

# Shared Tasks in NLPCC 2018

## Call for Participation

The conference on Natural Language Processing and Chinese Computing (NLPCC) is the annual conference of CCF TCCI (Technical Committee of Chinese Information, China Computer Federation). The NLPCC conferences have been successfully held in Beijing (2012), Chongqing (2013), Shenzhen (2014), Nanchang (2015), Kunming (2016) and Dalian (2017). This year's NLPCC conference will be held in Hohhot on August 26 - 30, 2018.

NLPCC 2018 will follow the NLPCC tradition of holding several shared tasks in natural language processing and Chinese computing. This year's shared tasks focus on both classic problems and newly emerging problems, including **Emotion Detection in Code-Switching Text, Grammatical Error Correction, Single Document Summarization, Spoken Language Understanding in Task-Oriented Dialogue Systems, Multi-Turn Human-Computer Conversations, Automatic Tagging of Zhihu Questions, Open Domain Question Answering, and User Profiling and Recommendation.**

Participants from both academia and industry are welcomed. Each group can participate in one or multiple tasks and members in each group can attend the NLPCC conference to present their techniques and results. The participants will be invited to submit papers to the main conference and the accepted papers will appear in the conference proceedings published by Springer LNCS.

### 1. Overview of the Shared Tasks

There are eight shared tasks in this year's NLPCC conference and the detailed description of each task can be found in the task guidelines to be released. Here we only give a brief overview of each task.

#### ➤ **Task 1 - 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 contain

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.

#### ➤ **Task 2 - Grammatical Error Correction**

With the expanding influence of China, learning Mandarin Chinese has grown in popularity around the world. Whereas the study of second language learning has started years ago, the specific research for CSL (Chinese as a Second Language) still has a long way to go. NLPCC 2018 Task 2 will be grammatical error correction for Chinese. The goal of the task is to develop techniques to automatically detect and correct errors made by writers of CSL. We will provide large-scale Chinese texts written by non-native speakers in which grammatical errors have been annotated and corrected by native speakers. Blind test data will be used to evaluate the outputs of the participating teams using a common scoring software and evaluation metric.

#### ➤ **Task 3 - Single Document Summarization**

This task provides a dataset for single document summarization of Chinese news articles, to evaluate and compare different document summarization techniques.

#### ➤ **Task 4 - Spoken Language Understanding in Task-Oriented Dialogue Systems**

This task aims to evaluate the Spoken Language Understanding (SLU), which includes intent classification and slot filling. We will provide a dataset generated from a commercial task-oriented dialogue system, with the noisy transcripts automatically recognized from spoken utterances and the corrected SLU results.

#### ➤ **Task 5 – Multi-Turn Human-Computer Conversations**

In this year's NLPCC intelligent conversation task, we focus on how to utilize contexts to conduct multi-turn human-computer conversations. The task contains two parts: (1) response retrieval, which means to find the original response given a particular query given the contexts and (2) response generation, which means a new utterance will be generated to respond the query. Both sub-tasks will be based on human-to-human conversation data in Chinese.

➤ **Task 6 - Automatic Tagging of Zhihu Questions**

The task aims to tag questions in Zhihu with relevant tags from a collection of predefined ones. Accurate tags can benefit several downstream applications such as recommendation and search of Zhihu questions.

➤ **Task 7 - Open Domain Question Answering**

In this year's NLPCC open domain QA share task, we focus on KNOWLEDGE and propose three sub-tasks, including (a) **knowledge-based question answering (KBQA)**, (b) **knowledge-based question generation (KBQG)**, and (c) **knowledge-based question understanding (KBQU)**. The task of KBQA is to answer natural language questions based on a given knowledge base. The task of KBQG is to generate natural language questions based on given knowledge base triples. The task of KBQU is to transform natural language questions into their corresponding logical forms. The first two sub-tasks are in Chinese, while the last sub-task is in English.

➤ **Task 8 - User Profiling and Recommendation**

In this year's NLPCC user modeling share task, we focus on two sub-tasks, including (a) **user tags prediction (UTP)**, and (b) **user following recommendation (UFR)**. The task of UTP is to predict which tags are related to a user. The task of UFR is to recommend users a user would like to follow. The two sub-tasks use social media data in China.

## 2. **How to Participate**

Please fill out the registration form and send it to the coordinator Fang Liu(刘芳) by email ([contact@nlpcc2018.info](mailto:contact@nlpcc2018.info)) before March 31, 2018.

If you have any question about the shared tasks, please do not hesitate to contact us by email.

## 3. **Important Dates**

2018/1/2: announcement of shared tasks and call for participation;

2018/3/1: release of detailed task guidelines & sample data release;

2018/3/31: registration deadline;  
2018/4/23: test data release;  
2018/4/30: participants' results submission deadline;  
2018/5/15: evaluation results release and call for system reports and conference papers;  
2018/6/5: conference paper submission deadline (only for shared tasks);  
2018/6/17: conference paper accept/reject decision meeting;  
2018/6/20: conference paper accept/reject notification;  
2018/6/30: camera-ready paper submission deadline;  
2018/8/26~30: NLPCC 2018 main conference;

#### **4. Shared Task Organizers (in alphabetical order)**

Yunbo Cao, Tencent Dingdang  
Nan Duan, Microsoft Research Asia  
Minlie Huang, Tsinghua University  
Daren Li, Zhihu  
Lei Li, Toutiao Lab  
Shoushan Li, Soochow University  
Yajuan Lv, Baidu  
Lin Pu, Zhihu  
Weiwei Sun, Peking University  
Xiaojun Wan, Peking University  
Zhongqing Wang, Soochow University  
Rui Yan, Peking University  
Fuzheng Zhang, Microsoft Research Asia  
Xuemin Zhao, Tencent Dingdang