GEM 2021

The 1st Workshop on Natural Language Generation, Evaluation, and Metrics

Proceedings of the Workshop

August 5 - 6, 2021
Bangkok, Thailand (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 978-1-954085-67-1

Message from the Organizing Committee

The first Workshop on Natural Language Generation, Evaluation, and Metrics (GEM) was held on August 6, colocated with ACL 2021. The GEM workshop is endorsed by SIGGEN. The organization of GEM was started following discussions during Bird-of-a-Feather sessions at ACL 2020 in which a large share of the attending generation researchers agreed that we need a community-driven project focused on combining advancements in data, models, and (automatic and human) evaluation to measure progress in natural language generation (NLG).

The focus of the GEM workshop was in the shared task for the associated benchmark which was created by the entire program committee. In addition to making an in-depth evaluation of generation models possible, GEM also aims to make generation research more inclusive of additional languages by being designed to be extended with newly created datasets and by prioritizing inclusion of datasets that target languages beyond English. Preliminary results of the shared task were announced at the workshop.

In addition to four reports of shared task participants, we also received 14 research papers of which 11 were accepted for presentation at the workshop. We further invited the authors of 12 Findings of the ACL papers to present, leading to a total of 27 presentations.

Asli Celikyilmaz gave an invited keynote and participated in one of the two panel discussions on responsible progress in NLG. The other panelists were Anya Belz, Hady Elsahar, Seraphina Goldfarb-Tarrant, He He, Mike Lewis, Lisa Li, Wang Lu, and Ehud Reiter.

We would like to thank the members of the Program Committee for their timely reviews. We also would like to thank the participants of the shared task and all volunteers who helped with the evaluations.

Antoine Bosselut, Esin Durmus, Varun Prashant Gangal, Sebastian Gehrmann, Yacine Jernite, Laura Perez-Beltrachini, Samira Shaikh, and Wei Xu

Organizers

Organizing Committee

Antoine Bosselut, Esin Durmus, Varun Prashant Gangal, Sebastian Gehrmann, Yacine Jernite, Laura Perez-Beltrachini, Samira Shaikh, and Wei Xu

Program Committee

Tosin Adewumi, Karmanya Aggarwal, Pawan Sasanka Ammanamanchi, Aremu Anuoluwapo, Antoine Bosselut, Khyathi Raghavi Chandu, Miruna Clinciu, Dipanjan Das, Kaustubh D. Dhole, Wanyu Du, Esin Durmus, Ondřej Dušek, Chris Emezue, Varun Gangal, Cristina Garbacea, Sebastian Gehrmann, Tatsunori Hashimoto, Yufang Hou, Yacine Jernite, Harsh Jhamtani, Yangfeng Ji, Shailza Jolly, Mihir Kale, Dhruv Kumar, Faisal Ladhak, Aman Madaan, Mounica Maddela, Khyati Mahajan, Saad Mahamood, Bodhisattwa Prasad Majumder, Pedro Henrique Martins, Angelina McMillan-Major, Simon Mille, Emiel van Miltenburg, Moin Nadeem, Shashi Narayan, Vitaly Nikolaev, Rubungo Andre Niyongabo, Salomey Osei, Ankur Parikh, Laura Perez-Beltrachini, Niranjan Ramesh Rao, Vikas Raunak, Juan Diego Rodriguez, Sashank Santhanam, João Sedoc, Thibault Sellam, Samira Shaikh, Anastasia Shimorina, Marco Antonio Sobrevilla Cabezudo, Hendrik Strobelt, Nishant Subramani, Wei Xu, Diyi Yang, Akhila Yerukola, Jiawei Zhou

Invited Speaker

Asli Celikyilmaz

Invited Panelists

Anya Belz, Asli Celikyilmaz, Hady Elsahar, Seraphina Goldfarb-Tarrant, He He, Mike Lewis, Lisa Li, Wang Lu, and Ehud Reiter

Table of Contents

Conference Program

Friday, August 6, 2021

11:30–12:00 Opening Remarks and Overview of the Virtual Platform

12:00–12:55 Poster Session

Flesch-Kincaid is Not a Text Simplification Evaluation Metric
Teerapaun Tanprasert and David Kauchak

Human Perception in Natural Language Generation
Lorenzo De Mattei, Huiyuan Lai, Felice Dell'Orletta and Malvina Nissim

Semantic Similarity Based Evaluation for Abstractive News Summarization
Figen Beken Fikri, Kemal Oflazer and Berrin Yanikoglu

Shades of BLEU, Flavours of Success: The Case of MultiWOZ
Tomáš Nekvinda and Ondřej Dušek

13:00–13:45 Panel Session

13:00–13:45 *Panel*

Friday, August 6, 2021 (continued)

14:00–15:00 Talk Session

14:00–14:15 *Personalized Response Generation with Tensor Factorization*
Zhenghui Wang, Lingxiao Luo and Diyi Yang

14:15–14:30 *A Review of Human Evaluation for Style Transfer*
Eleftheria Briakou, Sweta Agrawal, Ke Zhang, Joel Tetreault and Marine Carpuat

14:30–14:45 *GOT: Testing for Originality in Natural Language Generation*
Jennifer Brooks and Abdou Youssef

14:45–15:00 *Evaluating Text Generation from Discourse Representation Structures*
Chunliu Wang, Rik van Noord, Arianna Bisazza and Johan Bos

15:00–15:55 Poster Session

Human Evaluation of Creative NLG Systems: An Interdisciplinary Survey on Recent Papers
Mika Hämäläinen and Khalid Alnajjar

16:00–16:50 Keynote Session

16:00–16:50 *Keynote*
Asli Celikyilmaz

Friday, August 6, 2021 (continued)

17:00–17:45 Panel Session

17:00–17:45 *Panel*

18:00–19:00 GEM Overview Session

18:00–18:15 *The GEM Benchmark: Natural Language Generation, its Evaluation and Metrics*
Sebastian Gehrmann, Tosin Adewumi, Karmanya Aggarwal, Pawan Sasanka Ammanamanchi, Anuoluwapo Aremu, Antoine Bosselut, Khyathi Raghavi Chandu, Miruna-Adriana Clinciu, Dipanjan Das, Kaustubh Dhole, Wanyu Du, Esin Durmus, Ondřej Dušek, Chris Chinenye Emezue, Varun Gangal, Cristina Garbacea, Tatsunori Hashimoto, Yufang Hou, Yacine Jernite, Harsh Jhamtani, Yangfeng Ji, Shailza Jolly, Mihir Kale, Dhruv Kumar, Faisal Ladhak, Aman Madaan, Mounica Maddela, Khyati Mahajan, Saad Mahamood, Bodhisattwa Prasad Majumder, Pedro Henrique Martins, Angelina McMillan-Major, Simon Mille, Emiel van Miltenburg, Moin Nadeem, Shashi Narayan, Vitaly Nikolaev, Andre Niyongabo Rubungo, Salomey Osei, Ankur Parikh, Laura Perez-Beltrachini, Niranjan Ramesh Rao, Vikas Raunak, Juan Diego Rodriguez, Sashank Santhanam, João Sedoc, Thibault Sellam, Samira Shaikh, Anastasia Shimorina, Marco Antonio Sobrevilla Cabezudo, Hendrik Strobelt, Nishant Subramani, Wei Xu, Diyi Yang, Akhila Yerukola and Jiawei Zhou

18:15–18:30 *Reusable Templates and Guides For Documenting Datasets and Models for Natural Language Processing and Generation: A Case Study of the HuggingFace and GEM Data and Model Cards*
Angelina McMillan-Major, Salomey Osei, Juan Diego Rodriguez, Pawan Sasanka Ammanamanchi, Sebastian Gehrmann and Yacine Jernite

19:00–20:00 GEM Systems Session

19:00–19:15 *Structure-to-Text Generation with Self-Training, Acceptability Classifiers and Context-Conditioning for the GEM Shared Task*
Shreyan Bakshi, Soumya Batra, Peyman Heidari, Ankit Arun, Shashank Jain and Michael White

19:15–19:30 *NUIG-DSI's submission to The GEM Benchmark 2021*
Nivranshu Pasricha, Mihael Arcan and Paul Buitelaar

19:30–19:45 *SimpleNER Sentence Simplification System for GEM 2021*
K V Aditya Srivatsa, Monil Gokani and Manish Shrivastava

19:45–20:00 *System Description for the CommonGen task with the POINTER model*
Anna Shvets

Friday, August 6, 2021 (continued)

20:00–20:55 Poster Session

Decoding Methods for Neural Narrative Generation
Alexandra DeLucia, Aaron Mueller, Xiang Lisa Li and João Sedoc