ACL 2018

Machine Reading for Question Answering

Proceedings of the Workshop

July 19, 2018
Melbourne, Australia

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-948087-39-1

Introduction

Machine Reading for Question Answering (MRQA) has become an important testbed for evaluating how well computer systems understand human language, as well as a crucial technology for industry applications such as search engines and dialog systems. The research community has recently created a multitude of large-scale datasets over text sources such as Wikipedia (WikiReading, SQuAD, WikiHop), news and other articles (CNN/Daily Mail, NewsQA, RACE), fictional stories (MCTest, CBT, NarrativeQA), and general web sources (MS MARCO, TriviaQA, SearchQA). These new datasets have in turn inspired an even wider array of new question answering systems.

Despite this rapid progress, there is yet much to understand about these datasets and systems. While model performance is rapidly improving in domain for each dataset, generalization suffers when models are evaluated on new domains and datasets. Moreover, current model development focuses primarily on improving the test accuracy of models trained on in-domain data. Focusing solely on accuracy obscures other important desiderata, including model interpretability, robustness to distributional shift, ability to abstain from answering when there is no adequate answer, and adequate modeling of inference (e.g., entailment and multi-sentence reasoning). Similarly, the diversity of recent datasets call for an analysis of various natural language phenomena (coreference, paraphrase, entailment, multi-hop reasoning) these datasets present.

The goal of this workshop is to gather researchers to address and discuss recent research on MRQA systems and datasets. Currently, reading comprehension models are commonly presented at various venues such as ACL, EMNLP, NIPS, ICML, ICLR; this workshop will benefit the community by serving as a central venue for discussions.

The program features 11 new papers and 5 cross-submissions from related areas, to be presented as both posters and talks. We are also excited to host remarkable invited speakers, including Phil Blunsom, Antoine Bordes, Jianfeng Gao, Hannaneh Hajishirzi, Sebastian Riedel, Richard Socher.

We thank the program committee, the ACL workshop chairs, the invited speakers, our sponsors Facebook and Naver and our steering committee: Antoine Bordes, Percy Liang, Luke Zettlemoyer.

Eunsol Choi, Minjoon Seo, Danqi Chen, Robin Jia, Jonathan Berant

Organizing Committee:

Eunsol Choi, University of Washington
Minjoon Seo, NAVER & University of Washington
Danqi Chen, Stanford University
Robin Jia, Stanford University
Jonathan Berant, Tel-Aviv University

Program Committee:

Yoav Artzi
Danish Contractor
Rajarshi Das
Bhuwan Dhingra
Xinya Du
Matt Gardner
Mor Geva
Kevin Gimpel
Luheng He
Jonathan Herzig
Mohit Iyyer
Mandar Joshi
Dongyeop Kang
Ni Lao
Kenton Lee
Nasrin Mostafazadeh
Karthik Narasimhan
Rodrigo Nogueira
Panupong (Ice) Pasupat
Hoifung Poon
Siva Reddy
Xiang Ren
Tim Rocktäschel
Shimon Salant
Swabha Swayamdipta
Kristina Toutanova
Adam Trischler
Shuohang Wang
Tong Wang
Johannes Welbl
Caiming Xiong
Victor Zhong

Table of Contents

Ruminating Reader: Reasoning with Gated Multi-hop Attention
Yichen Gong and Samuel Bowman 1

Systematic Error Analysis of the Stanford Question Answering Dataset
Marc-Antoine Rondeau and T. J. Hazen 12

A Multi-Stage Memory Augmented Neural Network for Machine Reading Comprehension
Seunghak Yu, Sathish Reddy Indurthi, Seohyun Back and Haejun Lee 21

Tackling Adversarial Examples in QA via Answer Sentence Selection
Yuanhang Ren, Ye Du and Di Wang 31

DuReader: a Chinese Machine Reading Comprehension Dataset from Real-world Applications
Wei He, Kai Liu, Jing Liu, Yajuan Lyu, Shiqi Zhao, Xinyan Xiao, Yuan Liu, Yizhong Wang, Hua Wu, Qiaoqiao She, Xuan Liu, Tian Wu and Haifeng Wang 37

Robust and Scalable Differentiable Neural Computer for Question Answering
Jörg Franke, Jan Niehues and Alex Waibel 47

A Systematic Classification of Knowledge, Reasoning, and Context within the ARC Dataset
Michael Boratko, Harshit Padigela, Divyendra Mikkilineni, Pritish Yuvraj, Rajarshi Das, Andrew McCallum, Maria Chang, Achille Fokoue-Nkoutche, Pavan Kapanipathi, Nicholas Mattei, Ryan Musa, Kartik Talamadupula and Michael Witbrock 60

RECIPE: Applying Open Domain Question Answering to Privacy Policies
Yan Shvartzshanider, Ananth Balashankar, Thomas Wies and Lakshminarayanan Subramanian . 71

Neural Models for Key Phrase Extraction and Question Generation
Sandeep Subramanian, Tong Wang, Xingdi Yuan, Saizheng Zhang, Adam Trischler and Yoshua Bengio 78

Comparative Analysis of Neural QA models on SQuAD
Soumya Wadhwa, Khyathi Chandu and Eric Nyberg 89

Adaptations of ROUGE and BLEU to Better Evaluate Machine Reading Comprehension Task
An Yang, Kai Liu, Jing Liu, Yajuan Lyu and Sujian Li 98

Workshop Program

Thursday, July 19, 2018

8:45–9:00 *Opening Remarks*

Session 1

9:00–9:35 ***Invited Talk: Phil Blunsom, University of Oxford / Deepmind***

9:35–10:10 ***Invited Talk: Sebastian Riedel, University College London***

10:10–10:30 ***Best Paper Talk: A Systematic Classification of Knowledge, Reasoning, and Context within the ARC Dataset***

10:30–11:00 *Morning coffee break*

Session 2

11:00–11:35 ***Invited Talk: Richard Socher, Salesforce Research***

11:35–12:10 ***Invited Talk: Jianfeng Gao, Microsoft Research***

12:10–13:45 *Lunch*

Thursday, July 19, 2018 (continued)

Session 3

13:45–14:20 ***Invited Talk: Antoine Bordes, Facebook AI Research***

14:20–15:30 ***Poster Session (with a spotlight presentation)***

14:20-15:30 *Ruminating Reader: Reasoning with Gated Multi-hop Attention*
Yichen Gong and Samuel Bowman

14:20-15:30 *Systematic Error Analysis of the Stanford Question Answering Dataset*
Marc-Antoine Rondeau and T. J. Hazen

14:20-15:30 *A Multi-Stage Memory Augmented Neural Network for Machine Reading Comprehension*
Seunghak Yu, Sathish Reddy Indurthi, Seohyun Back and Haejun Lee

14:20-15:30 *Tackling Adversarial Examples in QA via Answer Sentence Selection*
Yuanhang Ren, Ye Du and Di Wang

14:20-15:30 *DuReader: a Chinese Machine Reading Comprehension Dataset from Real-world Applications*
Wei He, Kai Liu, Jing Liu, Yajuan Lyu, Shiqi Zhao, Xinyan Xiao, Yuan Liu, Yizhong Wang, Hua Wu, Qiaoqiao She, Xuan Liu, Tian Wu and Haifeng Wang

14:20-15:30 *Robust and Scalable Differentiable Neural Computer for Question Answering*
Jörg Franke, Jan Niehues and Alex Waibel

14:20-15:30 *A Systematic Classification of Knowledge, Reasoning, and Context within the ARC Dataset*
Michael Boratko, Harshit Padigela, Divyendra Mikkilineni, Pritish Yuvraj, Rajarshi Das, Andrew McCallum, Maria Chang, Achille Fokoue-Nkoutche, Pavan Kapanipathi, Nicholas Mattei, Ryan Musa, Kartik Talamadupula and Michael Witbrock

14:20-15:30 *RECIPE: Applying Open Domain Question Answering to Privacy Policies*
Yan Shvartzshanider, Ananth Balashankar, Thomas Wies and Lakshminarayanan Subramanian

14:20-15:30 *Neural Models for Key Phrase Extraction and Question Generation*
Sandeep Subramanian, Tong Wang, Xingdi Yuan, Saizheng Zhang, Adam Trischler and Yoshua Bengio

Thursday, July 19, 2018 (continued)

14:20-15:30 *Comparative Analysis of Neural QA models on SQuAD*
Soumya Wadhwa, Khyathi Chandu and Eric Nyberg

14:20-15:30 *Adaptations of ROUGE and BLEU to Better Evaluate Machine Reading Comprehension Task*
An Yang, Kai Liu, Jing Liu, Yajuan Lyu and Sujian Li

15:30–16:00 *Afternoon coffee break*

16:00–17:00 ***Panel Discussion***