NLPBT 2020

NLP Beyond Text

Proceedings of the First International Workshop on Natural Language Processing Beyond Text

November 20, 2020
Online

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 978-1-952148-84-2

Introduction

Humans interact with each other through several means (e.g., voice, gestures, written text, facial expressions, etc.) and a natural human-machine interaction system should preserve the same modality. However, traditional Natural Language Processing (NLP) focuses on analyzing textual input to solve language understanding and reasoning tasks, and other modalities are only partially targeted. This workshop aims to promote research in the area of Multi/Cross-Modal NLP, i.e., studying computational approaches exploiting the different modalities humans adopt to communicate. In particular, the focus of this workshop is (i) studying how to bridge the gap between NLP on spoken and written language and (ii) exploring how NLU models can be empowered by jointly analyzing multiple input sources, including language (spoken or written), vision (gestures and expressions) and acoustic (paralingustic) modalities. The former comes from the observation that voice-based interaction, which is typical of conversational agents, poses new challenges to NLU. The latter aims to address the way humans acquire and use language. Usually, it happens in a perceptually rich environment, where they communicate using modalities that go beyond language itself. Therefore, extending NLP to modalities beyond written text is a fundamental step in allowing AI systems to reach human-like capabilities.

Organizers:

Giuseppe Castellucci, Amazon
Simone Filice, Amazon
Soujanya Poria, Singapore University of Technology and Design
Erik Cambria, Nanyang Technological University
Lucia Specia, University of Sheffield

Program Committee:

Sawsan Alqahtani, George Washington University
Udit Arora, New York University
Loïc Barrault, University of Sheffield
Emanuele Bastianelli, Philips
Raffaella Bernardi, University of Trento
Fethi Bougares, University of Le Mans
Ozan Caglayan, Imperial College London
Marcus Collins, Amazon
Danilo Croce, University of Roma Tor Vergata
Jean-Benoit Delbrouck, Stanford University
Asif Ekbal, Indian Institute of Technology Patna
Marina Fomicheva, University of Sheffield
Alexander Gelbukh, Instituto Politécnico Nacional
Md Kamrul Hasan, Bangladesh University of Engineering and Technology
Sudipta Kar, Amazon
Penny Karanasou, Amazon
Aman Khullar, IIIT Hyderabad
Samira Korani, Sharif University
Paul Pu Liang, CMU
Jindřich Libovický, Charles University
Shervin Malmasi, Amazon
Elman Mansimov, New York University
Wasifur Rahman, University of Rochester
Akhilesh Ravi, Indian Institute of Technology Gandhinagar
Anna Rohrbach, UC Berkeley
Salvatore Romeo, Amazon
Ramon Sanabria, University of Edinburgh
Sonal Sannigrahi, Ecole Polytechnique
Ranjan Satapathy, Nanyang Technological University
Tejas Srinivasan, Carnegie Mellon University
Umut Sulubacak, University of Helsinki
Shuai Tang, University of California, San Diego
Noé Tits, University of Mons
Kohei Uehara, University of Tokyo
Andrea Vanzo, Sinequanon
Chao Wang, University of Southern California

Invited Speaker:

Loïc Barrault, University of Sheffield

Table of Contents

Program

November 20, 2020

15:00 - 15:10 UTC **Opening NLPBT**

15:10 - 16:10 UTC **Session: Session A**

15:10 UTC *Modulated Fusion using Transformer for Linguistic-Acoustic Emotion Recognition*
Jean-Benoit Delbrouck, Noé Tits and Stéphane Dupont

15:30 UTC *Multimodal Speech Recognition with Unstructured Audio Masking*
Tejas Srinivasan, Ramon Sanabria, Florian Metze and Desmond Elliott

15:50 UTC *EMNLP Findings Paper 1*
TBA

16:10 - 17:00 UTC **Keynote**

16:10 UTC *A Vision on (Simultaneous) Multimodal Machine Translation*
Loïc Barrault

17:00 - 18:30 UTC **Session: Poster Session**

Building a Bridge: A Method for Image-Text Sarcasm Detection Without Pretraining on Image-Text Data
Xinyu Wang, Xiaowen Sun, Tan Yang and Hongbo Wang

A Benchmark for Structured Procedural Knowledge Extraction from Cooking Videos
Frank F. Xu, Lei Ji, Botian Shi, Junyi Du, Graham Neubig, Yonatan Bisk and Nan Duan

A Multi-Modal English-Italian Parallel Corpus for End-to-End Speech-to-Text Machine Translation
Giuseppe Della Corte and Sara Stymne

Unsupervised Keyword Extraction for Full-Sentence VQA
Kohei Uehara and Tatsuya Harada

November 20, 2020 (continued)

18:30 - 19:50 UTC	**Session: Session B**
18:30 UTC	*MAST: Multimodal Abstractive Summarization with Trimodal Hierarchical Attention* Aman Khullar and Udit Arora
18:50 UTC	*Towards End-to-End In-Image Neural Machine Translation* Elman Mansimov, Mitchell Stern, Mia Chen, Orhan Firat, Jakob Uszkoreit and Puneet Jain
19:10 UTC	*EMNLP Findings Paper 2* TBA
19:30 UTC	*Reasoning Over History: Context Aware Visual Dialog* Muhammad Shah, Shikib Mehri and Tejas Srinivasan
19:50 - 20:00 UTC	**Closing NLPBT**

The EMNLP Findings papers will be announced on https://sites.google.com/view/nlpbt-2020