WUST CRF-Based System at NTCIR-13 ECA Task

Maofu Liu, Linxu Xia, Zhenlian Zhang, Yang Fu

School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065, China

liumaofu@wust.edu.cn

Introduction

◆ In NTCIR-13, Emotion Cause Analysis (ECA) is an evaluation task aiming to detect the emotion cause description from the text, which is a common problem shared widely among researchers of Natural Language Processing (NLP) and information access.

◆ We regard the task that the emotion cause detection at clause as a clause-level binary text classification problem. The clauses will be classified according to containing emotion cause or not.

◆ In our system, we apply the Conditional Random Field (CRF) classification model to identifying the emotion cause description with a series of features, such as basic word features, POS features, distance features, and contextual features.

System Architecture

◆ Our system consist of three main modules, i.e. data preprocessing, feature extraction, and CRF classifier.
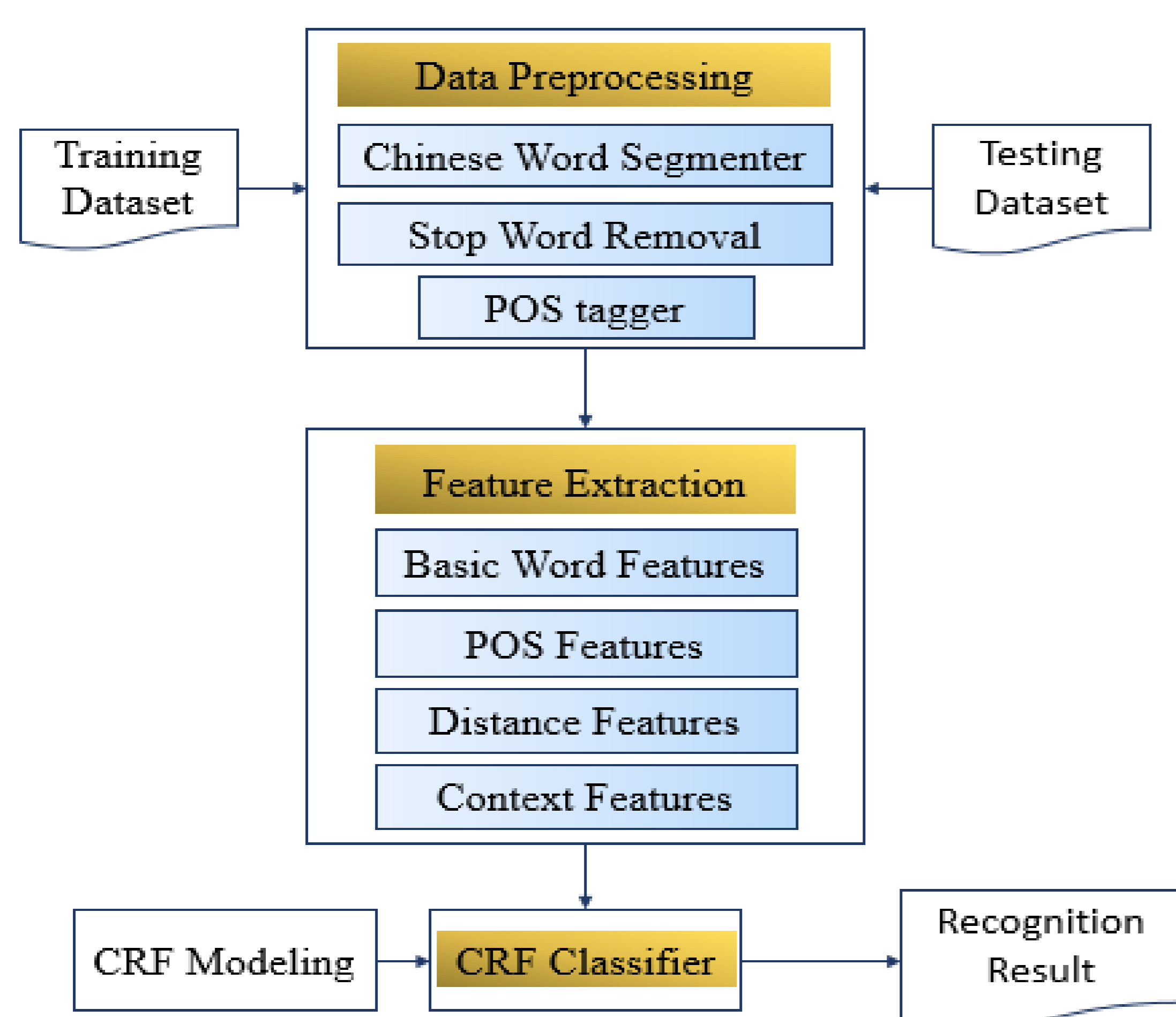
1. Data preprocessing

◆ In the data preprocessing, the system mainly implements the Chinese word segmentation, which removes the stop words according to the stop word list and part of speech (POS) tagging. We choose Jieba Chinese word segmenter to segment the Chinese word.

2. Feature extraction

◆ In this subsection, we divide feature sets into basic features and contextual features.

◆ The basic features of the sentence include nouns, verbs and their numbers. Only nouns and verbs are considered as the emotion cause, which is mainly composed of noun phrases and verbal phrases.

◆ The contextual features mainly contain the syntactic ones of the previous sentence and the next sentence, and they can be represented by feature templates.

◆ Generally speaking, the emotion cause can be found around emotional words, so the distance feature is an important location feature.

◆ In addition, the categories of emotional words can also provide information for the classification model, so we also take the classification features of emotional words into consideration.

3. CRF Classifier

◆ the CRF is a supervised learning model with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. The following Figure 1 describes the schematic diagram of CRF classification model in our system.
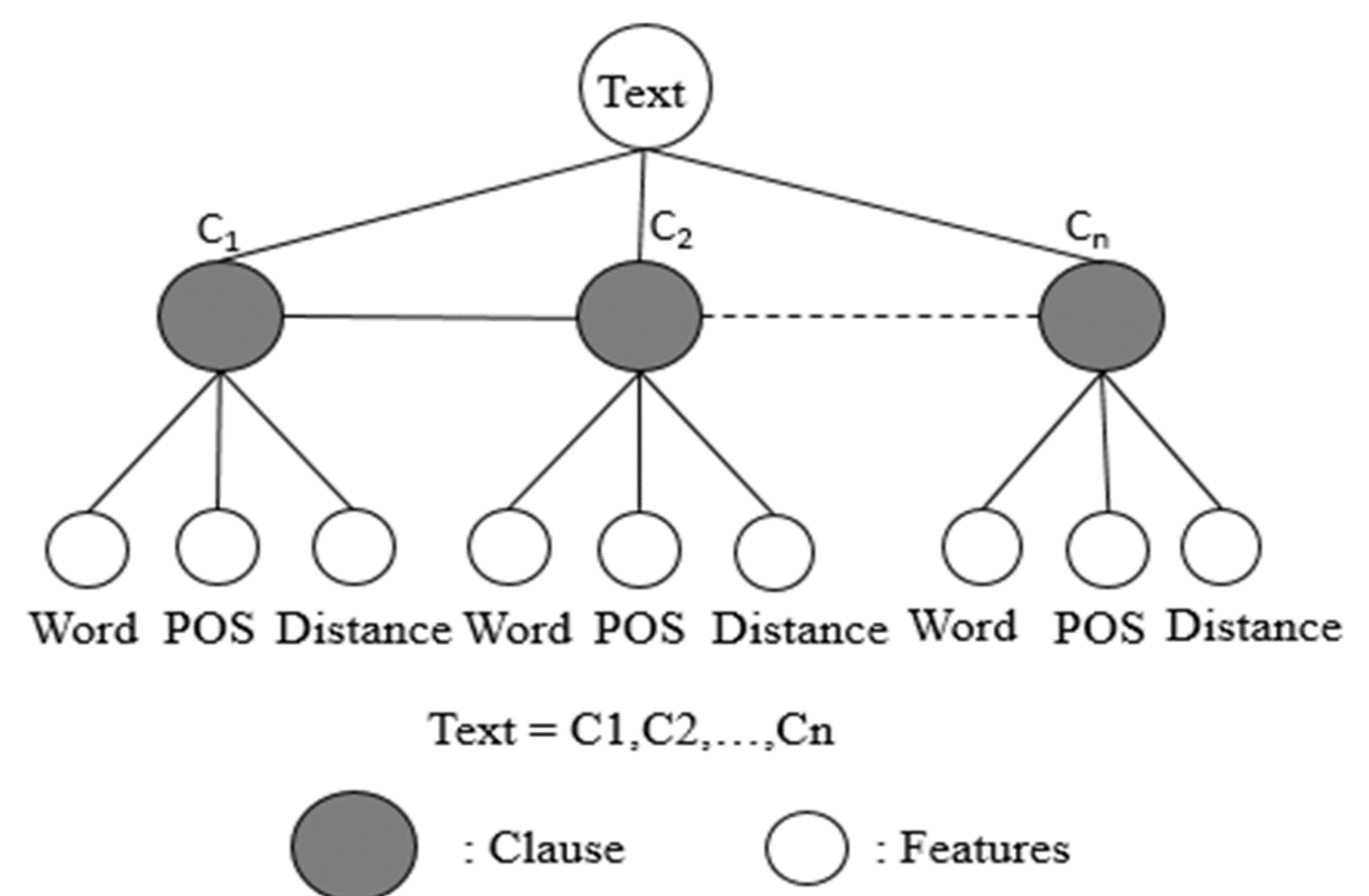

Figure 1. Schematic diagram of CRF classification model

◆ The specific formula is shown in formula (1)

$$P_\lambda(Y|W) = \frac{1}{Z(W)}\exp\left(\sum_{t\in T}\sum_k \lambda_k f_k(y_{t-1}, W, t)\right) \quad (1)$$

Experiments

◆ We submitted one system result to NTCIR-13 of emotion cause detection at clause level. The official evaluation results of performance are listed in the Table 1.

Table 1. Performances of the CRF model

Team	P	R	F
WUST	0.6930	0.6399	0.6654

From the Table 1, we can see that the CRF model has achieved good performance in the emotion cause description detection.

Conclusions

◆ In our work, we construct the classification model based on conditional random field to recognize emotion cause in Chinese text pair using the hybrid features, including words, POS, distance and context.

◆ In the future, we will introduce more features into the Emotion Cause Analysis, such as coherent and semantic features. The other methods, like neural network based on attention model, will be also applied to detecting the emotion cause in our future work.