A Study on Sentence Multi-emotions Analysis from Different Perspectives

Yunong Wu

A Thesis submitted to the University of Tokushima in partial fulfillment of the requirements for the degree of Doctor of Philosophy

March, 2014

Department of Information Science and Intelligent Systems
Graduate School of Advanced Technology and Science
The University of Tokushima

Acknowledgment

First of all, I would like to express my deepest gratitude to my supervisor Kenji Kita, who is a professor of the Department of Information Science and Intelligent Systems, the University of Tokushima. Thanks to his constitute encouragement, kind-hearted help, patient guidance, I not only obtain a lot of inspiration in my research, but also achieve a satisfied result in my research.

Secondly, I would like show my heartfelt gratitude to assistant Kazuyuki Matsumoto, who have given me a lot of advices and help and precious suggestions for my research, and assistant Kazuyuki Matsumoto had help me to revise my thesis.

Thirdly, I also would like to express sincere gratefulness to the members of committee: Professor Masami Shishibori and Fuji Ren, for spending their valuable time on reading my thesis and their insightful comments.

Finally, many thanks to all of the members in Kita-lab for their help and support all the times. Especially for Xin Kang, he gave me many valuable suggestions on my experiments and the great encouragement when I met some troubles in the research.

Contents

List of Figures

List of Tables

Abstract

Emotion prediction has been a core task in affective computing, which aims at finding the thorough human mental states by analyzing peoples activities. In this paper, we focus on predicting emotions in the public online blogs from different people, by extracting as many reasonable emotions for each blog sentence as possible. Concretely, we consider three different perspectives for analyzing the multiple emotions in a sentence: 1. predict sentence emotions by examining the emotion related topics in a global sense; 2. predict the sentence emotions from the context-sensitive word emotions; 3. predict sentence emotions by considering the emotional significance in the local bag-of-words. We build different probabilistic models from each perspective, to separately generate the sentence emotion probabilities. We then integrate these probabilistic models to jointly predict the emotion probabilities. Because the component models are based on different emotional assumptions with distinct features, the integrated predictions should predict emotions from more general perspectives and therefore yield the better results. In the experiment, we employ different evaluation criteria to compare the multi-emotion predictions from the single and the integrated modems. Compared to the results in the baseline model, our Bi-integrated model achieves a 5.64% higher Micro F1 and a 6.47% higher Macro F1 scores, respectively. Moreover, our Tri-integrated model acquires a 5.84% higher Micro F1 and a 6.26% higher Macro F1 scores than the baseline results, which have proved our assumption, and suggested interesting features in the different emotion perspectives.

Chapter 1

Introduction

There has been a growing number of studies in the field of affective information computing among the artificial intelligence community [41], as making machines to perceive the humans mental states could greatly help finding the potential interested buyers for the online markets or understanding the public opinions towards some public events. And with the rise of social websites around the world, large amount of data has been available for such functions to be realized.

In practice, however, because emotions are hidden mental states, they can only be predicted by analyzing human activities like the facial expression, the tone of voice, and the mental thoughts expressed by words. In this paper, we describe three probabilistic models for analyzing human emotions from the written words in blog articles, under the consideration of emotion prediction from different levels of language features: the global topical feature, the context sensitive word-emotion feature, and the local bag-of-word feature. We also leverage these probabilistic emotion-prediction results to jointly predict the text emotions.

Some text emotion analysis assumes two or three classes (positive, negative, and neural sometimes) [34] for a piece of text, such as a phrase or a sentence. However, the limitation of such emotion-polarity classification is obvious, since the positive and negative polarities are too simple to encode the true human feelings. Therefore, a lot of text emotion analysis has focused on the fine-grained emotion analysis, in which the text emotions are categorized in several categories, such as Ekmans six basic emotions [16][17]. These emotion categories are considered to be abstracted enough to represent most common human emotions in the real world. Most current research considers the text emotion analysis as a multi-class classification problem, by choosing one emotion label from the multiple emotion categories. In most cases, such methods are suitable for emotion prediction in the simple text pieces such as words and phrases. However, as we know, the complex emotions like Love & Expect or Anger & Surprise also frequently arise in longer pieces of text such as sentences and paragraphs. Moreover, some emotion words express multiple emotions, like the word (pleasant surprise) which conveys Joy and Surprise at the same time. Therefore, the sentences containing such words also express multiple emotions.

In fact, a recent study [23] indicates the complexity of the text emotions grows with the length of the text. In this study, we focus on the sentence emotion analysis as sentences are often considered to be the basic text pieces with self-contained semantic meaning, and we think of it as a multi-label classification problem in which case every reasonable emotion label in the sentences would be identified.

Human emotions have been considered as the private mental states. To understand emotions in the texts, we have to predict from different perspectives. For example, when talking about disasters such as earthquake, tsunami, or influenza, the emotions behind are most likely to be Sorrow, Hate, and Surprise. We consider these emotions as the topic emotions. Further, there are a great number of emotional words, which help directly expressing the sentence emotions. In sentence emotion analysis, such word emotions are also an informative aspect. Moreover, in some cases we can understand the sentence emotions even if there is no obvious emotional word. This suggests that except the emotional words, the existence or combination of other words could also indicate emotions in a sentence. Based on the above perspectives, in this study we intend to develop three levels of language features to analyze the sentence emotions: the global topics, the context-sensitive word emotions, and the local bag-of-words.

By following the previous discussion, it is reasonable to assume that in every long text piece (e.g. a sentence) there exist some emotion-related topics hidden behind the words. We define the emotion-related topics as the word clusters, each of which represents a specific word distribution while at the same time associated with a particular emotional label. By implicitly learning the word and topic distributions while explicitly learning the correlation between topics and emotions from a training corpus, we expect to train a

Labeled Latent Dirichlet Allocation (L-LDA) [44] model that help to predict the emotions in sentences more exactly.

Besides the topic emotion, in a sentence, there might be emotion-informative words which allow us to predict the writers emotions. In this study, we consider the emotions of these words as context-sensitive, because many factors such as negative modifications, contrast conjunctions, as well as the surrounding words could affect the emotion expression of an emotional word. A context-sensitive model for the word emotion prediction can be learned from the emotion corpus Ren-CECps [42] with the Conditional Random Fields (CRF) algorithm [25]. We follow the previous experiment [56] to predict the sentence emotions by accumulating the word emotions.

Besides the above emotion indicators, our observation of emotion recognition by human beings suggests that the ordinary words could also express emotions. In this case, for each particular emotion category and a set of words in a sentence, we consider a sigmoid function to calculate the probability of the existence of k_{th} emotion given the linear combination of word observations in this sentence. This leads to the Logistic Regression (LGR) [31] classification for each emotion category, with the output as the probability of containing k_{th} emotion in this sentence. And in order to predict the existence of multiple emotions, we could simply combine the results from these classifiers.

Finally, because we use probabilistic models to predict sentence emotions as discussed above, the probabilistic results from each model could interpret the existence of the sentence emotions from a particular perspective. In this case, we can take advantage of these probabilistic interpretations by the Bi-integration and Tri-integration of these results to better understand the sentence emotions. We also carefully select a set of probability thresholds for each emotion category based on a validation set for the integration models as well as the single models, which ensures a reasonable balance between the precisions and recalls of emotion prediction from all these models. Besides the probabilistic emotion predictions, we take the Support Vector Machines (SVM) [49] classifier into the baseline group. Six evaluation methods are employed to examine Multi-emotion annotation effects of results given by the basic three probabilistic machine learning methods as well as the Bi-integration and Tri-integration. The promising results prove that our Bi-integration and Tri-integration prediction are much better than the single machine learning methods based emotion annotation, which are also express that multi-emotion can be better captured from the text through the different perspectives.

For our proposed methods being much better understood, the system of our methods is demonstrated as Figure 1.1. In this study, we choose Ren-CECps as our data source, which is a well manually annotated Chinese emotion corpus. We divide the sentence extracted from the Ren-CECps into three parts, which are training set, validation set and test set represented by the red diamond box in the system chart. We get three different models including LGR model, L-LDA model and CRF model by using the corresponding machine learning methods respectively on the training set, and then utilize three models on the validation set to calculate the threshold values for confirming the existence of multi-emotion in sentence. There is one more threshold value should be calculated for CRF model, since we only can obtain the word emotions in the form of 9-dimensional emotion vector $\{p(No-emotion)$, $p(Joy)$, $p(Love)$, $p(Expectation)$, $p(Surprise)$, $p(Anxiety)$, $p(Sorrow)$, $p(Anger)$, $p(Hate)\}$, in which each element is a probability value that represents the possibility of the corresponding emotion existing in the sentence, given by CRF. We use the word-specified threshold to filter $No-emotion$ words out from the sentence for building the sentence emotion vector through the factor product which is accumulated by all the word emotion vectors without $No-emotion$ entry. After that, We integrate three sentence emotion vectors given by LGR, L-LDA and CRF respectively into a new emotion vector by multiply the corresponding entry in the vector. For judging the existence of sentence emotions, sentence-specified threshold is calculated from validation set to ensure this duty. The detailed illustration is in the Chapter 4.

1.1 Thesis Organization

The rest of this paper is organized as follows:

Chapter 2 describes Background of related researches about affective information computing. The general algorithms like machine learning methods are also introduced in this chapter.

Chapter 3 introduces the related works on affective information computing in recent years, which are emotion classification, emotion recognition, emotion extraction and so on

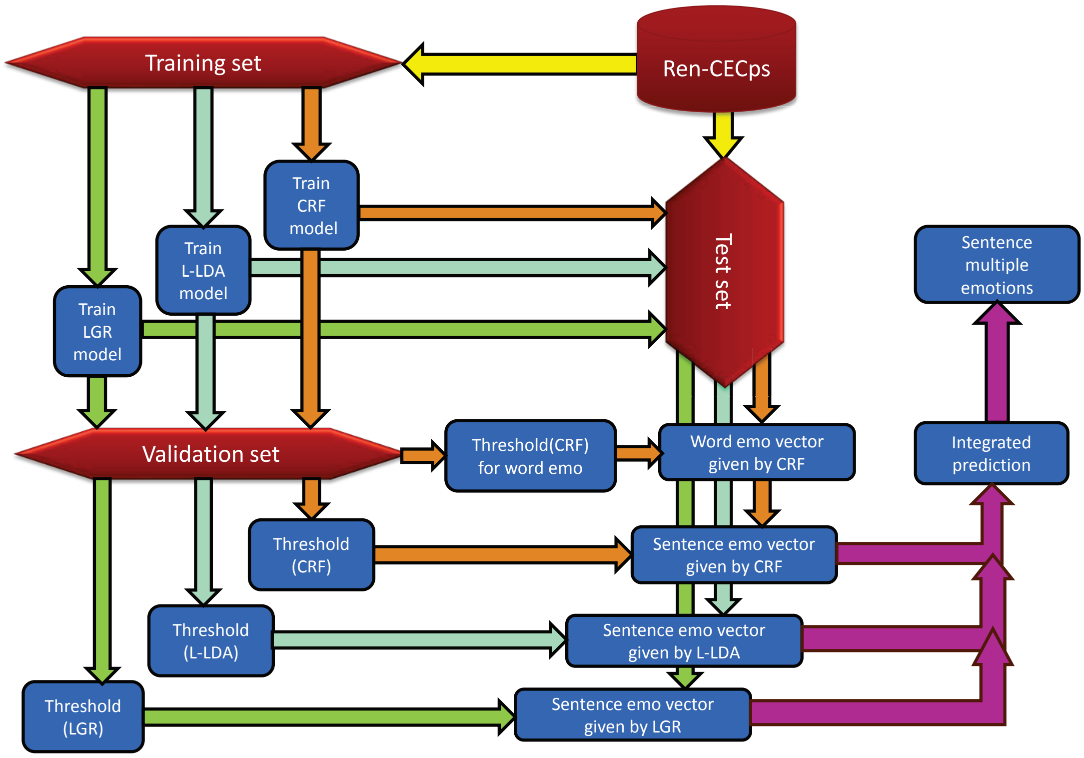

Figure 1.1: A joint sentence multi-emotion analysis system constructed by the different perspectives.

in different textual levels including word level, sentence level, paragraph level and document level. We make a comparison between these researches, and capture their advantages. Final, we show merit of our methods.

In Chapter 4, we introduce the data source named Ren-CECps which is a well annotated Chinese emotion corpus including 1487 documents, with 11,255 paragraphs, 35,096 sentences, and 878,164 words extracted from the Chinese portal website. According to the different characteristics of those machine learning methods, we adopt different emotion components as feature sets for training model respectively. Three machine learning methods are utilized for multi-emotions analysis including emotion annotation, emotion topics extraction and sentence emotion analysis.

In Chapter 5, we demonstrate experiment data source used in each experiment with different machine learning method and the procedures of sentence multi-emotion analysis in detail. We make important assumptions under which the emotions of documents can be reasonably computed through the emotional and statistical information of the words contained in the documents. Sentence emotion vector, in which each element is probability value indicating the possibility of corresponding emotion expressed in the sentence, are able to be calculated by the combination of the result given by three different machine learning method separately. For confirmation of existence for each emotion categories in the sentence emotion vector, we select a group of thresholds for each emotion categories which are computed from the training set. We make a comparison between the probability value and corresponding threshold value, and make sure the existence of emotion when the predicted probability value is bigger than the threshold value.

In Chapter 6, experiment results will be evaluated by a series of evaluation criteria including Precision, Recall, F-score measure, Accuracy, Hamming Loss, Micro F-score and Macro F-score. We analyse the result of sentence multi-emotion in detail.

Finally in Chapter 7, we conclude the merits and demerits of our find-grained sentence multi-emotion analysis by different perspectives, and list some works to be studied in the future, which could help to improve the system performance.

Chapter 2

Background

In this chapter, firstly, we introduce the definition of textual emotion and the form of emotion expression through the text. Then the specific emotion categories are illustrated in detail by different criterion. We give some examples to demonstrate emotions existing in different textual levels including word, sentence, paragraph and document level. Concretely, the common emotion components are also introduced in different textual levels. Since at most of the time, multiple emotion appear in the texts simultaneously, we propose that multi-emotions should be paid much more attention more than the single emotion, which will better describe the emotion states from the textual information. The chief emotion is also decided by the comparison of the corresponding emotion with probability value from the multiple emotions in the text. Next, we introduce the background of emotion analysis, including emotion classification and emotion prediction, which has become a popular subject in natural language processing studies. Finally, we introduce the common emotion analysis methods containing two branches which are unsupervised machine learning method and supervised machine learning method. The difference between them and the merit as well as the demerit are given with some examples.

2.1 Emotion Polarity and Category

People like classify polarity into two categories including positive and negative (or three categories sometimes with extra neutral). This would be great helpful in some fields in which polarity classification is enough to judge attitude from the given medium information like movie reviews [40], product comments, opinion of the public events and so on. For further understanding emotion states of writers, more fine-grained emotion classification is under consideration,like six emotion categories *Happy*, *Sad*, *Anger*, *Fear*, *Disgust*, and *Surprise*. proposed [18], or more emotion categories defined in [27] as *Awesome*, *Heartwarming*, *Surprising*, *Sad*, *Useful*, *Happy*, *Boring*, and *Anger*, ten emotion categories including *Happiness*, *Pleasantness*, *Relief*, *Fear*, *Sadness*, *Disappointment*, *Unpleasantness*, *Loneliness*, *Anxiety*, and *Anger* in [51], and even more emotion categories (i.e., 132 different emotions) in [32][33].

Ren-CECps is a well manually annotated Chinese emotion corpus, containing 1,487 Chinese blog articles with totally 35,096 sentences extracted from the portal website,as our data source to perform the experiment on emotion analysis. The eight basic emotion categories consisting of *Joy*, *Love*, *Expectation*, *Surprise*, *Anxiety*, *Sorrow*, *Anger* and *Hate* are assigned to the different length of textual pieces, such like word, sentence, paragraph and document. All of them are annotated with single or multiple emotions. There are three polarities classifications, which are positive, negative and neutral, are employed in the corpus. Exception of word, sentence, paragraph and document are classified into one of these classifications. We select Ren-CECps [43] as our experimental data source for multi-emotion analysis. For demonstrating polarity and emotion category more exactly, we give some example sentences as follows;

Polarity

Sentence 1: He felt warm standing in the sunshine(Polarity:positive).

In the sentence 1, there are two positive words, that are “warm” and “sunshine”. The word “warm” is used to described some heart-warming scene or comfortable surrounding, while word “sunshine” is also a positive word to be used frequently. Therefore, this sentence is assigned with positive label by annotator in the corpus.

Sentence 2: Liu Mei seemed to be very depressed when she knew that she failed in the exam(Polarity:negative).

In the sentence 2, both “failed” and “depressed” are negative words. The word “fail” means somebody doing something is unsuccessful or something can not going to be well. And the word “depress” shows negative mental state of person. So, this sentence is assigned with negative label by annotator in the corpus.

Sentence 3: Tomorrow is a national holiday(Polarity:neutral).

In the sentence 3, there is no word that can convey any emotions, so this sentence is assigned with neutral label.

Emotion

Sentence 1: Lee felt *relax*(Joy) only when he was home.

In the sentence 1, there is only a emotional word "relax" conveying the emotion of *Joy*. Thus, the sentence conveys one emotion of *Joy* in accordance with the word emotion.

Sentence 2: He can't help *laughing*(Love|Joy) when he saw the notification of the acceptance(Joy).

In the sentence above, the word "laughing" is annotated with emotions of *Love* and *Joy*, while word "acceptance"is annotated with one emotion of *Joy*. In general, sentence emotions coincide with word emotions, because words are the basic components of the sentence. So, in this case, the sentence can express two emotions of *Love* and *Joy*.

2.2 Emotion in Different Literary Forms and Textual Levels

Recent works on emotion analysis are carried out on different literary forms like movie review, product comment and news [51] [5]. It is reasonable why researchers choose movie review as experimental data object. There are large scale on-line such movie reviews collection that can be easily extracted. Moreover, these reviews are experimentally convenient because they can be used directly without manual label. The viewers summarize their overall sentiment with a rating indicator, hence, they are well used for experiment on polarity classification. [40] selected reviews concerned with stars or some numerical value from Internet Movie Database (IMDb) archive of the *rec.arts.movies.reviews* newsgroup, and concentrated on discriminating between positive and negative.

The domain of product comment is also one of the best object chosen for emotion analysis, since they are convenient access and exist in a variety of forms on the web: customers on Amazon give the feedback to show their satisfactory with commodity or complaint about delivery, which are constituted of words rich in emotions [19] [21]; and collected professional comment on web that specialize in a certain kind of product(like dazhong dianping wang at url:www.dianping.com/) [15]; and some users also write some comments on products in their social blogs (like facebook) [32] or mic-blogs (like twitter) [37] in which real voice and thought are written. [15] implemented some experiments on semantic classification of the product comments.

Accompanied with development of human interaction in web social net, social computing [35] come to a hot issue. Research on the news of social net is a good way to explore human behavior through their expressions like written news. Because the written news is the most visible and prominent clues, human express their emotions directly or indirectly through their interaction news. [5] collected 1000 news sentence from the social net to perform different experiment to compare the multiple emotion classification between machine learning and human.

There are four different length of textual level including *word*, *sentence*, *paragraph* and *document*. No matter which literary form of texts are utilized for emotion analysis, there are at least one textual level among four different length should be under consideration. For word level, experiment on single emotion recognition are often performed, while experiment about multiple emotions recognition is difficult to execute. Because, it's not easy for human to understand multiple emotions without context, much less to recognize the emotion automatically by machine.

However, multiple emotion are relatively simple to be analysed in the other three textual levels. In most of the case, words are viewed as the basic component of sentence, paragraph and document. When there are several emotional words existing is the sentence or other different length of textual level, sentence expresses multiple emotions in accordance with word emotions. Nevertheless, multiple emotions of long piece of text are not completely same to joint the all emotions of words, particularly emotional words express emotions in opposite polarity.

Consequently, we propose the integrated prediction from three different perspectives, in which three different machine learning methods are performed to build sentence emotion vectors separately, finally we create a new sentence emotion vector by the combination of three emotion vectors. In this study, we explore sentence emotions from bag-of-words by Logistic Regression, from context based emotional words by Conditional Random Field, from emotion-related topics by Labeled Latent Dirichlet Allocation.

2.3 Multi-emotions and Chief Emotion

In recent years, most of the works on emotion analysis are concentrated on either single emotion or multiple emotions [4]. Especially when we use the probabilistic models to solve such issues that we try to find the weight for every emotion category. Every emotion is assigned with a probability value which is a possibility that emotion arise in the sentence. We are able to estimate the contribution that each emotion makes. When we just focus on chief sentence emotion, we regard the emotion with the highest probability value as the chief emotion of sentence. For the multi-emotion of sentence, we can not simply decide the number of emotions which sentence may convey. If the multi-emotions arising in sentence are directly chosen by comparing the corresponding probability value of each emotion, it would cause the emotion missing in such situation when the emotion words convey emotion in different polarity with sentence, these emotion words would be assigned with the relative low probability value.

Therefore, we propose figure out nine threshold values for no-emotion words and eight emotion categories including *Joy*, *Love*, *Expectation*, *Surprise*, *Anxiety*, *Sorrow*, *Anger*, and *Hate*. Multi-emotions of sentence are decided after the comparison between the calculated probability value of each emotion and the corresponding threshold value. When the emotion gets a higher calculated probability value than the corresponding threshold, we think this emotion can be conveyed in the sentence. In this study, we concentrate on the chief emotion and multi-emotion analysis in the sentence.

Under the subject of word emotion classification, most researchers are concentrating on the different methods of exploring a variety of emotion lexicons. Such studies include [20], in which machine learning techniques are also utilized. However, at most of the time, emotion lexicons could not meet the demand from the real-world texts, since the emotion lexicons are static while in the real-world the word emotions could always change in different contexts. Therefore, such studies based on emotion lexicon often suffer from insufficient or misleading emotion features.

On the other hand, we notice that most previous researches on sentence emotion analysis have employed the approaches such as keyword matching, affective lexicon, a knowledge base method and so on. However, these keyword-matching based methods are generally too simple to count the emotion information contained in the sentence, especially when the emotions of matched words differ from the whole emotion type of the sentence. In other words, the sentence emotion analysis, which depends on single the knowledge base, would not be suitable for the complex emotion situations in the real world. Therefore, we explore an automatic sentence emotion analysis system, which takes the dynamically annotated emotional keywords as features and explores the statistics parameters extracted from large sets of corpus to help correctly accumulate the sentence emotion values from emotions of words.

2.4 Emotion Analysis Method

Under the subject of emotion analysis, generally, rule-based rationalist methods and statistic-based empirical methods are implemented to address such problem. Both of them are used to construct the language models with high frequency for emotion analysis. At the beginning period of affective information computing, rule-based rationalist methods are implemented as fundamental methods for emotion analysis, since it's relatively easy to summarize the simple rules. With the development of machine learning methods, the statistic-based empirical methods achieve the dominance gradually because of their merits that the language models are able to adjust parameters at anytime with respect to the practical training data.

2.4.1 Rule-based rationalist methods

There are some merits of rule-based rationalist methods, so as to many researchers employed them at their previous works. We conclude three advantages as follows,

Mer1t 1 The rule-based methods is well known as exact description for linguistic rule. These rules are practical for ability of relations description and generation.

Mer2t 2 The rule-based rationalist methods are easy to understand, clear expression and description. Many linguistic phenomenon can be expressed directly by the constitution and components of language model.

Mer3t 3 Moreover, the rule-base rationalist methods are undirected in essence. The language models, which are explored by rule-based rationalist methods, can be used for analysis and generation.

Thus, the language models are bi-direction in many case. Because of these merits, the rule-based rationalist methods are extensive used in the field of language knowledge, especially highly implemented in syntax and semantic branches.

However, there are also several demerits that we should pay our attention.

Demer1t 1 These language models have bad robustness, and crash easily. When we use experimental data which are in the different language rules, some errors appear and cause a lot of bad influence on language model, so as to make the language model out of work.

Demer2t 2 In addition, before rule summarization, we need to hire some linguistic experts to work together and conduct the knowledge-intensive research that is work intensity.

Demer3t 3 Another point, which affect development of language model, is that the rationalist language models are not able to be developed automatically by the machine learning. For this reason, the language models are not capable of automatic generalization by using computer. The system explored based on rationalist methods of natural language processing have many constraint to be updated, a little change would cause the ripple effect, and it's difficult to get rid of the bad influence.

Therefore, the rule-based rationalist methods make a worse effect than statistic-based methods in many practical computing. On account of the language models, explored by statistic-based empirical methods, are able to be optimized at anytime in accordance with practical training situation, however, those rationalist methods couldn't get any adjustment due to their essence even they should be in many practical cases.

2.4.2 Statistic-based empirical methods

Compared with the rule-based rationalist methods, the statistic-based empirical methods are more expert in address the natural language processing, the reasons are illustrated as follows,

Mer1t 1 we can build the effective statistic language models from training data semi-automatically or automatically.

Mer2t 2 The performance of statistic-based empirical methods depend on the scale of training data in a large extent. The more training data we have, the better performance the system enhance. Thus, the statistic language models are simply improved by expanding the scale of training data (corpus).

Mer3t 3 We can integrate the statistic-based empirical methods with the rule-based rationalist methods to solve the constraint problem in dealing with language models for the purpose of optimizing the language processing system continuously.

Mer4t 4 The statistic-based methods are also probable to be implemented in certain situation such as s slight distinction and fuzzy conception like words $little$, $muchmore$, which are only handled by fuzzy logic in traditional linguistics.

Despite of so many merits the statistic-based empirical methods have, we can't ignore limitations.

Demer1t 1 When we implement some experiments by using language models explored from statistic-based empirical methods. The running time is proportional to the number of classification in language models. No matter in training data or testing data, with the increase in number of classification categories, system efficiency significantly decreases.

Demer2t 2 Moreover, under the technology of corpus construction, gathering domain-specific data for construct the language models based on statistic methods is time consuming and laborious, even hardly avoid the mistakes.

Demer3t 3 The efficiency of language models are significant positive correlation to scale, representativeness, correctness, processing depth of corpus. That is to say, quality of training data in corpus makes a large influence on effect of language models explored by statistic-based empirical methods. But if only concentrate on the scale of corpus, it would lead to another problem which is data sparsity, for which we can employ the smoothing to handle.

In this study, we implement sentence multi-emotion analysis by using three machine learning methods including Labeled Latent Dirichlet Allocation, Conditional Random Field, and Logistic Regression, which are statistic-based empirical methods. Despite that these machine learning methods have so many merits compared to rule-based rationalist methods, we have to face the disadvantages. In order to overcome these disadvantages, we propose some problem-focus solutions.

The scale of data source

To deal with this problem, we extract 31,058 sentences from Chinese emotion corpus(Ren-CECps) to implement experiments. In the recent related works, many researchers only used 1,000 or 2,000 sentence to verify their thought since collecting domain-specific data is really a time-consuming and laborious task. These researchers could conduct their experiment successfully, therefore, we have confidence that our methods would achieve the satisfied results because we have a great number of experimental data source compared to other related works.

The quality of data source

Because in this study, we implement experiment by machine learning methods which are statistic-based empirical methods, we must carefully select sentences as our data source. As our data source is extracted from Ren-CECps, and it is a manually annotated Chinese emotion corpus, we think this corpus must be a good choice.

According to the corpus introduction, at the first annotation period, three annotator are asked to annotate the documents separately, and calculate their annotation agreement in three level including document level, paragraph level and sentence level, the detailed value as shown in table 2.1. Seeing the agreement value in the table, three agreements are approximately equal to each other and the average agreement achieve 0.764, which is

	Agreement
document level	0.831
paragraph level	0.705
sentence level	0.756
Average	0.764

Table 2.1: The agreement on different textual level.

regarded as relatively high in the corpus evaluation criterion. At the second annotation, the other five annotators are asked to check the errors and re-annotate the documents which they thought should be fixed.

Thus, we consider the corpus is in good quality. Our experimental data are extracted from this corpus, we are sure that our experiments will get the outstanding results.

The processing depth of data source

The third difficult point we care is the processing depth of data source, since this is also key point concerned with effect of language models. The Chinese emotion corpus (Ren-CECps) is manually annotated in multiple layers including document layer, paragraph layer, sentence layer and word layer.

For document layer, multiple emotion tags of nine emotion tags (eight emotion categories and No-emotion tag) are assigned to documents. The corresponding emotion intensity which is a float value with two decimal places ranging from 0.0 to 1.0. The key words are given which can summarize the documents.

For paragraph layer, annotators have the same annotation scheme as document layer, and summarize the document with some key words.

For sentence layer, emotion annotation scheme is same as the document layer. In the most of cases, the sentence have multiple emotion tags, single emotion occasionally, seeing in Figure 2.1. The annotator assign additional polarity tags (positive tag, negative tag and neutral tag) to the sentence. If the sentences are judged as neutral sentences, there are no emotion annotation steps for them. Modification relations annotation are also under consideration. The rhetorical devices and scopes are tagged manually. There are three modification relations are focused, that are Conjunction modification, Degree modification and Negative modification.

For word layer, nine emotion tags are assigned to words with corresponding emotion

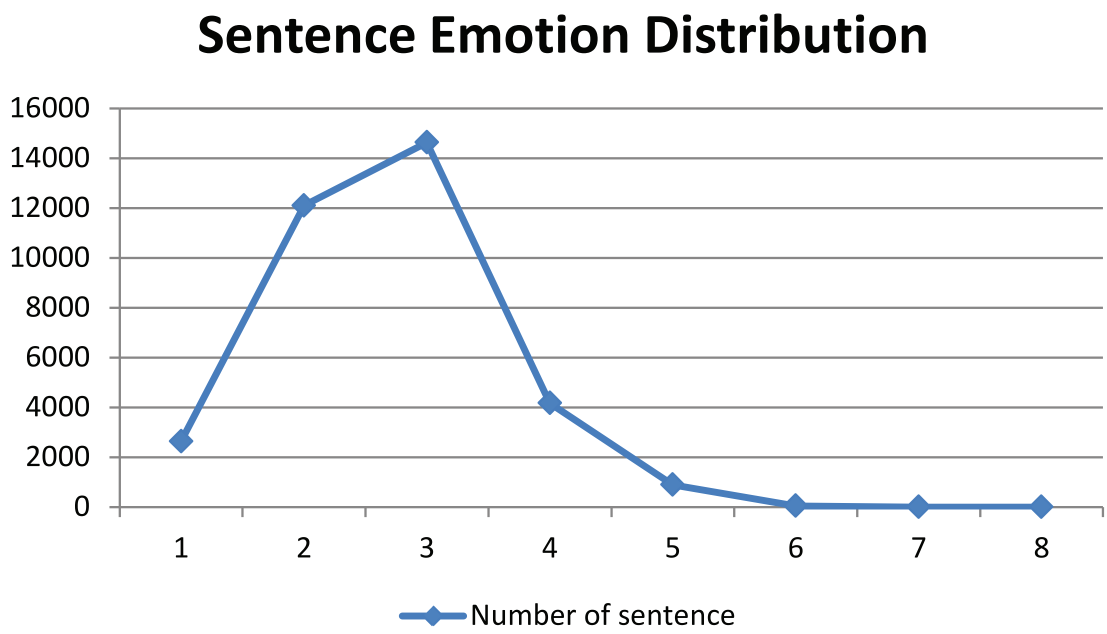

Figure 2.1: The emotion distribution of sentences.

intensities. Several words express multiple emotions less than three. The emotion distributions of words is shown as Figure 2.2.

Words as emotion components take the important role in the experiment on emotion analysis. We can capture more or less emotion information directly from them, since they are also the basic components of whole text. In this study, we view words as the components, as they are the parts of sentence. At most of the time, words convey the same polarity as sentence, even same emotions as sentence. In addition, word-related components also can be used as feature to solve the issue on affective information computing. So, in the experiments, we propose three kinds of feature which word feature, part-of-speech feature, and modification feature. For consideration above, we implement Conditional Random Field to recognize the word emotions in sentence, and predict the sentence multi-emotions by the factor product of word emotions of words within sentence.

Using words as basic components, but we try to directly predict sentence multi-emotion in a another perspective. We concentrate the contribution made by words within sentence, thus only local bag-of-words are employed for sentence multi-emotion prediction.

In the recent works, researchers conducted some experiments based on topics [45], and proved that topics in document (sentence) concerned with text analysis [45]. For

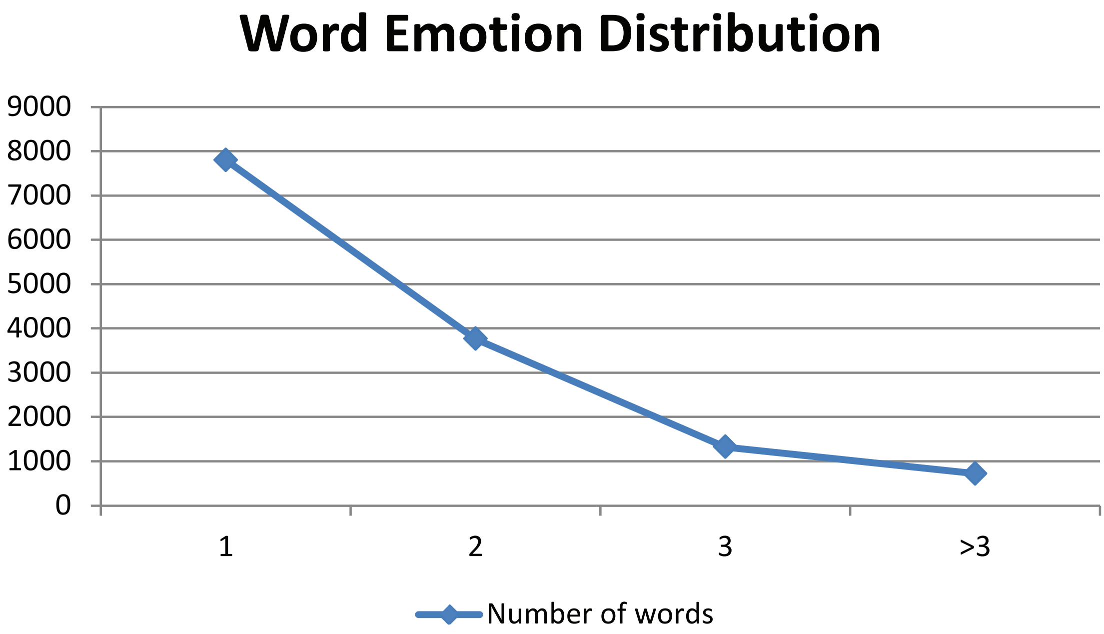

Figure 2.2: The emotion distribution of words.

consideration of sentence emotion expression surrounding emotion topics basic on the linguistics, we plan to explore the emotion-related topics in the sentence at first, and then predict sentence multi-emotion by using these emotion-related topics. Latent Dirichlet Allocation (LDA) is a well known topic extraction implementation, which is a unsupervised machine learning methods. It counts the appearance of words, and calculate the relations between words and topics in the form of corresponding probabilities. Because we propose to predict sentence multi-emotion in this study, and know that there are not so many words in sentence that would lead to the bad influence on results of topic extraction.

Therefore, we make a decision to use another topic language model name Labeled Latent Dirichlet Allocation. Labeled Latent Dirichlet Allocation (L-LDA) is the development based on LDA, which is supervised machine learning method. In L-LDA, the label and topic are fixed together one to one, if we take emotion tags as label in L-LDA, we sure that we can get emotion-related topics. Finally, using emotion-related topics to predict sentence multi-emotion is our third perspective in this study.

Chapter 3

Related Works

In the fields of affective information computing [46], emotion classification and emotion prediction are drawing more and more attention. No matter emotion classification or emotion prediction, coarse and fine-grained classification are taken into account. Recent researches focused on emotion analysis mainly in two branches including coarse classification [8] [6] [29] [14] and fine-grained classification [3]. The former worked on two or three classifications (positive, negative and neutral sometimes) based on different literary forms like blogs [30] [57] [23], movie [40] or comment reviews [48] [36], twitters [50][1], opinion mining [28] and so on. The later was concentrated on subtler emotion categories which was much closer to the writer inter emotion states. However, almost those related works were executed on single emotion analysis which can't convey the real emotion completely since writers expressed multiple emotions at most of the time. Kang et al. [24] implemented a Hierarchical Bayesian Network to analyze the complex emotions and topic. Das et al. [13] recognized multi-emotion from merge of words emotion and word distribution. In this study, we propose to analyze the multi-emotion in three different perspectives, which can better interpret multi-emotion than their works. The promising result proves that our integrated prediction gets closer to the writers real emotion states.

3.1 Coarse-grained Emotion Classification

One of the early study of coarse-grained emotion classification was made by Melville et al. [30], by combining lexical knowledge with text classification algorithms. We would like to discuss Melville's study because its probability pooling method for emotion prediction is most related to our work. In Melville's study, only the emotion polarity i.e. positive emotion and negative emotion were considered with respect to the blog articles. The emotion polarity was considered as following the multinomial distribution

$$P(e|S) = \frac{P(e)\prod_i P(w_i|e)}{P(S)}, \tag{3.1}$$

where e represented the emotion polarity and S was the piece of text containing words w_i. Since the probability of text piece $P(S)$ was constant with respect to emotion polarity $P(e|S)$, the emotion polarity was chosen by maxizing the product of prior probability $P(e)$ and the likelihood $P(w_i|e)$ as

$$e = argmax_e P(e) \prod_i P(w_i|e). \tag{3.2}$$

In Eq. 3.2, the emotion polarity was inferred by combining the information from two sources (or experts) through pooling distributions. The first expert was learned from the labeled training data, which contained the labeled sentiment examples based on the Lotus blogs of 34 and 111 examples of the positive and negative posts as well as the political candidate blogs of 49 positive posts and 58 negative posts. The second expert was based on a generative model which generated the emotion explanatory words based on a lexicon from the emotion polarities in the text pieces.

More concretely, from the labeled training data, the frequencies of words and emotion polarities were observed to calculate the conditional probability $P(w_i|e)$ as well as the prior emotion polarity probability $P(e)$. The construction of the generative model was more subtle. In the generative model, the conditional probability of each word w_i given the emotion polarity e i.e. $P(w_i|e)$ was considered following the multinomial distribution, by making four major assumptions. Given a vocabulary consisting of p positive words, n negative words, and u polarity-unknown words, the conditional probability $p(w_i|e)$ can be specified with the followings

$$
\begin{aligned}
P(w_+|+) = P(w_-|-) &= \frac{1}{p+n}, \\
P(w_+|-) = P(w_-|+) &= \frac{1}{p+n} \times \frac{1}{r}, \\
P(w_u|+) &= \frac{n(1-1/r)}{(p+n)u}, \\
P(w_u|-) &= \frac{p(1-1/r)}{(p+n)u},
\end{aligned}
\tag{3.3}
$$

where $+$ and $-$ indicated the sign of emotion polarity for both words and the text pieces, and r was referred as the polarity level which controls the probability ratio of words and text pieces associated with the same polarity to those associated with the opposite polarities:

$$
r = \frac{P(w_+|+)}{P(w_-|+)} = \frac{P(w_-|-)}{P(w_+|-)}. \tag{3.4}
$$

To derive the above probabilities in Eq. 3.3, the author made four assumptions in the generative model as follows. First, given the polarity e of the text pieces, there is no difference in observing different words with the same polarity. This assumption simplifies the multinomial distribution $P(w_i|e)$ by decreasing the category number from $p+n+u$ in the vocabulary to only three i.e. "+", "-", and "u". Second, the probabilities with respect to "+" and "-" should be symmetric, i.e. $P(w_+|+) = P(w_-|-)$ and $P(w_-|+) = P(w_+|-)$, which further simplifies the probability calculation. Third, the conditional probabilities for words associate with the same emotion polarity as text pieces should be greater than those with the opposite polarities, which is controlled by the probability ratio r as $0 < 1/r \leq 1$. Finally, the conditional probability $P(w_i|e)$ should be normalized to 1 as $\sum_i^{p+n+u} P(w_i|e) = 1$.

With the first expert learned from the labeled blog data, the prior emotion polarity probability $P(e)$ and conditional probability $P(w_i|e)$ with respect to each word in the vocabulary could be learned. With the second expert based on a generative model, the conditional probability $P(w_i|e)$ with respect to the words in the emotion polarity lexicon could be derived based on four major assumptions. To combine the information from two experts, the author employed the pooling methods, especially for the conditional probability $P(w_i|e)$. Specifically, the linear opinion pool aggregates the conditional probabilities through a linear combination

$$P(w_i|e) = \alpha_1 P_1(w_i|e) + \alpha_2 P_2(w_i|e), \tag{3.5}$$

where subscripts i.e. 1, 2 indicate the expert index, and α_1 and α_2 are the weighting parameters to ensure the probability sums to one, and the logarithm opinion pool aggregates the logarithm of the conditional probabilities

$$\log P(w_i|e) = \log Z + \alpha_1 \log P_1(w_i|e) + \alpha_2 \log P_2(w_i|e), \tag{3.6}$$

where α_1 and α_2 are the weighting parameters and Z is the normalization factor.

The author reported emotion polarity prediction accuracies on the blog articles with respect to Lotus, Movies, and Politics, respectively with 10-fold cross-validation, and found the linear opinion pooling rendered the best average accuracies for all blog domains. The author also suggested that the emotion lexicon could be very useful for emotion polarity prediction, as the prior knowledge. By combining the probabilities from the emotion lexicon as the background knowledge and the probabilities from the labeled examples with a supervised classification algorithm, the author could substantially improve the prediction of emotion polarities for blog articles for different domains.

However, there are some key issues needed further explored in Melville's work. First, the conditional probability $P_2(w_i|e)$ was constructed from on a lexicon, with very strong assumptions on the distribution of the emotional words and the pseudo text pieces that contained these words. These assumptions could ignore the important factors, such as the difference in different words in expressing the emotion polarities (against the first assumption), the unbalanced probabilities for different emotion polarities (against the second assumption), the variability of emotional words in the text pieces (against the third assumption), and most importantly the context dependencies in the word emotion distribution, which is not considered in this work. Moreover, the positive and negative classification of text emotions could provide information for analyzing the online business or the public opinions towards politics, but these information was not subtle enough to further analyze the mental states of the online customers or the public's emotional information.

3.2 Fine-grained Emotion Classification

Unlike the coarse-grained emotion classification, which predicts the emotion polarities of positive and negative in text pieces, the fine-grained emotion classification focus on very specific emotion categories, like Joy, Love, Expectation, Surprise, Anxiety, Sorrow, Anger, and Hate [42] for the text pieces.

The recent studies of the fine-grained emotion classification can be further specified according to their assumptions of the text emotions into two directions: the multi class single label emotion classification, represented by Das et al. [11][12][13], and the multi class multi label emotion classification, represented by Kang et al. [22][24][47]. Both directions are closely related to our emotion classification work.

In the work of Das et al. [11], specific emotion categories including Happiness, Sadness, Anger, Fear, Surprise, and Disgust based on the Ekman's basic emotions [16] were employed as the emotion classes for both words and sentences. The author assumed each text piece (for both the words and the sentences) with only one emotion label from the above six emotion categories. Thus, the emotion classification problem could be converted to a multi class text classification problem.

The word emotions were firstly learned through a CRF (Conditional Random Field) model, based on a blog corpus of 1300 labeled sentences. Emotion labels with the corresponding probability "senses" are generated by the CRF model, based on 10 active features including the part-of-speech tags in words, the first observed sentence for each topic, the existence of words in the Senti-WordNet, the reduplication for Bengali words, the identification of the question words, the identification of foreign words, some special punctuation symbols, the identification of quoted sentences, the negative words, and the emoticons. These features in the CRF model provides information in emotion prediction for both the current words and the context words. The sentence emotion label was then predicted by accumulating the "senses" of word emotions predicted by the CRF model, with either the Senti-WordNet based weighting or the corpus statistics based weighting. The author reported the promising accuracies of the word emotion and sentence emotion classifications, but the experiment setting was not quite clear.

In the recent work of Das et al. [12][13], the author also employed multiple emotion classifiers including CRF (Conditional Random Field), SVM (Support Vector Machines) [7], and FC (Fuzzy Classifier) to vote to identify the emotions in the text pieces. By performing the information gain based pruning (IGBP) on the development set, Das et al. could extract features that are important for the emotion prediction, emotion topic analysis, and emotion holder analysis. Experiment results suggested that combining the results from different emotion classifiers could also improve the multi class single label emotion classification [4].

The other direction of the fine-grained emotion classification, carried out by Kang et al. [22][24][47] was focused on the recognition of multiple emotion labels from the text pieces. Emotions were considered as the underlying explanatory factor for the word distributions in a blog corpus, and could explain the generation of words in a document together with the latent topic factors. The author proposed Bayesian models to analyze emotion distributions with respect to words and documents, by exploring as many emotion labels as possible.

In the emotion and topic analysis for words [22][24], Kang et al. constructed the generative model, by allowing the emotions and topics as the factors which could decide the generation of words in a document. The generation was encoded in a "V" structure, with the emotion variable and topic variable as the two parents of the word variable. In the Bayesian model, the assignments of the parents could depend on each other if the value of the child variable is visible, which means the emotions and topics could explain for each other and together explain the observation of words in a document.

Each emotion label was represented by a binary-valued variable e. And for each word w there existed K emotion variables e_k, where K was the number of emotion categories and k was the index of the emotion label. Each emotion variable e_k was assumed following the Bernoulli distribution $e_k \sim Ber(\theta_k)$, with θ_k as the proportion parameter. To model the variance of emotion distribution among documents, the model further assumed the proportion parameter θ_k as a random variable, and drew the sample value of θ_k through a Beta distribution $\theta_k \sim Beta(\beta_k)$, in which β_k was the concentration parameter counting the occurrence of the emotion label k from a labeled data set.

With careful derivation of the posterior probabilities of the emotion variables e_k and the topic variable, this model could iteratively generate samples of emotion labels and topics for words in many documents in an unlabeled corpus. The generated emotion and topic samples were then used in the Gibbs sampling inference algorithm to fit the model. After a few iterations, the Gibbs sampling inference could converge to the stable state, and the word emotions e_k as well as word topics z could predicted by maximizing the corresponding posterior probabilities $p(e_k|w, e_{-k}, z)$ and $p(z|w, e)$, respectively.

In the multi class multi label document emotion analysis [47], the authors built hierarchical Bayesian model to generate the labels for document emotions as well as the topics for documents and words. Particularly, document emotions and topics were made two parents of the word variables in this model, which also allowed the document emotions and topics to affect the selection of each other, and together to affect the word selection in the documents.

The experiment results based on a Chinese emotion corpus, for both word emotion prediction and document emotion prediction, suggested that multi class multi label emotion prediction could be significantly improved by incorporating the topics as the assistant explanatory factor in a Bayesian model, and the incorporation of Beta distribution for the proportional variables θ_k could reduce the side effect of unbalanced emotion observations in the labeled training data.

To conclude, the fine-grained emotion prediction could recognize more subtle emotion categories in the texts, and even predicted multiple emotion labels for the text pieces. The study of Das et al. also suggested that combination of multiple experts might improve the multi class emotion prediction. However, the combination of multiple experts by voting in Das's study was not enough to explore the significance of emotion predictions from different experts. To consider the information from multiple experts, the probability pooling method could be more reasonable than majority voting. And since we are pooling probabilities for multiple emotion classes and for multiple emotion labels, the existing pooling techniques are not suitable any more. Considering the importance of the topic factor in emotion prediction, we decide to also incorporate the topic information in a generative model but with a different probabilistic assumption. Similar to the emotion topic analysis in [22][24][47], we assume topics and emotions are explanatory factors of the word distributions. However, because the variation topics could be too much to cause over-fitting in some extent in the model, we would like to make further restrictions on the topic variation by assuming the direct selection of topics by the emotions, which means the emotion labels would have a one to one relation with the topics in the generation process. These probabilistic assumptions correspond to the supervised topic model, i.e. the Labeled-LDA model.

Chapter 4

Experiment Methods

In this chapter, we propose three distinct probabilistic models for the emotion analysis in sentences, based on the different perspectives of emotion prediction by human beings. We divide the Chinese emotion corpus Ren-CECps into three sets including Training set, Validation set and Test set. Then, three different models are calculated by L-LDA, LGR and CRF respectively on training set. We use these three models on validation set to compute the different thresholds for different algorithms. Sentence emotion vector can be directly predicted by executing LGR and L-LDA model on Test set, while indirectly predicted by performing the accumulation from word emotion vector using CRF model on Test set. Finally, we integrate three sentence emotion vectors to predict sentence multi-emotions, which is demonstrated in Figure 4.2.

4.1 Data Source

Ren-CECps consists of 1,487 Chinese blog articles with totally 35,096 sentences, which are collect from the Chinese portal webs such like Sina-blog, sohu-blog, 163-blog, qzone and so on. These articles are segmented, and annotated manually in four different textual levels, that are document level, paragraph level, sentence level and word level with eight basic emotion tags: *Expect*, *Joy*, *Love*, *Surprise*, *Anxiety*, *Sorrow*, *Anger* and *Hate*.

In most of time, multiple emotion tags are assigned to them, since some emotions overlap and are expressed at the same time. Three polarities including *positive*, *negative* and *neutral* are also assigned to the document, paragraph and sentence level. The basic

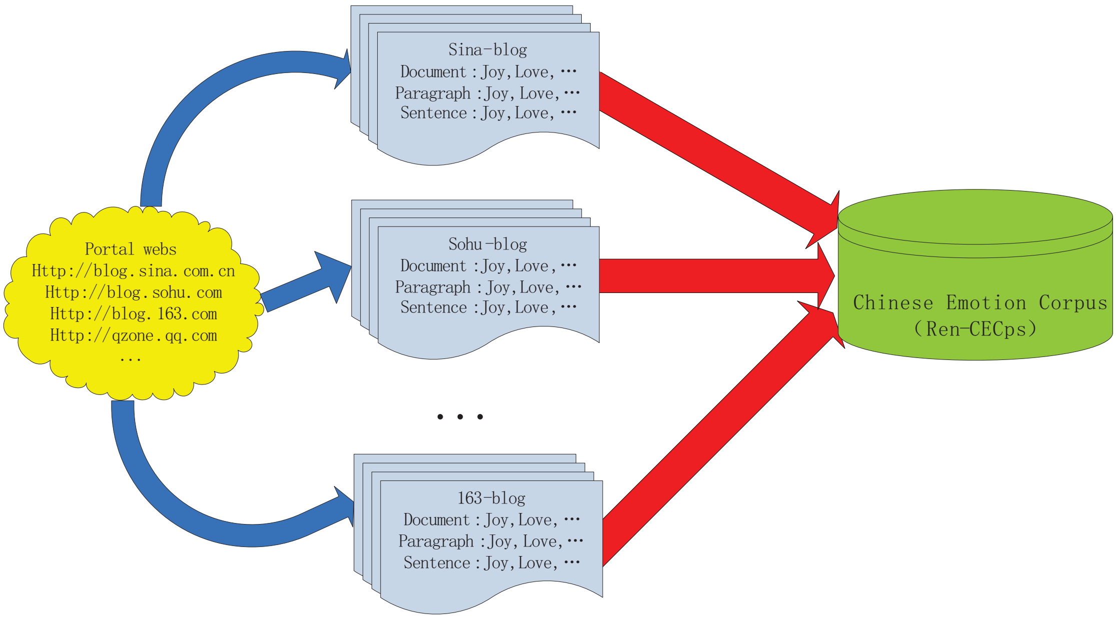

Figure 4.1: Ren-CECps: Chinese emotion corpus.

emotion elements containing emotional words and emotional phrase in the Ren-CECps. Both of them are given emotion tags with proper emotion intensity. Emotion intensities (from 0.1 to 1.0) of each emotion category are given. Besides emotional words, the modifiers including degree words, negative words and conjunctions are also annotated in this corpus, as demonstrated in the figure 4.1.

4.1.1 Emotional words

Some words can always express the emotions directly, we can easy to get to know the emotion states of writer by reading the word-itself. While some few indicate emotions indirectly depending on the context [10], that is to say, we can't capture the emotion states of writer easily, but we still can find useful information which would convey the latent emotion through adjust sentences. Here, we give some examples about the words that indicate emotions directly and indirectly.

For examples:

Indicate emotion directly

Sentence 1: I hope Yan could be such a person when she grows up.

	Num of emotional words	*Num of emotional phrases*
Expect	6,999	482
Joy	12,564	958
Love	29,418	2,147
Surprise	1,359	151
Anxiety	2,565	363
Sorrow	8,995	1,060
Anger	2,565	363
Hate	6,399	820

Table 4.1: The counts of emotional words and emotional phrases in each category.

In the sentence 1, the word "*hope*" conveys a kind of emotion as "*Expectation*" directly.

Sentence 2: His attitude is disgusting.

In the sentence 2, the word "*disgusting*" expresses the emotion as "*Hate*" directly also.

Indicate emotion indirectly

Sentence 3: If winter comes, can spring be far behind?

In the sentence 3, the word "*spring*" doesn't convey any emotion in usual, but in this context, it indicates the emotion with respect to the "*Expect*" according to the context.

Sentence 4: He felt the power when he saw the rising sun.

In the sentence 4, the word pairs "rising sun", which should be a single word in Chinese, are also the normal words without any emotions, but in the specific context, they contain some positive information that conveys the emotion of "*Expect*".

We compare some emotional words in the corpus, and find that there are many emotional words are annotated with multiple emotions in the corpus. In this study, we choose the emotional type, with respect to the highest emotion intensity, as the chief emotion of the word.

There are 13,621 different Emotional words annotated in Ren-CECps, and the emotional words occurred totally 71,199 times. The counts of emotional words from Ren-CECps in each category are shown in Table 4.1.

4.1.2 Emotional phrases

Emotional phrases are composition of words so as to convey emotions.

For examples,

Example 1: thank goodness

This phrase represents the emotion of $grateful(joy)$, while each single word could not express this emotion.

Example 2: good and evil

This phrase includes two words, each word convey opposite sentiment. The phrase can emphasize one of sentiment when this phrase is used in the context.

In Ren-CECps, some phrases are annotated with two or three kinds of sentiments. We use the same method as emotional words to determine the chief emotion of emotional phrases.

There are 2,963 different Emotional phrases annotated in Ren-CECps, and the occurrences of emotional phrases count 6,615 times. The counts of emotional phrases among Ren-CECps in each category are shown in Table 4.1.

4.1.3 POS (part-of-speech) tags

We are interested in the POS tags of all words, because words of some POS tags such as adjective and verb are more likely to convey emotion [9], such as the adjective *pretty* may express emotion of love, and the verb *dislike* can express the feeling of hate. Generally, most of adjective and verb can express more or less sentiments. Besides adjective and verb, there are totally 41 kinds of POS tags in our Data Source.

4.1.4 Degree words

Degree words perform the function of changing the intensity of emotions. High frequency of occurrence of degree word, such as *very*, *almost*, *alittle*, has been observed in Ren-CECps.

There are two examples show as follows:

Example 1:(very) happy

Example 2:(a little) regret

In the first example, the word *happy* is modified by the degree word *very*, which always increases the emotion intensity of the emotion words or emotional phrases. In the second example, the word *regret* is modified by the degree word *alittle*, which always decreases the emotion intensity of the emotion words or emotional phrases.

There are 1,039 different degree words annotated in Ren-CECps, the occurrences of degree words count 16,713 times, among which, 8,294 degree words have modified emotional words or emotional phrases.

4.1.5 Negative words

In Chinese articles in Ren-CECps, negative words, such as *no*, *cannot* and *donot*, appear with high-frequency. With negative words in the sentences, meaning of sentences can be reversed, however the emotion types may be changed or not.

Here are the examples:

Sentence 1: He hopes for snow.

Sentence 2: He doesn't hope for snow.

In comparison between two sentences above, we find that they express the opposite meaning, while two sentences convey the same emotion of expectation through the word *hope*.

There are 645 different negative words annotated in Ren-CECps, the negative words count for totally 13,750 times, among which, 3,668 negative words have modified emotional words or emotional phrases.

4.1.6 Conjunctions

People like using conjunctions in the complex sentences. On the one hand, conjunctions are used to join simple sub-sentences into a long sentence. On the other hand, the occurrence of some special conjunctions may signify the change of emotion intensity for sub-sentences.

For examples:

Sentence 1: For promotion, he ignored the family and betrayed friends.

Sentence 2: For promotion, he not only ignored the family, but also betrayed friends.

Compared to the first sentence, sentence 2 uses conjunctions *notonly...butalso...* to

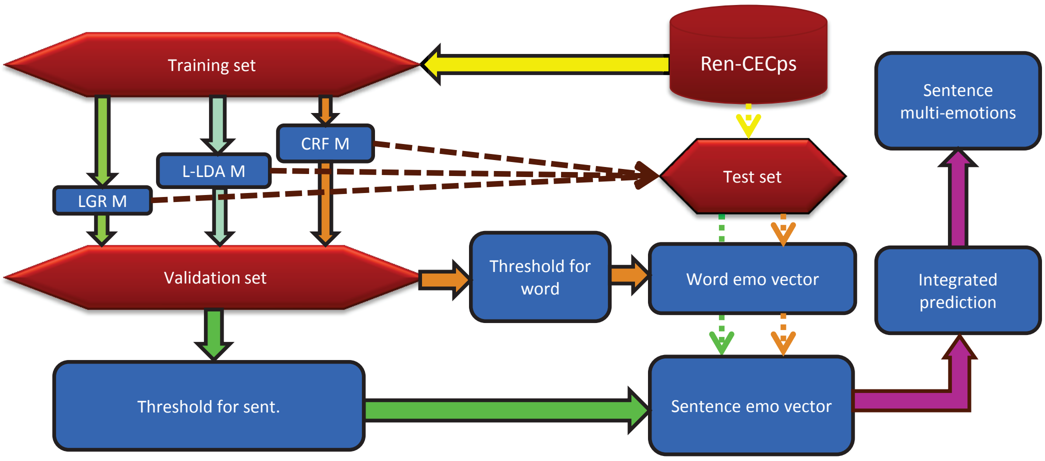

Figure 4.2: Multi-emotion analysis system constructed by the different perspectives.

express stronger critical emotion. There are 297 different conjunctions annotated in Ren-CECps, the total of conjunctions occur 12,900 times.

4.2 Predicting Sentence Multi-emotion

We propose three distinct probabilistic models for the emotion analysis in sentences, based on the different perspectives of emotion prediction by human beings. From each perspective, we employ a specific language feature to represent the aspect that we would consider for emotion prediction among the text. Specifically, these language features include the emotion-related global topics, the context-sensitive word emotions, and the local bag-of-words. We also integrate the probabilistic results generated from these models to improve the emotion prediction results, which is demonstrated in Figure 4.2.

4.2.1 Predicting sentence emotions from emotion-related topics

We have observed from large set of emotion-annotated texts that, at most of the time a document or a sentence describes some events that are centered on one single or multiple topics. Moreover, some topics could indicate similar emotions. For example, people tend to feel Sorrow, Hate, and Surprise when talking about disasters, and show emotions of Love and Joy when home and children are mentioned. Some experiments [26][39][38][2] conducted for sentiment analysis based on topics proved that there existed some correlations between topics and emotions [38] [2]. We expect to find their correlations with different emotion categories, even from some semantically less interpretable topics in a broad sense. Following this intuition, we employ the L-LDA model, which is suitable to learn the topic distribution in sentences and associate the topics explicitly with emotion labels in these sentences.

The L-LDA model for the emotion prediction in sentences can be formally depicted with a probabilistic graph model in Figure 4.3. Given a corpus of S sentences and N_s words in each sentence, we can associate each sentence s with some binary emotion labels Λ_{s1}, Λ_{s2}, $\cdots$,Λ_{sK}, with $\Lambda_{sK} = 1$ indicating the existence of emotion k in sentence s. The word topics and sentence emotion labels share the same categories (K) and the same semantic meaning in each category. The L-LDA model is a generative model, which means we can learn the model parameters by generating the sentences and their emotion labels and predict the sentence emotions by maximizing the posterior probabilities of these emotion labels.

Concretely, the square plate with letter S on the bottom-right corner indicates the total number of S sentences in the given corpus. Each sentence s consists of N_s words, which is denoted by the inside square plate N_s. We use nodes w and z to represent a word and the words topic. Shading on node w indicates that the values of these word variables are observed, while the values of the other unshaded nodes are latent. Given a topic assignment $z_{si} = k$ for word i in sentence s, the corresponding word variable w_{si} follows a Multinomial distribution $w_{si} \sim Mult(\beta_k)$ with a parameter vector β_k as the probabilities for word selection. We put β in a square plate K to indicate the word distribution with respect to each topic/label category. The topic variable z_{si} follows the Multinomial distribution $z_{si} \sim Mult(\theta_s)$, with a parameter vector θ_s as the probabilities for topic selection. Both β and θ are considered as random variables, in the L-LDA model, following Dirichlet distributions $\beta_k \sim Dir(\eta)$ and $\theta_s \sim Dir(\alpha \cdot \Lambda_s)$, with parameters η and α as the concentration parameters. These parameters indicate the prior knowledge about concentrations in the word selection and the topic selection. Through the a product $\alpha \cdot \Lambda_s$), the L-LDA model could restrict the candidate topics in sentence s to the existing emotion labels k for which $\Lambda_{sk} = 1$. Finally, the emotion labels in the L-LDA model follow the Bernoulli distribution $\Lambda_{sk} \sim Bernoulli(\sigma_k)$, where σ_k denotes the probability of having the emotion k in a sentence.

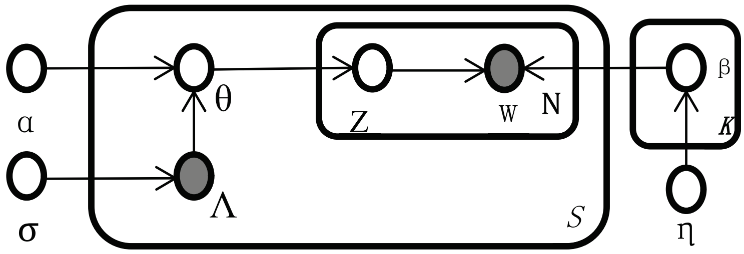

Figure 4.3: Graphical model of L-LDA is implemented for topic emotion extraction in sentence.

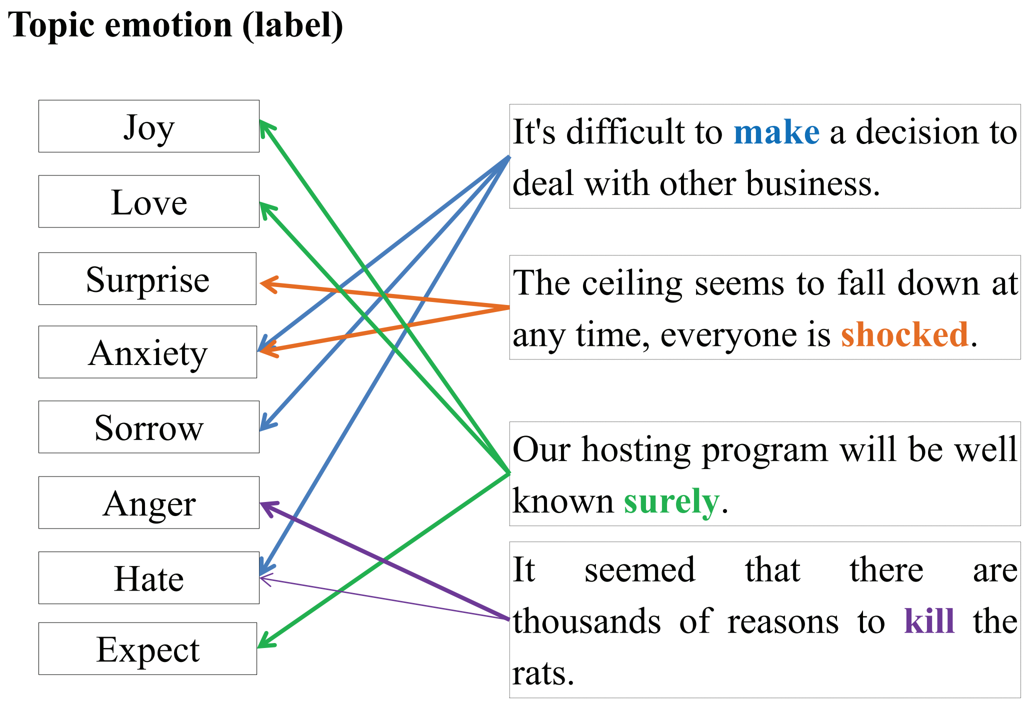

Figure 4.4: The correspondence between multi-emotion topics and sentences.

Figure 4.4 demonstrates the emotion-related topics learned from an L-LDA model. From the generative point of view, each sentence on the right column is generated by the corresponding emotion-related topics (or emotion labels) on the left column. We use different colors to specify the emotion-related topics and the generated words in each topic. Another important generation probability for the emotion-related topics in a sentence is stored in variable ϕ in the L-LDA model, which is a $K - dimensional$ vector indicating the conditional probability of observing each topic z_{sk} (corresponding to emotion k) given the words in sentence s as:

$$\begin{aligned} p^L(E_s|w_s) &= p(z_s|w_s) \\ &= \theta_s \\ &= [\theta_{s1}, \theta_{s2}, \ldots, \theta_{sK}] \end{aligned} \tag{4.1}$$

4.2.2 Predicting sentence emotions from emotional words

Words have been considered as the building blocks of sentences in natural language processing. They determine the syntactic structure of a sentence, and compose the predicates and the arguments in sentences semantic meanings. As for the text emotion analysis, we consider them as the sentence emotion indicators. And because the word emotion predictions are taken into account as the context information in our study, we expect these word emotions reveal the writers emotions in sentences.

Context-sensitive word emotion prediction

Instead of building a large fixed word-emotion lexicon as in the early studies, we consider the word emotions as random variables and try to maximize the conditional probability of the emotion assignment to a sequence of emotion variables given the word observations in a sentence. We use the CRF algorithm to predict the sequence of word emotions in this study. The merit of our method relies on two aspects. It brings variability in the word emotion prediction, by taking the negative modifications, contrast conjunctions, and degree adverbs into account to predict the most probable emotion assignment. And because CRF is a probabilistic model, it generates the marginal probability for each word emotion assignment, which is more informative than some emotion labels in a word-emotion lexicon.

Concretely, we consider $K + 1$ labels for the word emotions, with K distinct emotion labels and 1 no-emotion label. Therefore, for each word w_i we have $p(\hat{e}_i)$ which is a K+1-dimensional vector indicating the probabilities of assigning emotion and no-emotion label to this word. It has to be noticed that because most words in the real world as well as in the corpus are considered and labeled as no-emotion, the learning algorithm could be confused by the imbalanced labels in the training set. It might cause the prediction of a high recall score for the no-emotion label yet lower recalls for the emotion labels. And because the target of word emotion prediction is to help predicting the sentence emotions, it is of higher priority to find a reasonable amount of emotion indicative words with their emotions. Therefore, we select a threshold on the no-emotion prediction based on a validation set from the emotion corpus, in order to balance the recall and precision of no-emotion label prediction and consequently to assist the emotional word prediction.

The final word emotion prediction consists of the emotion indicative words w_{si} in sentence s and the corresponding K probabilities $p(e_{sik}$ for each emotion category after normalization. We use a factor ϕ_{si} to denote the normalized emotion distribution for word w_{si} as

$$\begin{aligned} \phi_{si} &= [p(e_{si1}), p(e_{si2}), \cdots, p(e_{siK})] \\ &= \frac{[p(\hat{e}_{si1}), p(\hat{e}_{si2}), \cdots, p(\hat{e}_{siK})]}{1 - p(\hat{e}_{si0})} \end{aligned} \tag{4.2}$$

where $p(\hat{e}_{siK})$ is the marginal probability of emotion k for word w_{si} directly from the CRF prediction, and $p(\hat{e}_{si0}$ is the probability of assigning no-emotion to word $w_{si}.p(e_{sik})$ is the probability of emotion k after ruling out the no-emotion label.

Sentence emotion accumulation

Previous study of text emotion analysis found that emotions can be accumulated from the lower-level texts to the higher-level. This suggests that the high-level sentence emotions could be accumulated from the low-level word emotions if the word emotion is reasonably predicted within the context.

In practice, accumulation can be implemented in different ways. Because we have the probabilistic predictions of the word emotions in ϕ, it would be convenient to accumulate the sentence emotion factor through a factor product of these word emotion factors:

$$\Phi_s = \exp(\frac{1}{N_s}\sum_{i=1}^{N_s} \ln \phi_{si}) \tag{4.3}$$

In Equation 4.3, Φ_{sk} and ϕ_{sik} are the k^{th} entries of the emotion factors for sentence s and the inside word i, respectively. N_s counts the total number of emotional words in sentence s. We introduce a normalization term $\frac{1}{N_s}$ in this factor product to ensure the comparability of the sentence emotion factors Φ_s among sentences with different number (N_s) of emotional words. It has to be noticed that with this emotion accumulation, we are summarizing the pieces of emotional information, for each emotion entry k, from the inside words for a sentence. Each entry ϕ_{sk} in the sentence emotion factor corresponds to the accumulated likelihood of having the emotion k in sentence s, which is assumed to be proportional to the probability of $p(E_{sk})$. We can more explicitly write the sentence emotion factor as Equation 4.4.

$$\Phi_s \propto [p(E_{s1}), p(E_{s2}), \cdots, p(E_{sK})] \tag{4.4}$$

With these accumulated sentence emotion factors, we can select the entry thresholds t_k^* for each emotion category k in the sentence emotion factor ϕ_s, based on a validation set, to expect the prediction for each emotion k with a promising F1 score. The procedure is depicted in Fig. 4. For each k, we iteratively assign a threshold variable t_k with three decimal places ranging from (0, 1). A comparison of the emotion probabilities in ϕ_{sk} with the threshold t_k is made for each sentence $s \in \{1, 2, , S\}$ in the validation set. With the observation of emotion labels y_{sk} for each sentence s and emotion entry k in the validation set, we can count the number of true positive (tp_k), true negative (tn_k), false positive (fp_k), false negative (fn_k) predictions with respect to each threshold t_k, and calculate the corresponding $F1_k$ scores. This threshold selection procedure saves the highest $F1score$ in $F1_k^*$ and returns the corresponding best threshold in t_k^*. With the emotion thresholds in t^*, the sentence emotions could be confirmed as Equation 4.5.

$$\begin{aligned} E_s &= 1\{\Phi_s > t^*\} \\ &= [1\{(E_{s1}) > t_1^*\}, 1\{(E_{s2}) > t_2^*\}, \cdots, 1\{(E_{sK}) > t_K^*\}] \end{aligned} \tag{4.5}$$

with the corresponding probability of

$$\begin{aligned} p^C(E_s|w_s) &\propto \Phi_s \\ &= [\Phi_{s1}, \Phi_{s2}, \cdots, \Phi_{sK}] \end{aligned} \tag{4.6}$$

In Equation 4.6, the expression 1 examines the statement in the brackets, and takes the value of either 1 if the statement is true or 0 otherwise. Therefore, for sentence emotion prediction in Equation 4.6, we have $E_{sk} = 1$ if the k^{th} entry in emotion factor ϕ_{sk} is greater than the threshold t_k^*.

4.2.3 Predicting sentence emotions from local bag-of-words

In this section, we introduce sentence emotion prediction through a linear probabilistic model. We assume the sentence emotions are independent binary-valued random variables, each (E_{sk}) follows a Bernoulli distribution with the success probability ϕ_{sk}:

$$p(E_{sk} = 1) = \theta_{sk} \tag{4.7}$$

In this model, we only consider the correlation between the sentence emotion label and the words within the sentence. For each emotion label, we would like to evaluate the corresponding emotional significance for each word in the vocabulary (V). This emotional significance is denoted by a weight parameter $_k$, which is a $|V| - dimensional$ weight vector. If we employ a binary valued word vector v_s to represent the existence of each word from vocabulary V in sentence s, then the overall emotional significance with respect to the emotion label k would be $\omega_k^T \nu_s$. And we use the sigmoid function to transform this emotional significance to the success probability as:

$$\begin{aligned} h_{sk} &= g(\omega_k^T \nu_s) \\ &= \frac{1}{1 + \exp(-\omega_k^T \nu_s)} \end{aligned} \tag{4.8}$$

The parameter value in emotional significance ω_k for each emotion category k can be learned through a standard LGR algorithm by minimizing the squared error in a training set with a regularization term. In practice, we select a subset of the emotion corpus Ren-CECps to train the LGR classifier and adjust the hyper-parameters such as the regularization type and the regularization weight based on a separate validation set of the emotion corpus. Note that for each emotion category, the LGR classifier generates the probabilities of assigning sentence emotion variables $E_{sk} = 1$ for all s. And to predict multiple sentence emotions, we need to learn K such LGR classifiers for each emotion category, which turns to be a multi-label LGR classifier. The probability of emotion assignment to a sentence s can be represented as:

$$\begin{aligned} p^G(E_s|w_s) &= h_s \\ &= [h_{s1}, h_{s2}, \cdots, h_{sK}] \end{aligned} \tag{4.9}$$

4.2.4 Joint prediction

We have introduced three probabilistic models for the multi-emotion prediction from sentences. Given a sentence, our models could predict the possibilities of emotion assignments in each emotion category, based on the global emotion-related topics, the context-sensitive word emotions, and the emotional significance of ordinary words. Intuitively, we can expect these models to recognize the sentence emotions from different perspectives as human do in emotion prediction, since previous experiment [55] succeeded in attempting joint prediction. Therefore, we could make improvement in sentence emotion prediction by integrating the probabilistic results from three distinct models.

To obtain the optimal prediction, we firstly integrate every two probabilistic results by multiplying every entry of the probability factors through the factor product:

$$p^{LC}(E_s|w_s) \propto p^L(E_s|w_s)p^C(E_s|w_s) \tag{4.10}$$

$$p^{LG}(E_s|w_s) \propto p^L(E_s|w_s)p^G(E_s|w_s) \tag{4.11}$$

$$p^{CG}(E_s|w_s) \propto p^C(E_s|w_s)p^G(E_s|w_s) \tag{4.12}$$

To make the probabilistic values in integrated emotion factors sensible, we need to set a series of thresholds for each emotion category by balancing the precision and recall scores of the emotion assignments through a validation set. This procedure is similar to the threshold selection illustrated in Figure 4.5.

Finally, to integrate all probabilistic results, we employ the similar factor product to three probability factors:

$$p^{LCG}(E_s|w_s) \propto p^L(E_s|w_s)p^C(E_s|w_s)p^G(E_s|w_s) \tag{4.13}$$

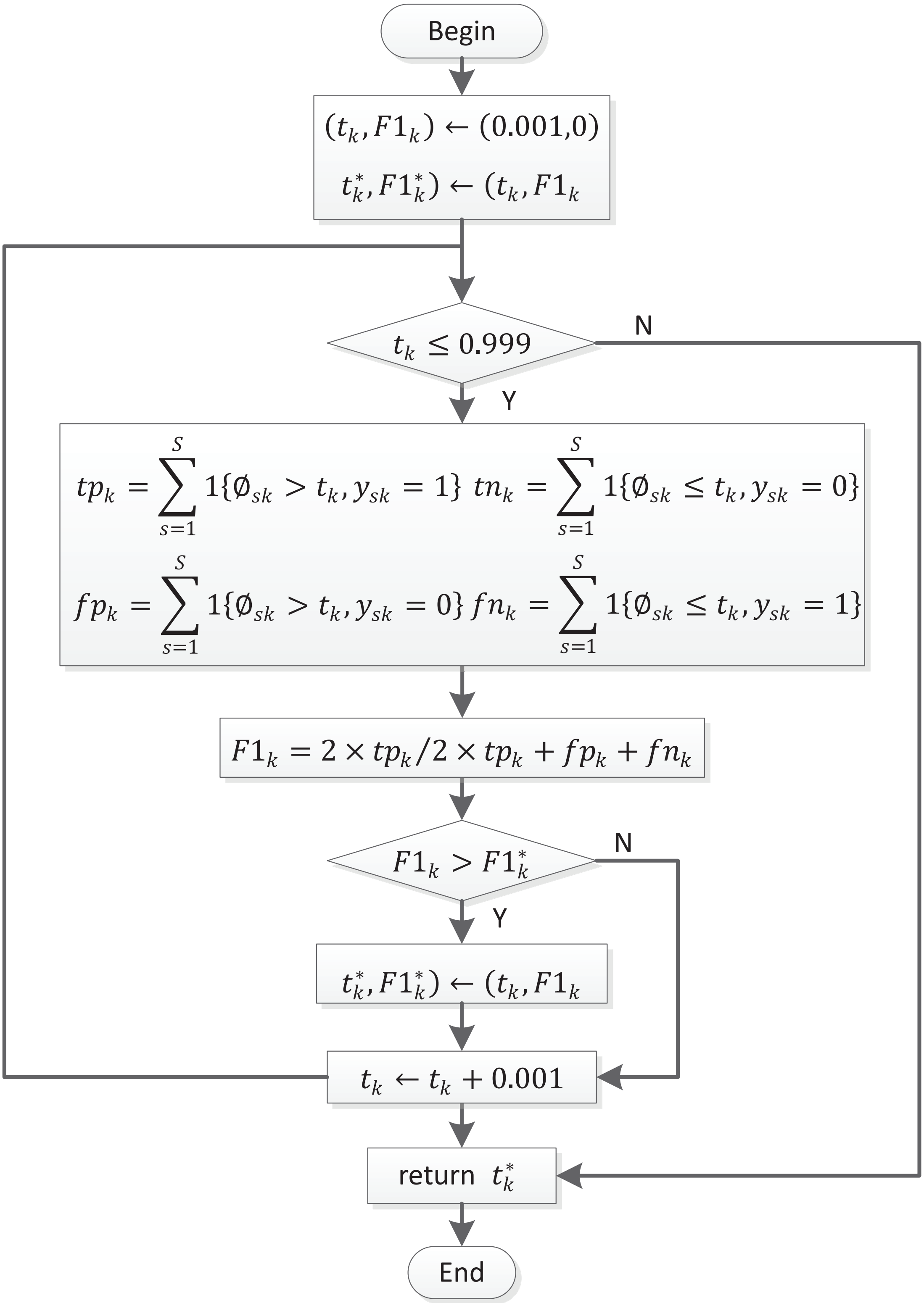

Figure 4.5: Threshold selection for confirming eight emotion categories and no-emotion.

Chapter 5

Experiments

We intend to predict sentence multi-emotion in three different perspectives by employing three probability models. Different experiments are implemented on single and the integrated methods including Bi-integration or Tri-integration. To examine the statistics significance of the results, we perform these experiments with 5-fold cross-validation. Five evaluation indicators are chosen to verify the performances.

We choose Ren-CECps, which is a manually annotated emotion corpus, for the experiment of sentence multi-emotion analysis. This corpus consists of 35096 sentences which are collected from the online Blog sites, and covers $K = 8$ common emotion categories (*Expect*, *Joy*, *Love*, *Surprise*, *Anxiety*, *Sorrow*, *Anger*, and *Hate*). In the corpus, sentences and words have been assigned with one or more emotion labels by considering the writers real emotional states. To ensure the emotion models are learned from the meaningful sentences, we filer some extremely short sentences (eg. with a single word) and the sentences only composed of punctuations, with 31058 sentences left.

For each fold of the cross-validation experiments, the corpus is divided into three subsets, with 18630 sentences for training models, 6214 sentences for selecting hyper-parameters and the emotion thresholds for each model, and 6214 sentences for evaluating the results of emotion prediction.

We separately train L-LDA, CRF, and LGR models for sentence emotion prediction based on the training set. Besides, we employ a multi-label Support Vector Machines (SVM) classifier for the sentence emotion prediction as a base line. The SVM classification is similar to the LGR algorithm by independently predicting in each emotion category, except that the SVM model only generates binary results to denote the emotion existence while the LGR model gives the more informative probabilities for emotion prediction.

We also select the hyper-parameters and thresholds of emotion probabilities in these models, by balancing the precision and recall scores of emotion prediction, based on the validation set. These hyper-parameters include the upper bound of the term frequencies and lower bound of the term-document frequencies in the L-LDA model, the balance between over-fitting and under-fitting in the CRF model, as well as the regularization type and the regularization term weight in the LGR and SVM models.

Finally, the probabilistic prediction results from the L-LDA, CRF, and LGR models are combined, as illustrated in section 4.2, to jointly predict the sentence emotions. The probability thresholds in each emotion entries in probabilistic outputs p^{LC}, p^{LG}, p^{CG}, and p^{LCG} are also selected to predict the sentence emotions.

5.1 Experiment1: sentence emotion prediction from emotion-related topics by L-LDA

5.1.1 Feature sets selection for L-LDA

As Labeled Latent Dirichlet Allocation (L-LDA) is a supervised machine learning method, we should add the label manually annotated label to predict the topic. We concentrate the task of emotion prediction, so we take the emotion tags of the corpus as the labels. Since each emotion maps one label in the L-LDA model, we call the predicted label as emotion-related label. In this experiment, we choose the **Standford Topic Modeling Toolbox** explored by the Standford Natural Language Processing Group (which can be found at url: http://nlp.stanford.edu/downloads/tmt/tmt-0.4/).

We use the words within sentence as the feature to perform the experiment, and filter the same words in one sentence to ensure every word is unique in sentence. Because the number of words in a sentence is relative less and some words used with the high frequency, we set two constraints to pre-process the sentences for guaranteeing topics inference more effective. The sentence are edited out, if the number of words less 5. In addition, we build a word vocabulary with the corresponding frequency of appearance for each word. We

	Example of L-LDA output architecture	
$cet-2.xml-3-3$	Expect	0.24831855170568798
	Love	0.009947655975201046
	Anxiety	0.38545729200623563
	Sorrow	0.24567989604558996
	Surprise	0.01135646424430853
	Hate	0.06602723482880334
	Anger	0.026320577272228102
	Joy	0.006892327921945441

Table 5.1: Example of L-LDA output architecture.

delete the words with the ten highest frequency in sentence, and construct the input data architecture as follows;

$$input = sent-id, \quad emo_0\ emo_1\ \cdots emo_7, \quad word_0\ word_1\ \cdots word_n$$

There are three columns in one sequence of the input data architecture, we separate them with comma. The first column is sentence id, and second emotion-related topic(here, emotion label), and the last words. In order to understand this data architecture more intuitively, we give some examples following.

$$input = cet-2.xml-1-2, \quad Expect\ Love, \quad it\ must\ be\ abundant\ noble\ kind-hearted$$

$$input = cet-18.xml-9-1, \quad Love, \quad mature\ man\ always\ old$$

As shown examples above, there are two emotion labels in the first sentence, and one in the second sentence.

Because we only interest in the relationship between the labels and the sentence, we only pay our attention to the probability of emotion-related topic. We take the example in table 5.1 to illustrate.

The emotion-related topic and corresponding probability are given by the L-LDA, the most likely to be expressed are *Hate*, *Anxiety* and *Sorrow* with high probabilities in this sentence, while *Love*, *Joy* and *Surprise* are expressed in a low probabilities.

5.2 Experiment2: sentence emotion prediction from emotional words by CRF

5.2.1 Feature sets selection for CRF

We want to predict the sentence multi-emotion under the consideration of contextual information, so we perform Conditional Random Field (CRF) to annotated word emotions to build the word emotion vector. We use the open source implementation of Conditional Random Field named **CRF++:Yet Another CRF toolkit**, which can be found at url: http://crfpp.googlecode.com/svn/trunk/doc/index.html.

There are five feature sets in experiment on emotional words annotation. We adopt n-gram models where n could be 1, 2 or 3 for words F_w, corresponding POS tags F_{pos}, conjunctions modification F_{cm}, negative words modification F_{nm} and degree words modification F_{dm}.

$$
\begin{aligned}
F_w &= \{1-gram(Word)\} \cup \{2-gram(Word)\} \cup \{3-gram(Word)\} \\
F_{pos} &= \{1-gram(pos)\} \cup \{2-gram(pos)\} \cup \{3-gram(pos)\} \\
F_{cm} &= \{1-gram(conj-mod)\} \cup \{2-gram(conj-mod)\} \\
&\quad \cup \{3-gram(conj-mod)\} \\
F_{nm} &= \{1-gram(neg-mod)\} \cup \{2-gram(neg-mod)\} \\
&\quad \cup \{3-gram(neg-mod)\} \\
F_{dm} &= \{1-gram(deg-mod)\} \cup \{2-gram(deg-mod)\} \\
&\quad \cup \{3-gram(deg-mod)\}
\end{aligned}
$$

where $conj-mod$, “deg-mod”, “neg-mod” mean the modification relations between words and conjunctions, words and degree words, words and negative words respectively.

To illustrate these feature sets more specifically, we give a simple example as is shown in Figure 5.1 for the short sub-sentence “Though he isnot very happy”. In this sentence, we have a conjunction “though”, a negative word “isnot” as well as a degree word “very”. The modification relations between words and these modifiers are represented by the dotted

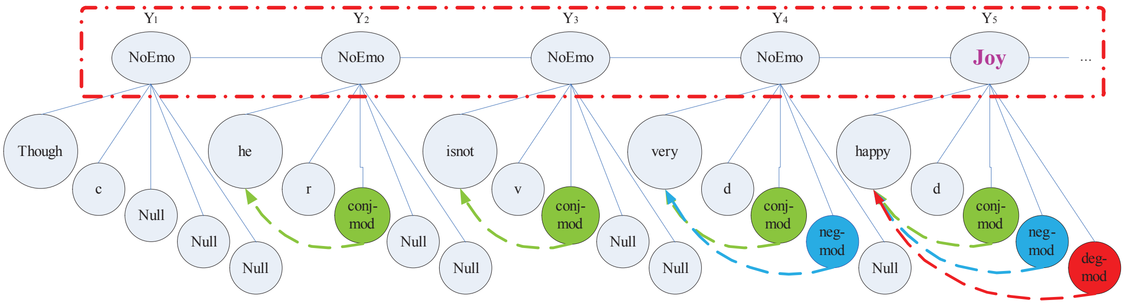

Figure 5.1: Modification relations between words and corresponding Modifiers.

arrows in Figure 5.1.

Here, we enumerate tri-gram model of the above five feature sets for the state node $Y3$ and $Y4$ in Figure 5.1 as below:

state node $Y3$

$$
\begin{aligned}
F_w &= \{though, he, isnot\} \cup \{he, isnot, very\} \\
&\quad \cup \{isnot, very, happy\} \\
F_{pos} &= \{c, r, v\} \cup \{r, v, d\} \cup \{v, d, d\} \\
F_{cm} &= \{Null, conj-mod, conj-mod\} \cup \{conj-mod, conj-mod, conj-mod\} \\
&\quad \cup \{conj-mod, conj-mod, conj-mod\} \\
F_{nm} &= \{Null, Null, Null\} \cup \{Null, Null, neg-mod\} \\
&\quad \cup \{Null, neg-mod, neg-mod\} \\
F_{dm} &= \{Null, Null, Null\} \cup \{Null, Null, Null\} \\
&\quad \cup \{Null, Null, deg-mod\}
\end{aligned}
$$

state node $Y4$

$ExperimentNo.$	$CombinationNo.$
$CRF-based-exp-No1$	F_w,F_{pos}
$CRF-based-exp-No2$	F_w,F_{pos},F_{cm}
$CRF-based-exp-No3$	F_w,F_{pos},F_{dm}
$CRF-based-exp-No4$	F_w,F_{pos},F_{nm}
$CRF-based-exp-No5$	F_w,F_{pos},F_{cm},F_{dm}
$CRF-based-exp-No6$	F_w,F_{pos},F_{cm},F_{nm}
$CRF-based-exp-No7$	F_w,F_{pos},F_{dm},F_{nm}
$CRF-based-exp-No8$	F_w,F_{pos},F_{cm},F_{dm},F_{nm}

Table 5.2: Feature sets assigned for each CRF-based emotional word annotation experiments

$$F_w = \{he, isnot, very\} \cup \{isnot, very, happy\} \cup \{\}$$

$$F_{pos} = \{r, v, d\} \cup \{v, d, d\} \cup \{\}$$

$$F_{cm} = \{conj-mod, conj-mod, conj-mod\} \cup \{conj-mod, conj-mod, conj-mod\} \cup \{\}$$

$$F_{nm} = \{Null, Null, neg-mod\} \cup \{Null, neg-mod, neg-mod\} \cup \{\}$$

$$F_{dm} = \{Null, Null, Null\} \cup \{Null, Null, deg-mod\} \cup \{\}$$

5.2.2 Feature sets for each CRF-based experiments

In order to explore the best feature sets for experiments on emotional word annotation, we try several combination of different feature sets,which are word, part-of-speech, conjunction modification, degree word modification and negative word modification.

The detailed assignment are shown as table 5.2:

5.2.3 Results of each CRF-based experiments

In this subsection, the best combination of feature sets is decided by comparison of F1-score given by the corresponding experiment.

As it's shown in Figure 5.2. $CRF-based-exp-No8$, adopts all five feature sets, gets the highest F-score, which proves the effectiveness of modification features for emotional keywords annotation. Therefore, we use all the feature sets including word, part-of-speech, conjunction modification, negative modification and degree modification in the emotional words recognition to build the word emotion vector.

Finally, we create the sentence emotion vector by multiplying the equivalent entries of

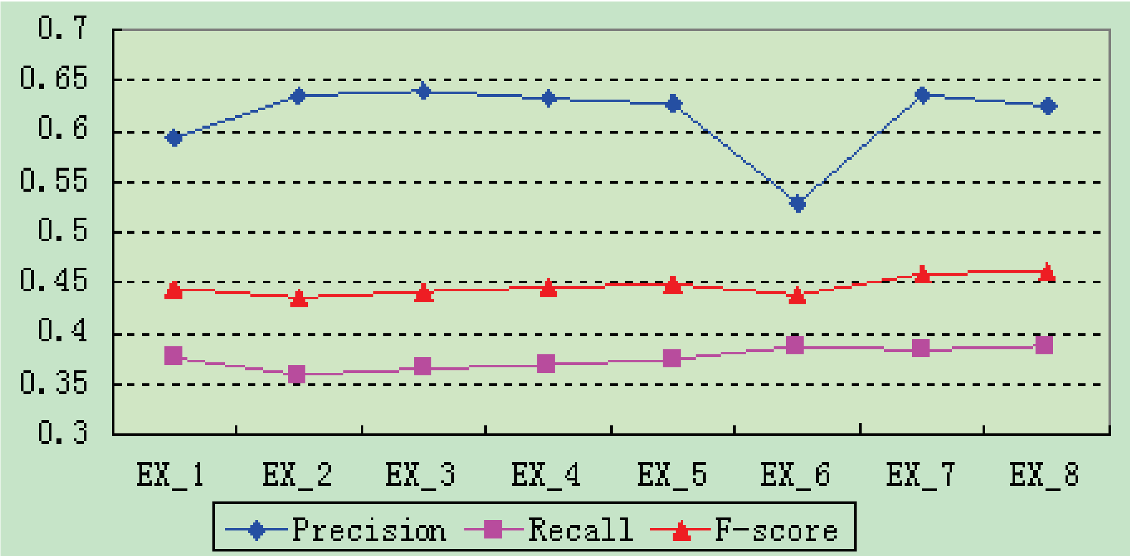

Figure 5.2: Average Precision, Recall and F-score of Eight Experiments.

word emotion vectors. The data form like table 5.1.

5.3 Experiment3: sentence emotion prediction from bag-of-words by LGR

5.3.1 Feature sets selection for LGR

For Taking the local bag-of-words information into account, we use the Logistical Regression to classify sentence into eight emotion categories. We also choose a open source toolkit to perform this function, which is called scikit-learn downloaded at http://scikit-learn.org/stable/index.html.

There are several machine learning methods supplied in this toolkit, however, we only use LGR this time. Logistical Regression is a binary machine learning method, so it only can be used for address binary classification. To deal with multi-class problem, using one-verse-all policy is a better way. We plan to train eight different classifiers for getting eight probabilities, which represent the possibilities of the positive label existing in sentence.

Finally, we build the sentence emotion vector by combination of probabilities correspond to the positive label. In other words, each element in sentence emotion vector is a probability value with respect to each emotion category.

5.4 Experiment4: sentence emotion prediction by integration

To predict sentence multi-emotion, we propose to create the new sentence emotion vector by combination of sentence vectors calculated by L-LDA, CRF and LGR respectively. At the first round, we attempt implement Bi-integration by the combination of any two emotion vector out of three. At the second round, we attempt the Tri-integration. The element in new eight dimensional emotion vector is still a probability calculated by multiplying the same entry in the old emotion vector. To make sure the existence of each emotion in sentence, we introduce a sentence-specific threshold for each emotion category, which is a float value in three decimal places. If the probability value is bigger than the threshold value, we think that the corresponding emotion exist in sentence.

Chapter 6

Evaluating the Single and Integrated Prediction

In this study, we propose the probabilistic models to predict multi-emotions in sentences, by allowing multiple emotion labels assigned for each sentence. For this multi-label classification problem [53] [52] [58] [54], we have to employ some overall evaluation methods to evaluate the synthetic performance in emotion prediction.

6.1 Evaluation methods

Six multi-label evaluation methods, including $HammingLoss$ [15], $Accuracy$, $Precision$, $Recall$ [16], $MicroF_{score}$ and $MacroF_{score}$, are employed to thoroughly analyze the emotion classification results. Specifically, we use Hamming Loss to evaluate the error rate of emotion prediction regardless of the emotion categories. The details of the evaluation methods are illustrated below:

6.1.1 Hamming Loss

The average percentage of misclassified labels.

$$\text{hloss}(H) = \frac{1}{S} \prod_{s=1}^{S} \frac{|E_s \bigoplus \tilde{E}_s|}{K} \tag{6.1}$$

where H denotes the emotion prediction model, while E_s and $\tilde{E}_s$ are the predicted emotion labels and the true emotion labels in sentence s. $\oplus$ is the exclusive OR operator. S counts the number of sentences in the test set, and $K = 8$ in this case is the number of emotion categories.

6.1.2 Accuracy

The average percentage of correctly classified labels among all the correctly and incorrectly classified labels.

$$\text{Precision}(H) = \frac{1}{S}\sum_{s=1}^{S}\frac{|E_s \wedge \tilde{E}_s|}{|E_s|} \tag{6.2}$$

6.1.3 Precision

The average percentage of correctly classified labels among all the predicted labels.

$$\text{Precision}(H) = \frac{1}{S}\sum_{s=1}^{S}\frac{|E_s \wedge \tilde{E}_s|}{|E_s|} \tag{6.3}$$

6.1.4 Recall

The average percentage of correctly classified labels among all the true labels.

$$\text{Recall}(H) = \frac{1}{S}\sum_{s=1}^{S}\frac{|E_s \wedge \tilde{E}_s|}{|\tilde{E}_s|} \tag{6.4}$$

6.1.5 $MicroF_{score}$ and $MacroF_{score}$

The averaged measure of precision and recall, for multiple emotions analysis. When calculating F_{score} for 2-label (binary) classification problems, we need to count number of correctly predicted positive labels (tp), the number of incorrectly predicted positive labels (fp), the number of correctly predicted negative labels (tn), and the number of incorrectly predicted negative labels (fn). And the formula for the F_{score} is

$$\begin{aligned}
\Phi^{(1)} &= \left(\Phi_1^{(1)}, \cdots, \Phi_k^{(1)}\right) \\
\Phi^{(2)} &= \left(\Phi_1^{(2)}, \cdots, \Phi_k^{(2)}\right) \\
\Phi^{(1)}\Phi^{(2)} &= \left(\Phi_1^{(1)}\Phi_1^{(2)}, \cdots, \Phi_k^{(1)}\Phi_k^{(2)}\right)
\end{aligned} \tag{6.5}$$

When calculating $MicroF_{score}$ and $MacroF_{score}$ for the multi-label classification problems, we have for each label k a set of counts as $(tp_k,\ fp_k,\ tn_k,\ fn_k)$. The $MicroF_{score}$ calculates the $F1_{score}$ of the positive predictions, by ignoring the predicted emotion labels:

$$\text{MicroF1} = \frac{2 \times tp_*}{2 \times tp_* + fp_* + fn_*} \tag{6.6}$$

where tp_*, fp_*, and fn_* count the total numbers of true positive predictions, false positive predictions, and false negative predictions regardless of the emotion label.

$$F_{score} = \frac{2 \times tp}{2 \times tp + fp + fn} \tag{6.7}$$

The $MacroF_{score}$ calculates the average of $F1_{scores}$ for each emotion label:

$$\text{MacroF1} = \frac{1}{K} \sum_{k=1}^{K} \frac{2 \times tp_k}{2 \times tp_k + fp_k + fn_k} \tag{6.8}$$

where tp_k, fp_k, and fn_k denote the counts of true positive predictions, false positive predictions, and false negative predictions for each emotion label k, while K is the number of emotion categories.

6.2 Result analysis

With these evaluation criteria, we are able to examine the emotion prediction results from each single probabilistic model as well as the integrated models. The evaluation scores for single and integrated models, from the 5-fold cross-validation experiments, are shown in the box plots from Figure 6.1 $\sim$ 6.9, with the maximum, 1st quartile, median, 3rd quartile, and minimum evaluation scores plotted in separate boxes. We abbreviate the names of the ingredient models with single characters, i.e., "C" for CRF, "G" for LGR, and "D" for L-LDA.

Seeing the accuracy in the Figure 6.1, the Tri-integration achieve the best results among all the methods. "C-G" gets the best result in the Bi-integration, while "LGR" gets the best one in the single methods. The plot box of CRF is bigger than others, which suggests that the score of each sub-experiment differ in magnitude in the 5-fold cross-validation, and also prove that the word emotions recognition significantly depends

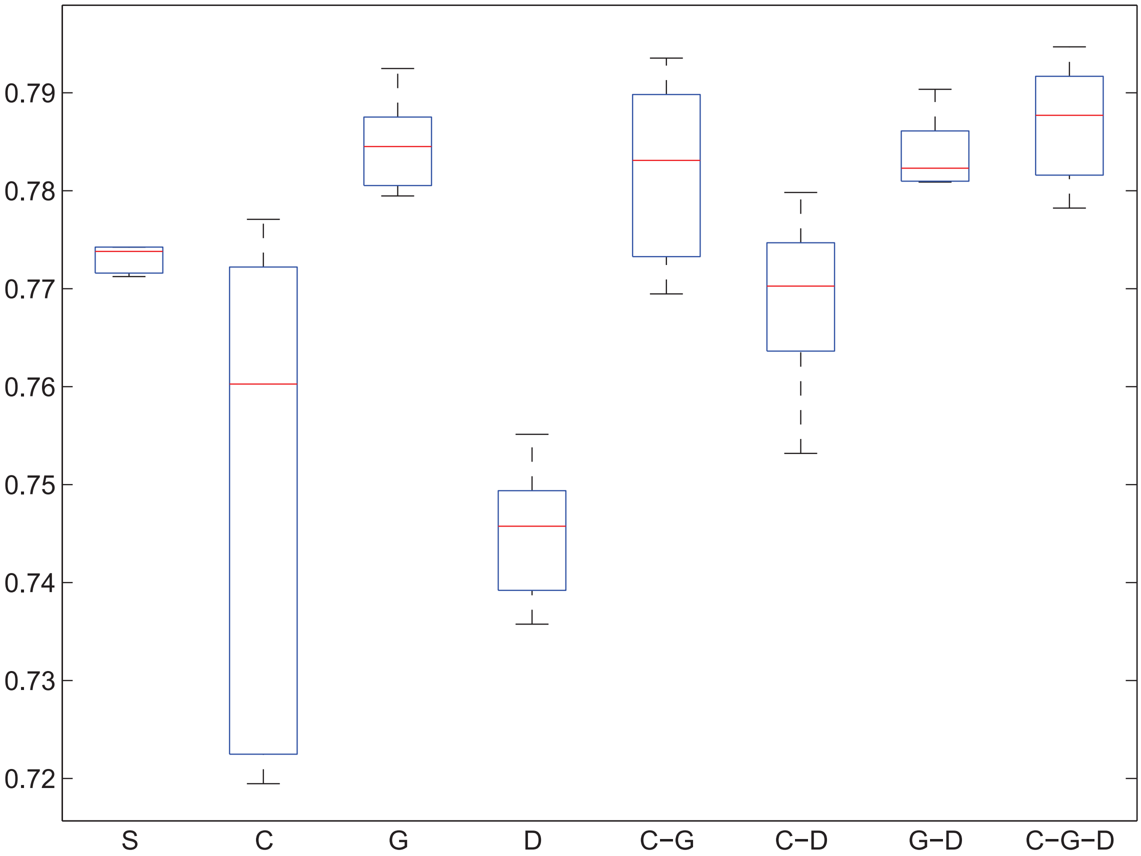

Figure 6.1: Accuracy for single and integrated models from 5-fold cross validation experiment.

on training data, and when the grate change occurs in training data, the result of word emotion recognition also makes a great change. In other words, words emotions make a great influence on sentence multi-emotion. The plot boxes of Bi-integration using CRF show the same pattern.

As shown in Figure 6.2, our Tri-integration get the best result with the lowest score, which means that there are least misclassified labels existing. The Bi-integration get the medium results compared to the single methods. There are most misclassified labels in the prediction by CRF, we think that predicting sentence emotions indirectly for word emotions is reason for causing such great distinction. Like the accuracy, there is also a broad distinction in the plot box of CRF in the 5-fold cross-validation, which suggests that there still a certain development should be make in the factor product of word emotions.

In the comparison of plot boxes of $MacroF_{score}$ in Figure 6.3, the Tri-integration gets the best place, while Bi-integration gets second place overall, which prove the effectiveness of our integration methods. In the single method, SVM gets worst score. No matter

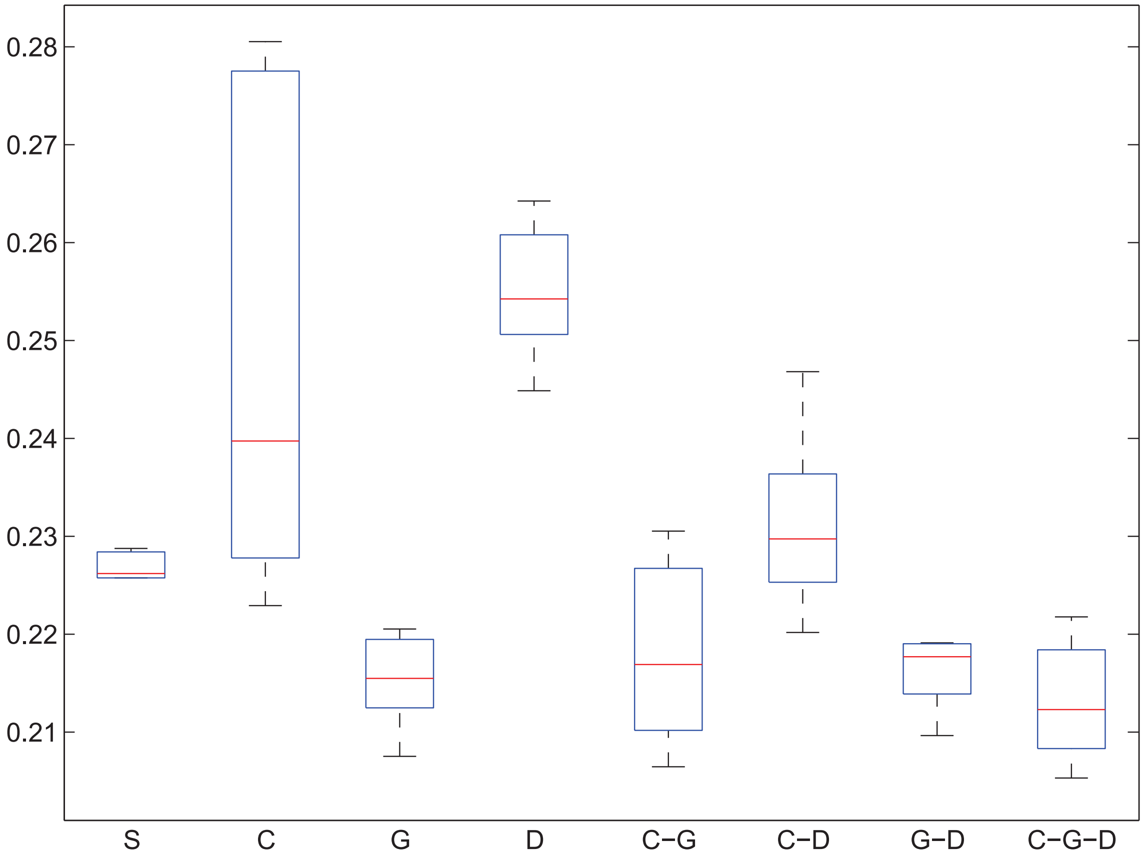

Figure 6.2: Hamming Loss for single and integrated models from 5-fold cross validation experiment.

integration methods or single methods, there are no apparent distinctions in the 5-fold cross-validation. That is to say, all the methods perform well across overall sets of data. Even emotion *surprise* (which have the least number in the data source) is predicted with the similar probability as same to other emotion categories.

As shown in Figure 6.4, the Bi-integration perform best among all the methods including Tri-integration. In the single methods, LGR also gets the similar score as same as the score of the Bi-integration and Tri-integration. But SVM gets worst score, it seems that SVM isn't able to distinct major categories ($Love$ and $Joyinthisstudy$) in data source. CRF still shows a bad performance with a great fluctuation when the different training data is using for sentence multi-emotion prediction. The reason we think is that because the language model based on CRF predicts sentence emotions by calculating the contextual relation, when we use 5-fold cross-validation experimental data to conduct the experiment by CRF, the contextual information is sufficient, so the a great fluctuation

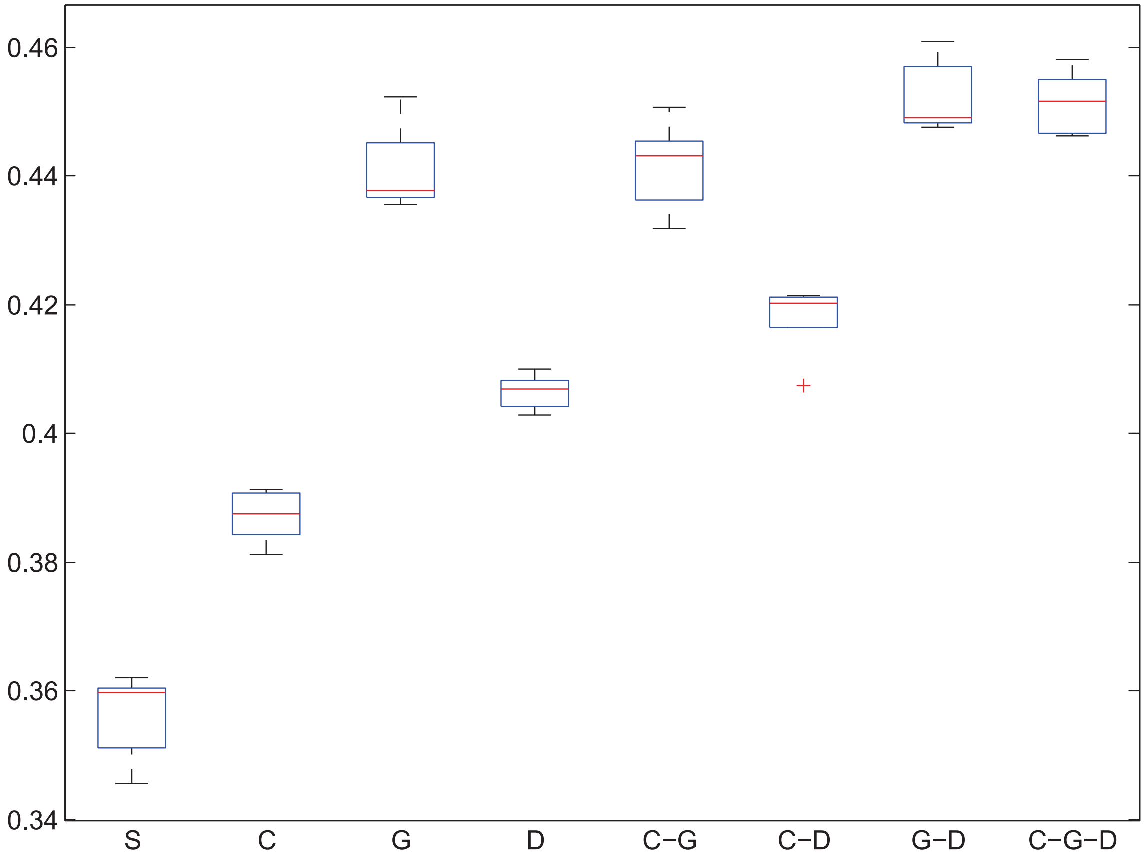

Figure 6.3: MacroF score for single and integrated models from 5-fold cross validation experiment.

occurs in 5-fold cross-validation.

To summarize the macro precision scores for sentence emotion prediction from the 5-fold cross-validation, we plot these macro precision scores in the box-plots for the single and integrated models in Figure 6.5. The red bars in the box-plots correspond to the median scores, which could best represent the evaluation of macro precisions. We find the integrated model "C-G-D" achieves the highest median macro precision among all the macro precision scores from the single and integrated models, and we also find the integrated models "C-G", "C-D", and "G-D" render higher median macro precision score than the single models of CRF, LGR, and L-LDA, respectively. These comparisons suggest that integrating the separate predictions from different probabilistic models could effectively improve the precision for sentence emotion prediction for different emotion categories. The length of each box corresponds to the variance of macro precision scores in the 5-fold cross-validation. We find the integrated model "C-G-D" also achieves smaller

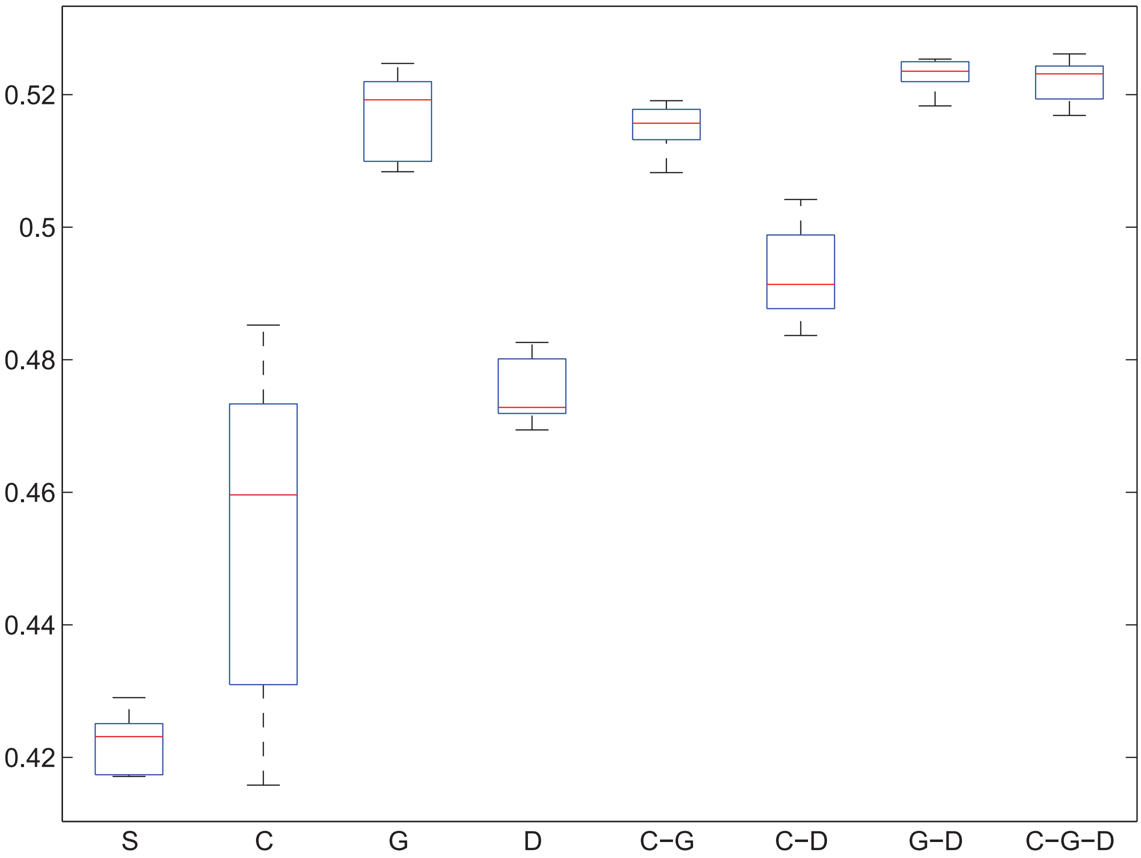

Figure 6.4: MicroF score for single and integrated models from 5-fold cross validation experiment.

variance in macro precision than “C-G”, which suggests that the integration also makes the emotion prediction more stable.

Figure 6.6 shows the summarization of the micro precision scores for sentence emotion prediction from the 5-fold cross-validation. Because micro averaging ignores the emotion category and only considers the correctness of predictions, we can use micro precision to evaluate the overall accuracies for these emotion classifiers. Again, we find the integrated model “C-G-D” achieves the highest median micro precision, and the other integrated models “C-G”, “C-D”, and “G-D” also achieve substantially improvements in their micro precisions. These comparisons indicate that the integrated models could generate more accurate sentence emotion predictions, even without considering the emotion categories. The length of the boxes for the integrated models are also smaller than the box length for single models, which also indicates the more stable predictions in the integrated models.

We plot the macro recalls from the 5-fold cross-validation for the single and integrated

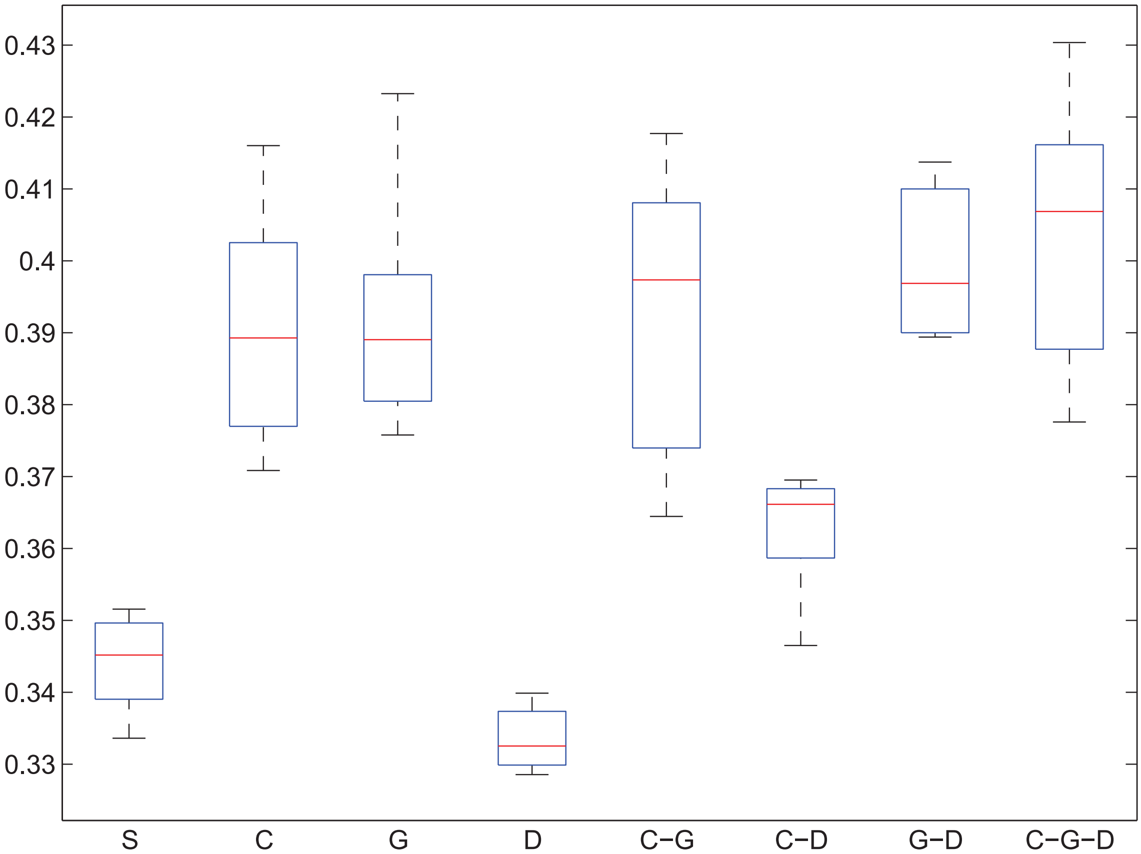

Figure 6.5: Macro precision for single and integrated models from 5-fold cross validation experiment.

models in Figure 6.7. The macro recall evaluates the averaged ratio of correctly predicted emotions among the number of original emotions, for different emotion categories. We find that the integrated model "G-D" achieves the highest median macro recall score among all the integrated and single emotion prediction models, and the integrated model "C-G-D" gets a very close macro recall score to the "G-D" model, and the single model L-LDA also achieves very promising macro recall score. The comparison indicates that the integrated model, by incorporating the predictions from single models, also renders the averaged macro recall scores. However, we can still get improvements by integrating models, as the "G-D" model. The length of macro recall box-plots suggests the integrated model "G-D" achieves the most stable emotion prediction among all the integrated and single emotion prediction models.

Finally, in Figure 6.8, we plot the micro recall scores from the 5-fold cross-validation for the single and integrated models. We can observe an evident improvement of the

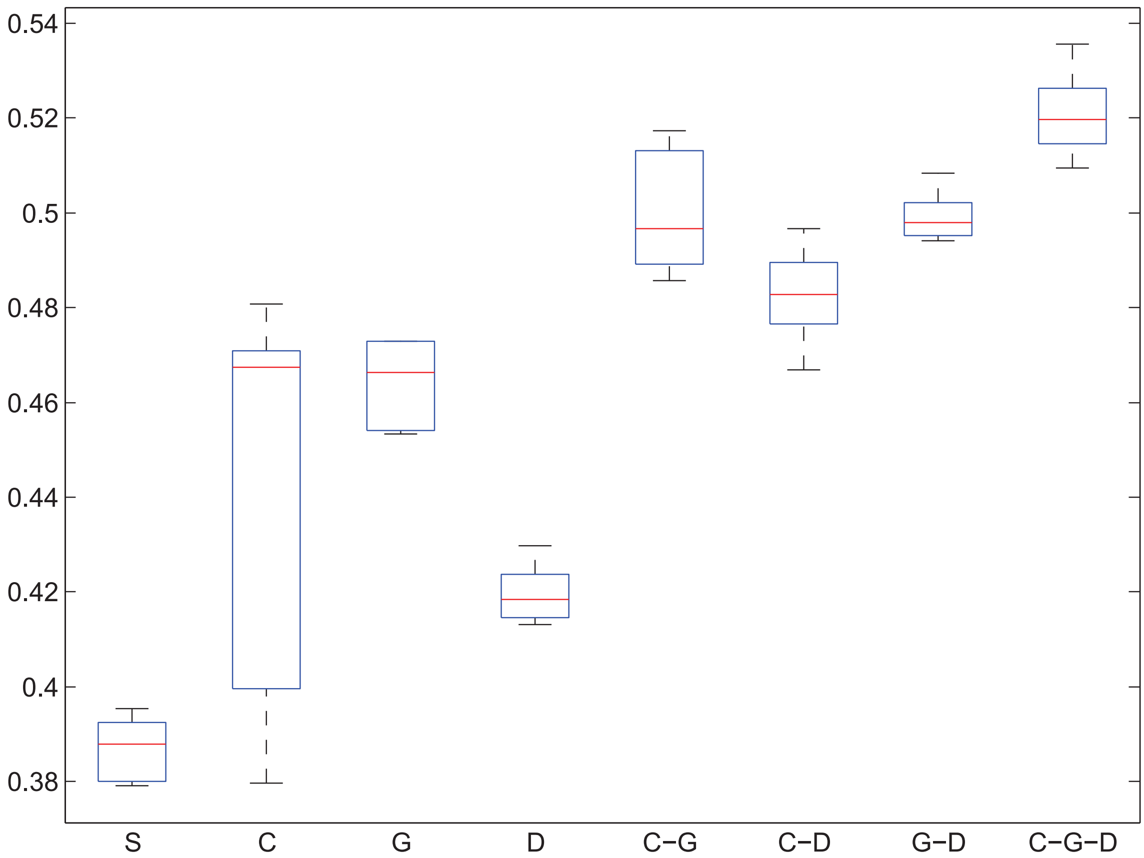

Figure 6.6: Micro precision for single and integrated models from 5-fold cross validation experiment.

median micro recall score by combining the single models to the integrated models. Like the macro recall score plot, the “G-D” integrated model achieves the highest median micro recall among all models, and the “C-G-D” integrated model also achieves a very closed median micro recall. Unlike the macro recall score plot, the “C-D” integrated model achieves the highest stability in the micro recall for emotion prediction. By looking into the length for different box-plots, we are able to conclude that the integration of multiple models could generate more stable micro recall scores than the single models.

The detailed median evaluations are reported in Table 6.1.

We first evaluate the emotion prediction with single models. The LGR model gets the best Hamming loss score of 21.55%, indicating that by just considering the local bag-of-word feature in a sentence, a multi-label probabilistic model could on average correctly make predictions for around 6.3 emotion labels out of the $K = 8$ labels. Compared to the conflicts in human annotations, all the reported error rates are acceptable. The Hamming

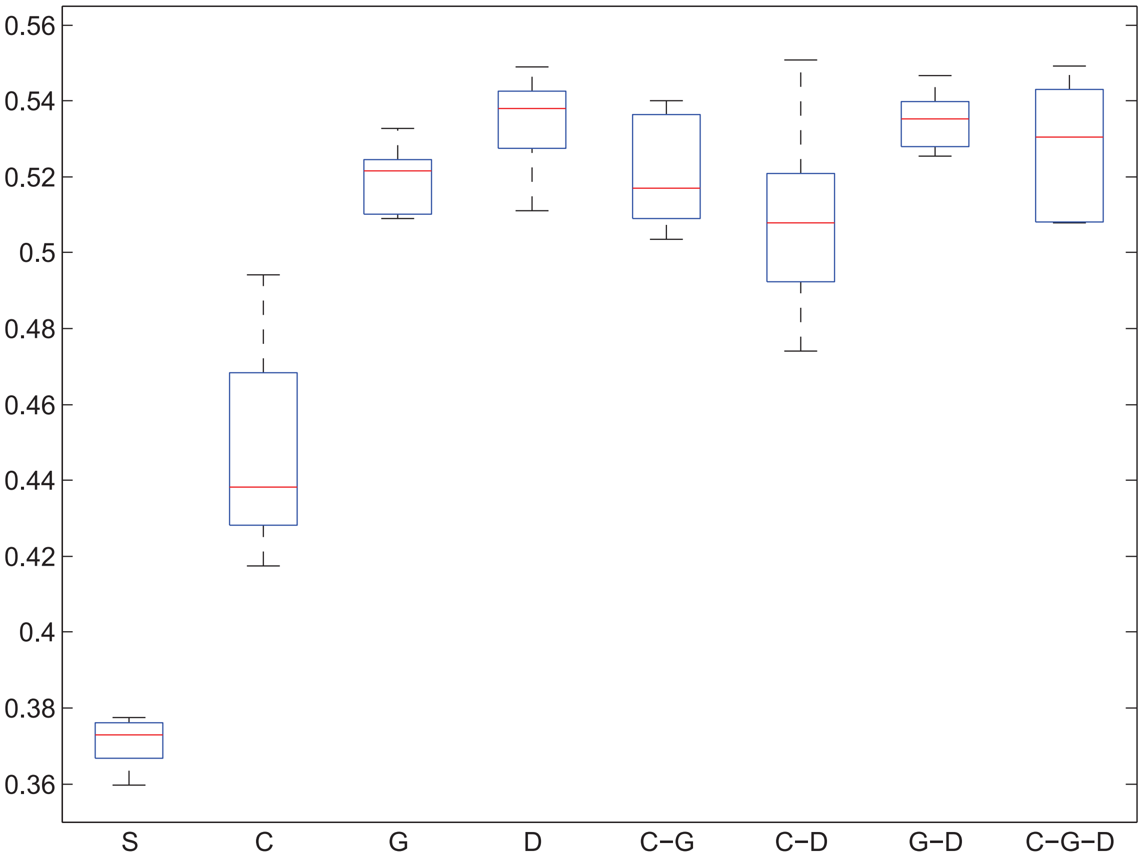

Figure 6.7: Macro recall for single and integrated models from 5-fold cross validation experiment.

loss evaluates the error rate for both positive and negative predictions, while the averaged Precision and Recall scores suggest the quality of the positively predicted emotion labels in each sentence. The CRF model achieves the highest average Precision of 46.74%, and the L-LDA model gets the best average Recall of 63.44%.

These results suggest that the context-sensitive word emotions from a CRF model help predicting emotions with fewer positive errors, while the emotion-related topics help the L-LDA model to recognize more subtle emotions in the sentences. We use the micro

SVM	CRF	LGR	L-LDA	C-G	C-D	G-D	C-G-D	
hloss	22.62	23.97	21.55	25.42	21.69	22.98	21.77	21.23
Precision	38.80	46.74	46.63	41.85	49.66	48.28	49.79	51.97
Recall	44.92	56.86	62.81	63.44	64.40	63.06	66.47	64.87
$Micro_{F1}$	42.32	45.95	51.93	47.27	51.56	49.13	52.35	52.32
$Macro_{F1}$	35.97	38.76	43.77	40.70	44.31	42.02	44.91	45.16

Table 6.1: Result of single, Bi-integration and Tri-integration by the different evaluation methods.

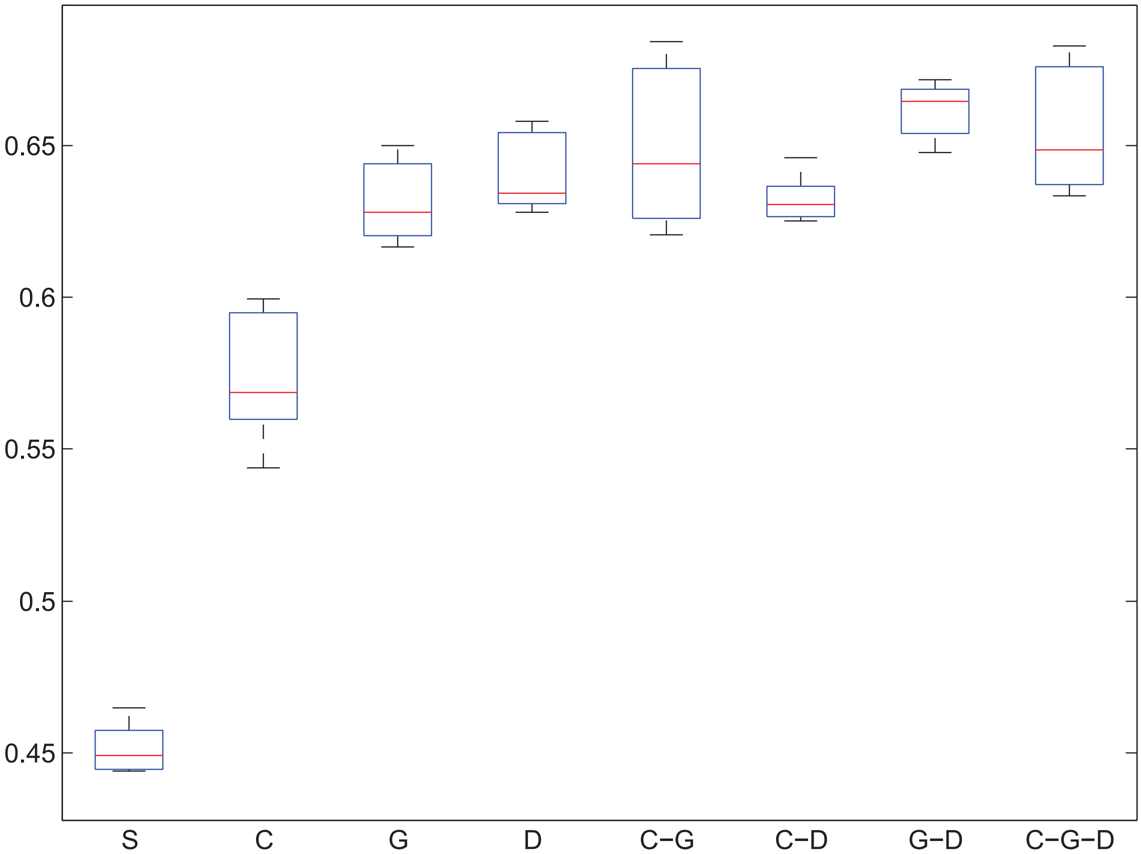

Figure 6.8: Micro recall for single and integrated models from 5-fold cross validation experiment.

and macro-averaged F1 scores to evaluate the overall performance of emotion prediction, by either distinguishing the emotion categories or not. The LGR model achieves the highest $MicroF_{score}$ (51.93%) and $MacroF_{score}$ (43.77%) among all the single models. We also notice that the $MicroF_{score}$ are around 7%-8% higher than the $MacroF_{score}$ for the sentence emotion prediction from different models, which implies that there might be distinguishable gaps in the F1 scores for different emotion categories. In Figure 6.9, we plot these F1 scores from the single and integrated models. All emotion classification models render higher average F1 scores in Love, Anxiety, and Sorrow, but get much lower average F1 scores in Anger and Surprise. This observation further suggests that the difficulty in predictions for different emotion categories varies significantly.

By integrating the probabilistic results from the single models, we construct a series of integration models (including Bi-integration and Tri-integration) for the multi-emotion prediction. Table 6.1 shows the evaluation scores of these models. All the Bi-integrated

Emotions	Word emotion count			Sentence emotion count		
	Valid	*Test*	*Train*	*Valid*	*Test*	*Train*
Sorrow	2014	2068	5780	1236	1237	3570
Surprise	257	320	885	166	152	518
Joy	2723	2766	8118	981	947	2916
Expect	1553	1503	4527	700	711	2181
Anxiety	3051	3155	9323	1501	1565	4568
Hate	1418	1405	4306	530	492	1538
Love	6135	549	18485	1799	1808	1019
Anger	559	6236	1699	329	336	5417
Total	17710	18002	53123	7242	7248	21727

Table 6.2: The occurrence of emotions in each category in the three sets.

models "C-G", "C-D", and "G-D" acquire better performance compared to the component models, by considering the Precision, Recall, $MicroF_{score}$ and $MacroF_{score}$. And by integrating the three models together, denoted as "C-G-D", we get even better emotion predictions. Finally, compared to the base line SVM model, our models achieve much better performance.

To further examine the results of multi-emotion predictions and compare the difficulty in predicting different emotion categories, we report the F1 score of emotion classification for each emotion label in Figure 6.9.

For both single and integrated models, we find that the F1 scores for Love, Anxiety, and Sorrow predictions are higher than other emotion predictions, while the Surprise and Anger predictions yield the lowest F1 scores among all the emotion predictions. The difference of classification difficulties among the basic emotions lies in the nature of the human emotions. For one reason, a great number of emotional words and phrases convey two or more negative emotions at the same time, which makes it impossible even for human to annotate. Take the following sentence as an example:

Sentence :I can't stand him cheating on me.

One corpus annotator labelled the word "cheating" with emotion of "Anger", while others labeled with "Hate" or both of them.

For example, by analyzing the emotion distribution among sentences and words in the emotion corpus, we find a significant imbalance among the number of emotion labels. Table 6.2 counts the occurrences word emotions and sentence emotions in the Training (Train) set, the Validation (Valid.) set, and the Test set, respectively, in which the words

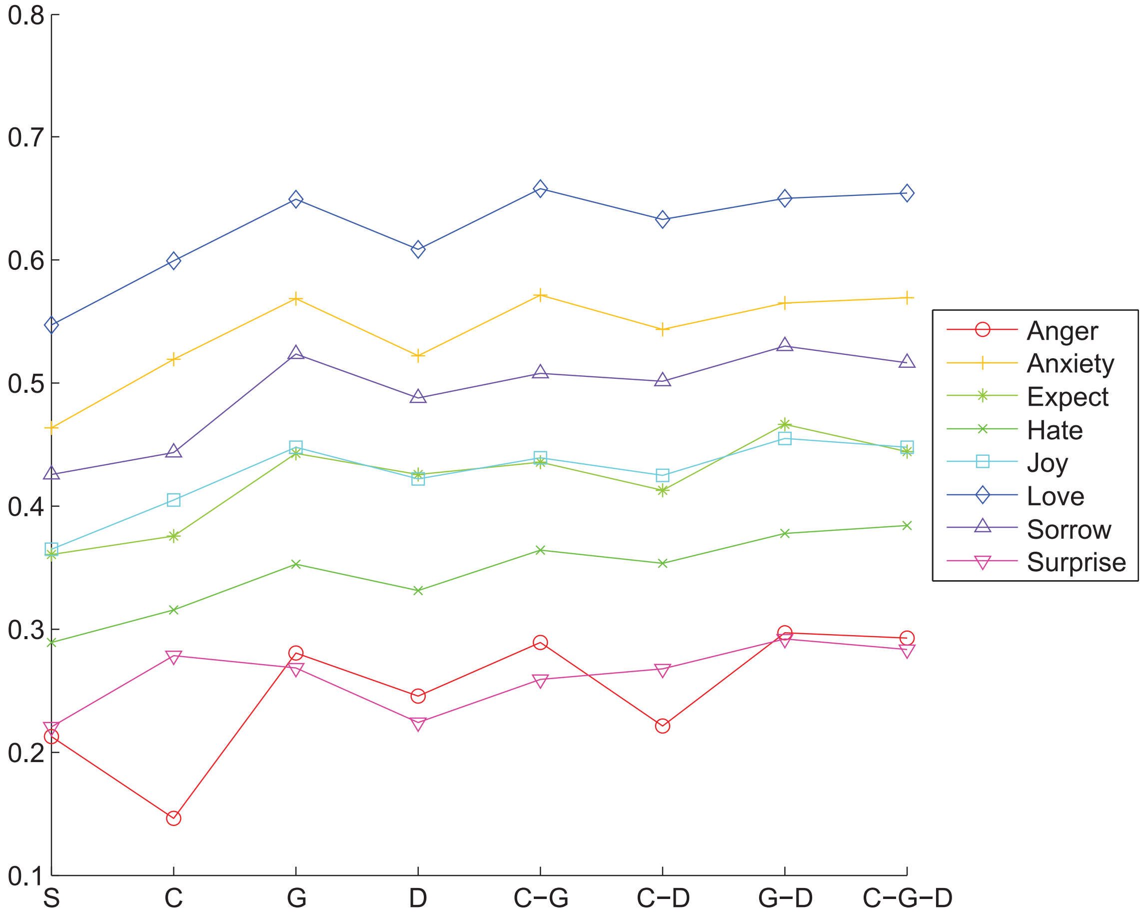

Figure 6.9: F1 score of each emotion category for single and integrated models from 5-fold cross validation experiment.

and sentences can express multiple emotions at the same time. It can be inferred that some emotions like Surprise and Anger are less frequently expressed. In this case, increasing the number of corresponding emotional samples would help improving the prediction results. Finally, by a horizontal comparison among the F1 score from different models, we also observe a rising trend of F1 score from single models to the Bi-integration models and to the Tri-integration model.

Chapter 7

Conclusion and Future Work

7.1 Summary of Sentence Multi-emotion Analysis

In this study, we propose three probabilistic models to predict the multiple emotion labels in sentences. Each model considers the emotion prediction from a particular perspective. We device the data source (Ren-CECps) into three sets which are training set, test set and validation set. We train language models on training set, and select threshold for word and sentence respectively on validation set, and finally evaluate effectiveness of the single methods and our Bi-integration and Tri-integration.

In the first perspective, we propose to predict sentence multi-emotion under consideration of the emotion-related topics. Since there are not so many words in the sentence, we have to pay attention to the effectiveness of topic extracted from sentences. Moreover, we try to find the relation between topic and emotion. Therefore, instead of Latent Dirichlet Allocation (LDA) which is a unsupervised machine learning method, we determine to choose the developed method of LDA named Labeled Latent Dirichlet Allocation (L-LDA) which is a supervised machine learning method to construct topic model. We map each topic to emotion tags, and succeed in construct the relation between emotions and sentence topics. The L-LDA model concludes the emotion labels by learning a series of emotion-related topics from a training corpus. Finally, we construct the sentence emotion vector in which each element is probability value representing the possibility of corresponding emotion existing in sentence.

In the second perspective, we predict sentence multi-emotion under the consideration of contextual information. Conditional Random Field (CRF) is best choice for us to employ, because the CRF annotate each token by calculate the relation of adjust sentences. Firstly, we use the CRF to recognize the words emotion in sentence. Each word is assigned with nine emotion tags (including eight emotion categories and No-emotion tag) with the corresponding probability. We view No-emotion tags in sentence are the confusing items, so we select threshold to filter them on the validation set, and construct the word emotion vector. For building the sentence emotion vector, we accumulate the sentence emotions from the context sensitive word emotions by multiplying relative entry of word emotion vector of words within sentence, and normalize them by smoothing method for keeping them comparable, since every sentence has different length, in other words, the number of words are different in every sentence.

In the third perspective, we explore contribution made from the local bag-of-words. We implement multi-emotion prediction by using Logistic Regression (LGR). Since the LGR is the binary classification model, we have to train eight LGR classifiers by one-verse-all. We build the sentence emotion vector by combination of the probability of each positive label in binary classification.

We implement the Bi-integration and Tri-integration by multiplying the same entry of the emotion vectors calculated from L-LDA, CRF and LGR respectively. To determine the multi-emotion in sentence, we compute emotion-specific threshold for each emotion category on validation set. If the probability is bigger than the corresponding emotion-specific threshold value, we think the emotions exist in sentence.

We employ different evaluation methods to compare and examine the multi-emotion prediction results from single and integrated models. The Hamming loss scores indicate an acceptable error rate in our emotion prediction results compared to the consistence in human annotations. The average Precision, Recall, and the Micro- and Macro-averaged F1 scores evaluate the quality of the positively predicted emotions in the sentences. By comparing these scores among single models, we observe that the word emotion based CRF model generates fewer false positive errors compared to the others (higher Precision), while the emotion-related L-LDA model recognizes a lot more subtle emotions (higher Recall) and the highest F1 scores. By evaluating the predictions from the integrated models, we get even better performance. The integration of each two models achieve much higher Precision, Recall, and averaged F1 scores compared to the component models. The integration of three models achieves the best Precision, Recall, and Micro-averaged F1 scores.

Besides, we further analyze the detailed classification results for each emotion category. The comparison of the F1 scores for emotion prediction in different emotion categories suggests that there exist significant differences in predicting different emotions, regardless of the models we employ. We predict from an observation of the emotion corpus Ren-CECps that the inherently lower expression frequencies in some emotions could be an important reason of the reduced performance in these emotion categories. Nevertheless, the promising results prove that our integrated methods (including Bi-integration and Tri-integration) achieve better performance than the single methods.

After the box plots figure, we realize that the result of CRF based sentence multi-emotion make a great fluctuation in the 5-fold cross-validation. Two probable points we summarized are the imbalance of emotion distribution and direct use of factor product. We implement 5-fold cross-validation by dividing data source into five parts randomly. Three parts for training, the rest two for validation and test respectively. The sentences from documents in Chinese emotion corpus (Ren-CECps) are collected manually by different annotators, so the documents in corpus differ in many kinds of linguistic styles. Some documents tell story, some documents describe the daily news, while some documents share writers's feeling. For this reason, there is a great distinction of the number of emotions and emotion categories between documents in various linguistic styles. On the other hand, we build the sentence emotion vector by factor product from the accumulation of word emotions of all emotional words in sentence. We didn't use any weight for emotion categories, just considered that each emotion category made the same contribution to the sentence emotions. In fact, as we know, sentence has polarity, namely positive or negative. If we take the polarity into account, the result may get further development.

7.2 Future Work

Predicting multiple emotions from the text is still a new and challenging task. In this study, we have investigated the local bag-of-words, context-sensitive information and global language features to solve this problem. In the future work, we will introduce some emotion-specific weight under the consideration of sentence polarity and explore some method to deal with the imbalance of emotion distribution. Moreover, We believe that the text emotion recognition is a deep-level language understanding, and the human emotions can be better predicted by relating the natural language processing techniques, the widely existed common sense knowledge, and the psychology studies in the future.

Bibliography

[1] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau. Sentiment analysis of twitter data. In *Proceedings of the Workshop on Languages in Social Media*, pages 30–38. Association for Computational Linguistics, 2011.

[2] J. Allan, J. G. Carbonell, G. Doddington, J. Yamron, and Y. Yang. Topic detection and tracking pilot study final report. 1998.

[3] S. Aman and S. Szpakowicz. Identifying expressions of emotion in text. In *Text, Speech and Dialogue*, pages 196–205. Springer, 2007.

[4] P. Bhowmick, A. Basu, and P. Mitra. Reader perspective emotion analysis in text through ensemble based multi-label classification framework. *Computer and Information Science*, 2(4):64–74, 2009.

[5] P. K. Bhowmick, A. Basu, and P. Mitra. Classifying emotion in news sentences: When machine classification meets human classification. *International Journal on Computer Science and Engineering*, 2(1):98–108, 2010.

[6] L. L. Bo Pang. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. pages 271–278, 2004.

[7] C. Burges. *Advances in kernel methods: support vector learning.* The MIT press, 1999.

[8] F. Chaumartin. Upar7: A knowledge-based system for headline sentiment tagging. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pages 422–425, 2007.

[9] P. Chesley, B. Vincent, L. Xu, and R. K. Srihari. Using verbs and adjectives to automatically classify blog sentiment. *Training*, 580(263):233, 2006.

[10] G. L. Clore, A. Ortony, and M. A. Foss. The psychological foundations of the affective lexicon. *Journal of personality and social psychology*, 53(4):751, 1987.

[11] D. Das and S. Bandyopadhyay. Word to sentence level emotion tagging for bengali blogs. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pages 149–152. Association for Computational Linguistics, 2009.

[12] D. Das and S. Bandyopadhyay. Extracting emotion topics from blog sentences: use of voting from multi-engine supervised classifiers. In *Proceedings of the 2nd international workshop on Search and mining user-generated contents*, pages 119–126. ACM, 2010.

[13] D. Das and S. Bandyopadhyay. Holder and topic based analysis of emotions on blog texts: A case study for bengali. In *Mining Social Networks and Security Informatics*, pages 127–143. Springer, 2013.

[14] S. Dasgupta and V. Ng. Mine the easy, classify the hard: a semi-supervised approach to automatic sentiment classification. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2*, pages 701–709. Association for Computational Linguistics, 2009.

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

[16] P. Ekman. An argument for basic emotions. *Cognition & Emotion*, 6(3-4):169–200, 1992.

[17] P. Ekman. All emotions are basic. *The nature of emotion: Fundamental questions*, pages 15–19, 1994.

[18] P. Ekman et al. *Universals and cultural differences in facial expressions of emotion.* University of Nebraska Press Lincoln, 1971.

[19] M. Gamon. Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis. In *Proceedings of the 20th international conference on Computational Linguistics*, page 841. Association for Computational Linguistics, 2004.

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

[21] M. Hu and B. Liu. Mining opinion features in customer reviews. In *Proceedings of the National Conference on Artificial Intelligence*, pages 755–760. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2004.

[22] X. Kang and F. Ren. Sampling latent emotions and topics in a hierarchical bayesian network. In *2011 International Conference on Natural Language Processing and Knowledge Engineering (NLP-KE)*, pages 37–42. IEEE, 2011.

[23] X. Kang, F. Ren, and Y. Wu. Bottom up: Exploring word emotions for chinese sentence chief sentiment classification. In *Natural Language Processing and Knowledge Engineering (NLP-KE), 2010 International Conference on*, pages 1–5. IEEE, 2010.

[24] R. F. Kang Xin. Predicting complex word emotions and topics through a hierarchical bayesian network. *China Communications*, 9(3):99, 2012.

[25] J. Lafferty, A. McCallum, and F. C. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. 2001.

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

[27] K. Lin, C. Yang, and H. Chen. Emotion classification of online news articles from the reader's perspective. In *Web Intelligence and Intelligent Agent Technology, 2008. WI-IAT'08. IEEE/WIC/ACM International Conference on*, volume 1, pages 220–226. IEEE, 2008.

[28] I. Maks and P. Vossen. A verb lexicon model for deep sentiment analysis and opinion mining applications. *ACL HLT 2011*, page 10, 2011.

[29] S. Matsumoto, H. Takamura, and M. Okumura. Sentiment classification using word sub-sequences and dependency sub-trees. *Advances in Knowledge Discovery and Data Mining*, pages 21–32, 2005.

[30] P. Melville, W. Gryc, and R. D. Lawrence. Sentiment analysis of blogs by combining lexical knowledge with text classification. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1275–1284. ACM, 2009.

[31] S. Menard. *Applied logistic regression analysis.* Number 106. Sage, 2002.

[32] G. Mishne. Experiments with mood classification in blog posts. In *Proceedings of ACM SIGIR 2005 Workshop on Stylistic Analysis of Text for Information Access*, page 19. Citeseer, 2005.

[33] G. Mishne and M. De Rijke. Capturing global mood levels using blog posts. In *AAAI 2006 Spring symposium on computational approaches to analysing weblogs*, pages 145–152, 2006.

[34] S. Mohammad and T. Yang. Tracking sentiment in mail: How genders differ on emotional axes. *ACL HLT 2011*, page 70, 2011.

[35] A. Neviarouskaya, H. Prendinger, and M. Ishizuka. Textual affect sensing for sociable and expressive online communication. *Affective Computing and Intelligent Interaction*, pages 218–229, 2007.

[36] V. Ng, S. Dasgupta, and S. Arifin. Examining the role of linguistic knowledge sources in the automatic identification and classification of reviews. In *Proceedings of the COLING/ACL on Main conference poster sessions*, pages 611–618. Association for Computational Linguistics, 2006.

[37] B. R. B. B. R. R. O'Connor and N. A. Smith. From tweets to polls: linking text sentiment to publid opinion time series. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*, pages 122–129. AAAI Press, 2010.

[38] N. O'Hare, M. Davy, A. Bermingham, P. Ferguson, P. Sheridan, C. Gurrin, and A. F. Smeaton. Topic-dependent sentiment analysis of financial blogs. In *Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*, pages 9–16. ACM, 2009.

[39] B. Pang and L. Lee. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2):1–135, 2008.

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

[41] R. W. Picard. Affective computing: challenges. *International Journal of Human-Computer Studies*, 59(1):55–64, 2003.

[42] C. Quan and F. Ren. Construction of a blog emotion corpus for chinese emotional expression analysis. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3-Volume 3*, pages 1446–1454. Association for Computational Linguistics, 2009.

[43] C. Quan and F. Ren. A blog emotion corpus for emotional expression analysis in chinese. *Computer Speech & Language*, 24(4):726–749, 2010.

[44] D. Ramage, D. Hall, R. Nallapati, and C. D. Manning. Labeled lda: A supervised topic model for credit attribution in multi-labeled corpora. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*, pages 248–256. Association for Computational Linguistics, 2009.

[45] D. Ramage, C. Manning, and S. Dumais. Partially labeled topic models for interpretable text mining. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 457–465. ACM, 2011.

[46] F. Ren. Affective information processing and recognizing human emotion. *Electronic notes in theoretical computer science*, 225:39–50, 2009.

[47] F. Ren and X. Kang. Employing hierarchical bayesian networks in simple and complex emotion topic analysis. *Computer Speech & Language*, 2012.

[48] S. Siersdorfer, S. Chelaru, W. Nejdl, and J. San Pedro. How useful are your comments?: analyzing and predicting youtube comments and comment ratings. In *Proceedings of the 19th international conference on World wide web*, pages 891–900. ACM, 2010.

[49] J. A. Suykens and J. Vandewalle. Least squares support vector machine classifiers. *Neural processing letters*, 9(3):293–300, 1999.

[50] M. Thelwall, K. Buckley, and G. Paltoglou. Sentiment in twitter events. *Journal of the American Society for Information Science and Technology*, 62(2):406–418, 2011.

[51] R. Tokuhisa, K. Inui, and Y. Matsumoto. Emotion classification using massive examples extracted from the web. In *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*, pages 881–888. Association for Computational Linguistics, 2008.

[52] G. Tsoumakas and I. Katakis. Multi-label classification. *International Journal of Data Warehousing & Mining*, 3(3):1–13, 2007.

[53] G. Tsoumakas and I. Katakis. Multi-label classification: An overview. *International Journal of Data Warehousing and Mining (IJDWM)*, 3(3):1–13, 2007.

[54] K. T. G. Tsoumakas, G. Kalliris, and I. Vlahavas. Multi-label classification of music into emotions. In *ISMIR 2008: Proceedings of the 9th International Conference of Music Information Retrieval*, page 325. Lulu. com, 2008.

[55] Y. Wu, K. Kita, K. Matsumoto, and X. Kang. A joint prediction model for multiple emotions analysis in sentences. In *Computational Linguistics and Intelligent Text Processing*, pages 149–160. Springer, 2013.

[56] Y. Wu, K. Kita, F. Ren, K. Matsumoto, and X. Kang. Exploring emotional words for chinese document chief emotion analysis. In *PACLIC*, pages 597–606, 2011.

[57] C. Yang, K. Lin, and H. Chen. Emotion classification using web blog corpora. In *Web Intelligence, IEEE/WIC/ACM International Conference on*, pages 275–278. Ieee, 2007.

[58] M.-L. Zhang and Z.-H. Zhou. Ml-knn: A lazy learning approach to multi-label learning. *Pattern Recognition*, 40(7):2038–2048, 2007.