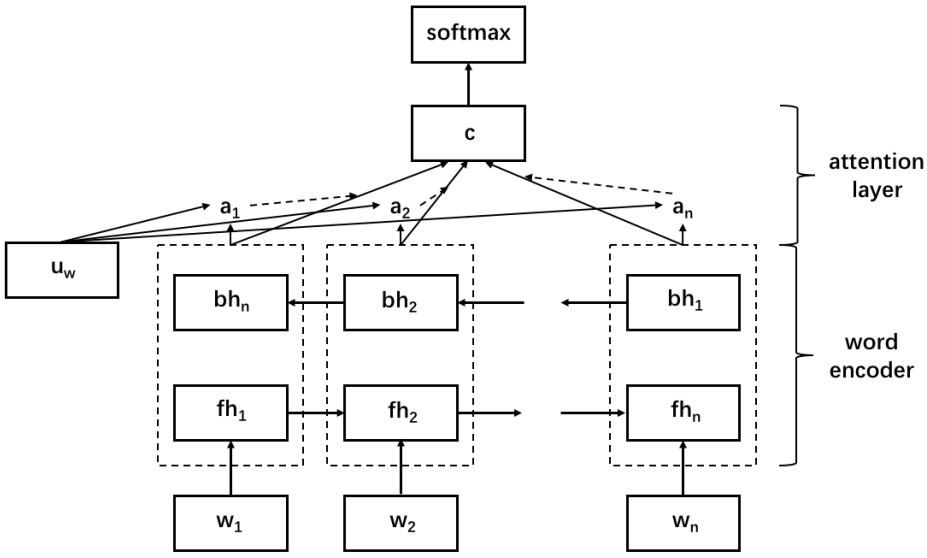

FIGURE 8.1: The architecture of AttBiLSTM model.

Word Encoder

Considering that the entries of Weibo are sentences of less than 140 words, in contrast to related work of (Yang et al., 2016), in my research I focus on sentence-level social media sentiment classification. Assuming that a sentence contains n words $(w_1, w_2, ..., w_n)$, w_k denotes the kth word in a sentence and n presents the length of a sentence, every word is

embedded into a d-dimensional vector which is called word embedding (Bengio et al., 2003). Then, an embedding matrix $M^{n \times d}$ is generated by word embedding layer, where n is the length of the sentence and d is the embedding size. Finally, the matrix is applied as input for the bidirectional LSTM networks.

Bidirectional LSTM networks are well-suited to classifying, processing, and making predictions based on time series data because there can be lags of unknown duration between important events in a time series. Bidirectional networks outperform unidirectional ones, and Long Short Term Memory (LSTM) is much faster, and also more accurate than both standard Recurrent Neural Networks (RNNs) or time-windowed Multilayer Perceptrons (MLPs) (Graves and Schmidhuber, 2005).

An LSTM network computes a mapping from an input sequence $x = (x_1, ..., x_\mathrm{T})$ to an output sequence $y = (y_1, ..., y_\mathrm{T})$ by calculating the network unit activations using the following equations iteratively from $t = 1$ to T. The equations of the LSTM cell are as follows (Hochreiter and Schmidhuber, 1997):

$$X = \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} \tag{8.1}$$

$$f_t = \sigma(W_f \cdot X + b_f) \tag{8.2}$$

$$i_t = \sigma(W_i \cdot X + b_i) \tag{8.3}$$

$$o_t = \sigma(W_o \cdot X + b_o) \tag{8.4}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot tanh(W_c \cdot X + b_c) \tag{8.5}$$

$$h_t = o_t \odot tanh(c_t) \tag{8.6}$$

where the W terms denote weight matrices, W_i, W_f, W_o are diagonal weight matrices for peephole connections, the b terms denote bias vectors (b_i is the input gate bias vector), σ is the logistic sigmoid function, and i, f, o and c are respectively the input gate, forget gate, output gate and cell activation vectors, $\odot$ is the element-wise product of the vectors, the cell input and cell output activation functions, generally (as well as in my research) $tanh$. x_t denotes the word embedding of the input of LSTM cell and h_t is the vector of the hidden state. The specific schematic is illustrated in Figure 8.2.

The bidirectional LSTM networks contain two independent LSTMs, which acquire annotations of words by merging information from two directions of a sentence (Graves and Schmidhuber, 2005). Specifically, at the time step t, the forward LSTM calculates the hidden state fh_t based on the previous hidden fh_{t-1} state and the input vector x_t, while

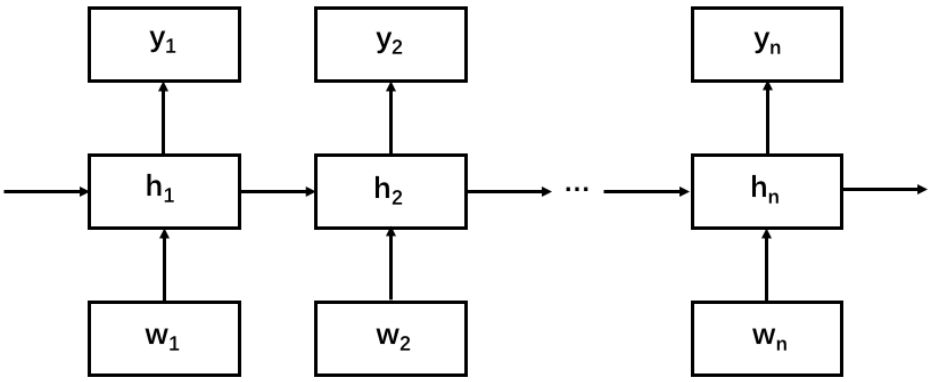

FIGURE 8.2: The architecture of LSTM model.

the backward LSTM calculates the hidden state bh_t based on the opposite hidden state bh_{t-1} and the input vector x_t. Finally, the vectors of two directions are concatenated as the final hidden state. The two LSTM neural network parameters in bidirectional LSTM networks are independent of each other, and they share the same word embeddings of the sentence. The final output h_t of the bidirectional LSTM model at time step t is defined as following:

$$h_t = \left[fh_t, bh_t \right] \tag{8.7}$$

Attention Layer

My proposed model projects a raw Weibo post into a vector representation, on which I build a classifier to perform sentiment classification. In this subsection, I introduce how I build the sentence-level vector progressively from word

vectors by using the attention structure. Not all words contribute equally to the representation of the Weibo entry meaning. Hence, I introduce attention mechanism to extract words which are important to the meaning of the post and show how I calculate the total of the representation of those informative words to form a sentence vector.

As I mentioned above, the bidirectional LSTM networks produce a hidden h_t state at each time step. I first feed the word annotation h_t through a one-layer MLP to learn a hidden representation u_t. Then I measure the importance of the word as the similarity of u_t with a word-level context vector u_w and get a normalized importance weight α_t through a softmax function. Secondly, I compute the sentence vector c as a weighted sum of the word annotations based on the weights. The context vector u_w can be perceived as a high-level representation of a fixed query of "the informative word" over the words like those used in memory networks (Sukhbaatar, Weston, and Fergus, 2015). The equations are described as follows:

$$u_t = tanh(W_w h_t + b_w) \tag{8.8}$$

$$\alpha_t = \frac{exp(u_t^\top u_w)}{\sum_t exp(u_t^\top u_w)} \tag{8.9}$$

$$c = \sum_t \alpha_t h_t \tag{8.10}$$

Softmax Layer

The word context vector u_w is randomly initialized and jointly learned during the training process. The outputs of softmax layer $\hat{y}$ are the probabilities of each category. The softmax function is defined as (Bridle, 1990; Merity et al., 2016):

$$\hat{y} = softmax(W_c c + b_c) \tag{8.11}$$

where W_c is the weight and b_c is bias, both of them calculated during the model training process.

8.2 Experiments

8.2.1 Preprocessing

Initializing word vectors with those obtained from an unsupervised neural language model is a popular method to improve performance in the absence of a large supervised training set. For my experiment, I crawled a large dataset of 7.6 million posts from Weibo API from May 2015 to July 2017 for calculating word embeddings. Firstly, I filtered images and videos and treated them as noise. Secondly, I added the Chinese Internet slang lexicon and Chinese emoji lexicon into the dictionary of Jieba[1], a python package for Chinese text segmentation. Followed by the text segmentation with Jieba, the output of tokens is fitted into the word2vec model (Mikolov et al., 2013) for training. The output of

[1] https://github.com/fxsjy/jieba

the vectors has a dimensionality of 300, and then they were trained through the continuous skip-gram model.

In the next step, I collected 4,000 Weibo posts containing ambiguous (🙂, 🐱, 😂, 🤷‍♀️, 😏, 🤔, 😤, 🙄) emojis, ensuring each entry has only one emoji of given type (cases with more emojis of the same type were allowed). To use these posts as my training data, I asked three Chinese native speakers (each of them being a Weibo user for more than eight years) to annotate posts using four category labels: "positive," "negative," "optimistic humorous," and "pessimistic humorous." After one annotator labeled polarities of all posts, two other native speakers confirmed the correctness of his annotations. Whenever there was a disagreement, all evaluators decided the final polarity through discussion. I verbally confirmed with all annotators that they understood the meaning of "optimistic humorous" and "pessimistic humorous" categories.

I trained the AttBiLSTM model with 10 epochs to discover that the performance achieved the highest value when the dropout rate was 0.25, and the batch size was 64. The validity of the model was examined by 10-fold cross-validation.

8.2.2 Performance Test

Using the trained word2vec model, I passed word vectors of training data into the AttBiLSTM model to train the proposed model. I collected and annotated 180 Weibo entries

TABLE 8.1: Results of two categories sentiment classification by AttBiLSTM method only.

Categories	Evaluation	Results
Positive	Precision	69.07%
	Recall	82.76%
	F1-score	74.71%
Negative	Precision	78.05%
	Recall	50.00%
	F1-score	60.95%

TABLE 8.2: Results of two categories sentiment classification by AttBiLSTM considering Internet slang and emojis lexicons.

Categories	Evaluation	Results
Positive	Precision	82.35%
	Recall	84.48%
	F1-score	83.40%
Negative	Precision	73.77%
	Recall	70.31%
	F1-score	71.99%

with the eight emojis mentioned above as a testing set, deleting images and videos. Then I used the proposed method to calculate the probability of each category and confirmed the precision, recall, and F1-score.

I compared the results of a) two-categories sentiment classification by the AttBiLSTM method only; b) two-categories sentiment classification by the AttBiLSTM considering Internet slang and emojis lexicons; c) four-categories sentiment classification by AttBiLSTM method only, with my proposed method d) four-categories sentiment classification by AttBiLSTM considering Internet slang and emojis lexicons. In two-categories sentiment classification, "optimistic humorous" and "pessimistic humorous" labels were treated

TABLE 8.3: Results of four categories sentiment classification by AttBiLSTM method only.

Categories	Evaluation	Results
Positive	Precision	71.15%
	Recall	70.00%
	F1-score	70.57%
Negative	Precision	68.89%
	Recall	67.39%
	F1-score	68.13%
Optimistic Humorous	Precision	72.72%
	Recall	60.61%
	F1-score	66.18%
Pessimistic Humorous	Precision	39.29%
	Recall	61.11%
	F1-score	47.83%

as "positive" and "negative".

Results of two categories sentiment classification are shown in Table 8.1 and 8.2, respectively. Table 8.3 and 8.4 show the results of four-categories sentiment classification methods. The results proved that:

1) both slang and emojis lexicons can improve sentiment analysis results by recognizing humorous entries which are challenging to polarize;

2) adding new "optimistic humorous" and "pessimistic humorous" categories can improve the result of bi-polarity sentiment prediction;

3) my proposed methods can obtain the best performance for humor detecting and traditional bi-polarity sentiment analysis.

Limited to small annotated data, the precision of the "negative" and "pessimistic humorous" was relatively low, but

TABLE 8.4: Results of my proposed method

Categories	Evaluation	Results
Positive	Precision	89.79%
	Recall	88.00%
	F1-score	88.89%*
Negative	Precision	78.57%
	Recall	71.74%
	F1-score	74.99%*
Optimistic Humorous	Precision	79.71%
	Recall	83.33%
	F1-score	81.48%*
Pessimistic Humorous	Precision	65.00%
	Recall	72.22%
	F1-score	68.42%*

$^{*}p < 0.05$

by adding "optimistic humorous" and "pessimistic humorous" categories and considering Internet slang and emojis, the F1-score of each category outperformed previous method. My proposed four-categories sentiment analysis approach has improved the performance showing that low-cost, small-scale data labeling can outperform widely used state-of-the-art when emoji and slang information is added to the learning process.

8.3 Considerations

In my proposed approach, I paid more attention to emojis and Internet slang in humorous microblogs. I investigated how adding these features influence the previously proposed methods for recognizing humorous posts which are problematic when it comes to semantic analysis. Figure 8.3 presents an example of a microblog that was correctly

classified by my proposed method as “optimistic humorous" while the baseline recognized it incorrectly as a negative one.

Post: 裁判：我当初就是因为没有对手，才选择做裁判的 🤷

Pinyin: *Cai pan: Wo dang chu jiu shi yin wei mei you dui shou, cai xuan ze zuo cai pan de* 🤷

Segmentation: 裁判/：/我/当初/就是/因为/没有/对手/，/才/选择/做/裁判/的/[摊手]

Translation: Referee: Because there was no opponent who could beat me, I chose to become a referee 🤷

FIGURE 8.3: Example of correct classification of humorous post.

This post and similar entries were usually posted as a comment a GIF or video showing a referee who displays her or his skills in basketball by performing a slam dunk. This entry seems to express an implied humorous nuance of an exaggerated surprise when the poster saw how good the referee was. Because this expression is accompanied by 🤷 emoji, it improves the performance of classification and predicts the implicit humorous meaning.

As a solution of problems of humor detection encountered in previous research (Li et al., 2018), the fine-grained sentiment classification method I proposed can detect the emotions of Weibo posts more clearly. In Figure 8.4, I show an example of a microblog which was correctly classified by my proposed method as “optimistic humorous," while

Post: 吃饱了就有力气减肥了 😆😆
Pinyin: *Chi bao le jiu you li qi jian fei le* 😆😆
Segmentation: 吃饱了/就/有/力气/减肥/了/[阴险]/[阴险]
Translation: When your stomach is full, you get the strength to reduce weight 😆😆

FIGURE 8.4: Another example of correct classification of humorous post.

the baseline recognized it as a positive one. As the evaluators agreed, it seems that this user wrote a joke just for fun, and my proposed method correctly recognized this kind of emotion.

Error analysis showed that the results of detecting negative emotions and "pessimistic humorous" emotions in my proposed method were still relatively low, which is closely related to the difficulty in recognizing sarcasm and irony. I plan to train more deep learning models and increase the amount of data to improve the results of pessimistic humorous detecting in the future.

Furthermore, some posts were wrongly predicted due to new slang missing from both the parser's dictionary and my slang lexcion which brought clearly negative impact on the results. In the research of (Ptaszynski et al., 2016), the authors pointed that a typical cause of gradual decrease of performance of systems dealing with Internet language has been the fact, that Internet slang has been constantly changing. This point is also reflected in my research; in Figure 8.5

Post: 名人英文金句翻译与当下流行语神对应, 励志心灵鸡汤秒变毒鸡汤 🐶
Pinyin: *Ming ren ying wen jin ju fan yi yu dang xia liu xing yu shen dui ying, li zhi xin ling ji tang miao bian du ji tang* 🐶
Segmentation: 名人/英文/金句/翻译/与/当下/流行语/神/对应/, /励志/心灵鸡汤/秒/变毒/鸡汤/[doge]
Translation: Translation of English familiar quotations corresponds to the current Internet slang, chicken soup becomes poisonous chicken soup in seconds 🐶

FIGURE 8.5: An example of an "optimistic humorous" post misclassified as "pessimistic humorous".

I show an example of a post misclassified as "pessimistic humorous" category, but annotated as "optimistic humorous" by annotators. Slang expressions as *miao bian* ("changing in seconds") and *du ji tang* ("poisonous chicken soup") were parsed incorrectly, and one shifted character caused mis-recognition by the segmentation tool. *Du ji tang* is a slang word transformed from *ji tang* which means "anti-motivational quotes" (for example: "Some are born great, some achieve greatness, and some wind up like you" or "I'm not lazy, I am just highly motivated to do nothing"). As the abbreviation of "Chicken Soup for the Soul" book series, *ji tang* is used to express the meaning of "motivational quotes" in recent years. Abundant new words similar to *du ji tang* are emerging on social media every year. Adding new phrases to slang lexicon is costly, and it is not

enough to keep up with the speed of Internet slang evolution. To deal this phenomenon, a character level contextualized word embedding method (for example, pre-trained Chinese word embedding model by BERT) could be considered in the next stage of this research.

Chapter 9

Conclusions of the Thesis

In this study, I proposed HEMOS (Humor-EMOji-Slang) system for fine-grained sentiment classification. I collected 576 frequent Chinese Internet slang expressions and created a slang lexicon; then, I converted the 109 Weibo emojis into textual features creating a Chinese emoji lexicon. I also analyzed sentiment polarity of Weibo emojis. I performed series of experiments to verify the validity of both lexicons and emoji polarities. Furthermore, with new "optimistic humorous type" and "pessimistic humorous type" added, I created the basis for a new, four-level sentiment classification of Weibo posts. I applied both lexicons to my novel deep learning approach, namely attention-based bi-directional long short-term memory recurrent neural network (AttBiLSTM) for more fine-grained sentiment analysis of Chinese social media. My experimental results show that the HEMOS system can significantly improve the performance for predicting sentiment polarity on Weibo.

In order to achieve an even more effective Chinese sentiment analysis method, I am going to increase the amount

of labeled data to solve the problem of visibly lower results for "pessimistic humorous" and "negative" categories in the proposed fine-grained classification.

In this study, I exclude images and videos. However, in further research, it would be interesting to add images to the data source. I assume this may enhance the text sentiment analysis since stickers and memes also carry emotions. To utilize such additional information, an image processing phase must be added during the preprocessing stage.

Moreover, during the data labeling phrase, I found that, compared with regular users, there is a high occurrence of posts with specific emojis that are used by spammers (users that spread malicious links or commercial content). Dealing with this problem could be an interesting research topic, and my methods could be useful for differentiating regular users from spammers.

My ultimate goal is to investigate how much the newly introduced emotion-related features are beneficial for sentiment analysis by feeding them to a deep learning model, which should allow us to construct a high-quality sentiment recognizer for a wider spectrum of sentiment in the Chinese language.

Bibliography

Aldunate, Nerea and Roberto González-Ibáñez (2017). "An integrated review of emoticons in computer-mediated communication". In: *Frontiers in Psychology* 7, p. 2061.

Attardo, Salvatore (2010). *Linguistic theories of humor*. Vol. 1. Walter de Gruyter.

Batty, Magali and Margot J Taylor (2003). "Early processing of the six basic facial emotional expressions". In: *Cognitive Brain Research* 17.3, pp. 613–620.

Bengio, Yoshua et al. (2003). "A neural probabilistic language model". In: *Journal of machine learning research* 3.Feb, pp. 1137–1155.

Bridle, John S (1990). "Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition". In: *Neurocomputing*. Springer, pp. 227–236.

Chen, Yuxiao et al. (2018). "Twitter Sentiment Analysis via Bi-sense Emoji Embedding and Attention-based LSTM". In: *2018 ACM Multimedia Conference on Multimedia Conference*. ACM, pp. 117–125.

Chen, Zhao et al. (2015). "Combining convolution neural network and word sentiment sequence features for Chinese text sentiment analysis". In: *Journal of Chinese Information Processing*, 29(6): 172–178.

Cortes, Corinna and Vladimir Vapnik (1995). "Support-vector networks". In: *Machine learning* 20.3, pp. 273–297.

Dhaka, Deepali and Monica Mehrotra (2019). "Cross-Domain Spam Detection in Social Media: A Survey". In: *International Conference on Emerging Technologies in Computer Engineering*. Springer, pp. 98–112.

Dhande, Lina L and Girish K Patnaik (2014). "Analyzing sentiment of movie review data using Naive Bayes neural classifier". In: *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)* 3.4, pp. 313–320.

Ekman, Paul and Wallace V Friesen (1971). "Constants across cultures in the face and emotion." In: *Journal of personality and social psychology* 17.2, p. 124.

Felbo, Bjarke et al. (2017). "Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm". In: *arXiv preprint arXiv:1708.00524*.

Furnham, Adrian (1984). "Tourism and culture shock". In: *Annals of Tourism Research* 11.1, pp. 41–57.

Graves, Alex and Jürgen Schmidhuber (2005). "Framewise phoneme classification with bidirectional LSTM and other neural network architectures". In: *Neural Networks* 18.5-6, pp. 602–610.

Guibon, Gaël, Magalie Ochs, and Patrice Bellot (2016). "From emojis to sentiment analysis". In: *WACAI 2016*.

Ho, Tin Kam (1995). "Random decision forests". In: *Document analysis and recognition, 1995., proceedings of the third international conference on*. Vol. 1. IEEE, pp. 278–282.

Hochreiter, Sepp and Jürgen Schmidhuber (1997). "Long short-term memory". In: *Neural Computation* 9.8, pp. 1735–1780.

Jones, Graham M and Bambi B Schieffelin (2009). "Talking text and talking back:"My BFF Jill" from boob tube to YouTube". In: *Journal of Computer-Mediated Communication* 14.4, pp. 1050–1079.

Jones, Jonathon D (2005). *A complete analysis of Plato's philosophy of humor*. http://www.jonathonjones.com/papers/plato.pdf.

Kavanaugh, Andrea L et al. (2012). "Social media use by government: From the routine to the critical". In: *Government Information Quarterly* 29.4, pp. 480–491.

Khan, Aurangzeb et al. (2010). "A review of machine learning algorithms for text-documents classification". In: *Journal of advances in information technology* 1.1, pp. 4–20.

Kim, Yoon (2014). "Convolutional neural networks for sentence classification". In: *arXiv preprint arXiv:1408.5882*.

Kulkarni, Vivek and William Yang Wang (2017). "TFW, DamnGina, Juvie, and Hotsie-Totsie: On the Linguistic and Social Aspects of Internet Slang". In: *arXiv preprint arXiv:1712.08291*.

Li, Da, Rafal Rzepka, and Kenji Araki (2018). "Preliminary Analysis of Weibo Emojis for Sentiment Analysis of Chinese Social Media, Proceedings". In: *The 32th Annual Conference of the Japanese Society for Artificial Intelligence*, 1J3–04.

Li, Da et al. (2018). "Emoticon-Aware Recurrent Neural Network Model for Chinese Sentiment Analysis". In: *The Ninth IEEE International Conference on Awareness Science and Technology (iCAST 2018)*, pp. 161–166.

— (2019a). "A Novel Machine Learning-based Sentiment Analysis Method for Chinese Social Media Considering Chinese Slang Lexicon and Emoticons". In: *The AAAI-19 Workshop on Affective Content Analysis, AffCon 2019*, CEUR WS Vol–2328, paper 10.

— (2019b). "Emoji-Aware Attention-based Bi-directional GRU Network Model for Chinese Sentiment Analysis". In: *Joint Proceedings of the Workshops on Linguistic and Cognitive Approaches to Dialog Agents (LaCATODA 2019) and on Bridging the Gap Between Human and Automated Reasoning (BtG 2019) co-located with 28th International Joint Conference on Artificial Intelligence (IJCAI 2019)*, CEUR WS Vol–2452, paper 2.

Li, Ruijing et al. (2014). "A method of polarity computation of Chinese sentiment words based on gaussian distribution". In: *International Conference on Intelligent Text Processing and Computational Linguistics*. Springer, pp. 53–61.

Liu, Shuhua Monica and Jiun-Hung Chen (2015). "A multi-label classification based approach for sentiment classification". In: *Expert Systems with Applications* 42.3, pp. 1083–1093.

Manuel, K, Kishore Varma Indukuri, and P Radha Krishna (2010). "Analyzing internet slang for sentiment mining". In: *2010 second Vaagdevi*

international conference on information Technology for Real World Problems. IEEE, pp. 9–11.

Martin, Rod A (2006). *The psychology of humor*. Elsevier.

Martin, Rod A and Thomas Ford (2018). *The psychology of humor: An integrative approach*. Academic Press.

Mathews, Lindsay (2016). *Role of Humor in Emotion Regulation: Differential Effects of Adaptive and Maladaptive Forms of Humor*. CUNY Academic Works.

Merity, Stephen et al. (2016). "Pointer sentinel mixture models". In: *arXiv preprint arXiv:1609.07843*.

Mikolov, Tomas et al. (2013). "Efficient estimation of word representations in vector space". In: *arXiv preprint arXiv:1301.3781*.

Novak, Petra Kralj et al. (2015). "Sentiment of emojis". In: *PlOS ONE* 10.12, e0144296.

Peng, Haiyun, Erik Cambria, and Amir Hussain (2017). "A review of sentiment analysis research in Chinese language". In: *Cognitive Computation* 9.4, pp. 423–435.

Ptaszynski, Michal et al. (2016). "Sustainable cyberbullying detection with category-maximized relevance of harmful phrases and double-filtered automatic optimization". In: *International Journal of Child-Computer Interaction* 8, pp. 15–30.

Raskin, Victor (2012). *Semantic mechanisms of humor*. Vol. 24. Springer Science & Business Media.

Reyes, Antonio, Paolo Rosso, and Davide Buscaldi (2012). "From humor recognition to irony detection: The figurative language of social media". In: *Data & Knowledge Engineering* 74, pp. 1–12.

Rish, Irina et al. (2001). "An empirical study of the naive Bayes classifier". In: *IJCAI 2001 workshop on empirical methods in artificial intelligence*. Vol. 3. 22. IBM New York, pp. 41–46.

Ruch, Willibald (1993). *Exhilaration and humor*. Vol. 1. The Guilford Press, pp. 605–616.

Rzepka, Rafal, Noriyuki Okumura, and Michal Ptaszynski (2017). "Worlds Linking Faces – Meaning and Possibilities of Contemporary Pictograms

(in Japanese)". In: *Journal of the Japanese Society for Artificial Intelligence*, pp. 350–355.

Sharma, Priyanka and Manavjeet Kaur (2013). "Classification in pattern recognition: A review". In: *International Journal of Advanced Research in Computer Science and Software Engineering* 3.4.

Sharma, Raja (2011). *Comedy in New Light-Literary Studies*. Lulu. com.

Soliman, Taysir Hassan et al. (2014). "Sentiment analysis of Arabic slang comments on facebook". In: *International Journal of Computers & Technology* 12.5, pp. 3470–3478.

Sukhbaatar, Sainbayar, Jason Weston, Rob Fergus, et al. (2015). "End-to-end memory networks". In: *Advances in neural information processing systems*, pp. 2440–2448.

Tan, Songbo and Jin Zhang (2008). "An empirical study of sentiment analysis for Chinese documents". In: *Expert Systems with applications* 34.4, pp. 2622–2629.

Tang, Wenbing, Zuohua Ding, and Mengchu Zhou (2019). "A spammer identification method for class imbalanced weibo datasets". In: *IEEE Access* 7, pp. 29193–29201.

Vinodhini, G and RM Chandrasekaran (2012). "Sentiment analysis and opinion mining: a survey". In: *International Journal* 2.6, pp. 282–292.

Wang, Xinyu et al. (2013). "A depression detection model based on sentiment analysis in micro-blog social network". In: *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, pp. 201–213.

Weinberger, Kilian Q and Lawrence K Saul (2009). "Distance metric learning for large margin nearest neighbor classification". In: *Journal of Machine Learning Research* 10.Feb, pp. 207–244.

Wu, Liang, Fred Morstatter, and Huan Liu (2016). "Slangsd: Building and using a sentiment dictionary of slang words for short-text sentiment classification". In: *arXiv preprint arXiv:1608.05129*.

Yang, Zichao et al. (2016). "Hierarchical attention networks for document classification". In: *Proceedings of the 2016 Conference of the North*

American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 1480–1489.

Yu, Hsiang-Fu, Fang-Lan Huang, and Chih-Jen Lin (2011). "Dual coordinate descent methods for logistic regression and maximum entropy models". In: *Machine Learning* 85.1-2, pp. 41–75.

Yue, Xiao Dong (2010). *Exploration of Chinese humor: Historical review, empirical findings, and critical reflections*.

Yue, Xiaodong (2014). "The attitudes towards humor of Confucian, Buddhist, and Taoist culture". In: *Psychological Exploration*, pp. 1–5.

Zagibalov, Taras and John Carroll (2008). "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, pp. 1073–1080.

Zhang, Changli et al. (2009). "Sentiment analysis of Chinese documents: From sentence to document level". In: *Journal of the American Society for Information Science and Technology* 60.12, pp. 2474–2487.

Zhao, Jichang et al. (2012). "Moodlens: an emoticon-based sentiment analysis system for chinese tweets". In: *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1528–1531.

Zhao, Peijun et al. (2018). "Analyzing and Predicting Emoji Usages in Social Media". In: *Companion of the The Web Conference 2018 on The Web Conference 2018*. International World Wide Web Conferences Steering Committee, pp. 327–334.

Zhuo, Shaojian, Xing Wu, and Xiangfeng Luo (2014). "Chinese text sentiment analysis based on fuzzy semantic model". In: *Cognitive Informatics & Cognitive Computing (ICCI* CC), 2014 IEEE 13th International Conference on*. IEEE, pp. 535–540.

Acknowledgements

I would like to express my gratitude to the following people:

Prof. Kenji Araki

for his unending encouragement and support throughout all my years in the Language Media Laboratory;

Assistant Prof. Rafał Rzepka

for providing continuous support and guidance on all my research projects and endeavors;

Prof. Michal Ptaszynski

for providing valuable advice on my research projects;

Members of the review committee:

Prof. Yuji Sakamoto, Prof. Miki Haseyama and Prof. Toshihiko Ito

for their insight and help in refining this thesis;

My wife Mengmeng Wang

for her invaluable support;

Colleagues from the Language Media Laboratory

for all the interesting discussions and mutual support;

and

All my friends, both in real life and online

for their encouragement and unswerving belief in my abilities.