

## SemEval-2010 Task 11: Event detection in Chinese news sentences

Qiang Zhou

Tsinghua University, Beijing 100084, P. R. China

zq-lxd@mail.tsinghua.edu.cn

The goal of the task is to detect and analyze the event contents in real world Chinese news texts. It consists of finding key verbs or verb phrases to describe these events in the Chinese sentences after word segmentation and part-of-speech tagging, selecting suitable situation descriptions for them, and anchoring different situation arguments with suitable syntactic chunks in the sentence. Three main sub-tasks are as follows: (1) Target verb WSD; (2) Sentence SRL; (3) Event detection.

We will select 100 high-frequency Chinese target verbs for this task. Among them, 30 verbs have multiple senses and 70 verbs have single sense. Each target verb will be assigned more than 50 annotated sentences to consist of training and test sets. Each annotated sentence will have following event information: (1) word segmentation and POS tags; (2) the target verb (or verb phrase) and its position in the sentence; (3) the event description (situation description formula or natural explanation text) of the target verb (or verb phrase) in the context of the sentences; (4) the chunks annotated with suitable syntactic constituent tags, functional tags and event argument role tags. The training and test set will be extracted from the data set with ratio 8:2.

For the WSD subtask, we give two evaluation measures: WSD-Micro-Accuracy and WSD-Macro-Accuracy. The correct conditions are: the selected situation description formula and natural explanation text of the target verbs will be same

with the gold-standard codes. We evaluated 27 multiple-sense target verbs in the test set.

For the SRL subtask, we give three evaluation measures: Chunk-Precision, Chunk-Recall, and Chunk-F-measure. The correct conditions are: the recognized chunks should have the same boundaries, syntactic constituent and functional tags, and situation argument tags with the gold-standard argument chunks of the key verbs or verb phrases. We only select the key argument chunks (with semantic role tags: x, y, z, L or O) for evaluation.

For the event detection subtask, we give two evaluation measures: Event-Micro-Accuracy and Event-Macro-Accuracy. The correct conditions are: (1) The event situation description formula and natural explanation text of the target verb should be same with the gold-standard ones; (2) All the argument chunks of the event descriptions should be same with the gold-standard ones; (3) The number of the recognized argument chunks should be same with the gold-standard one.

8 participants downloaded the training and test data. Only 3 participants uploaded the final results. Among them, 1 participant (User ID = 156) submitted 4 results and 1 participant (User ID = 485) submitted 2 results. So we received 7 uploaded results for evaluation. The mean elaboration time of the test data is about 30 hours. The following is the evaluation result table. All the results are ranked with Event-Macro-Accuracy.

User ID	System ID	WSD-Micro-A	WSD-Macro-A	Chunk-P	Chunk-R	Chunk-F	Event-Micro-A	Event-Macro-A	Rank
485	480-a	87.54	89.59	80.91	77.91	79.38	52.12	53.76	1
485	480-b	87.24	89.18	80.91	76.95	78.88	50.59	52.05	2
303	109	73.00	70.64	63.50	57.39	60.29	22.85	23.05	3
156	348	79.23	82.18	58.33	53.32	55.71	20.05	20.23	4
156	350	77.74	81.42	58.33	53.32	55.71	20.05	20.22	5
156	347	81.30	83.81	58.33	53.32	55.71	20.33	20.19	6
156	349	79.82	82.58	58.33	53.32	55.71	20.05	20.14	7

The results show the event detection task is still an open problem for exploring in the Chinese language.