NLPCC 2014 Shared Tasks Guidelines

Emotion Analysis in Chinese Weibo Texts

1. Introduction

This task aims to evaluate the techniques for analyzing the emotions in Weibo (Chinese twitter) texts. The bakeoff includes three sub-tasks, namely Weibo emotion classification (required), emotion sentence identification and classification (required), and emotion expression extraction (optional).

2. The Description of Sub-tasks:

2.1 Weibo Emotion classification

This sub-task aims to determine whether a Weibo text has emotion and classify the Weibo according to its emotion category. The target emotion category contains Anger, Disgust, Fear, Happiness, Like, Sadness and Surprise. Notice that one document may contain more than one kind of emotion. In this sub-task, participants are required to submit two dominant emotions for each piece of text.

(a) Data Description

The training data and testing data is in xml format, for instance:

<weibo id="1">

<sentence id="1">今天虽然依旧阴天，可是天气真的很暖和。

</sentence>

<sentence id="2">还是很开心的，保持个好心态，好运气会来的。

</sentence>

</weibo>

(b) Submission Format

Each team should submit one file containing all the results for this sub-task. Each line corresponds to one weibo text. It should include the following information and use '\t' as the separator:

Sub-task-ID: the id of sub-task, Weibo emotion classification is 1

System-ID: the name of the system

run-tag: the tag of submission, each team could submit 2 result at most.

run-type: the tag of close testing (label as C) or open testing (label as O), more details can be found in section 4.

weibo-ID: id of Weibo

emotion-tag: the tag of emotion, "Y" for documents which contains emotion and "N" for the rest.

emotion_1-type: the main emotion of the document: anger、disgust、fear、happiness、like、sadness or surprise. If no emotion exists, label it as "none".

emotion_2-type: the secondary emotion of document.

For the above example, the submission should be:

1system_name 1 C 1 Y happiness none

(c) Evaluation Metric

In the emotion classification sub task, evaluation metric is based on precision, recall and F-measure. For the top-2 emotion classification, we take average precision in multi-label classification as metric:

$$\text{Average}_{\text{Precision}}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{|Y_i|}\sum_{k=1}^{|Y_i|}\frac{\left|\left(emotion_k \in Y_i \middle| rank(x_i, emotion_k) \leq rank(x_i, y)\right)\right|}{rank(x_i, emotion_k)}$$

Here $rank(x_i, y)$ is the position of emotion y label in documentx_i. $|Y_i|$is the number of emotions in document x_i

2.2 Emotion Sentence Identification and Classification

This sub-task aims to identify the emotional sentences in each Weibo document and judge the emotion category for each sentence. Since one sentence may contain more than one emotion, each sentence should be labeled with 2 dominant emotions in it.

(a) Submission Format

Each team should submit one file containing all the results for this sub-task. Each line corresponds to one weibo sentence. It should include the following information and use '\t' as the separator:

ID: the id of task, emotion sentence classification is 2

System-ID: the name of the system

run-tag: the tag of submission, each team could submit 4 result at most, i.e. 2 open test and 2 close test.

run-type: the tag to distinguish between close test (C) and open test (O),more details can be found in section 4.

weibo-id: id of document

sentence-id: id of sentence.

emotion-tag: the tag of emotion, "Y" for sentences which contains emotion and "N" for the rest.

emotion_1-type: the main emotion of sentence: anger、disgust、fear、happiness、like、sadness、surprise. If no emotion exists, label it as "none".

emotion_2-type: the secondary emotion in the sentence.

For the following instance, the main emotion is happiness and the secondary emotion is sadness, such that the submitting format is:

2 system_name 1 C 2 1 Y happiness sadness

<weibo id="2">

<sentence id="1">没心没肺的时候，往往我们感到最开心，而后来，幸福，也都带着忧伤。</sentence >

</weibo>

(b) Evaluation Metric

In the recognition of emotion sentence sub task, metric is based on precision, recall and F-measure. For the top-2 emotion classification, we take average precision in multi-label classification as metric:

$$
\text{Average}_{\text{Precision}}
= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{|Y_i|}\sum_{k=1}^{|Y_i|}\frac{\left|\left(emotion_k \in Y_i \middle| rank(x_i, emotion_k) \leq rank(x_i, y)\right)\right|}{rank(x_i, emotion_k)}
$$

Here $rank(x_i, y)$ is the position of emotion y label in sentence x_i. $|Y_i|$ is the number of emotions in document x_i

2.3 Emotion expression extraction

In this sub task, participants should extract the emotion expression of each sentence. Notice that this sub task is optional and independent of the above tasks. Each sentence contains **at most two emotion expression**, the data is in xml format.

For instance:

<weibo id="3">

<sentence id="1">总想安装一个听歌的软件，终于找到了！赞一个

</sentence >

<sentence id="2">我觉得好开心！好激动！</sentence >

</weibo>

Sentence 1 contains one emotion expression "赞一个". In Sentence 2, the emotion expression is "好开心" and "好激动". Notice that the answer must be matched exactly. In this instance, the answer is "好开心", the submission of "开心" will be regarded as wrong. In micro-blog the icon is used commonly to express the emotion. However, this subtask focuses on the technique of text mining and information extraction, such that we **only evaluate the emotion expression in text.**

(a) Submission Format

Each team should submit one file containing all the results for this sub-task. Each line corresponds to one weibo sentence. It should include the following information and use '\t' as the separator:

ID: the id of sub task, emotion expression extraction sub-task is 3

System-ID: the name of the system

run-tag: the tag of submission, each team could submit 4 result at most, 2 open test and 2 close test.

run-type: run-type: the tag to distinguish between close test (C) and open test (O), more details can be found in section 4.

weibo-id: id of document

sentence-id: id of sentence.

sentence-id: ID of sentence.

emotion-expression_1: the 1st expression of emotion. If the sentence does

not contain emotion expression, it is labeled as "null".

emotion-expression_2: the 2nd expression of emotion. If the sentence contains only one emotion expression, it is labeled as "null".

In the above instance, the submission should be:

3 system_name 1 C 3 1 赞一个 null

3 system_name 1 C 3 2 好开心好激动

(b) Evaluation Metric

In the emotion expression extraction sub task, the evaluation metric is based on precision, recall on each sentence i:

$$\text{Precision}_i = \frac{\#system\_correct(expression \neq null)}{\#system\_proposed(expression \neq null)}$$

$$\text{Recall}_i = \frac{\#system\_correct(expression \neq null)}{\#gold(expression \neq null)}$$

Here, #gold is the manually labeled number of emotion expression in sentence i, #system_correct is the matched number in submission of emotion expression in sentence i, #system_proposed is the number of submission of emotion expression in sentence i.

Based on sentence precision, we proposed **sentence level** and **document level** evaluation metric:

Sentence level:

$$\text{Precision}_{sentence} = \frac{\sum_i Precision_i}{\#system\_proposed\_sentence(expression \neq null)}$$

$$\text{Recall}_{sentence} = \frac{\sum_i recall_i}{\#gold\_sentence(expression \neq null)}$$

$$\text{F} - \text{measure}_{sentence} = \frac{2 \times Precision_{sentence} \times Recall_{sentence}}{Precision_{sentence} + Recall_{sentence}}$$

Here, #gold_sentence is number of sentence contains emotion expression in manually notation. #system_proposed_sentence is the number of sentence contains emotion expression in submission.

Document level:

$$\text{Precision}_{document} = \sum_{d \in D} \frac{\sum_{i \in d} \frac{Precision_i}{\#system\_proposed\_in\_d(expression \neq null)}}{\#system\_proposed\_documents(expression \neq null)}$$

$$\text{Recall}_{document} = \sum_{d \in D} \frac{\sum_{i \in d} \frac{Recall_i}{\#gold\_in\_d(expression \neq null)}}{\#gold\_documents(expression \neq null)}$$

$$\text{F} - \text{measure}_{document} = \frac{2 \times Precision_{document} \times Recall_{document}}{Precision_{document} + Recall_{document}}$$

Here, #gold_documents is number of document contains emotion expression in

manually notation. #gold_in_d is the number of sentence contains emotion expression in manually notation in document *d*. #system_proposed_sentence is the number of document contains emotion expression in submission. #system_proposed_in_d is the number of sentence contains emotion expression in manually notation in document *d*.

3. Data

In this evaluation, the data are collected from Sina Weibo, in xml format, Unicode encoding (utf-16).

4. Evaluation Arrangement

The evaluation is off-line conducted. Each team should utilize fully automatical method in emotion analysis. In order to keep the objectivity of evaluation, there are open test and close test in evaluation. Here, open test **do not** limit the NLP tools, extra training data or emotion knowledge related resources. As for the close test, we restrict the NLP tools (Stanford parser), emotion knowledge resource (emotion words noumenon by Dalian University of Technology, knowledge base of emotion common sense by Harbin Institute of Technology), and training data (contains NLP&CC 2014 and NLP&CC 2013 emotion analysis in Chinese Weibo training data). Extra resources or tools are not allowed except Chinese segmentation and POS tagging tools.