NAACL HLT 2018

Storytelling

Proceedings of the First Workshop

June 5, 2018
New Orleans, Louisiana

©2018 The Association for Computational Linguistics

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 978-1-948087-23-0

Introduction

Welcome to the first *ACL workshop on Storytelling!

Human storytelling has existed for as far back as we can trace, predating writing. Humans have used stories for entertainment, education, cultural preservation; to convey experiences, history, lessons, morals; and to share the human experience.

Part of grounding artificial intelligence work in human experience can involve the generation, understanding, and sharing of stories. This workshop highlights the diverse work being done in storytelling and AI across different fields.

Papers at this workshop are multi-disciplinary, including work on neural, pipeline, and linguistic approaches to understanding and creating stories.

We are also pleased to host a Visual Storytelling challenge, highlighting different methods for automatically generating stories given a set of images; and an invited talk from Nasrin Mostafazadeh on communicating about events through storytelling.

Enjoy the workshop!

Workshop website: `http://www.visionandlanguage.net/workshop2018`

Visual Storytelling Challenge

This workshop hosts the first *ACL Visual Storytelling challenge. The Visual Storytelling challenge provided teams with the VIST visual storytelling dataset[1] to generate stories from sequences of five images, and was hosted on EvalAI[2]. Submissions were evaluated using both automatic metrics and human evaluation. The winner was chosen based on the best performance across the human evaluations for **focus**, **structure and coherence**, **detail**, how **visually grounded** the stories were, how **shareable** the stories were, and whether the stories sounded like they were **written by a human**. More details can be found on the workshop challenge website at: `http://www.visionandlanguage.net/workshop2018/#challenge`

Four teams submitted, from all over the world:

DG-DLMX
Diana González-Rico, Gibran Fuentes-Pineda
Institute for Research in Applied Mathematics and Systems (IIMAS), Universidad Nacional Autónoma de México (UNAM)

NLPSA 501
Chao-Chun Hsu, Szu-Min Chen, Ming-Hsun Hsieh, Lun-Wei Ku
Academia Sinica, Taiwan

UCSB-NLP
Xin Wang, Wenhu Chen, Yuan-Fang Wang, and William Yang Wang
University of California, Santa Barbara, USA

SnuBiVtt
Min-Oh Heo, Taehyeong Kim, Kyung-Wha Park, Seonil Son, Byoung-Tak Zhang
Seoul National University

Automatic evaluation was based on METEOR. Results on the public and private test (newly collected) are listed below.

Team	Public Test Set	Private Test Set (New)
DG-DLMX	0.3088	0.3100
NLPSA501	0.3171	0.3184
UCSB-NLP	0.3179	0.3170
SnuBiVtt	0.2982	0.3066

The winning team was decided by a group of human judges recruited on Amazon Mechanical Turk. Workers were asked to rate 200 randomly selected stories on the following six aspects, using a 5-point Likert scale from Strongly Disagree to Strongly Agree:

- Focus ("This story is focused."): Each story should have a focus, i.e., each sentence of the story should contain information that is "naturally" relevant to the rest of the story, nothing should be random or irrelevant.
- Structure and Coherence ("The story is coherent."): The story should be well-structured, grammatical and well-organized. The story shouldn't be a bunch of related information, but should build from sentence to sentence into a coherent body of readable text describing the set of images.
- "I Would Share": If these were my photos, I would like using a story like this to share my experience with my friends.
- "Written by a Human": This story sounds like it was written by a human.

[1] `http://visionandlanguage.net/VIST/`
[2] `https://evalai.cloudcv.org/`

- "Visually Grounded": This story directly reflects concrete entities in the photos.
- “Detailed”: This story provides an appropriate level of details.

These yielded the following final results:

Team	Focused	Coherent	Willing to Share	Written by A Human	Visually Grounded	Detailed	Total Score
DG-DLMX	3.347	3.278	2.871	3.222	2.886	2.893	18.498
SnuBiVtt (Late)	3.548	3.524	3.075	3.589	3.236	3.323	20.295
NLPSA501	3.111	2.870	2.769	2.870	3.072	2.881	17.574
UCSB-NLP	3.236	3.065	2.767	3.029	3.032	2.867	17.995
Human (Public Test Set)	4.025	3.975	3.772	4.003	3.965	3.857	23.596

The SnuBiVtt team outperformed all other teams on all aspects. However, they technically can not win this VIST Challenge due to the late submission.

The team with the highest overall score within the submission deadline is DG-DLMX, making DG-DLMX the winning team!

Thanks to everyone who participated!

Organizers:

Margaret Mitchell, Google Research
Ting-Hao 'Kenneth' Huang, Carnegie Mellon University
Francis Ferraro, University of Maryland, Baltimore County
Ishan Misra, Carnegie Mellon University

Program Committee:

Elizabeth Clark
David Elson
Marjan Ghazvininejad
Andrew Gordon
Gunhee Kim
Boyang Li
Stephanie Lukin
Joao Magalhaes
Lara Martin
Saif Mohammad
Nasrin Mostafazadeh
Mark Riedl
Luowei Zhou

Invited Speaker:

Nasrin Mostafazadeh, Elemental Cognition

Event-centric Context Modeling: The Case of Story Comprehension and Story Generation

Building AI systems that can process natural language input, comprehend it, and generate an engaging and contextually relevant output in response, has been one of the longest-running goals in AI. In human-human communications, a major trigger to our meaningful communications are “events" and how they cause/enable future events. We often communicate through telling stories in the form of related series of events.

In this talk, I present my work on language processing in terms of events and how they interact with each other in time. Mainly through the lens of storytelling, I will focus on story comprehension and collaborative story generation, with a major emphasis on commonsense reasoning and narrative knowledge as showcased in the Story Cloze Test framework. Through different use cases, I will highlight the importance of establishing a contentful context and modeling multimodal contexts (such as visual and textual) in various AI tasks.

Table of Contents

Conference Program

Tuesday, June 5, 2018

10:00–10:10 ***Opening Remarks***

10:15–11:00 *Invited Talk by Nasrin Mostafazedeh*

11:00–11:25 *Learning to Listen: Critically Considering the Role of AI in Human Storytelling and Character Creation*
Anna Kasunic and Geoff Kaufman

11:25–11:50 *Linguistic Features of Helpfulness in Automated Support for Creative Writing*
Melissa Roemmele and Andrew Gordon

11:50–13:45 ***Lunch***

13:45–14:10 *A Pipeline for Creative Visual Storytelling*
Stephanie Lukin, Reginald Hobbs and Clare Voss

14:10–14:35 *Telling Stories with Soundtracks: An Empirical Analysis of Music in Film*
Jon Gillick and David Bamman

14:35–15:00 *Towards Controllable Story Generation*
Nanyun Peng, Marjan Ghazvininejad, Jonathan May and Kevin Knight

15:00–15:20 ***Break***

15:20–16:00 *Storytelling Challenge*

16:00–16:25 *An Encoder-decoder Approach to Predicting Causal Relations in Stories*
Melissa Roemmele and Andrew Gordon

16:25–16:50 *Neural Event Extraction from Movies Description*
Alex Tozzo, Dejan Jovanovic and Mohamed Amer

Tuesday, June 5, 2018 (continued)

16:50–17:00 ***Closing Remarks***