

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 has 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 and 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—Bilingual information, code-switching, emotion analysis, 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 [33], [45], [53]. Previous research has focused solely on analyzing emotions in monolingual text [12], [31]. 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 **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 different 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

Manuscript received February 26, 2016; revised June 4, 2016, August 20, 2016, and October 6, 2016; accepted November 19, 2016. Date of publication December 7, 2016; date of current version January 16, 2017. This work was supported in part by an Early Career Scheme project sponsored by the Research Grants Council of Hong Kong (PolyU 5593/13H) and a PolyU Faculty Research Grant (No. 1-ZVEK), in part by the Key Project of the National Natural Science Foundation of China under Grant 61331011, and in part by the National Natural Science Foundation of China under Grants 61273320 and 61375073. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Haizhou Li.

Z. Wang was with The Hong Kong Polytechnic University, Hong Kong. He is now with the School of Computer Science and Technology, Soochow University, Suzhou 215006, China (e-mail: wangzq.antony@gmail.com).

S. Y. M. Lee is with the Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong (e-mail: ymlee@polyu.edu.hk).

S. Li, and G. Zhou are with the School of Computer Science and Technology, Soochow University, Suzhou 215006, China (e-mail: lishoushan@suda.edu.cn; gdzhou@suda.edu.cn).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TASLP.2016.2637280

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 II, we give an overview of related work. In Section III, we introduce the corpus and present some statistics of the data. In Section IV, we propose the factor graph model to predict emotions with both bilingual and emotional information. Section V reports the details of our experimental results. Finally, we conclude and discuss some future work in Section VI.

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 [40], [68] and dominates the literature, fine-grained feature-based sentiment analysis aims to extract objective attributes, such as holder and target from a review [25], [34].

Subsequently, a great deal of work has taken other important features into account. For instance, [54] examined the relation between subjectivity and polarity classification; [35] identified three types of valence shifters: negations, intensifiers, and diminishers; and [78] 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 [7], [40] and semi-supervised learning [14], [39].

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) [49], [77]. The earliest research on emotion has focused on the representation and processing of emotion in facial expressions and body language [4], [16]. More recently, there have been mounting researches on the neurobiological basis of emotion [24], [52] and how emotion is

linked with other aspects of human cognition [9], [62], [63]. 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, [58] automatically built a word-emotion mapping dictionary for social emotion detection and [80] 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, [45] used a co-training framework to infer the news from readers' comments and writers' emotions collectively; [79] used class sequential rules for emotion classification of micro-blog texts by regarding each post as a data sequence; and [41] 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, [18] proposed an ensemble learning based framework to learn the emotion and sentiment collectively, while [74] 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, [47] 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 [8], communication accommodation theory [20], the markedness model [50], and the conversational analysis model [6].

Code-switched documents have also received considerable attention in the NLP community [2], [19]. Several studies have focused on identification and analysis, including mining translations in code-switched documents [43], predicting code-switched points [65], identifying code-switched tokens [44], adding code-switched support to language models [42], and developing POS tagging for code-switching text [27].

A related 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 [11], information retrieval [17], text classification [3], and so on. There are a number of studies on predicting polarity through multilingual text. [72] 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 [73] incorporated the unlabeled data in the target language into the same classification method with co-training to improve the classification performance. [76] 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, [46] 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 [46], [48] 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 ([5]; Schumann, 1999), as well as bilingual performance and language choice [55], [64]. For example, there are a number of factors that may influence the use of emotion vocabulary, such as sociocultural competence, gender, and topic [15].

A few works focus on sentiment analysis in code-switching text [61], [69]. 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; [86]). Most of the work on the factor graph model focuses on social network analysis. For example, Tang *et al.* (2011a) and [86] 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. [26] employed a factor graph model to discover triadic closure patterns in dynamic networks with both network topology and demography information. [82] 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, [81] 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. [75] 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, [23] used a random walk model in a large word relatedness

graph and estimated polarity for any given word. [34] built a context-aware relation graph of candidate tags and proposed a random walk algorithm to rank tags for suggestion. [38] 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 [32], 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:

- 1) *None* means the post does not contain any corresponding emotions. [E7] is an example in which no emotion is found.
- 2) *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.
- 3) *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.
- 4) *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 **muji** 买了点东西然后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 [30]. 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.

To verify the quality of the annotation, two human annotators were asked to annotate 1,000 posts. the inter-annotator

TABLE I
RESULTS OF AGREEMENT ANALYSIS

	Kappa score
Emotion	0.692
Caused Language	0.767

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 I 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:

$$Z_{e,p} \in \{0, 1\} \quad (1)$$

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 .

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 [66] and a belief propagation algorithm [26] 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

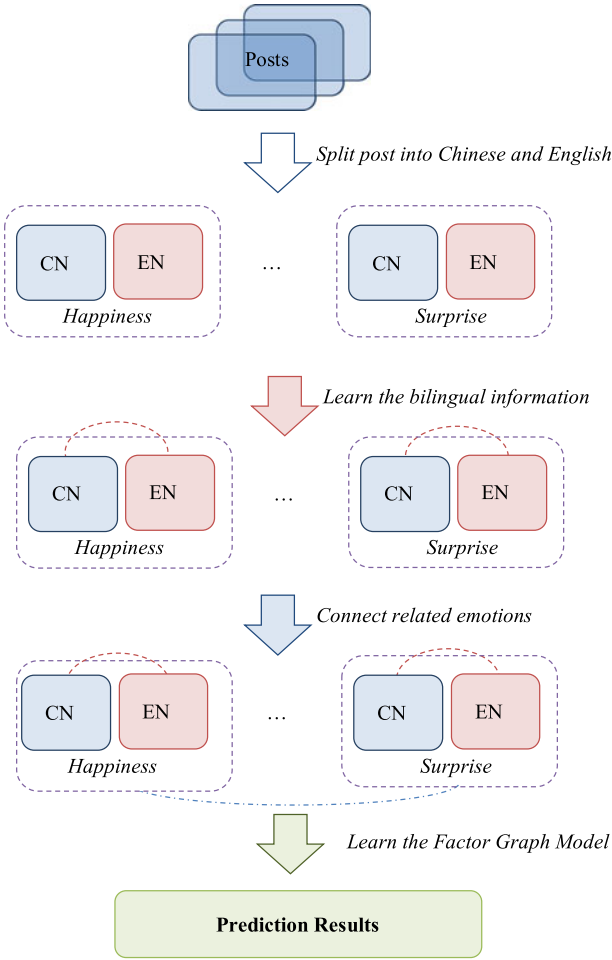


Fig. 1. Overview of our framework.

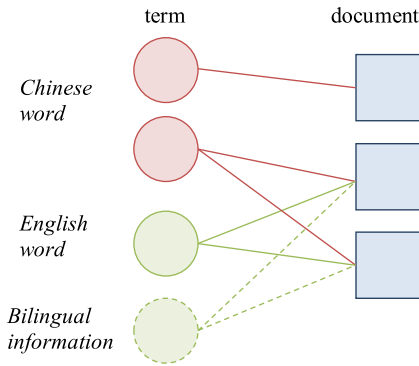


Fig. 2. Example of the bipartite graph.

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 .

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 [51]) with translation and language model being the only features, is used:

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

where

$$h_1(c, e) = \log(p_\gamma(c|e)) \quad (3)$$

is the translation model which is converted from the bilingual lexicon¹ and

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

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 dictionary³ 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 engine⁴ [68]. 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.

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

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

³TongYiCiLin (同义词林) is adopted as the Chinese synonym dictionary: <http://www.ttp-cloud.com/>

⁴We use BING as the search engine: <http://www.bing.com/>

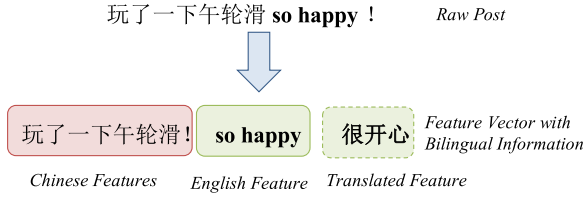


Fig. 3. Example of bilingual information.

The output will be generated only if it reaches the maximum probability as in the following [10], [83]:

$$c = \arg \max \prod p(w_c) \quad (5)$$

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:

$$P(Y|X, G) = \frac{P(X, G|Y) P(Y)}{P(X, G)} \quad (6)$$

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

$$P(Y|X, G) \propto P(X|Y) P(Y|G) \quad (7)$$

where $P(Y|G)$ represents the probability distribution of labels given the overall network structure and $P(X|Y)$ 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

$$P(Y|X, G) \propto P(Y|G) \prod_i P(x_i|y_i) \quad (8)$$

where $P(x_i|y_i)$ 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 $P(Y|G)$ and $P(x_i|y_i)$. In this study, we model them using a Markov random field. Consequently, $P(Y|G)$ and

$P(x_i|y_i)$ can be instantiated as follows [22]:

$$P(x_i|y_i) = \frac{1}{N_1} \exp \left\{ \sum_{j=1}^d \alpha_j f_j(x_{ij}, y_i) \right\} \quad (9)$$

$$P(Y|G) = \frac{1}{N_2} \exp \left\{ \sum_i \sum_{j \in NB(i)} g(i, j) \right\} \quad (10)$$

where N_1 and N_2 are normalization factors. Eq. (9) indicates that we define an attribute function $f_j(x_{ij}, y_i)$ for each post associated with the emotion. Eq. (10) 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 value associated with each post i by each emotion e . We define the local textual attribute as a feature [21]. We can accumulate all the attribute functions and obtain local entropy for a text:

$$\frac{1}{N_1} \exp \left(\sum_i \sum_k \alpha_k f_k(x_{ik}, y_i) \right) \quad (11)$$

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 IV.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:

$$g(y_i, y_j) = \exp \left\{ \beta_{ij} (y_i - y_j)^2 \right\} \quad (12)$$

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.

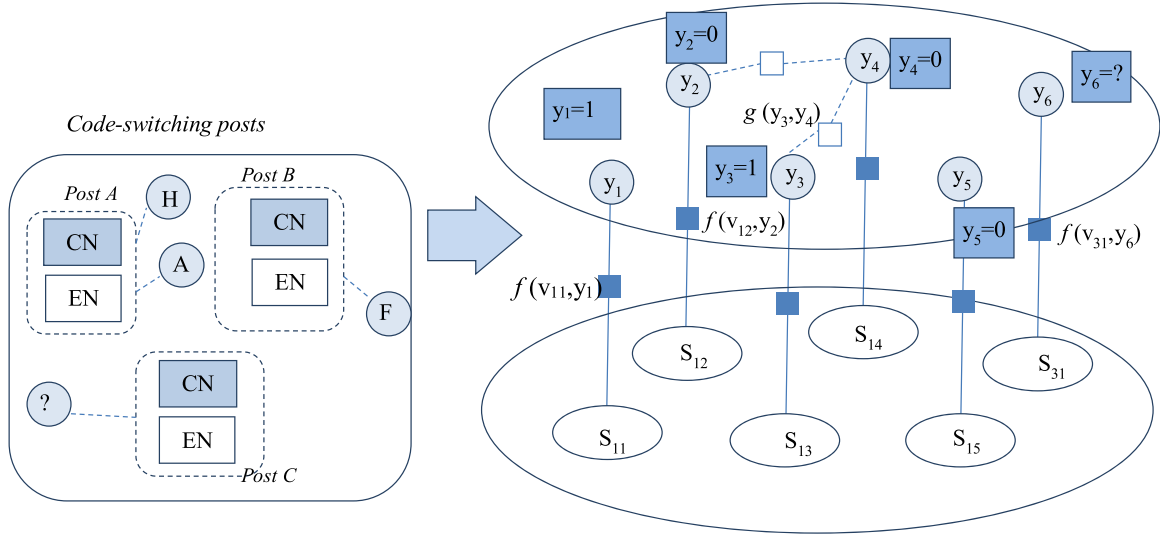


Fig. 4. Graph Representation of the Factor Graph Model.

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 $L(\theta)$.

$$\theta^* = \arg \max L(\theta) \quad (13)$$

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):

$$\frac{L(\theta)}{\beta_k} = E[g(i, j)] + E_{P_{\beta_k}(Y|X, G)}[g(i, j)] \quad (14)$$

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 $P_{\beta_k}(Y|X, G)$ 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.

V. EXPERIMENTS

In this section, we first show some statistics on the dataset and then introduce the experimental settings. Finally, we evaluate the

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., $P(y_i x_i, G)$ 2) Perform LBP to calculate the marginal distribution of each variable, i.e., $P(y_i, y_j X_{(i,j)}, G)$ 3) Calculate the gradient of β_k according to Eq. 15 (for α 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.

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 $p(e_i | l_j)$ 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

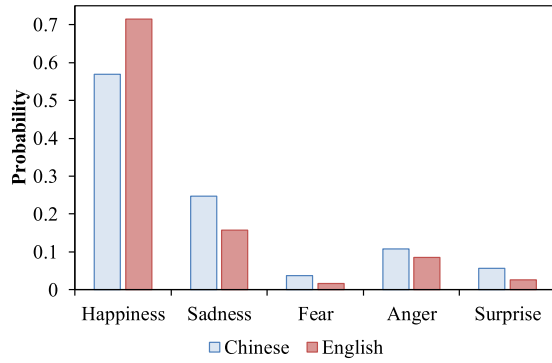


Fig. 6. Joint Distribution of Emotions and Caused Languages.

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

	#avg. word
Chinese	19.8
English	2.9

TABLE III
THE LANGUAGE CO-EXPRESSED PROBABILITY

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

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.

2) *Word Count Analysis*: Table II 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.

3) *Languages Co-Expressed Distribution*: Table III 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.

TABLE IV
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	—

[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 IV. The direction of the table is from left to right, for example $p(\text{sadness}|\text{happiness})$ is in row 2 and column 3, and the value is 0.060.

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 III, 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.

⁵<https://github.com/xpqiu/fnlp/>

TABLE V
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

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.

- 1) *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
- 2) *ME-CN* considers only the Chinese text of each post function as a feature to train a ME classification model.
- 3) *ME-EN* considers only the English text of each post function as a feature to train a ME classification model.
- 4) *ME-Bilingual* means using a Maximum Entropy model by the bilingual information in code-switching text for predicting emotion.
- 5) *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.
- 6) *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 V 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 with F1-Measure of the five emotions, respectively, in Fig. 7. From the figure we deduced that the improvement of JFGM with regard to the emotions

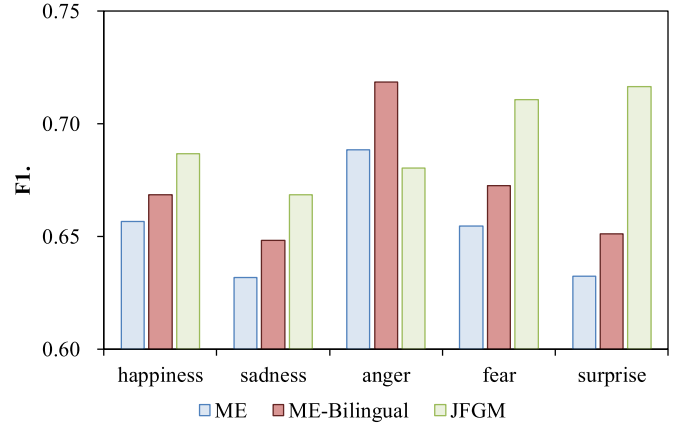


Fig. 7. Results on Different Emotions.

TABLE VI
EXAMPLE OUTPUT OF ME AND ME-CN

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

TABLE VII
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 ... (<i>It was hard for me to choose between this year's two must-buy items, so I bought both. I like it.</i>)	Anger	Happiness
我们又是时候 high 。 (<i>It is time for us to be happy.</i>)	None	Happiness
2011 的公事终于完成了。但是, 心情很 down 。 (<i>I've finished all my business for 2011. However, I feel depressed.</i>)	None	Sadness

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 III also support this outcome.

3) *Analysis of Usefulness for Monolingual Information*: Tables VI and VII 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

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

TABLE VIII
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

Table VI, 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 VII can be predicted via the English word 'high'.

4) *Analysis of Usefulness for Joint Model:* Table VIII 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.

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

The authors would like to thank Helena Yan Ping Lau for her excellent work on corpus analysis and helpful discussion.

They would also like to thank the anonymous reviewers for their helpful comments and feedbacks.

REFERENCES

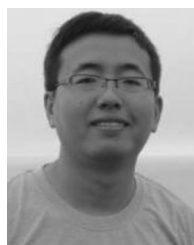
- [1] H. Adel, N. Vu, and T. Schultz, "Combination of recurrent neural networks and factored language models for code-switching language modeling," in *Proc. Annu. Meeting Assoc. Comput. Linguist.*, 2013, pp. 206–211.
- [2] H. Adel, N. Vu, K. Kirchhoff, D. Telaar, and T. Schultz, "Syntactic and semantic features for code-switching factored language models," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 3, pp. 431–440, Mar. 2015.
- [3] M. Amini, C. Goutte, and N. Usunier, "Combining coregularization and consensusbased self-training for multilingual text categorization," in *Proc. 33rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2010, pp. 475–482.
- [4] R. Andrew, "Evolution of facial expressions," *Science*, vol. 142, pp. 1034–1041, 1963.
- [5] J. Arnold, *Affect in Language Learning*. Cambridge, MA, USA: Cambridge Univ. Press, 1999.
- [6] P. Auer, *Code-Switching in Conversation*. Evanston, IL, USA: Routledge, 1999.
- [7] J. Blitzer, M. Bredze, and F. Pereira, "Biographies, Bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification," in *Proc. 45th Annu. Meeting Assoc. Comput. Linguist.*, 2007, pp. 187–205.
- [8] J. Blom and J. Gumperz, "Social meaning in linguistic structures: Code switching in northern norway," in *Directions in Sociolinguistics*. New York, NY, USA: Winston, 1972.
- [9] D. Bridge, J. Chiao, and K. Paller, "Emotional context at learning systematically biases memory for facial information," *Memory Cogn.*, vol. 38, pp. 125–133, 2010.
- [10] P. Brown *et al.*, "A statistical approach to machine translation," *Comput. Linguist.*, vol. 16, no. 2, pp. 79–85, 1990.
- [11] D. Burkett and D. Klein, "Two languages are better than one (for syntactic parsing)," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2008, pp. 877–886.
- [12] Y. Chen, S. Lee, S. Li, and C. Huang, "Emotion cause detection with linguistic constructions," in *Proc. 23rd Int. Conf. Comput. Linguist.*, 2010, pp. 179–187.
- [13] M. Clyne, *Community Languages: The Australian Experience*. Cambridge, U.K.: Cambridge Univ. Press, 1991.
- [14] S. Dasgupta and V. Ng, "Mine the easy and classify the hard: Experiments with automatic sentiment classification," in *Proc. Joint Conf. 47th Annu. Meeting Assoc. Comput. Linguist./4th Int. Joint Conf. Natural Lang. Process. AFNAP*, 2010, pp. 701–709.
- [15] J. Dewaele and A. Pavlenko, "Emotion vocabulary in interlanguage," *Lang. Learn.*, vol. 52, no. 2, pp. 265–324, 2002.
- [16] P. Ekman and W. V. Friesen, *Facial Action Coding System*. Palo Alto, CA, USA: Consulting Psychology Press, 1978.
- [17] W. Gao, J. Blitzer, M. Zhou, and K. Wong, "Exploiting Bilingual information to improve web search," in *Proc. 47th Annu. Meeting Assoc. Comput. Linguist./4th Int. Joint Conf. Natural Lang. Process. AFNAP*, 2009, pp. 1075–1083.
- [18] W. Gao, S. Li, S. Lee, G. Zhou, and C. Huang, "Joint learning on sentiment and emotion classification," in *Proc. 22nd ACM Int. Conf. Inf. Knowl. Manage.*, 2013, pp. 1505–1508.
- [19] D. Garrette, H. Alpert-Abrams, T. Berg-Kirkpatrick, and D. Klein, "Unsupervised code-switching for multilingual historical document transcription," in *Proc. Annu. Conf. North Amer. Chapter Assoc. Comput. Linguist.*, 2015, pp. 1036–1041.
- [20] H. Giles and R. Clair, *Language and Social Psychology*. London, U.K.: Blackwell, 1979.
- [21] J. Gumperz, *Discourse Strategies*. New York, NY, USA: Cambridge Univ. Press, 1982.
- [22] J. Hammersley and P. Clifford, "Markov field on finite graphs and lattices," 1971, to be published.
- [23] A. Hassan and D. Radev, "Identifying text polarity using random walks," in *Proc. 47th Annu. Meeting Assoc. Comput. Linguist.*, 2010, pp. 395–403.
- [24] P. Hervé, A. Razafimandimby, M. Vigneau, B. Mazoyer, and N. Tzourio-Mazoyer, "Disentangling the brain networks supporting affective speech comprehension," *NeuroImage*, vol. 61, no. 4, pp. 1255–1267, 2012.
- [25] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proc. 10th SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2004, 168–177.

- [26] H. Huang, J. Tang, S. Wu, L. Liu, and X. Fu, "Mining triadic closure patterns in social networks," in *Proc. 23rd Int. Conf. World Wide Web*, 2014, pp. 499–504.
- [27] A. Jamatia, B. Gambäck, and A. Das, "Part-of-speech tagging for code-mixed english-hindi twitter and facebook chat messages," in *Proc. Recent Adv. Natural Lang. Process.*, 2015, pp. 239–248.
- [28] F. Kong, H. Ng, and G. Zhou, "A constituent-based approach to argument labeling with joint inference in discourse parsing," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2014, pp. 68–77.
- [29] J. Lafferty, A. McCallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proc. 18th Inter. Conf. Mach. Learn.*, 2014, pp. 282–289.
- [30] S. Lee and Z. Wang, "Emotion in code-switching texts: Corpus construction and analysis," in *Proc. 8th SIGHAN Workshop Chin. Lang. Process.*, 2015, pp. 91–99.
- [31] S. Lee, H. Zhang, and C. Huang, "An event-based emotion corpus," in *Proc. Chin. Lexical Semant. Workshop*, 2013, pp. 635–644.
- [32] S. Lee, Y. Chen, C. Huang, and S. Li, "Detecting emotion causes with a linguistic rule-based approach," *Comput. Intell.*, vol. 29, no. 3, pp. 390–416, 2013.
- [33] S. Lee, S. Li, and C. Huang, "Annotating events in an emotion corpus," in *Proc. 9th Int. Conf. Lang. Res. Eval.*, 2014.
- [34] S. Li, R. Wang and G. Zhou, "Opinion target extraction via shallow semantic parsing," in *Proc. 26th AAAI Conf. Artif. Intell.*, 2012, pp. 1671–1677.
- [35] A. Kennedy and D. Inpken, "Sentiment classification of movie reviews using contextual valence shifters," *Comput. Intell.*, vol. 22, no. 2, pp. 110–125, 2006.
- [36] D. Li, "Cantonese-english code-switching research in Hong Kong: A Y2K review," *World Englishes*, vol. 19, no. 3, pp. 305–322, 2000.
- [37] H. Li, Z. Liu, and M. Sun, "Random walks on context-aware relation graphs for ranking social tags," in *Proc. Int. Conf. Comput. Linguist.*, 2012, pp. 653–662.
- [38] J. Li, J. Zhu, and B. Zhang, "Discriminative deep random walk for network classification," in *Proc. 54th Annu. Meeting Assoc. Comput. Linguist.*, 2016, pp. 1004–1013.
- [39] S. Li, C. Huang, G. Zhou, and S. Lee "Employing personal/impersonal views in supervised and semi-supervised sentiment classification," in *Proc. 48th Annu. Meeting Assoc. Comput. Linguist.*, 2010, pp. 414–423.
- [40] S. Li, Y. Xue, Z. Wang, and G. Zhou, "Active learning for cross-domain sentiment classification," in *Proc. 23rd Int. Joint. Conf. Artif. Intell.*, 2013, pp. 2127–2133.
- [41] S. Li, L. Huang, R. Wang, and G. Zhou, "Sentence-level emotion classification with label and context dependence," in *Proc. 53th Annu. Meeting Assoc. Comput. Linguist.*, 2015, pp. 1045–1053.
- [42] Y. Li and P. Fung, "Code-switch language model with inversion constraints for mixed language speech recognition," in *Proc. Int. Conf. Comput. Linguist.*, 2012, pp. 1671–1680.
- [43] W. Ling, G. Xiang, C. Dyer, A. Black, and I. Trancoso, "Microblogs as parallel corpora," in *Proc. 51th Annu. Meeting Assoc. Comput. Linguist.*, 2013.
- [44] C. Lignos and M. Marcus, "Toward web-scale analysis of code-switching," in *Proc. Annu. Meeting Linguist. Soc. Amer.*, 2013.
- [45] H. Liu, S. Li, G. Zhou, C. Huang, and P. Li, "Joint modeling of news reader's and comment writer's emotions," in *Proc. 51th Annu. Meeting Assoc. Comput. Linguist.*, 2013, pp. 511–515.
- [46] B. Lu, C. Tan, C. Cardie and B. Tsou, "Joint bilingual sentiment classification with unlabeled parallel corpora," in *Proc. 49th Annu. Meeting Assoc. Comput. Linguist.*, 2011, pp. 320–330.
- [47] K. Luo, Z. Deng, L. Wei, and H. Yu, "JEAM: A novel model for cross-domain sentiment classification based on emotion analysis," in *Proc. Empirical Methods Natural Lang. Process.*, 2015, pp. 2503–2508.
- [48] X. Meng, F. Wei, X. Liu, M. Zhou, G. Xu, and H. Wang, "Cross-lingual mixture model for sentiment classification," in *Proc. 54th Annu. Meeting Assoc. Comput. Linguist.*, 2012, pp. 1403–1412.
- [49] R. Mihalcea and H. Liu, "A corpus-based approach to finding happiness," in *Proc. AAAI Spring Symp. Comput. Approaches Weblogs*, 2006.
- [50] C. Myers-Scotton, *Duelling Language: Grammatical Structure in Code-Switching*. Oxford, U.K.: Clarendon, 1997.
- [51] F. Och and H. Ney, "Discriminative training and maximum entropy models for statistical machine translation," in *Proc. 40th Annu. Meeting Assoc. Comput. Linguist.*, 2002, pp. 295–302.
- [52] I. Olson, A. Plotzker, and Y. Ezzayat, "The enigmatic temporal poles: A review of findings on social and emotional processing," *Brain*, vol. 130, pp. 1718–1731, 2007.
- [53] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," in *Proc. Empirical Methods Natural Lang. Process.*, 2002, pp. 79–86.
- [54] B. Pang and L. Lee, "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts," in *Proc. 42nd Annu. Meeting Assoc. Comput. Linguist.*, 2004, Art no. 271.
- [55] A. Pavlenko, "Structural and conceptual equivalence in acquisition and use of emotion words in a second language," *Mental Lexicon*, vol. 3, no. 1, pp. 91–120, 2008.
- [56] N. Peng, Y. Wang, and M. Dredze, "Learning polylingual topic models from code-switched social media documents," in *Proc. 52nd Annu. Meeting Assoc. Comput. Linguist.*, 2014, pp. 674–679.
- [57] C. Quan and F. Ren, "Construction of a blog emotion corpus for Chinese emotional expression analysis," in *Proc. Empirical Methods Natural Lang. Process.*, 2009, pp. 1446–1454.
- [58] Y. Rao, X. Quan, W. Liu, Q. Li, and M. Chen, "Building word-emotion mapping dictionary for online news," in *Proc. SDAD 1st Int. Workshop Sentiment Discovery Affective Data*, 2012.
- [59] D. Roth and W. Yih, "Integer linear programming inference for conditional random fields," in *Proc. 22nd Int. Conf. Mach. Learn.*, 2005, pp. 736–743.
- [60] V. Sindhwani and P. Melville, "Document-word co-regularization for semi-supervised sentiment analysis," in *Proc. Int. Conf. Data Mining*, 2008, pp. 1025–1030.
- [61] D. Sitaram, S. Murthy, D. Ray, D. Sharma, and K. Dhar, "Sentiment analysis of mixed language employing Hindi-English code switching," in *Proc. Int. Conf. Mach. Learn. Cybern.*, 2015, pp. 271–276.
- [62] C. Smith and R. Lazarus, "Appraisal components, core relational themes, and the emotions," *Cogn. Emotion*, vol. 7, pp. 233–269, 1993.
- [63] C. Smith and L. Kirby, "Toward delivering on the promise of appraisal theory," in *Appraisal Processes in Emotion: Theory, Methods, Research*. Oxford, U.K.: Oxford Univ. Press, 2001.
- [64] R. Schrauf, "Bilingual autobiographical memory: Experimental studies and clinical cases," *Culture Psychol.*, vol. 6, no. 4, pp. 387–417, 2000.
- [65] T. Solorio and Y. Liu, "Learning to predict code-switching points," in *Proc. Empirical Methods Natural Lang. Process.*, 2008, pp. 973–981.
- [66] J. Tang, T. Lou, and J. Kleinberg, "Inferring social ties across heterogeneous networks," in *Proc. 5th ACM Int. Conf. Web Search data Mining*, 2012, pp. 743–752.
- [67] X. Tang, X. Wan, and X. Zhang, "Cross-language context-aware citation recommendation in scientific articles," in *Proc. 37th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2014, pp. 817–826.
- [68] P. Turney, "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews," in *Proc. 40th Annu. Meeting Assoc. Comput. Linguist.*, 2002, pp. 417–424.
- [69] D. Vilares, M. Alonso, and C. Gómez-Rodríguez, "Sentiment analysis on monolingual, multilingual and code-switching twitter corpora," in *Proc. 6th Workshop Comput. Approaches Subjectivity, Sentiment Soc. Media Anal.*, 2015, pp. 2–8.
- [70] S. Volkova, W. Dolan, and T. Wilson, "CLex: A lexicon for exploring color, concept and emotion associations in language," in *Proc. 13th Conf. Eur. Chapter Assoc. Comput. Linguist.*, 2012, pp. 306–314.
- [71] G. Xu, X. Meng, and H. Wang, "Build Chinese emotion lexicons using a graph-based algorithm and multiple resources," in *Proc. 23rd Int. Conf. Comput. Linguist.*, 2010, pp. 1209–1217.
- [72] X. Wan, "Using bilingual knowledge and ensemble techniques for unsupervised Chinese sentiment analysis," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2008, pp. 553–561.
- [73] X. Wan, "Co-training for cross-lingual sentiment classification," in *Proc. Joint Conf. 47th Annu. Meeting Assoc. Comput. Linguist./4th Int. Joint Conf. Natural Lang. Process. AFNAP*, 2009, pp. 235–243.
- [74] R. Wang, S. Li, G. Zhou, and H. Shi, "Joint sentiment and emotion classification with integer linear programming," in *Proc. Database Syst. Adv. Appl.*, 2015, pp. 259–265.
- [75] Z. Wang, S. Li, H. Shi, and G. Zhou, "Skill inference with personal and skill connections," in *Proc. 25th Int. Conf. Comput. Linguist.*, 2014, pp. 520–529.
- [76] B. Wei and C. Pal, "Cross lingual adaptation an experiment on sentiment classifications," in *Proc. Annu. Meeting Assoc. Comput. Linguist.*, 2010, pp. 258–262.
- [77] J. Wiebe, T. Wilson, and C. Cardie, "Annotating expressions of opinions and emotions in language," *Lang. Resour. Eval.*, vol. 39, nos. 2/3, pp. 165–210, 2005.
- [78] T. Wilson, J. Wiebe, and P. Hoffman, "Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis," *Comput. Linguist.*, vol. 35, no. 3, pp. 399–433, 2009.

- [79] S. Wen and X. Wan, "Emotion classification in microblog texts using class sequential rules," in *Proc. 28th AAAI Conf. Artif. Intell.*, 2014, pp. 187–193.
- [80] M. Yang, B. Peng, Z. Chen, D. Zhu, and K. Chow, "A topic model for building fine-grained domain-specific emotion lexicon," in *Proc. 52nd Annu. Meeting Assoc. Comput. Linguist.*, 2014, pp. 421–426.
- [81] Z. Yang, K. Cai, J. Tang, L. Zhang, Z. Su, and J. Li, "Social context summarization," in *Proc. 34th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2011, pp. 255–264.
- [82] Y. Zhang, J. Tang, Z. Yang, J. Pei, and P. Yu, "COSNET: Connecting heterogeneous social networks with local and global consistency," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2015, pp. 1485–1494.
- [83] H. Zhao, Y. Song, C. Kit, and G. Zhou, "Cross language dependency parsing using a bilingual lexicon," in *Proc. 47th Annu. Meeting Assoc. Comput. Linguist.*, 2009, pp. 55–63.
- [84] G. Zhou and K. Fang, "Global learning of noun phrase anaphoricity in coreference resolution via label propagation," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2009, pp. 978–986.
- [85] X. Zhu and Z. Ghahramani, "Learning from labeled and unlabeled data with label propagation," Carnegie Mellon Univ. Pittsburgh, PA, USA, Tech. Rep. CMU-CALD-02-107, 2002.
- [86] H. Zhuang, J. Tang, W. Tang, T. Lou, A. Chin, and X. Wang, "Actively learning to infer social ties," *Proc. Data Mining Knowl. Discovery*, vol. 25, no. 2, pp. 270–297, 2012.



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



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



Zhongqing Wang received the Ph.D. degree in 2016 from Soochow University, Suzhou, China. He is a Postdoctoral Research Fellow at Singapore University of Technology Design, Singapore. 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.



Guodong Zhou received the 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, Soochow, China. His research interests include information retrieval, natural language processing.