

NLPCC 2018 Shared Task Guideline:

Emotion Detection in Code-switching Text

This task aims to evaluate the techniques of automatic classification of emotion in code-switching text. Different from monolingual text, code-switching text contains more than one language, and the emotion can be expressed by either monolingual (e.g., 这个 show 真好看, 今天感觉很 happy) or bilingual form (e.g., 嗓子 hold 不住了啊). Hence, the challenges are: 1) how to integrate both monolingual and bilingual forms to detect emotion, and 2) how to bridge the gap to between two languages.

Note that, there may be more than one emotion in a post. Hence, our task is a multi-class classification task: you should predict all possible emotion for each post.

All kinds of strategies are welcome, including the traditional corpus-based distributional method, dictionary-based semantic computation, as well as the recently developed word embedding methods and deep learning strategies. Also, the participating system is encouraged to use any external resources.

1. Data

In this task, we provide training, development, and testing data. Each post in the dataset contains two language text (Chinese and English), we call such post as code-switched text. Following are some examples of code-switched posts with different emotions.

[E1] 婚礼上新娘大秀鼓技, **high** 翻全场! (*happiness*)
(The bride played the drums at the wedding, everyone was very **high**!)

[E2] 美好假期已经开始, **have a nice time**. (*happiness*)
(The happy vacation has begun, **have a nice time**.)

[E3] 上了一天的课, 嗓子 **hold** 不住了啊。 (*sadness*)
(I have been teaching the whole day, my throat **can't take it anymore**.)

For each post, five basic emotions were annotated, namely *happiness*, *sadness*, *fear*, *anger* and *surprise*. The distribution of different emotions in training and development data is on the following table.

	train	dev
happiness	0.304	0.302
sadness	0.181	0.165
anger	0.095	0.115
fear	0.108	0.117
surprise	0.109	0.126

Table 1: Distribution of different emotions

Note that, we already segment Chinese text of each post, you do not need to segment by yourself.

2. Evaluation

The format of submitted results should be same as training data, we show an example on the following. In addition, we also provide evaluation tool to evaluate the performance of your system and check the format before you submit the result.

```
<Tweet id="0">
  <Happiness>
    T
  </Happiness>
  <Sadness>
    F
  </Sadness>
  <Anger>
    F
  </Anger>
  <Fear>
    F
  </Fear>
  <Surprise>
    F
  </Surprise>
  <Content>
    今天 晚上 参加 beauty forever 的 周年 晚会 ， 感谢 annie 和
    eddie 的 邀请 ， 真 未 想到 会 拿 上 一 等 大 奖 ， 全 场 一
    百 多 人 我 成 为 幸 运 儿 ， 我 要 把 奖 金 去 助 养 儿 童 ， 把 快
    乐 分 享 给 有 需 要 的 孩 子 ， 真 hold 不 住 我 的 快 乐 ！
  </Content>
</Tweet>
```

Figure 1: Example of data format

We will compute precision (P), recall (R), and F1-Score for each emotion separately, and will compute the macro averaged P, R and F1 with all emotions. Our official scoring metric is macro-averaged F1-Score.

3. Baseline

We implement a simple baseline on training data, and test on the development data. The performance is shown on the following table. We use SVM¹ as the classification method, and unigram as features.

¹ <http://svmlight.joachims.org/>

	P.	R.	F1.
happiness	0.610	0.691	0.648
sadness	0.307	0.650	0.417
anger	0.299	0.667	0.413
fear	0.256	0.788	0.386
surprise	0.142	0.391	0.208

Table 2: Performance of baseline

4. Contact Information

For any questions about this shared task, please contact Zhongqing Wang from Soochow University. Email: wangzq@suda.edu.cn

Reference

- [1] Sophia Lee, and Zhongqing Wang. Emotion in Code-switching Texts: Corpus Construction and Analysis. In *Proceeding of SIGHAN-2015*.
- [2] Zhongqing Wang, Yue Zhang, Sophia Lee, Shoushan Li, and Guodong Zhou. A Bilingual Attention Network for Code-switched Emotion Prediction. In *Proceeding of COLING-2016*.
- [3] Zhongqing Wang, Sophia Lee, Shoushan Li, and Guodong Zhou. Emotion Analysis in Code-Switching Text with Joint Factor Graph Model. *Transactions on Audio, Speech and Language Processing*, 2017, 03.