The following publication Z. Wang, S. Y. M. Lee, S. Li and G. Zhou, "Emotion Analysis in Code-Switching Text With Joint Factor Graph Model," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 25, no. 3, pp. 469-480, March 2017 is available at https://dx.doi.org/10.1109/TASLP.2016.2637280.

Emotion Analysis in Code-Switching Text with Joint Factor Graph Model

Zhongqing Wang, Sophia Yat Mei Lee, Shoushan Li, and Guodong Zhou

Abstract—**Previous research on emotions analysis have placed much emphasis in monolingual instead of bilingual text. However, emotions on social media platforms are often found in bilingual or code-switching posts. Different from monolingual text, emotions in code-switching text can be expressed in both monolingual or bilingual forms. Moreover, more than one emotion can be expressed within a single post; yet they tend to be related in some ways which offers some implications. It is thus necessary to consider the correlation between different emotions. In this paper, a joint factor graph model is proposed to address this issue. In particular, attribute functions of the factor graph model are utilized to learn both monolingual and bilingual information from each post, factor functions are used to explore the relationship among different emotions, and a belief propagation algorithm is employed to learn and predict the model. Empirical studies demonstrate the importance of emotion analysis in code-switching text and the effectiveness of our proposed joint learning model.**

Index Terms—**Emotion Analysis, Code-switching, Bilingual Information, Factor Graph Model**

I. Introduction

Due to the popularity of opinion-rich resources (e.g. online review sites, forums, and micro-blog websites), emotion analysis in text has become highly significant in obtaining useful information for studies on social media (Pang et al., 2002; Liu et al., 2013; Lee et al., 2014). Previous research has focused solely on analyzing emotions in monolingual text (Chen et al., 2010; Lee et al., 2013a). However, code-switching posts are common found in social media. Emotions expressed in these posts are presented in both monolingual and bilingual ways. [E1-E3] are three examples of code-switching posts on Weibo.com that contain both Chinese and English words. While [E1] expresses the *happiness* emotion in English and [E2] expresses the *sadness* emotion in both Chinese and English, the *sadness* emotion in [E3] is expressed through a mixed Chinese-English phrase (hold 不住, ‘*cannot take it*’).

From [E1-E3], we can see that it is much more difficult to detect emotions in the code-switching text than in the monolingual ones, since emotions in code-switching posts can be expressed in either one (i.e. [E1]) or two languages (i.e. [E2]). Consider [E2] where the English Text ‘*I feel so lonely*’ conveys the emotion of *sadness* which is indeed triggered by the previous text in Chinese. Moreover, some emotions are expressed in two languages simultaneously, for example, the fixed expression “hold 不住” in [E3] carries the emotion of *sadness*. In addition, our statistics show that 14.2% of posts contain more than one emotion. For example, in [E4] Ken feels *happy* to hear that other people are calling his son ‘mini ken’, while his wife (the blogger) feels *sad* that their baby is less and less like her. The two emotions identified in the post are closely related. It is not uncommon to find posts with contrasting emotions or an emotion triggering another emotion.

[E1] 玩了一下午轮滑 **so happy**！
(*I went rollerblading the whole afternoon,* ***so happy****!*)

[E2] 这个七夕，麻麻出去打麻将了，外面轰隆隆打着雷，刚刚看完过去很久的《分手合约》。真的很伤感。**Today, I feel so lonely.**
(Mum went out playing Mahjong on Chinese Valentine’s Day. It is thundering outside, and I just finished watching <A Wedding Invitation>. It’s really sad. ***Today, I feel so lonely.****)*

[E3] 求月假快点到来吧 我真的 **hold 不住**了啊！！
(*Wish the monthly holiday would come sooner, I* ***can’t take it anymore****.*)

[E4] 每个见到小魔王的人都说他是 **mini ken**，搞得 **ken** 每次都很 **happy**。自从不喝我奶之后，越来越不像我了。
(*Every time when people meet the little monster, they’d call him ‘****mini ken****’ and* ***Ken*** *would be so* ***happy*** *to hear about that. Ever since he stopped drinking my milk, he is less like me.*)

In code-switching posts, emotions could be expressed in more than one languages. Different emotions within a single post can be closely related to each other. Therefore, the major challenges of emotion prediction in code-switching texts are to explore bilingual information in each post and correlate dif-

This work was partly done during Zhongqing Wang’s visit at The Hong Kong Polytechnic University. This work is supported by an Early Career Scheme project (ECS) sponsored by the Research Grants Council of Hong Kong (No. PolyU 5593/13H) and a PolyU Faculty Research Grant (No. 1-ZVEK), and supported by the Key Project of the National Natural Science Foundation of China (No. 61331011) and the National Natural Science Foundation of China (No. 61273320, and No. 61375073).

Zhongqing Wang, Shoushan Li, and Guodong Zhou are with the School of Computer Science and Technology, Soochow University, Suzhou, China (e-mail: wangzq.antony@gmail.com; lishoushan@suda.edu.cn; gdzhou@suda.edu.cn).

Sophia Yat Mei Lee is with the Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong SAR, China (e-mail: ymlee@polyu.edu.hk).

Guodong Zhou is the corresponding author of our paper.

ferent emotions within a single post.

A more straightforward method to explore bilingual information is to transform them into monolingual text via machine translation. The problem is that since text in social media is relatively informal, additional work is required to improve the performance of the translation model and language model. Moreover, exploring relationships among emotions in a single post and learning both bilingual and emotional information collectively also pose extra challenges to emotion detection.

In this paper, we propose a uniform framework to predict emotions in code-switching text by considering both bilingual and emotional information. In particular, factor functions are proposed to connect different emotions within a single post. The attribute functions are proposed to address bilingual text. Lastly, a joint factor graph model is proposed to incorporate both factor functions and attribute functions, and the belief propagation algorithm is utilized to learn the model collectively. Evaluation of the dataset from *Weibo.com* indicates the importance of emotion analysis in code-switching text and the effectiveness of our proposed approach.

The remaining parts of the paper are organized as follows. In Section 2, we give an overview of related work. In Section 3, we introduce the corpus and present some statistics of the data. In Section 4, we propose the factor graph model to predict emotions with both bilingual and emotional information. Section 5 reports the details of our experimental results. Finally, we conclude and discuss some future work in Section 6.

II. RELATED WORK

In this section, we discuss important related work on emotion analysis, analysis on code-switching text, and the factor graph model.

A. Emotion Analysis

Our study is related to sentiment analysis in general. Previous studies can mainly be classified into either document-level or feature-level sentiment analysis. While coarse-grained document-level sentiment classification aims to predict the sentiment orientation (positive/negative) of a review (Turney, 2002; Li et al., 2013) and dominates the literature, fine-grained feature-based sentiment analysis aims to extract objective attributes, such as holder and target from a review (Hu & Liu, 2004; Li et al., 2012).

Subsequently, a great deal of work has taken other important features into account. For instance, Pang and Lee (2004) examined the relation between subjectivity and polarity classification; Kennedy and Inkpen (2006) identified three types of valence shifters: negations, intensifiers, and diminishers; and Wilson et al. (2009) explored the difference between prior and contextual polarity and recognized the importance of identifying neutral instances. Some studies deal with scenarios for sentiment classification that are more complicated, such as cross-domain adaptation (Blitzer et al., 2007; Li et al., 2013) and semi-supervised learning (Dasgupta & Ng 2009; Li et al., 2010).

The polarity-driven approach in sentiment analysis is often criticized for being too general to satisfy some real-world applications, such as product design. Thus, researchers began to explore more fine-grained affective information, such as emotions (e.g. happiness and sadness) (Wiebe et al., 2005; Mihalcea & Liu, 2006). The earliest research on emotion has focused on the representation and processing of emotion in facial expressions and body language (Andrew, 1963; Ekman & Friesen, 1978). More recently, there have been mounting researches on the neurobiological basis of emotion (Olson et al., 2007; Hervé et al., 2012) and how emotion is linked with other aspects of human cognition (Smith & Lazarus, 1993; Smith & Kirby, 2001; Bridge et al., 2010). Furthermore, emotion in text has also been analyzed recently; however, most of the work focused on analyzing emotions in monolingual text. Some of these studies focused on lexicon building. For example, Rao et al. (2012) automatically built a word-emotion mapping dictionary for social emotion detection and Yang et al. (2014) proposed a novel emotion-aware topic model to build a domain-specific lexicon. Further on, emotion classification is one of the important tasks in emotion analysis. For example, Liu et al. (2013) used a co-training framework to infer the news from readers' comments and writers' emotions collectively; Wen and Wan (2014) used class sequential rules for emotion classification of micro-blog texts by regarding each post as a data sequence; and Li et al. (2015) proposed a factor graph based framework to incorporate both label and context dependency for emotion classification.

Although most previous researches focused on sentiment or emotion analysis separately, there are some recent studies concerned the interaction between sentiment and emotion. For example, Gao et al. (2013) proposed an ensemble learning based framework to learn the emotion and sentiment collectively, while Wang et al. (2015) proposed an integer linear programming based approach to learn these two types of information more directly. Moreover, both emotion and sentiment can be used to help each other. For example, Luo et al. (2015) propose an EM based probabilistic model to incorporate emotion information into sentiment prediction process.

Despite a growing number of research on sentiment and emotion, little has been done on the analysis of emotion in code-switching contexts due to the complications arising from the difficulty of processing two languages simultaneously.

B. Research on Code-switching and Bilingual Text

Research on code-switching could be traced back to the 1970s. Several theories have been proposed to account for the motivations behind code-switching, such as diglossia (Blom & Gumperz, 1972), communication accommodation theory (Giles & Clair, 1979), the markedness model (Myers-Scotton, 1993), and the conversational analysis model (Auer, 1984).

Code-switched documents have also received considerable attention in the NLP community (Adel et al., 2015; Garrette et al., 2015). Several studies have focused on identification and analysis, including mining translations in code-switched documents (Ling et al., 2013), predicting code-switched points (Solorio & Liu, 2008), identifying code-switched tokens (Lignos & Marcus, 2013), adding code-switched support to language models (Li & Fung, 2012), and developing POS tagging for code-switching text (Jamatia et al., 2015).

Arelated research topic, multilingual natural language processing, has begun to draw more and more attention in the computational linguistic community due to its broad real-world applications. Relevant studies have been reported in different natural language processing tasks, such as parsing (Burkett et al., 2010), information retrieval (Gao et al., 2009), text classification (Amini et al., 2010), and so on. There are a number of studies on predicting polarity through multilingual text. Wan (2008) proposed an ensemble method to combine one classifier trained with labeled data from the source language and the other classifier trained with their translated data. Subsequently, Wan (2009) incorporated the unlabeled data in the target language into the same classification method with co-training to improve the classification performance. Wei and Pal (2010) regarded cross-lingual sentiment classification as a domain adaptation task and applied structural correspondence learning (SCL) to tackle this problem. Their approach achieves a better performance than the co-training algorithm. More recently, Lu et al. (2011) performed cross-lingual sentiment classification from a different perspective. Instead of using machine translation engines, they used a parallel corpus to help perform semi-supervised learning in both English and Chinese sentence-level sentiment classifications. Similar to Lu et al. (2011), Meng et al. (2012) also employed parallel corpus for cross-lingual sentiment classification. They also explored the case when no labeled data is available in the parallel corpus.

The research of emotions has also been linked to the field of bilingualism. Previous studies have demonstrated that emotion is closely related to second language learning and use (Arnold, 1999; Schumann, 1999), as well as bilingual performance and language choice (Schrauf, 2000; Pavlenko, 2008). For example, there are a number of factors that may influence the use of emotion vocabulary, such as sociocultural competence, gender, and topic (Dewaele & Palvenko, 2002).

A few works focus on sentiment analysis in code-switching text (Vilares et al., 2015; Sitaram et al., 2015). However, these studies always treated mixed code-switching text as a single document for classification, which is essentially the same as monolingual text. None of these studies used a joint model to learn the monolingual and bilingual information individually and collectively. The present study is the first attempt in analyzing emotion in Chinese-English code-switching text.

C. *Factor Graph Model*

The Factor Graph Model (FGM) is one kind of probabilistic graph model which has been proven effective in representing and optimizing the graph structure (Tang et al., 2011; Zhuang et al., 2012). Most of the work on the factor graph model focuses on social network analysis. For example, Tang et al. (2011a) and Zhuang et al. (2012) formalized the problem of social relationship learning into a semi-supervised framework and proposed the Partially-labeled Pairwise Factor Graph Model (PLP-FGM) for learning to infer the type of social ties. Huang et al. (2014) employed a factor graph model to discover triadic closure patterns in dynamic networks with both network topology and demography information. Zhang et al. (2015) proposed a factor graph based on connecting heterogeneous networks to identify users from multiple heterogeneous social networks by considering both local and global consistency. Apart from that there are also some text mining tasks using a factor graph model. For example, Yang et al. (2011b) generated summaries by modeling tweets and social contexts into a dual wing factor graph (DWFG), which utilized the mutual reinforcement between Web documents and their associated social contexts. Wang et al. (2014) predicted personal skills on a profile text using a factor graph model with both personal and skill connections.

The Markov Random Walk model is another type of graph model which also studies the relations between documents. As its name suggests, random walk chooses a certain vertex in the graph for the first step and then randomly migrates through the edges. There are numerous studies using a random walk algorithm in natural language processing tasks. For example, Hassan and Radev (2010) used a random walk model in a large word relatedness graph and estimated polarity for any given word. Li et al. (2012) built a context-aware relation graph of candidate tags and proposed a random walk algorithm to rank tags for suggestion. Li et al. (2016) proposed a discriminative deep random walk model to learn the latent space representations for capturing topological structure, which then was used for social network classification.

However, none of the related works have used graph model to predict emotions and explore the bilingual information in code-switching text.

III. Data Collection and Annotation

In this section, we introduce the annotation scheme and show some general data distribution.

A. *Annotation Scheme*

We retrieve our dataset from *Weibo.com*, one of the famous SNS (Social Network Service) websites in China. Code-switching posts are identified by employing encoding code for each character in the post. After removing posts that contain noise and advertisements, we extract 4,195 code-switching posts for emotion annotation.

[E5] and [E6] are examples of removed noise posts. [E5] does not convey any emotion and [E6] is an advertisement of the cosmetic brand ‘Benefit’. To remove these noise posts, we first annotated some noise post data manually, which was followed by the training of a noise detector to detect noise posts automatically (Li et al., 2014). Besides, we also remove short (less than 10 words) and redundancy posts.

[E5] 我参与了“你最喜欢的明星团体投票”，投给了“**SHINee**”这个选项。
(*I voted for “**SHINee**” in “Your Favorite Band Election”*)

[E6] **benefit** 贝玲妃 ~ 泡沫洁面膏 ~ 推荐的洗面奶
(***Benefit** Foam Cleanser is the cleansing milk I recommend.*)

Following Lee et al. (2013b), five basic emotions are annotated in each post, namely *happiness*, *sadness*, *fear*, *anger*, and *surprise*. Since emotions can be expressed through the two languages separately or collectively, we need to annotate four kinds of causal situations for each emotion, i.e. *None*, *English*, *Chinese*, and *Both*. The descriptions of these situations are illustrated below:

- *None* means the post does not contain any corresponding emotions. [E7] is an example in which no emotion is found.
- *Chinese* means the emotion in the post is exclusively expressed through Chinese text. As Weibo.com is a Chinese SNS Website, Chinese is the matrix language on this website, meaning that most posts express emotions through Chinese. [E8] serves as an example. The emotion of *happiness* is expressed through Chinese.
- *English* means the emotion in the post is exclusively expressed through English text. As English is the embedded language, fewer English words are found to express emotions in the posts. [E1] is an example which expresses the *happiness* emotion through English.
- *Both* means the emotions of the posts are expressed through both Chinese and English texts. For example, the *sadness* emotion in [E2] is expressed through both Chinese and English texts. Note that although [E9] contains more than one emotion: the *happiness* emotion is expressed through English and the *surprise* emotion is expressed through Chinese, we still need to annotate *happiness-en* and *surprise-cn*, respectively, since the causal situations are annotated for each emotion separately.

[E7] 歌剧院旁边很多 **coffee house**。

(*There are lots of* ***coffee houses*** *beside the opera*.)

[E8] 各位，**good morning**。我真的发现还是早上能睡懒觉的人最幸福 。

(*Everybody,* ***good morning****. I really think that the most fortunate people are those who can have a lie-in in the mornings*.)

[E9] 虽然你今天很 **warm heart**，但是昨天跟我说的话，真的让人很震撼。

(*Although you made me feel warm today, the words you said yesterday really shook me*.)

In the above descriptions, we find that the emotion can be expressed in an either monolingual (i.e. *Chinese* and *English*) or bilingual (i.e. *Both*) form. Hence, it is necessary to learn the prediction model with monolingual and bilingual ways collectively. Moreover, although some of the emotions are expressed explicitly (i.e. [E1]), which can be detected by a lexical based approach, many emotions are expressed implicitly. For example, [E10] expresses the *sadness* emotion with irony and [E11] expresses the *happiness* emotion without any cue words. Therefore, word representation with bilingual information is needed for capturing these implicit emotions in text. On the other hand, more than one emotion could be expressed in a post. For example, there are two emotions in [E8], the *happiness* emotion is expressed in English, while, the *surprise* emotion is expressed in Chinese. It is thus necessary to consider the emotional relations between different emotions.

[E10] 昨晚一夜没睡，早起直接飙酒，喝多上车回校，回校一睁眼过站，多么 **happy** 的一天。

(*I didn't sleep for the whole night yesterday, drank in the morning. I got drunk and went back to school by bus, and I missed my stop. Such a* ***happy*** *day*.)

[E11] 机缘巧合之下去 **muji** 买了点东西然后来才醒悟过来我这次论文的题目就是关于 **muji** 的。这叫什么？这就是 **fate** 呐！

(*Went to* ***Muji*** *by chance to get something and realized later that the topic of my paper this time is about* ***Muji****. What a coincidence! That is* ***fate****!*)

B. General Data Distribution

Out of 4,195 annotated posts, 2,312 posts are found to express emotions. 81.4% of emotional posts are expressed through Chinese text. Although English text contains relatively fewer words in each post, 43.5% of emotional posts are expressed through English (Lee and Wang, 2015). This tendency indicates that English is of vital importance in emotion expression even in code-switching contexts dominated by Chinese. More notably, there are overlaps between Chinese and English emotional posts since some emotional posts are posted in both Chinese and English.

TABLE 1
RESULTS OF AGREEMENT ANALYSIS

	Kappa score
Emotion	0.692
Caused Language	0.767

To verify the quality of the annotation, two human annotators were asked to annotate 1,000 posts. the inter-annotator agreement was calculated using Cohen's Kappa coefficient. *Caused Language* means that the language causes the emotion in the post. For example, in [E1] the caused language is English and in [E2] the caused languages are both Chinese and English. Table 1 shows the results of the agreement analysis. We find that the agreement is high, indicating that the quality of the annotation and scheme is effective. In addition, the agreement of the emotion annotation is lower than that of the caused language which is probably due to the fact that some posts express more than one emotion and the implicit emotion in some posts is hard to annotate.

IV. JOINT FACTOR GRAPH MODEL WITH BILINGUAL AND EMOTIONAL INFORMATION

Our goal is to predict emotions for each post. Formally, for a post p with the emotion e, we need an objective integer variable to define if the emotion e is expressed in the post as shown below:

$$Z_{e,p} \in \{0,1\} \tag{1}$$

If $Z_{e,p} = 1$, the post p expresses the emotion e and vice versa. Hence, we have five objective variables $Z_{e,p}$ for a post p with the five emotions (i.e. *happiness*, *sadness*, *fear*, *anger*, and *surprise*). For a post p with the corresponding emotion e, each objective variable $Z_{e,p}$ is associated with an attribute vector x_i and a label y_i ($y_i = Z_{e,p}$) indicating if the emotion e is expressed in the post p.

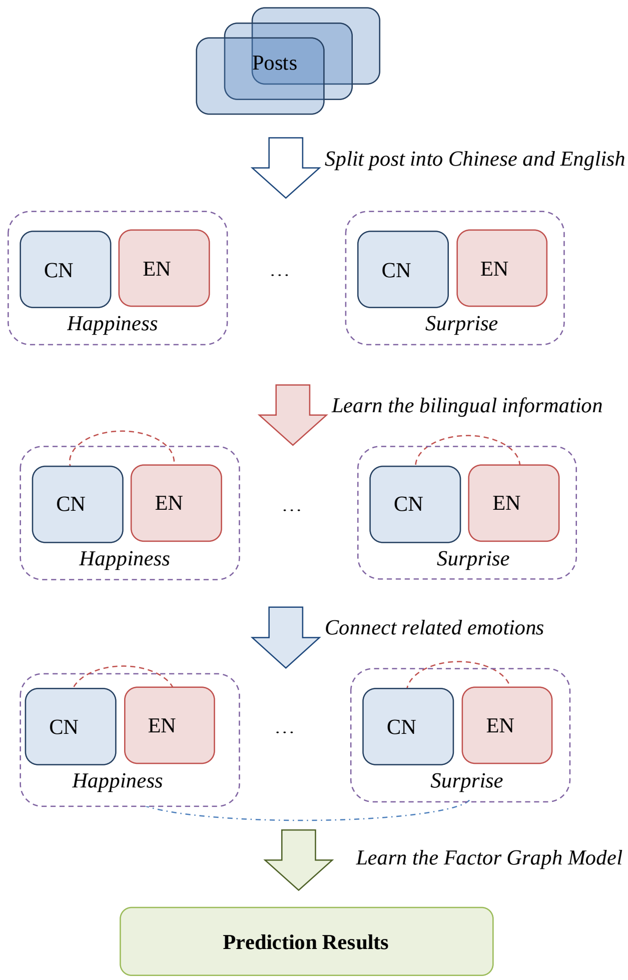


Fig.1. Overview of our framework

For inferring the objective variables $Z_{e,p}$ of each post p, a straightforward approach is to predict emotions of each post individually. However, the drawback of this approach is that it ignores both bilingual and emotional information. Thus, we need to explore these pieces of information for each post and learn the model collectively. In this study, we propose a joint factor graph model to incorporate both bilingual and emotional information for predicting emotions.

The overview of the proposed factor graph model is demonstrated in Fig.1. Firstly, we use a machine translation based approach to extract the bilingual information from the code-switching text, then we connect the related emotion in a single post. Finally, we use a factor graph model (Tang et al., 2012) and a belief propagation algorithm (Huang et al., 2014) to incorporate both bilingual and emotional information in the learning stage as soft constraints. In our framework, the bilingual information of each post is treated as the attribute function, the emotional connections between emotions are treated as the factor functions, the factor graph model is used to incorporate both the attribute functions and factor functions, and the belief propagation algorithm is used to learn and predict the model. Thus, the core issues of our approach are as follows: 1) exploring bilingual information in the code-switching text; 2) connecting the correlated emotions of each post by considering emotional information; 3) and using these two kinds of information to build the factor graph model. We will introduce these issues one by one in the following subsections.

A. Bilingual Information

The documents with both monolingual and bilingual information can be considered as a bipartite graph represented in Fig.2. In a bipartite graph, the nodes consist of two parts: documents and all terms extracted from the documents. An undirected edge (d_i, w_k) exists if and only if the document d_i contains the term w_k.

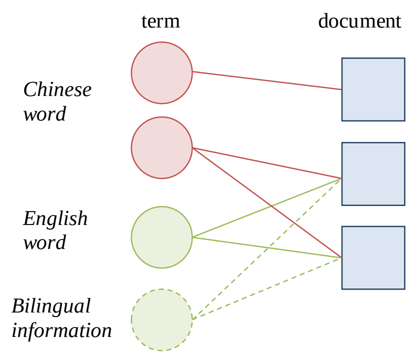


Fig. 2. Example of the bipartite graph

Note that there are three kinds of terms in the graph, i.e. Chinese words, English words, and bilingual information. Although Chinese words and English words cannot be connected directly, the label information between them would be propagated through the bipartite graph through word-document relations and bilingual information.

For using bilingual information, a word-by-word statistical machine translation strategy is adopted to translate English words into Chinese. For each word, only the most confident translations would be adopted. Although the word-by-word translation model is simple and traditional, we consider it meets our requirements. The numbers of English words in each post are limited, and the context of English words are always made up by Chinese words; therefore, a more powerful model, such as phrase-based translation or neural translation which would both need context for translation would not be effective for this task. For better clarity, a word-based decoding adopting a log-linear framework (see Och & Ney, 2002) with translation and language model being the only features, is used:

$$P(c \mid e) = \frac{\exp\left[\sum_{i=1}^{2} \lambda_i h_i(c,e)\right]}{\sum_c \exp\left[\sum_{i=1}^{2} \lambda_i h_i(c,e)\right]} \tag{2}$$

where

$$h_1(c,e) = \log(p_\gamma(c \mid e)) \tag{3}$$

is the translation model which is converted from the bilingual lexicon[1] and

$$h_2(c,e) = \log(p_{\theta_{LM}}(c)\, p_{\theta_{SYN}}(c)\, p_{\theta_{PMI}}(c)) \tag{4}$$

is the language model, and $p_{\theta_{LM}}(c)$ is the bigram language model which is trained from a large scale Weibo dataset[2]. As text in microblogs is informal, a synonym dictionary[3] as well as PMI based word correlation are used to enhance the language model for machine translation. $p_{\theta_{SYN}}(c)$ denotes the synonym similarity between translated words and their contexts; if the translated word and the context c is synonymous, $p_{\theta_{SYN}}(c) = 1$, other $p_{\theta_{SYN}}(c) = 0$. This is necessary since the sense of the translated words and their contexts are expected to be similar. $p_{\theta_{PMI}}(c)$ represents the PMI similarity between translated words and their context; while the PMI score is calculated by the individual and co-occurred hit count between translated words and context from the search engine[4] (Turney, 2002). This ensures that the translated words are closely associated with their context.

The candidate target sentences made up of a sequence of the optional target words are ranked by the language model. The output will be generated only if it reaches the maximum probability as in the following (Brown et al., 1990; Zhao et al., 2009):

$$c = \arg\max \prod p(w_c) \tag{5}$$

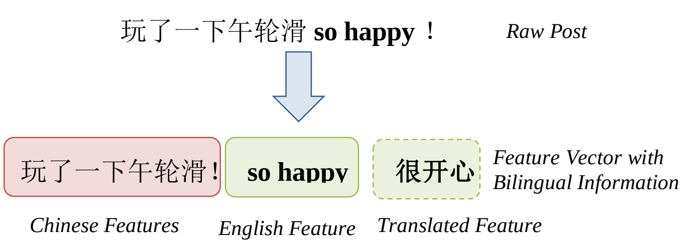


Fig. 3. Example of bilingual information

The example for constructing bilingual information is illustrated in Fig.3. To build a feature vector, the raw code-switching post is divided into three parts: Chinese features, English features, and translated features. The translated features are translated by the above machine translation based approach. All three types of features are merged together to build a uniform feature vector for constructing the attribute functions in the joint factor graph model.

B. Factor Graph Model Construction

After we explore the bilingual information in each post, we use a factor graph model to predict emotions with both bilingual and emotional information. At the same time, we connect different emotions in a single post by considering the emotional relations.

Formally, given a network G, for a post p with the corresponding emotion e, each objective variable $Z_{e,p}$ is associated with an attribute vector x_i and a label y_i ($y_i = Z_{e,p}$) indicating if the emotion e is expressed in the post. If $X = \{x_i\}$ and $Y = \{y_i\}$, we have the following formula:

$$P(Y \mid X,G) = \frac{P(X,G \mid Y)P(Y)}{P(X,G)} \tag{7}$$

This probabilistic formula indicates that labels depend on both local attributes associated with the overall structure of the network. According to Bayes' rule, we have

$$P(Y \mid X,G) \propto P(X \mid Y)P(Y \mid G) \tag{8}$$

where $P(Y \mid G)$ represents the probability distribution of labels given the overall network structure and $P(X \mid Y)$ denotes the probability distribution of generating attributes X associated with their labels Y. For simplicity's sake, we assume that the probability distribution of attributes given the label of each edge is conditionally independent and we can have

$$P(Y \mid X,G) \propto P(Y \mid G)\prod_i P(x_i \mid y_i) \tag{9}$$

where $P(x_i \mid y_i)$ is the probability of the generated attribute x_i given the label y_i. As a result, the issue to be solved is how to instantiate $P(Y \mid G)$ and $P(x_i \mid y_i)$. In this study, we model them using a Markov random field. Consequently, $P(Y \mid G)$ and $P(x_i \mid y_i)$ can be instantiated as follows (Hammersley and Clifford, 1971):

$$P(x_i \mid y_i) = \frac{1}{N_1}\exp\left\{\sum_{j=1}^{d} \alpha_j f_j(x_{ij}, y_i)\right\} \tag{10}$$

$$P(Y \mid G) = \frac{1}{N_2}\exp\left\{\sum_i \sum_{j \in NB(i)} g(i,j)\right\} \tag{11}$$

where N_1 and N_2 are normalization factors. Eq. (10) indicates that we define an attribute function $f_j(x_{ij}, y_i)$ for each post associated with the emotion. Eq. (11) represents the set of correlation factor functions $g(i,j)$ over each pair (i,j) in the network. $NB(i)$ denotes the set of neighbor nodes of i with the emotional connection in the same post. Next, we will briefly introduce possible ways to define the attribute functions $\{f(x_{ij}, y_i)\}_j$ and factor function $g(i,j)$.

Attribute functions $\{f(x_{ij}, y_i)\}_j$: This denotes the attribute

[1] *MDBG CC-CEDICT* is adopted as the bilingual lexicon: http://www.mdbg.net/chindict/chindict.php?page=cedict

[2] The large-scale *Weibo* dataset contains 2,716,197 posts in total.

[3] *TongYiCiLin (同义词林)* is adopted as the Chinese synonym dictionary: http://www.ltp-cloud.com/

[4] We use *BING* as the search engine: http://www.bing.com/

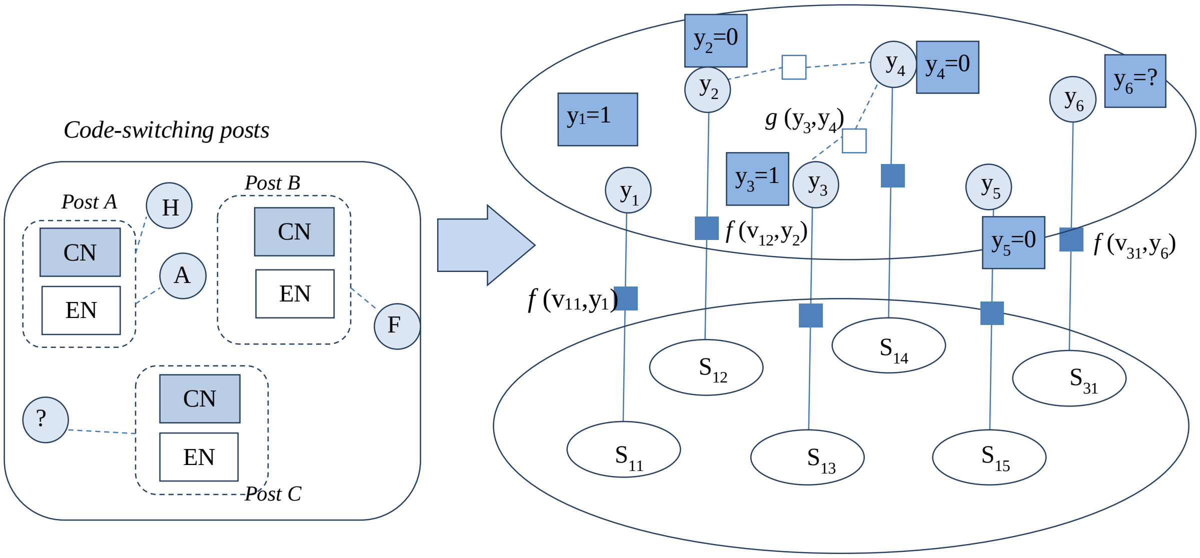


Fig 4: Graph Representation of the Factor Graph Model

value associated with each post i by each emotion e. We define the local textual attribute as a feature (Lafferty et al., 2001). We can accumulate all the attribute functions and obtain local entropy for a text:

$$\frac{1}{N_1}\exp\left(\sum_i\sum_k \alpha_k f_k\left(x_{ik}, y_i\right)\right) \tag{12}$$

Here α_j is the weight of the j^{th} attribute. We use both monolingual and bilingual information of each post to build the attribute function with each emotion. The bilingual information is generated by the word-by-word statistical machine translation strategy which has been mentioned in Subsection 4.A.

Factor function $g(y_i, y_j)$: We define the factor function through the pairwise network structure. That is, if nodes i and j have emotional connection, a factor function for this connection is defined as:

$$g\left(y_i, y_j\right) = \exp\left\{\beta_{ij}\left(y_i - y_j\right)^2\right\} \tag{13}$$

In particular, emotional connections link two nodes (i, j) in the same post p (e.g. $Z_{e_i,p}$ and $Z_{e_j,p}$) since we consider the emotions in the same post being related and affected by each other. In addition, β_k is the weight of the k^{th} factor function.

To better understand our model, an example is given in Fig. 4. In this example, there are six nodes from three posts, one node is used to detect one type of emotion for a post. For example, S_{11} is used to detect the *happiness* emotion of post i. Among them, five nodes are labeled (two are labeled with the category "1", i.e. $y = 1$, the others are labeled with the category "0", i.e. $y = 0$), and one node is unlabeled (they are represented by y=?). We have six attribute functions. For example, $f(v_{11}, y_i)$ denotes the set of local textual attribute functions of y_i. We also have two pairwise relationships (e.g. (y_2, y_4), (y_3, y_4)) based on the structure of the input network. For example, $g(y_3, y_4)$ denotes emotional connection between y_3 and y_4, while they share the emotional relationship in the left figure since they are from the same post.

C. *Learning and Prediction*

Given the factor graph model as constructed above, we use a Loopy Belief Propagation algorithm to learn the model and perform the prediction.

Learning the factor model is to find the best configuration for free parameters $\theta = (\{\alpha\}, \{\beta\})$ that maximizes the log likelihood objective function $L(\theta)$.

$$\theta^* = \arg\max L(\theta) \tag{14}$$

To solve the objective function, we adopt a gradient descent method. We use β (the weight of the factor function $g(y_i, y_j)$) as the example to explain how we learn the parameters (the algorithm also applies to tune α by simply replacing β with α). Specifically, we first write the gradient of each β_k with regard to the objective function (Eq. 14):

$$\frac{L(\theta)}{\beta_k} = E\left[g\left(i, j\right)\right] + E_{P_{\beta k}(Y|X,G)}\left[g\left(i, j\right)\right] \tag{15}$$

where $E[g(i, j)]$ is the expectation of factor function $g(i, j)$ given the data distribution and $E_{P_{\beta k}(Y|X,G)}[g(i, j)]$ is the expectation of factor function $g(i, j)$ under the distribution $P_{\beta k}(Y \mid X, G)$ given by the estimated model.

We approximate the marginal distribution $E_{P_{\beta k}(Y|X,G)}[g(i, j)]$ using LBP (Loopy Belief Propagation) (Tang et al., 2011). With the marginal probabilities the gradient can be obtained by summing overall triads. In this way, the algorithm essentially performs a transfer learning over the complete network. Finally, with the obtained gradient we update each parameter with a learning rate η. After the learning is complete, all unknown nodes are assigned with labels that maximize marginal probabilities. The learning algorithm is summarized in Fig. 5.

Input: Network G, Learning rate η
Output: Estimated parameters θ
Initialize $\theta \leftarrow 0$
Repeat

1) Perform LBP to calculate the marginal distribution of unknown variables, i.e., $P(y_i \mid x_i, G)$
2) Perform LBP to calculate the marginal distribution of each variable, i.e., $P(y_i, y_j \mid X_{(i,j)}, G)$
3) Calculate the gradient of β_k according to Eq. 15 (for a with a similar formula)
4) Update parameter θ with the learning rate η

$$\theta_{\text{new}} = \theta_{\text{old}} + \eta \frac{L(\theta)}{\theta}$$

Until *Convergence*

Fig.5. Learning Algorithm for the Joint Factor Graph Model

V. Experiments

In this section, we first show some statistics on the dataset and then introduce the experimental settings. Finally, we evaluate the performance of our proposed approach for predicting emotions in code-switching text.

A. Statistics and Analysis

In this subsection, we provide some statistics and analysis for illustrating the distributions between emotions and caused languages.

1) Joint Distribution of Emotions and Caused Languages

To understand the distribution of emotions and caused languages, we first calculate the joint distribution between emotions and caused languages, as in Fig. 6. The Y-axis of the figure presents the conditional probability $p(e_i \mid l_j)$ of a post expressing emotion e_i given that the caused language l_j.

Fig. 6 suggests that: 1) *happiness* occurs more frequently than other emotions; 2) people tend to use English to express *happiness* more than *sadness*; 3) the general distribution of emotions expressed through Chinese and English texts are similar; and 4) *fear* and *surprise* occur less frequently in English text than the other three emotions.

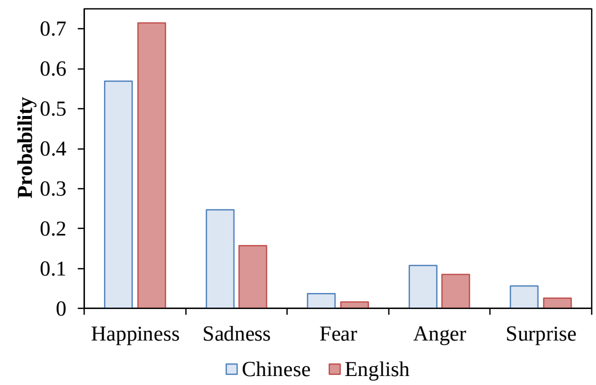


Fig. 6. Joint Distribution of Emotions and Caused Languages

2) Word Count Analysis

Table 2 shows the statistics on the average number of words in each language in a post. We notice that as our data all written by Chinese individuals, it is not surprising to find that the average number of Chinese words in a post is higher than English Moreover, the emotions expressed in English are mostly single words, e.g. '*happy*', '*high*', and '*surprise*'. Note that, as mentioned above, although the average number of Chinese words is higher than the English words, English is of vital importance for emotion expressions even in Chinese-dominated code-switching texts.

TABLE 2
STATISTICS ON AVERAGE NO. OF WORDS IN A POST

	#avg. word
Chinese	19.8
English	2.9

TABLE 3
THE LANGUAGE CO-EXPRESSED PROBABILITY

Emotion	Co-occurrence Probability
Happiness	0.159
Sadness	0.143
Fear	0.085
Anger	0.253
Surprise	0.080

3) Languages Co-expressed Distribution

Table 3 indicates the conditional probabilities of a post expressing emotion *e* by both Chinese and English texts. It is noticed that although there are a number of posts expressing emotions through Chinese or English separately, there are many posts expressing emotions via both Chinese and English texts collectively. Translation (e.g. [E12]), interjection (e.g. [E13]), and irony (e.g. [E9]) are the underlying reasons for emotions appearing in both Chinese and English in a single post.

[E12] stream flow, slowly away a few leaf, also taking the memory.溪水缓慢地流动着，带走了几片落叶，也带走了记忆。

(The Chinese text is translated by the blogger already)

[E13] 在家喝红酒，想加点冰，却发现冰都用完了。**sigh**~~（*Wanted to add some ice cubes when I was drinking red wine at home, but they have run out.* ***Sigh~~***）

4) Emotions Co-occurrence Distribution

We examined the conditional probabilities of a post expressing emotion e_i given that the post contains emotion e_j. The conditional probabilities are shown in Table 4. The direction of the table is from left to right, for example $p(sadness \mid hapiness)$ is in row 2 and column 3, and the value is 0.060.

TABLE 4
EMOTION CO-OCCURRENCE DISTRIBUTION

	Happiness	Sadness	Fear	Anger	Surprise
Happiness	-	0.060	0.016	0.025	0.019
Sadness	0.088	-	0.023	0.033	0.023
Fear	0.114	0.114	-	0.068	0.023
Anger	0.090	0.079	0.034	-	0.011
Surprise	0.086	0.071	0.014	0.043	-

Based on the statistical analysis, we conclude that more than one emotion can be expressed in a single post. The total percentage of posts expressing more than one emotion is 14.2%. Multiple events ([E14]) and different time frame ([E8]) are the circumstances for more than one emotion being mentioned in the same post.

[E14] 今天起得蛮早，因为下雪，所以很 **happy** 。但是在雪地里摔了一跤，真的好疼啊。

(*Woke up quite early today. Felt so* ***happy*** *as it's snowing. However, I fell down on the ground and I was in pain.*)

B. Experimental Settings

As discussed in Section 3, the data are collected from Weibo.com. We randomly select half of the annotated posts as the training data and the other half as the testing data. Besides, we use *FNLP* [5] for Chinese word segmentation and adopt F1-Measure (F1.) to measure the performance on emotion prediction.

To see whether the improvement in F1-Measure is statistically significant, we conduct significance tests using *t*-testing with *p*-values less than or equal to 0.01, in-between (0.01, 0.05), and bigger than 0.05, which imply significantly better, moderately better, and slightly better, respectively.

The CPU of the workstation for the experiment is Intel Q8300 with a 4G memory. The inference speed is high. However, since our proposed model is a graph model, the algorithm requires plenty of iterations to converge; the training time is about half an hour.

C. Experimental Results

In this subsection, we present selected experimental results and analyze our findings.

1) Comparison with Baselines

The purpose of the first experiment is to investigate whether our proposed approach was able to improve emotion prediction in code-switching text. For fair comparison, the following models are implemented.

- *Maximum Entropy (ME)* is the basic model which uses all the Chinese and English texts of each post as features to train a ME classification model[6].
- *ME-CN* considers only the Chinese text of each post function as a feature to train a ME classification model.
- *ME-EN* considers only the English text of each post function as a feature to train a ME classification model.
- *ME-Bilingual* means using a Maximum Entropy model by the bilingual information in code-switching text for predicting emotion.
- *FGM-Emotion* means only considering the relations between emotions on the factor graph model. Note that when the factor graph model does not consider the emotion relations, the graph model would regress to the maximum entropy model.
- *JFGM* is our proposed joint model which uses the factor graph model to learn and predict emotion in code-switching posts by considering both bilingual and emotional information collectively.

TABLE 5
COMPARISON WITH BASELINES

Method	Average F1.
ME	0.653
ME-CN	0.653
ME-EN	0.486
ME-Bilingual	0.672
FGM-Emotion	0.690
JFGM	**0.693**

Table 5 shows the experimental results of different models with average F1-Measure for all emotions. From the table, we can see that: 1) as English is the embedded language in our corpus, the result of ME-EN which only considers the English text is instable; 2) ME-Bilingual significantly outperforms the basic ME approach (*p*-value<0.01). This indicates the effectiveness of the bilingual information over the simple text information. 3) Emotion relation is effectively predicting emotions in code-switching text since the FGM-Emotion outperforms the basic ME model. 4) JFGM significantly outperforms the basic ME approach (*p*-value<0.01) indicating the effectiveness of both the bilingual and emotional information.

2) Results on Different Emotions

We perform the experimental results of different models

[5] https://github.com/xpqiu/fnlp/

[6] The ME algorithm is implemented with the *MALLET Toolkit*, http://mallet.cs.umass.edu

with F1-Measure of the five emotions, respectively, in Fig.10. From the figure we deduced that the improvement of JFGM with regard to the emotions *fear* and *surprise* is higher than the other three emotions. This indicates that these two emotions are easier affected by other emotions. The improvement of ME-Bilingual on the *anger* emotion is higher than on the other emotions. This shows that *anger* is expressed by the two languages collectively more frequently than other emotions. The statistics in Table 3 also support this outcome.

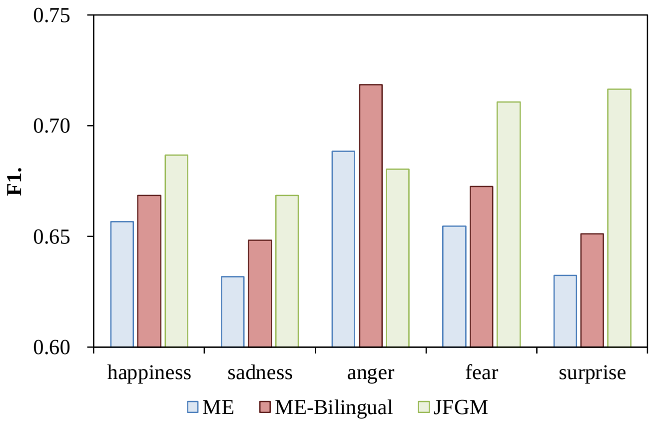


Fig 10: Results on Different Emotions

TABLE 6
EXAMPLE OUTPUT OF ME AND ME-CN

Post	ME	ME-CN
我的 **touch** 又挂彩了，苹果不给力 。 (*My **iPod Touch** broke down again, Apple needs to do better.*)	None	Sadness
今天运气不错，准点登机，也会准点起飞滴！**thanks god** ！ (*I'm so lucky today, we're boarding on time, maybe we'll take off on time too! **Thank God!***)	None	Happiness
飞机飞到济南去了，晕死我了...又 **delay** ，我要回家 (*The plane arrived in Ji'nan, I felt so dizzy...Another **delay**, I want to go home.*)	Anger	Sadness

TABLE 7
EXAMPLE OUTPUT OF ME AND ME-EN

Post	ME	ME-EN
今年必入单品元素，真的很难取舍， 两条都买了。**i like it** … (*It was hard for me to choose between this year's two must-buy items, so I bought both. **I like it**.*)	Anger	Happiness
我们又是时候 **high** 。 (*It is time for us to be **happy**.*)	None	Happiness
2011 的公事终于完成了。但是，心情很 **down** 。 (*I've finished all my business for 2011. However, I feel **depressed**.*)	None	Sadness

3) *Analysis of Usefulness for Monolingual Information*

Table 6 and Table 7 show some examples that illustrate the usefulness of Chinese and English information. The results of the monolingual classifier are correct, while the results of ME with mixed text are incorrect. This suggests that although the ME with mixed text outperforms the two monolingual classifiers in general, there are still some emotions that can be detected by a simple monolingual classifier. For example, the first post in Table 6, the English word '*touch*' does not express any emotion, one can only predict the emotion based on the Chinese text. In addition, the emotion of the second post in Table 7 can be predicted via the English word '*high*'.

4) *Analysis of Usefulness for Joint Model*

Table 8 shows some examples that illustrate the usefulness of our proposed JFGM model for emotion detection in code-switching text. The results of the proposed JFGM are correct; whereas the results of ME are incorrect. From the examples we can see that the JFGM model can learn the Chinese and English information collectively. For example, in the second post, the emotion is expressed by the mixed phrase "*hold 不住*", this kind of emotion is very hard to detect with a simple ME based approach.

TABLE 8
EXAMPLE OUTPUT OF ME AND JFGM

Post	ME	JFGM
睡不着，起来打锣鼓。**so happy**! (*I couldn't sleep, so I woke up and played DOTA. **So happy!***)	None	Happiness
感冒了一场， 我真是 **hold 不住**，头太痛了 (*I caught a cold, I **can't take it any more**. My head aches too much.*)	None	Sadness
天氣貌似不是很好耶 。。。**bad weather**! (*It seems that the weather is not very good, **bad weather!***)	Fear	Sadness

VI. CONCLUSIONS

In this paper, we addressed a novel task; namely, emotion detection in code-switching text. Previous researches have mainly focused on analyzing emotions in monolingual text. However, code-switching posts are common on social media platforms where emotions can be expressed in either monolingual or bilingual. The major challenges of emotion prediction in code-switching text are the need to explore the bilingual information of each post and connect different emotions in a single post. To address these challenges, we collected and extracted code-switching posts from Weibo.com, annotated the posts according to five basic emotions, finally we presented statistical and analysis results to illustrate the distribution between emotions and caused languages. Moreover, we proposed a uniform framework to predict emotions in code-switching text by considering both bilingual and emotional information. In particular, the factor functions were proposed to connect different emotions within a single post by considering the emotional relations. The attribute functions were proposed to connect the Chinese and English texts by considering the bilingual information between them. Finally, the factor graph model was proposed to incorporate both factor functions and attribute functions to a uniform framework to learn the emotion prediction model collectively. The empirical studies demonstrated that our model significantly improves emotions prediction in bilingual text.

ACKNOWLEDGEMENT

We gratefully acknowledge Helena Yan Ping Lau for her excellent work on corpus analysis and helpful discussion. We thank the anonymous reviewers for their helpful comments and feedbacks.

REFERENCES

[1] Adel H., N. Vu, and T. Schultz. 2013. Combination of Recurrent Neural Networks and Factored Language Models for Code-Switching Language Modeling. In Proceedings of ACL-13.

[2] Adel H., N. Vu, K. Kirchhoff, D. Telaar, and T. Schultz. 2015. Syntactic and Semantic Features for Code-Switching Factored Language Models. In IEEE/ACM Transactions on Audio, Speech, and Language Processing, 23(3), 431-440.

[3] Amini M., C. Goutte, and N. Usunier. 2010. Combining Coregularization and Consensusbased Self-training for Multilingual Text Categorization. In Proceeding of SIGIR-10.

[4] Andrew R. 1963. Evolution of Facial Expressions. Science, 142, 1034-1041.

[5] Arnold J. 1999. Affect in Language Learning. Cambridge, MA: CUP.

[6] Auer P. 1999. Code-Switching in Conversation. Routledge.

[7] Blitzer J., M. Bredze, and F. Pereira. 2007. Biographies, Bollywood, Boom-boxes and Blenders: Domain Adaptation for Sentiment Classification. In Proceedings of ACL-07.

[8] Blom J., and J. Gumperz. 1972. Social Meaning in Linguistic Structures: Code Switching in Northern Norway. Directions in Sociolinguistics. New York: Winston.

[9] Bridge D., J. Chiao, and K. Paller. 2010. Emotional Context at Learning Systematically Biases Memory for Facial Information. Memory & Cognition, 38, 125-133.

[10] Brown P., J. Cocke, S. Pietra, V. Pietra, F. Jelinek, J. Lafferty, R. Mercer, and P. Roossin. 1990. A Statistical Approach to Machine Translation. Computational Linguistics, 16(2):79–85.

[11] Burkett D., and D. Klein. 2008. Two Languages are Better than One (for Syntactic Parsing). In Proceedings of EMNLP-08.

[12] Chen Y., S. Lee, S. Li, and C. Huang. 2010. Emotion Cause Detection with Linguistic Constructions. In Proceeding of COLING-10.

[13] Clyne M. 1991. Community Languages: The Australian Experience. Cambridge: Cambridge University Press.

[14] Dasgupta S., and V. Ng. 2009. Mine the Easy and Classify the Hard: Experiments with Automatic Sentiment Classification. In Proceedings of ACL-IJCNLP-09.

[15] Dewaele J., and A. Pavlenko. 2002. Emotion Vocabulary in Interlanguage. Language Learning, 52 (2), 265-324.

[16] Ekman P., and W.V. Friesen. 1978. Facial Action Coding System. California: Consulting Psychology Press.

[17] Gao W., J. Blitzer, M. Zhou, and K. Wong. 2009. Exploiting Bilingual Information to Improve Web Search. In Proceedings of ACL/IJCNLP-09.

[18] Gao W., S. Li, S. Lee, G. Zhou, and C. Huang. 2013. Joint Learning on Sentiment and Emotion Classification. In Proceedings of CIKM 2013.

[19] Garrette D., H. Alpert-Abrams, T. Berg-Kirkpatrick, and D. Klein. 2015. Unsupervised Code-Switching for Multilingual Historical Document Transcription. In NAACL-15.

[20] Giles H., and R. Clair. 1979. Language and Social Psychology. London: Basil Blackwell.

[21] Gumperz, J. 1982. Discourse strategies. New York: Cambridge University Press.

[22] Hammersley J., and P. Clifford. 1971. Markov Field on Finite Graphs and Lattices, Unpublished manuscript.

[23] Hassan A., and D. Radev. 2010. Identifying Text Polarity Using Random Walks. In Proceedings of ACL-10.

[24] Hervé P., A. Razafimandimby, M. Vigneau, B. Mazoyer, and N. Tzourio-Mazoyer. 2012. Disentangling the Brain Networks Supporting Affective Speech Comprehension. NeuroImage, 61(4), 1255-1267.

[25] Hu M. and B. Liu. 2004. Mining and Summarizing Customer Reviews. In Proceedings of SIGKDD 04.

[26] Huang H., J. Tang, S. Wu, L Liu, and X Fu. 2014. Mining Triadic Closure Patterns in Social Networks. In Proceedings of WWW-14.

[27] Jamatia A., B. Gambäck, and A. Das. 2015. Part-of-Speech Tagging for Code-Mixed English-Hindi Twitter and Facebook Chat Messages. In Proceedings of RANLP-15.

[28] Kong F., H. Ng, and G. Zhou. 2014. A Constituent-Based Approach to Argument Labeling with Joint Inference in Discourse Parsing. In Proceedings of EMNLP-14

[29] Lafferty J., A. McCallum, and F. Pereira. 2001. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In Proceedings of ICML-01.

[30] Lee S., and Z. Wang. 2015. Emotion in Code-switching Texts: Corpus Construction and Analysis. In Proceeding of SIGHAN-2015.

[31] Lee S., H. Zhang, and C. Huang. 2013a. An Event-Based Emotion Corpus. In Proceedings of CLSW 2013.

[32] Lee S., Y. Chen, C. Huang, and S. Li. 2013b. Detecting Emotion Causes with a Linguistic Rule-Based Approach. Computational Intelligence, 29(3), 390-416.

[33] Lee S., S. Li, and C. Huang. 2014. Annotating Events in an Emotion Corpus. In Proceedings of LREC-14.

[34] Li S., R. Wang and G. Zhou. 2012. Opinion Target Extraction via Shallow Semantic Parsing. In Proceedings of AAAI-2012, 2012.

[35] Kennedy A., and D. Inpken. 2006. Sentiment Classification of Movie Reviews using Contextual Valence Shifters. Computational Intelligence, 22(2):110–125.

[36] Li D. 2000. Cantonese-English Code-switching Research in Hong Kong: A Y2K Review. World Englishes. 19(3), 305-322.

[37] Li H., Z. Liu, and M. Sun. 2012. Random Walks on Context-Aware Relation Graphs for Ranking Social Tags. In Proceedings of COLING-12.

[38] Li J., J. Zhu, and B. Zhang. 2016. Discriminative Deep Random Walk for Network Classification. In Proceedings of ACL-12.

[39] Li S., C. Huang, G. Zhou, and S. Lee 2010. Employing Personal/Impersonal Views in Supervised and Semi-supervised Sentiment Classification. In Proceedings of ACL-10.

[40] Li S., Y. Xue, Z. Wang, and G. Zhou. 2013. Active Learning for Cross-domain Sentiment Classification. In Proceeding of IJCAI-13.

[41] Li S., L. Huang, R. Wang, and G. Zhou. 2015. Sentence-level Emotion Classification with Label and Context Dependence. In Proceeding of ACL-15.

[42] Li Y., and P. Fung. 2012. Code-switch Language Model with Inversion Constraints for Mixed Language Speech Recognition. In Proceedings of COLING-12.

[43] Ling W., G. Xiang, C. Dyer, A. Black, and I. Trancoso. 2013. Microblogs as Parallel Corpora. In Proceedings of ACL-13.

[44] Lignos C., and M. Marcus. 2013. Toward Web-scale Analysis of Code-switching. In Proceedings of Annual Meeting of the Linguistic Society of America-2013.

[45] Liu H., S. Li, G. Zhou, C. Huang, and P. Li. 2013. Joint Modeling of News Reader's and Comment Writer's Emotions. In Proceedings of ACL-13, shorter.

[46] Lu B., C. Tan, C. Cardie and B. Tsou. 2011. Joint Bilingual Sentiment Classification with Unlabeled Parallel Corpora. In Proceedings of ACL-11.

[47] Luo K., Z. Deng, L. Wei, and H. Yu. 2015. JEAM: A Novel Model for Cross-Domain Sentiment Classification Based on Emotion Analysis. In Proceedings of ACL-15.

[48] Meng X., F. Wei, X. Liu, M. Zhou, G. Xu, and H. Wang. 2012. Cross-Lingual Mixture Model for Sentiment Classification. In Proceedings of ACL-12.

[49] Mihalcea R., and H. Liu. 2006. A Corpus-based Approach to Finding Happiness. In Proceedings of the AAAI Spring Symposium on Computational Approaches to Weblogs.

[50] Myers-Scotton C. 1997. Duelling Language: Grammatical Structure in Code-switching. Oxford: Clarendon.

[51] Och F., and H. Ney. 2002. Discriminative Training and Maximum Entropy Models for Statistical Machine Translation. In Proceedings of ACL-02.

[52] Olson I., A. Plotzker, and Y. Ezzyat. 2007. The Enigmatic Temporal Poles: A Review of Findings on Social and Emotional Processing. Brain.

[53] Pang B., L. Lee, and S. Vaithyanathan. 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques. In Proceedings of EMNLP-02.

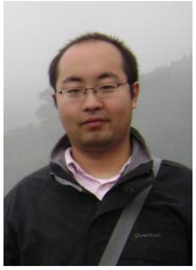
[54] Pang B., and L. Lee. 2004. A Sentimental Education: Sentiment Analysis using Subjectivity Summarization based on Minimum Cuts. In Proceedings of ACL-04.

[55] Pavlenko A. 2008. Structural and Conceptual Equivalence in Acquisition and Use of Emotion Words in a Second Language. Mental Lexicon, 3(1): 91-120.

[56] Peng N., Y. Wang, and M. Dredze. 2014. Learning Polylingual Topic Models from Code-Switched Social Media Documents. In Proceedings of ACL-14.

[57] Quan C., and F. Ren. 2009. Construction of a Blog Emotion Corpus for Chinese Emotional Expression Analysis. In Proceedings of EMNLP-09.

[58] Rao Y., X. Quan, W. Liu, Q. Li, and M. Chen. 2012. Building Word-emotion Mapping Dictionary for Online News. In Proceedings of SDAD 2012 The 1st International Workshop on Sentiment Discovery from Affective Data.
[59] Roth D., and W. Yih. 2005. Integer Linear Programming Inference for Conditional Random Fields. In Proceedings of ICML-2005.
[60] Sindhwani V. and P. Melville. 2008. Document-Word Co-Regularization for Semi-supervised Sentiment Analysis. In Proceedings of ICDM-08.
[61] Sitaram D., S. Murthy, D. Ray, D. Sharma and K. Dhar. 2015. Sentiment analysis of mixed language employing Hindi-English code switching. In Proceedings of International Conference on Machine Learning and Cybernetics (ICMLC).
[62] Smith C., and R. Lazarus. 1993. Appraisal Components, Core Relational Themes, and the Emotions. Cognition and Emotion, 7, 233–269.
[63] Smith C., and L. Kirby. 2001. Toward Delivering on the Promise of Appraisal Theory. Appraisal Processes in Emotion: Theory, Methods, Research. Oxford, UK: Oxford University Press.
[64] Schrauf R. 2000. Bilingual Autobiographical Memory: Experimental Studies and Clinical Cases. Culture and Psychology. 6 (4), 387-417.
[65] Solorio T., and Y. Liu. 2008. Learning to Predict Code-Switching Points. In Proceedings of EMNLP-08.
[66] Tang J., T Lou, and J Kleinberg. 2012. Inferring Social Ties across Heterogeneous Networks. In Proceedings of WSDM-12.
[67] Tang X., X. Wan, and X. Zhang. 2014. Cross-Language Context-aware Citation Recommendation in Scientific Articles. In Proceedings of SIGIR-14.
[68] Turney P. 2002. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In Proceedings of ACL-02.
[69] Vilares D., M. Alonso, and C. Goméz-Rodríguez. 2015. Sentiment Analysis on Monolingual, Multilingual and Code-Switching Twitter Corpora. In Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA 2015).
[70] Volkova S., W. Dolan, and T. Wilson. 2012. CLex: A Lexicon for Exploring Color, Concept and Emotion Associations in Language. In Proceedings of EACL-12.
[71] Xu G., X. Meng, and H. Wang. 2010. Build Chinese Emotion Lexicons Using A Graph-based Algorithm and Multiple Resources. In Proceeding of COLING-10.
[72] Wan X. 2008. Using Bilingual Knowledge and Ensemble Techniques for Unsupervised Chinese Sentiment Analysis. In Proceedings of ACL-2008.
[73] Wan X. 2009. Co-Training for Cross-Lingual Sentiment Classification. In Proceedings of ACL/IJCNLP-09.
[74] Wang R., S. Li, G. Zhou, and H. Shi. 2015. Joint Sentiment and Emotion Classification with Integer Linear Programming. In Proceeding of DASFAA-2015.
[75] Wang Z., S. Li, H. Shi, and G. Zhou. 2014. Skill Inference with Personal and Skill Connections. In Proceeding of COLING-14.
[76] Wei B., and C. Pal. 2010. Cross Lingual Adaptation an Experiment on Sentiment Classifications. In Proceedings of ACL-10.
[77] Wiebe J., T. Wilson, and C. Cardie. 2005. Annotating Expressions of Opinions and Emotions in Language. Language Resources and Evaluation, 39(2/3):165–210.
[78] Wilson T., J. Weibe, and P. Hoffman. 2009. Recognizing Contextual Polarity: An Exploration of Features for Phrase-level Sentiment Analysis. Computational Linguistics, 35(3):399–433.
[79] Wen S. and X. Wan. 2014. Emotion Classification in Microblog Texts Using Class Sequential Rules. In Proceedings of AAAI-14.
[80] Yang M., B. Peng, Z. Chen, D. Zhu, and K. Chow. 2014. A Topic Model for Building Fine-grained Domain-specific Emotion Lexicon. In Proceedings of ACL-14.
[81] Yang Z., K. Cai, J. Tang, L. Zhang, Z. Su and J. Li. 2011b. Social Context Summarization. In Proceeding of SIGIR-11.
[82] Zhang Y., J. Tang, Z. Yang, J. Pei, and P. Yu. 2015. COSNET: Connecting Heterogeneous Social Networks with Local and Global Consistency. In KDD-15.
[83] Zhao H., Y. Song, C. Kit, and G. Zhou. 2009. Cross Language Dependency Parsing using a Bilingual Lexicon. In Proceedings of ACL-09.
[84] Zhou G. and K. Fang. 2009. Global Learning of Noun Phrase Anaphoricity in Coreference Resolution via Label Propagation. In Proceedings of EMNLP-2009.
[85] Zhu X. and Z. Ghahramani. 2002. Learning from Labeled and Unlabeled Data with Label Propagation. CMU CALD Technical Report. CMU-CALD-02-107.
[86] Zhuang H, J. Tang, W. Tang, T. Lou, A. Chin, and X. Wang. 2012. Actively Learning to Infer Social Ties. In Proceedings of Data Mining and Knowledge Discovery (DMKD-12), vol.25 (2), pp. 270-297.



Zhongqing Wang received his Ph.D. degree in 2016 from Soochow University, China. He is a postdoctoral research fellow at Singapore University of Technology Design. He was a research associate at the Hong Kong Polytechnic University in 2015. His current research interests include natural language processing, sentiment analysis and social computing.

Shoushan Li received his Ph.D. degree in 2008 from National Laboratory of Pattern Recognition, CASIA, Beijing, China. He is a Full Professor in the School of Computer Science and Technology, Soochow University. His current research interests include natural language processing, social computing, and sentient analysis.

Sophia Yat Mei Lee received her Ph.D. degree in 2010 from The Hong Kong Polytechnic University. She is Assistant Professor at The Hong Kong Polytechnic University. Her current research interests include linguistic, natural language processing, emotion analysis.

Guodong Zhou received his Ph.D. degree in 1999 from the National University of Singapore, Singapore. He is a Full Professor in the School of Computer Science and Technology, and the Director of the Natural Language Processing Laboratory from Soochow University. His research interests include information retrieval, natural language processing.