



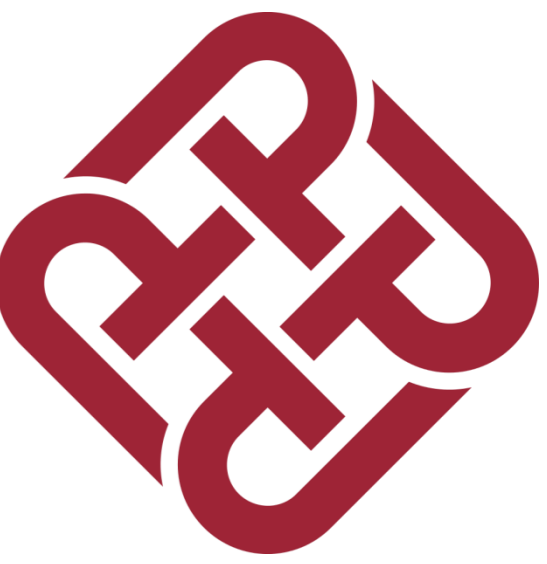
Emotion Detection in Code-switching Texts via Bilingual and Sentimental Information

Zhongqing Wang^{1,2}, Sophia Yat Mei Lee², Shoushan Li¹, and Guodong Zhou¹

¹Natural Language Processing Lab, Soochow University, Suzhou, China

²Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University

{wangzq.antony, sophiaym}@gmail.com, {lishoushan, gdzhou}@suda.edu.cn



Introduction

- Code-switching is commonly used in social media.
- Emotions in code-switching texts can be expressed in either *monolingual* or *bilingual* forms.
- Label propagation based approach is employed to learn and predict in the *bipartite graph* with both *bilingual* and *sentimental* information.

Examples of code-switching emotional posts

[E1] 我们已经自high起来了

(We are already getting hyper ourselves.)

Happiness (EN)

[E2] 最厌恶的一句话就是“爱情没有先来后到，不被爱的才是第三者”。shit!

(A quote, to my great disgust, is "There's no staking claims in a relationship based on who got there first - the one who isn't loved is the true third party." Shit!)

Anger (BOTH)

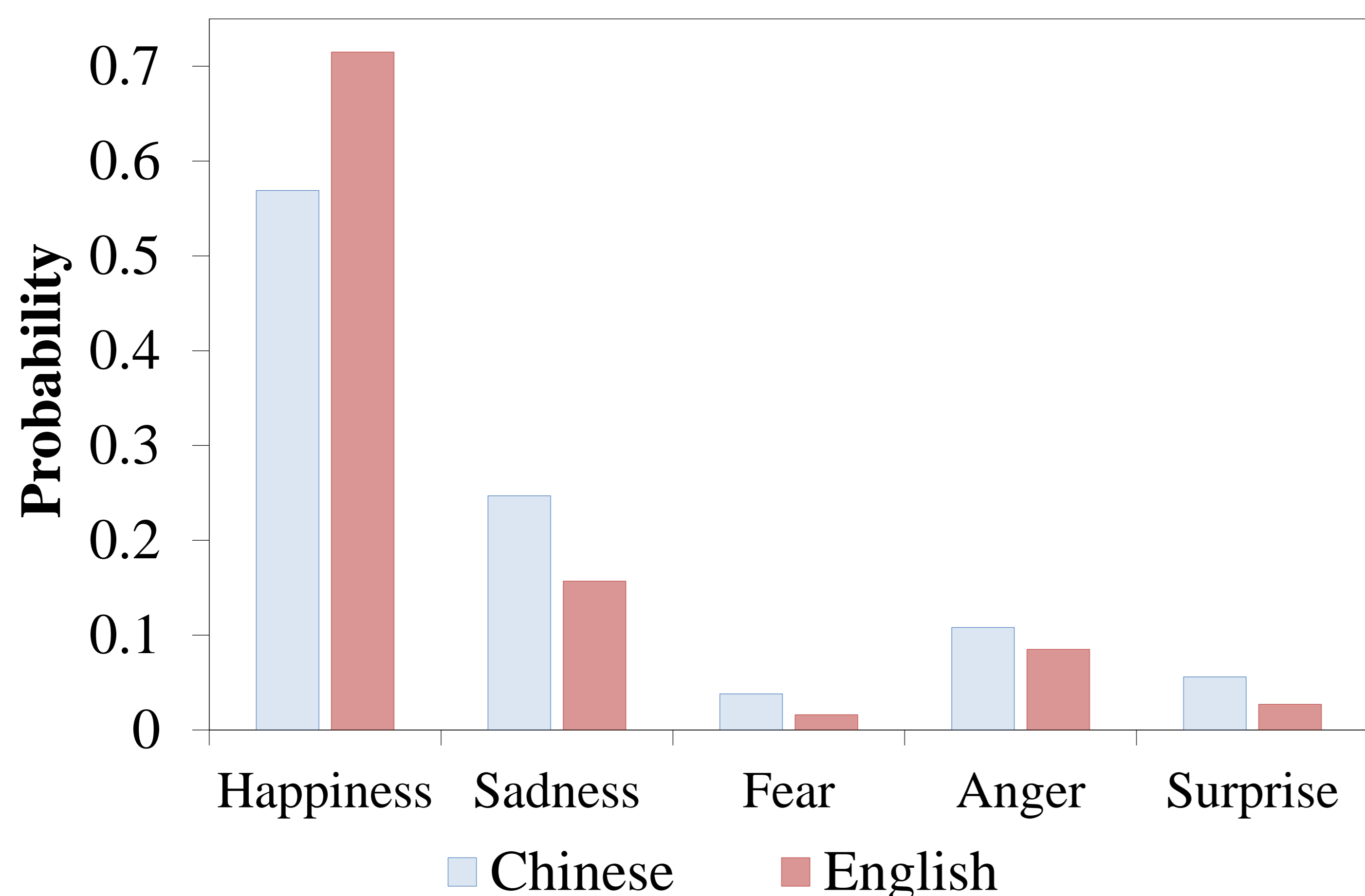
[E3] 这么个划重点法。。。窝们hold不住啊!!!

(The so-called "highlighting"...we can't hold it anymore.)

Fear(MIXED)

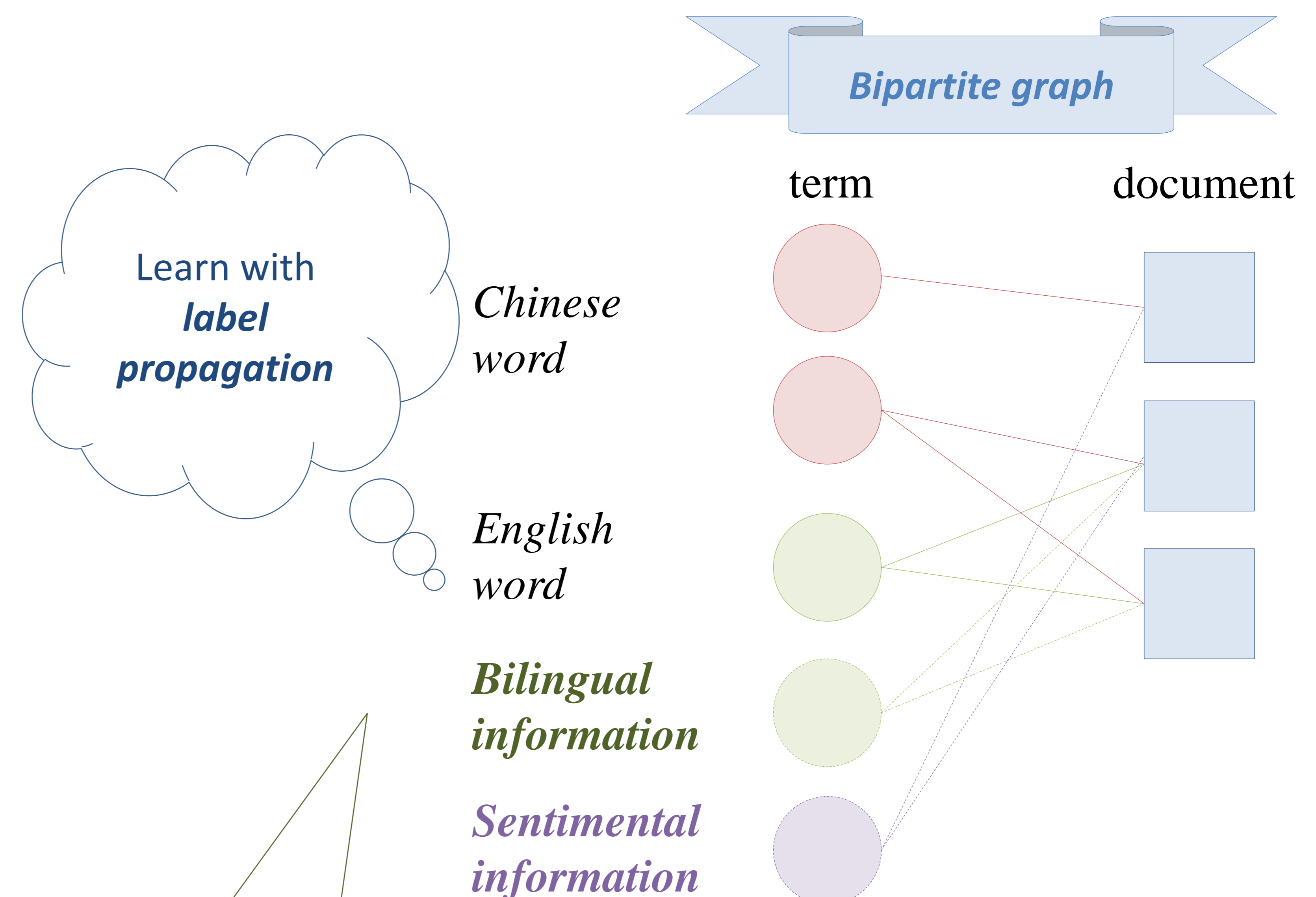
Data Collection

- We collect our data set from *Weibo.com*.
- we extract 4,195 code-switching posts for annotation.
- Five basic emotions are annotated, namely *happiness, sadness, fear, anger and surprise*.
- 81.4% of emotional posts are expressed through *Chinese*. 43.5% of emotion posts are caused by *English*.



Joint Distribution of Emotions and Caused Languages

LP-based Emotion Detection



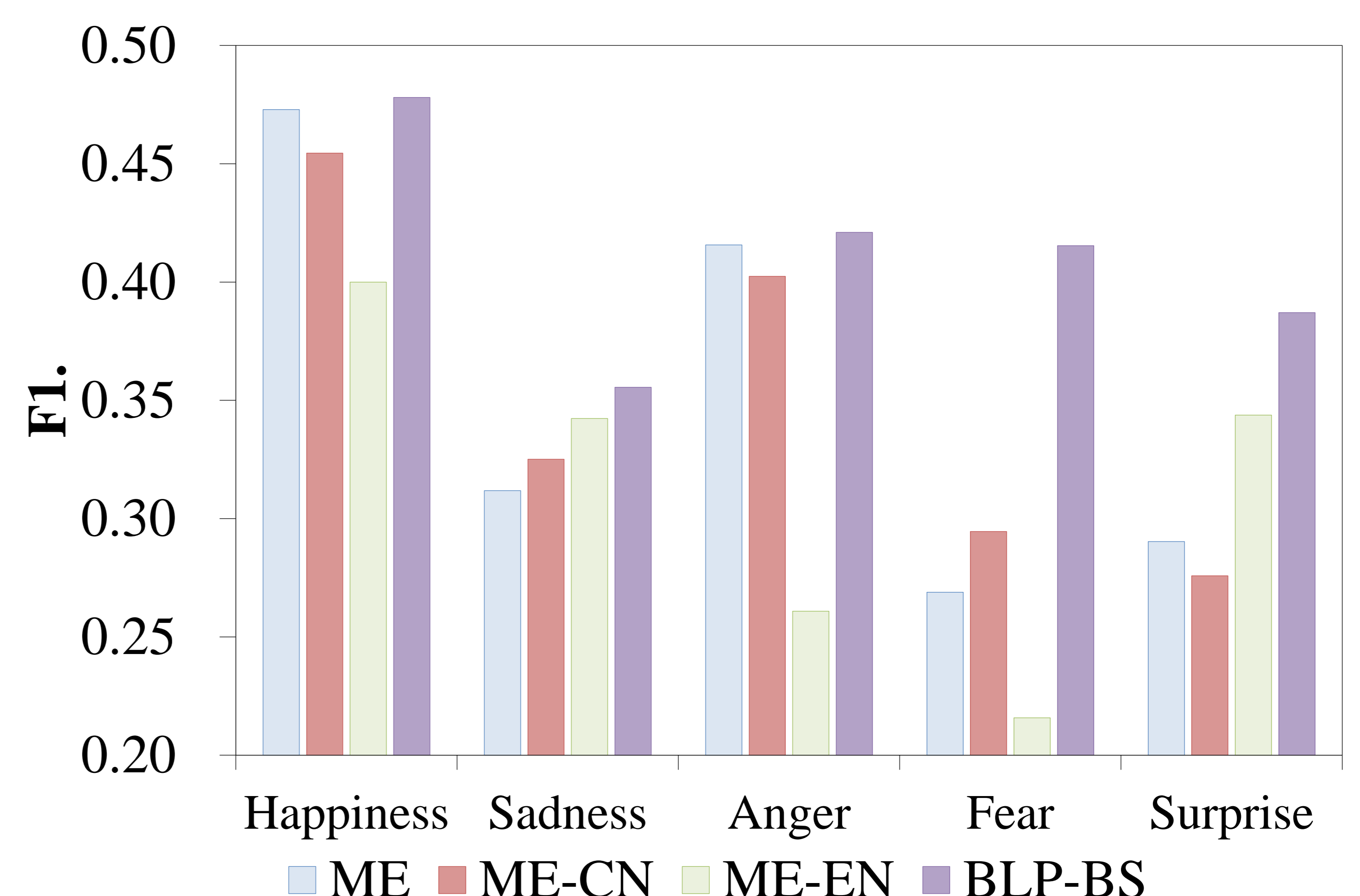
- Statistical machine translation is adopted to translate English words into Chinese.

- Synonym dictionary and PMI similar are used to enhance the language model for machine translation.

- Sentiment analysis is used to extract the polarity of both Chinese and English texts, and then connect words of similar polarity.

Experimental Result

- We randomly select *half* of the annotated posts as the training data and another *half* as the test data.
- F1-Measure (*F1.*) is adopted to measure the performance of each model in the respective emotions.



Results of emotion detection in code-switching texts