

EMNLP 2017

Workshop on New Frontiers in Summarization

Workshop Proceedings

September 7, 2017
Copenhagen, Denmark

©2017 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 978-1-945626-89-0

Introduction

Can intelligent systems be devised to create concise, fluent, and accurate summaries from vast amounts of data? Researchers have strived to achieve this goal in the past fifty years, starting from the seminal work of Luhn (1958) on automatic text summarization. Existing research includes the development of extractive and abstractive summarization technologies, evaluation metrics (e.g., ROUGE and Pyramid), as well as the construction of benchmark datasets and resources (e.g., annual competitions such as DUC (2001-2007), TAC (2008-2011), and TREC (2014-2016 on Microblog/Temporal Summarization)).

The goal for this workshop is to provide a research forum for cross-fertilization of ideas. We seek to bring together researchers from a diverse range of fields (e.g., summarization, visualization, language generation, cognitive and psycholinguistics) for discussion on key issues related to automatic summarization. This includes discussion on novel paradigms/frameworks, shared tasks of interest, information integration and presentation, applied research and applications, and possible future research foci. The workshop will pave the way towards building a cohesive research community, accelerating knowledge diffusion, developing new tools, datasets and resources that are in line with the needs of academia, industry, and government.

The topics of this workshop include:

- Abstractive and extractive summarization
- Language generation
- Multiple text genres (News, tweets, product reviews, meeting conversations, forums, lectures, student feedback, emails, medical records, books, research articles, etc)
- Multimodal Input: Information integration and aggregation across multiple modalities (text, speech, image, video)
- Multimodal Output: Summarization and visualization + interactive exploration
- Tailoring summaries to user queries or interests
- Semantic aspects of summarization (e.g. semantic representation, inference, validity)
- Development of new algorithms
- Development of new datasets and annotations
- Development of new evaluation metrics
- Cognitive or psycholinguistic aspects of summarization and visualization (e.g. perceived readability, usability, etc)

In total we received 23 valid submissions (withdrawns are excluded), including 14 long papers and 9 short papers. All papers underwent a rigorous double-blind review process. Among these, 13 papers (7 long, 6 short) are selected for acceptance to the workshop, resulting in an overall acceptance rate of about 57%. We appreciate the excellent reviews provided by the program committee members, and we are grateful to our invited speakers who enriched this workshop with their presentations and insights.

Lu, Giuseppe, Jackie, Fei

Organizers:

Lu Wang (Northeastern University, USA)
Giuseppe Carenini (University of British Columbia, Canada)
Jackie Chi Kit Cheung (McGill University, Canada)
Fei Liu (University of Central Florida, USA)

Program Committee:

Enrique Alfonseca (Google Research)
Asli Celikyilmaz (Microsoft Research)
Jianpeng Cheng (University of Edinburgh)
Greg Durrett (The University of Texas at Austin)
Michael Elhadad (Ben-Gurion University of the Negev)
Benoit Favre (Aix-Marseille University)
Katja Filippova (Google Research)
Wei Gao (Qatar Computing Research Institute)
Shafiq Joty (Qatar Computing Research Institute)
Mijail Kabadjov (University of Essex)
Mirella Lapata (University of Edinburgh)
Junyi Jessy Li (University of Pennsylvania)
Yang Liu (The University of Texas at Dallas)
Annie Louis (University of Essex)
Daniel Marcu (University of Southern California)
Gabriel Murray (University of the Fraser Valley)
Jun-Ping Ng (Amazon)
Hiroya Takamura (Tokyo Institute of Technology)
Simone Teufel (University of Cambridge)
Kapil Thadani (Yahoo Inc.)
Xiaojun Wan (Peking University)

Invited Speakers:

Katja Filippova (Google Research, Switzerland)
Andreas Kerren (Linnaeus University, Sweden)
Ani Nenkova (University of Pennsylvania, USA)

Table of Contents

<i>Video Highlights Detection and Summarization with Lag-Calibration based on Concept-Emotion Mapping of Crowdsourced Time-Sync Comments</i> Qing Ping and Chaomei Chen	1
<i>Multimedia Summary Generation from Online Conversations: Current Approaches and Future Directions</i> Enamul Hoque and Giuseppe Carenini	12
<i>Low-Resource Neural Headline Generation</i> Ottokar Tilk and Tanel Alumäe	20
<i>Towards Improving Abstractive Summarization via Entailment Generation</i> Ramakanth Pasunuru, Han Guo and Mohit Bansal	27
<i>Coarse-to-Fine Attention Models for Document Summarization</i> Jeffrey Ling and Alexander Rush	33
<i>Automatic Community Creation for Abstractive Spoken Conversations Summarization</i> Karan Singla, Evgeny Stepanov, Ali Orkan Bayer, Giuseppe Carenini and Giuseppe Riccardi	43
<i>Combining Graph Degeneracy and Submodularity for Unsupervised Extractive Summarization</i> Antoine Tixier, Polykarpos Meladianos and Michalis Vazirgiannis	48
<i>TL;DR: Mining Reddit to Learn Automatic Summarization</i> Michael Völske, Martin Potthast, Shahbaz Syed and Benno Stein	59
<i>Topic Model Stability for Hierarchical Summarization</i> John Miller and Kathleen McCoy	64
<i>Learning to Score System Summaries for Better Content Selection Evaluation.</i> Maxime Peyrard, Teresa Botschen and Iryna Gurevych	74
<i>Revisiting the Centroid-based Method: A Strong Baseline for Multi-Document Summarization</i> Demian Gholipour Ghalandari	85
<i>Reader-Aware Multi-Document Summarization: An Enhanced Model and The First Dataset</i> Piji Li, Lidong Bing and Wai Lam	91
<i>A Pilot Study of Domain Adaptation Effect for Neural Abstractive Summarization</i> Xinyu Hua and Lu Wang	100

Workshop Program

08:45–10:30 Morning Session 1

08:45–08:50 *Opening Remarks*

08:50–09:50 *Invited Talk*
Andreas Kerren

09:50–10:10 *Video Highlights Detection and Summarization with Lag-Calibration based on Concept-Emotion Mapping of Crowdsourced Time-Sync Comments*
Qing Ping and Chaomei Chen

10:10–10:30 *Multimedia Summary Generation from Online Conversations: Current Approaches and Future Directions*
Enamul Hoque and Giuseppe Carenini

10:30–11:00 *Break*

11:00–12:30 Morning Session 2

11:00–12:00 *Invited Talk*
Katja Filippova

12:00–12:15 *Low-Resource Neural Headline Generation*
Ottokar Tilk and Tanel Alumäe

12:15–12:30 *Towards Improving Abstractive Summarization via Entailment Generation*
Ramakanth Pasunuru, Han Guo and Mohit Bansal

12:30–14:00 *Lunch*

September 7, 2017 (continued)

14:00–15:30 Poster Session

Video Highlights Detection and Summarization with Lag-Calibration based on Concept-Emotion Mapping of Crowdsourced Time-Sync Comments
Qing Ping and Chaomei Chen

Coarse-to-Fine Attention Models for Document Summarization
Jeffrey Ling and Alexander Rush

Automatic Community Creation for Abstractive Spoken Conversations Summarization
Karan Singla, Evgeny Stepanov, Ali Orkan Bayer, Giuseppe Carenini and Giuseppe Riccardi

Multimedia Summary Generation from Online Conversations: Current Approaches and Future Directions
Enamul Hoque and Giuseppe Carenini

Combining Graph Degeneracy and Submodularity for Unsupervised Extractive Summarization
Antoine Tixier, Polykarpos Meladianos and Michalis Vazirgiannis

TL;DR: Mining Reddit to Learn Automatic Summarization
Michael Völske, Martin Potthast, Shahbaz Syed and Benno Stein

Low-Resource Neural Headline Generation
Ottokar Tilk and Tanel Alumäe

Topic Model Stability for Hierarchical Summarization
John Miller and Kathleen McCoy

Learning to Score System Summaries for Better Content Selection Evaluation.
Maxime Peyrard, Teresa Botschen and Iryna Gurevych

Revisiting the Centroid-based Method: A Strong Baseline for Multi-Document Summarization
Demian Gholipour Ghalandari

Towards Improving Abstractive Summarization via Entailment Generation
Ramakanth Pasunuru, Han Guo and Mohit Bansal

Reader-Aware Multi-Document Summarization: An Enhanced Model and The First Dataset
Piji Li, Lidong Bing and Wai Lam

A Pilot Study of Domain Adaptation Effect for Neural Abstractive Summarization
Xinyu Hua and Lu Wang

September 7, 2017 (continued)

15:30–17:15 Afternoon Session

15:30–16:30 *Invited Talk*
Ani Nenkova

16:30–16:50 *Reader-Aware Multi-Document Summarization: An Enhanced Model and The First Dataset*
Piji Li, Lidong Bing and Wai Lam

16:50–17:10 *Learning to Score System Summaries for Better Content Selection Evaluation.*
Maxime Peyrard, Teresa Botschen and Iryna Gurevych

17:10–17:15 Closing Remarks

