Overview of the NLPCC 2018 Shared Task: Open Domain QA

Nan Duan[(✉)]

Microsoft Research Asia, Beijing, China
nanduan@microsoft.com

Abstract. We give the overview of the open domain QA shared task in the NLPCC 2018. In this year, we release three sub-tasks including Chinese knowledge-based question answering (KBQA) task, Chinese knowledge-based question generation (KBQG) task, and English knowledge-based question understanding (KBQU) task. The evaluation results of final submissions from participating teams will be presented in the experimental part.

Keywords: Question answering · Question generation
Question understanding

1 Background

Question Answering (or QA) is a fundamental task in Artificial Intelligence, whose goal is to build a system that can automatically answer natural language questions. In the last decade, the development of QA techniques has been greatly promoted by both academic and industry fields, and many QA-related topics have been well studied by researchers from all over world.

In order to further advance QA-related research in China, we organize this open domain QA shared task series in the past several years via NLPCC, and in this year, we release following 3 sub-tasks: (1) Chinese Knowledge-based Question Answering (KBQA); (2) Chinese Knowledge-based Question Generation (KBQG); and (3) English Knowledge-based Question Understanding (KBQU). You can see that comparing to previous two shared tasks, we retain the KBQA task and add KBQG and KBQU as two new tasks. The reason of adding these two new tasks is that we think the capabilities of asking questions in a proactive way and understanding user utterances in a deep way are very important to building human-computer interaction engines, such as search engine, chitchat bot, and task bot.

2 Task Description

The NLPCC 2018 open domain QA shared task includes 2 sub-tasks for Chinese language: KBQA and KBQG, and 1 sub-task for English language: KBQU.

2.1 KBQA Task

For KBQA task, we provide a train set and a test set. In train set, both questions and their golden answers are provided. In test set, only questions are provided. The participating teams should predict an answer for each question in test set, based on a given large-scale Chinese KB. If no answer can be predicted for a given question, just set the value of <answer id="X"> to an empty string. The quality of a KBQA system will be evaluated by answer exact match. An example in train set is given below:

<question id="X">	你知道迪克牛仔是什么属相吗？
<answer id="X">	鼠

We provide a large-scale Chinese KB to participating teams, and it includes knowledge triples crawled from web. Each knowledge triple has the form: <Subject, Predicate, Object> , where 'Subject' denotes a subject entity, 'Predicate' denotes a relation, and 'Object' denotes an object entity. A sample of knowledge triples is given in Fig. 1, and the statistics of the Chinese KB is given in Table 1.

```
新还珠格格  | | entity.primaryName  ||| 新还珠格格
新还珠格格  | | 中文名  ||| 新还珠格格
新还珠格格  | | 外文名  ||| New my fair Princess
新还珠格格  | | 出品时间  ||| 2011年和2014年
新还珠格格  | | 出品公司  ||| 上海创翊文化传播有限公司
新还珠格格  | | 制片地区  ||| 中国大陆，中国台湾
新还珠格格  | | 拍摄地点  ||| 横店影视城
新还珠格格  | | 发行公司  ||| 上海创翊文化传播有限公司
新还珠格格  | | 首播时间  ||| 2011年7月16日
新还珠格格  | | 导演  ||| 李平，丁仰国
新还珠格格  | | 编剧  ||| 琼瑶，黄素媛
新还珠格格  | | 主演  ||| 李晟，海陆，张睿，李佳航，潘杰明，赵丽颖，邱心志，邓萃雯，刘雪华
新还珠格格  | | 集数  ||| 总共98集→第一部1至37集→第二部37至74集→第三部74至98集
新还珠格格  | | 每集长度  ||| 前三部：45分钟  第四部：48分钟
新还珠格格  | | 类型  ||| 古装，爱情，励志，喜剧
新还珠格格  | | 上映时间  ||| 前三部：2011年07月16日至2011年9月8日第四部：2016年暑期档
新还珠格格  | | 在线播放平台  ||| 芒果TV,PPTV,暴风影音，优酷，搜狐。
新还珠格格  | | 总策划  ||| 杨文红，苏晓
新还珠格格  | | 出品人  ||| 欧阳常林
新还珠格格  | | 总监制  ||| 魏文彬
新还珠格格  | | entity.description  ||| 《新还珠格格》翻拍自琼瑶经典之作《还珠格格》，由李晟、渀
```

Fig. 1. An example of the Chinese KB.

Table 1. Statistics of the Chinese KB.

# of Subject Entities	8,721,640
# of Triples	47,943,429
# of Averaged Triples per Subject Entity	5.5

2.2   KBQG Task

For KBQG task, we provide a train set and a test set. In train set, both triples and their golden questions are provided. In test set, only triples are provided. The participating teams should generate a natural language question for each triple in test set, and this generated question can be answered by the object entity of the given triple. The quality of a KBQG system will be evaluated by BLEU-4. An example in train set is given below:

<triple id="X">	微软 ⦀ 创始人 ⦀ 比尔盖茨
<question id="X">	是谁创办了微软公司？

2.3   KBQU Task

For KBQU task, we provide a train set and a test set. In train set, both questions and their golden logical forms are provided. In test set, only questions are provided. The participating teams should predicate a logical form for each question in test set. The quality of a KBQU system will be evaluated by logical form exact match. An example in train set is given below:

<question id="X">	what is fight songs of Maryland
<logical form id="X">	(lambda ?x (sports.team.fight_song Maryland ?x))

3   Evaluation Results

There are 19 submissions to the KBQA task, and Table 2 lists the evaluation results.

Table 2. Evaluation results of the KBQA task.

Organization	System name	Answer extract match
Central China Normal University	CCNU-319	0.3900
天津深思维科技有限公司	DeeperThought	0.4498
Chinese Academy of Sciences, Institute of Automation	Dream_on_Road	0.4676
Dalian University of Technology	DUTIR_9147	0.2832
East China Normal University	ECNU	0.5275
Guangdong University of Foreign Studies	GDUFSLEC	0.0971
China Academy of Engineering Physics	Lawe	0.3366
AI Lab, Lenovo Research	LEQAU 主系统	0.5647
AI Lab, Lenovo Research	LEQAU副系统a	0.5696

Table 2. (*continued*)

Organization	System name	Answer extract match
AI Lab, Lenovo Research	LEQAU副系统b	0.5502
NetDragon Websoft Inc	NDers	0.6294
Northeastern University	NEUQA	0.5825
Peking University	Pkult	0.4984
Southeast University	SEU- WDS- KBQA	0.6926
Soochow University	SUDA-HLT	0.4337
University of Science and Technology of China	USTC-NELSLIP	0.5647
Xi'an Jiao Tong University	XJBot	0.6294
Zhejiang University	Yiwise-KBQA	0.6359
Central China Normal University	zilean	0.5023

There are 9 submissions to the KBQG task, and Table 3 lists the evaluation results.

Table 3. Evaluation results of the KBQG task.

Organization	System name	Character BLEU-4
Southwest University	AQG	0.4723
Southwest University	AQG-PAC_greedy_relation_predict	0.4360
Southwest University	AQG-PAC_soft_relation_predict	0.4188
Southwest University	AQG-question_sentence&relation_predict	0.4141
Central China Normal University	CCNU-319	0.4131
Peking University	ICL-1	0.4781
Peking University	ICL-2	0.3820
Southeast University	LPAI	0.3647
Central China Normal University	unique AI group	0.3652

There are 3 submissions to the KBQU task, and Table 4 lists the evaluation results.

Table 4. Evaluation results of the KBQU task.

Organization	System name	Logical form exact match
AI Lab, Lenovo Research	LEQAU 主系统	0.3020
AI Lab, Lenovo Research	LEQAU 副系统a	0.3060
AI Lab, Lenovo Research	LEQAU 副系统b	0.1900

4   Conclusion

This paper briefly introduces the overview of this year's 3 open domain QA shared tasks. In the future, we plan to provide more QA datasets for Chinese QA field. In the future, we will build more datasets for QA research, such as multi-turn QA dataset and cross-lingual QA dataset.