

AAACL-IJCNLP 2022

**The 2nd Conference of the Asia-Pacific Chapter of the
Association for Computational Linguistics and the 12th
International Joint Conference on Natural Language
Processing**

Proceedings of the Student Research Workshop

November 20, 2022
Online

©2022 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 978-1-955917-56-8

Message from the General Chairs

Welcome to the ACL-IJCNLP 2022 Student Research Workshop (SRW)!

The ACL-IJCNLP 2022 SRW is held in conjunction with the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (ACL) and the 12th International Joint Conference on Natural Language Processing conference (IJCNLP).

The ACL-IJCNLP 2022 SRW provides a forum for student researchers who are investigating various areas related to Computational Linguistics and Natural Language Processing. The workshop provides an excellent opportunity for student participants to present their work and receive valuable feedback from the international research community as well as from selected panelists - experienced researchers, specifically assigned according to the topic of their work, who will prepare in-depth comments and questions in advance of the presentation. The workshop's goal is to aid students at multiple stages of their education: including undergraduate, master's, junior, and senior PhD students.

This year's submissions were organized into three different categories. Two of the three categories are general research papers and thesis proposals, following the tradition established by the previous SRWs:

- General research papers: Papers in this category can describe completed work, or work in progress with preliminary results. For these papers, the first author must be a current graduate or undergraduate student.
- Thesis proposals: This category is appropriate for advanced students who have decided on a thesis topic and wish to get feedback on their proposal and broader ideas for their continuing work.

In this year's workshop, we also introduced a special theme category, **Human-Centered NLP**:

- The rising prevalence of living along with artificial intelligence has brought lots of benefits to people's daily life. However, it also poses a challenge of building reliable, explainable, and empathic tools to provide better service, from essential natural language applications (e.g., machine translation, and text classification) to mental support. Especially during the current pandemic, people are more likely to feel alone and rely upon technology. This, together with the advances in the field of Natural Language Processing, has motivated the exploration of human-centered technology. In this special theme, we invite submissions that address diverse human-centered questions, particularly encouraging bringing together perspectives and methods from NLP and affective computing to improve individuals' lives physically and mentally. Topics of interest include (but are not limited to): (1) affective systems to understand human emotion and respond to their emotional feedback; (2) sentiment analysis in social media, e-commerce data, etc.; (3) human factors in the NLP evaluation system; (4) reliable and explainable NLP models; (5) ethics in NLP, including debiasing, detoxification, etc.

We received a total of 40 submissions: 38 general research papers, 1 thesis proposal, and 1 human-centered NLP paper. We accepted 14 general research papers, one thesis proposal, and one human-centered NLP paper, resulting in an overall acceptance rate of 40%. The decision-making process was competitive, but we were delighted that all accepted submissions have great creativity and make contributions to their fields. The accepted submissions are diverse not only in topics but also in terms of student demographics. In addition to the accepted ones, we wish each author of every submission the best of luck in their future endeavors.

Following previous SRWs, we also provided the pre-submission mentoring program for participants. The mentoring program offers students the opportunity to get feedback by a mentor prior to submitting their work for review. 12 papers participated in the pre-submission mentoring program. We would like to thank the 12 pre-submission mentors that spend their time and effort to help improve the work of the

student authors. We would also like to thank all members of the program committee for their in-depth, detailed review and constructive suggestions for each submission. We are especially grateful to all the emergency reviewers who provide timely support and submit their high-quality feedback. Carrying on the practice of the ACL-IJCNLP 2020 SRW, we also include an SRW Keynote and an SRW Best Paper Award in this workshop.

Preparing a workshop is never an easy business. Many thanks to our faculty advisors, Sebastian Ruder and Xiaojun Wan, who provided enormous help and inspiring suggestions through the preparation of this workshop. Special thanks to Boaz Shmueli and Yin Jou Huang, the co-chairs of the ACL-IJCNLP 2020 SRW, who shared their invaluable experience in preparing the workshop. We are also grateful to Yulan He for her constant and timely support. A huge shout out to Yanran Li, who agreed to give the SRW keynote. We sincerely appreciate all of the organizers of the ACL-IJCNLP conference for their effort. And of course, we would like to thank all the student authors and participants who submitted their work to the workshop. This workshop cannot be successful without any of them.

We hope you enjoy the ACL-IJCNLP 2022 SRW!

Organizing Committee

Student Research Workshop Co-Chairs

- Hanqi Yan, University of Warwick, UK
- Zonghan Yang, Tsinghua University, China

Faculty Advisors

- Sebastian Ruder
- Xiaojun Wan

Pre-submission Mentors

- François Charton
- Sedigheh Eslami
- Mor Geva
- Jian Guan
- Itika Gupta
- Yanran Li
- Bin Liang
- Yusheng Su
- Zhixing Tan
- Huiyuan Yang
- Xiaoqing Zheng

Program Committee

- Piush Aggarwal
- Evelin Amorim
- Ali Araabi
- Ehsaneddin Asgari
- Valerio Basile
- Anjali Bhavan

- Ronald Cardenas
- Jonathan P. Chang
- Xinchu Chen
- Sijie Cheng
- Qinyuan Cheng
- David Chiang
- Ganqu Cui
- Xiang Dai
- Min-Yuh Day
- Alok Debnath
- Louise Deléger
- Qingxiu Dong
- John Dougrez-Lewis
- Heba Elfardy
- Micha Elsner
- Matthew Engelhard
- Zheng Fang
- Qiaozi Gao
- Tirthankar Ghosal
- Udo Hahn
- Hardy Hardy
- Christopher Homan
- Dirk Hovy
- Shengding Hu
- Hen-Hsen Huang
- Kaiyu Huang
- Jeff Jacobs
- Masoud Jalili Sabet
- Ganesh Jawahar
- Zara Kancheva
- Philipp Koehn
- Varun Kumar

- Artuur Leeuwenberg
- Lei Li
- Jasy Suet Yan Liew
- Xiaokai Liu
- Xuebo Liu
- Valentin Malykh
- Xiaoyue Mi
- Tsvetomila Mihaylova
- Gosse Minnema
- Pavankumar Reddy Muddireddy
- Kenton Murray
- Masaaki Nagata
- Mitra Pabitra
- Endang Pamungkas
- Lucy Park
- Mah Parsa
- Ivaylo Radev
- Sree Harsha Ramesh
- Vikas Raunak
- Arijit Sehanobish
- Sunayana Sitaram
- Kevin Small
- Richard Sproat
- Yusheng Su
- Zhaoyue Sun
- Shabnam Tafreshi
- Liling Tan
- Keegan Te Taka
- Laure Thompson
- Yan Wang
- Xinyu Wang
- Yuzhong Wang

- Jun Wang
- Jiashuo Wang
- Shuo Wang
- Xing Wang
- Bonnie Webber
- Jheng-Long Wu
- Zhenglu Yang
- Zeyuan Yang
- Michael Yoder
- Yao Zhang
- Xinran Zhang
- Byoung-Tak Zhang
- Xikun Zhang
- Wenjia Zhang
- Linhai Zhang
- Ce Zheng
- Zhong Zhou
- Qinhong Zhou
- Pierre Zweigenbaum

Table of Contents

<i>Emotional Intensity Estimation based on Writer’s Personality</i> Haruya Suzuki, Sora Tarumoto, Tomoyuki Kajiwara, Takashi Ninomiya, Yuta Nakashima and Hajime Nagahara	1
<i>Bipartite-play Dialogue Collection for Practical Automatic Evaluation of Dialogue Systems</i> Shiki Sato, Yosuke Kishinami, Hiroaki Sugiyama, Reina Akama, Ryoko Tokuhisa and Jun Suzuki	8
<i>Toward Building a Language Model for Understanding Temporal Commonsense</i> Mayuko Kimura, Lis Kanashiro Pereira and Ichiro Kobayashi	17
<i>Optimal Summaries for Enabling a Smooth Handover in Chat-Oriented Dialogue</i> Sanae Yamashita and Ryuichiro Higashinaka	25
<i>MUTE: A Multimodal Dataset for Detecting Hateful Memes</i> Eftekhari Hossain, Omar Sharif and Mohammed Moshikul Hoque	32
<i>A Simple and Fast Strategy for Handling Rare Words in Neural Machine Translation</i> Nguyen-Hoang Minh-Cong, Vinh Thi Ngo and Van Vinh Nguyen	40
<i>C3PO: A Lightweight Copying Mechanism for Translating Pseudocode to Code</i> Vishruth Veerendranath, Vibha Masti, Prajwal Anagani and Mamatha HR	47
<i>Outlier-Aware Training for Improving Group Accuracy Disparities</i> Li-Kuang Chen, Canasai Kruengkrai and Junichi Yamagishi	54
<i>An Empirical Study on Topic Preservation in Multi-Document Summarization</i> Mong Yuan Sim, Wei Emma Zhang and Congbo Ma	61
<i>Detecting Urgency in Multilingual Medical SMS in Kenya</i> Narshion Ngao, Zeyu Wang, Lawrence Nderu, Tobias Mwalili, Tal August and Keshet Ronen . .	68
<i>Language over Labels: Contrastive Language Supervision Exceeds Purely Label-Supervised Classification Performance on Chest X-Rays</i> Anton Wiehe, Florian Schneider, Sebastian Blank, Xintong Wang, Hans-Peter Zorn and Christian Biemann	76
<i>Dynamic Topic Modeling by Clustering Embeddings from Pretrained Language Models: A Research Proposal</i> Anton Eklund, Mona Forsman and Frank Drewes	84
<i>Concreteness vs. Abstractness: A Selectional Preference Perspective</i> Tarun Tater, Diego Frassinelli and Sabine Schulte im Walde	92

Conference Program

Sunday, November 20, 2022

14:00–15:00 *Keynotes*

15:00–16:20 **Oral Session I**

15:00–15:20 *Emotional Intensity Estimation based on Writer's Personality*
Haruya Suzuki, Sora Tarumoto, Tomoyuki Kajiwara, Takashi Ninomiya, Yuta Nakashima and Hajime Nagahara

15:20–15:40 *Bipartite-play Dialogue Collection for Practical Automatic Evaluation of Dialogue Systems*
Shiki Sato, Yosuke Kishinami, Hiroaki Sugiyama, Reina Akama, Ryoko Tokuhisa and Jun Suzuki

15:40–16:00 *Toward Building a Language Model for Understanding Temporal Commonsense*
Mayuko Kimura, Lis Kanashiro Pereira and Ichiro Kobayashi

16:40–18:40 **Oral Session II**

17:00–17:20 *Optimal Summaries for Enabling a Smooth Handover in Chat-Oriented Dialogue*
Sanae Yamashita and Ryuichiro Higashinaka

17:20–17:40 *MUTE: A Multimodal Dataset for Detecting Hateful Memes*
Eftekhar Hossain, Omar Sharif and Mohammed Moshiul Hoque

18:00–18:20 *A Simple and Fast Strategy for Handling Rare Words in Neural Machine Translation*
Nguyen-Hoang Minh-Cong, Vinh Thi Ngo and Van Vinh Nguyen

18:20–18:40 *C3PO: A Lightweight Copying Mechanism for Translating Pseudocode to Code*
Vishruth Veerendranath, Vibha Masti, Prajwal Anagani and Mamatha HR

Sunday, November 20, 2022 (continued)

19:40–21:40 Oral Session III

- 19:40–20:00 *Outlier-Aware Training for Improving Group Accuracy Disparities*
Li-Kuang Chen, Canasai Kruengkrai and Junichi Yamagishi
- 20:00–20:20 *An Empirical Study on Topic Preservation in Multi-Document Summarization*
Mong Yuan Sim, Wei Emma Zhang and Congbo Ma
- 20:20–20:40 *Detecting Urgency in Multilingual Medical SMS in Kenya*
Narshion Ngao, Zeyu Wang, Lawrence Nderu, Tobias Mwalili, Tal August and Keshet Ronen
- 20:40–21:00 *Language over Labels: Contrastive Language Supervision Exceeds Purely Label-Supervised Classification Performance on Chest X-Rays*
Anton Wiehe, Florian Schneider, Sebastian Blank, Xintong Wang, Hans-Peter Zorn and Christian Biemann
- 21:00–21:20 *Dynamic Topic Modeling by Clustering Embeddings from Pretrained Language Models: A Research Proposal*
Anton Eklund, Mona Forsman and Frank Drewes
- 21:20–21:40 *Concreteness vs. Abstractness: A Selectional Preference Perspective*
Tarun Tater, Diego Frassinelli and Sabine Schulte im Walde

21:40–22:20 Poster Session

22:20–22:30 *Best Paper Award and Closing Remarks*