FuturED 2024

Workshop on the Future of Event Detection

Proceedings of the Workshop

November 15, 2024

Introduction

Welcome to the 1st Workshop on the Future of Event Detection (FuturED), held as part of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP). Our one-day event takes place on November 15, 2024, in a hybrid format, allowing for both virtual and in-person participation in Miami, Florida.

With the exponential growth of digital data, Event Detection (ED) has become a critical challenge in Natural Language Processing and Data Mining, with key applications in areas such as early warning systems, emergency response, situational awareness, public health monitoring, and understanding societal trends. Despite recent advances in Large Language Models (LLMs) and Generative AI, ED remains a challenging problem, especially when applied across diverse domains, low-resource languages, different data modalities, finer granularities, and extensive integrations. FuturED aims to serve as a forum for discussing the latest advancements in ED research and applications, and exploring how this field will evolve over the next twenty years. The workshop emphasizes bringing together researchers from interdisciplinary fields who have approached ED from various angles, both in theory and in practice, to foster a comprehensive vision of ED's future.

We accepted seven papers to the workshop, as well as two EMNLP Findings papers, covering a wide range of topics—from specialized ED methods to broader discussions about the field's progress, from text-only data to multimodal approaches, and from static learning scenarios to dynamic social network data streams. These papers highlight the scope and depth of ongoing research, while also shedding light on the challenges and potential future directions for ED. Along with the oral presentations, FuturED features two keynote talks from Professor Heng Ji of the University of Illinois Urbana-Champaign and Dr. Lise St. Denis of the University of Colorado Boulder. Their talks will address key challenges and innovations in ED while offering a forward-looking perspective on its future, particularly in the era of LLMs and their applications across various domains.

The FuturED Organizing Committee would like to extend our sincere thanks to our keynote speakers for their inspiring talks, the authors for their valuable contributions, and the Program Committee members for their hard work. We are also grateful to the EMNLP 2024 Workshop Chairs for their support. We hope this workshop and the discussions that arise will provide valuable insights for future ED research and pave the way for further advancements in the field.

FuturED 2024 Organizers

Organizing Committee

Program Chairs

Joel Tetreault, Dataminr Inc.
Thien Huu Nguyen, University of Oregon
Hemank Lamba, Dataminr Inc.
Amanda Lee Hughes, Brigham Young University

Program Committee

Program Committee

Milad Alshomary, Leibniz Universität Hannover
Kfir Bar, Reichman University
Philip Blair, Blair Software
Elizabeth Boschee, Information Sciences Institute
Florian Boudin, Nantes University
Allaa Boutaleb, Sorbonne University / RTE France
Susan Windisch Brown, University of Colorado at Boulder
Cody Buntain, University of Maryland
Aoife Cahill, Dataminr
Tommaso Caselli, Rijksuniversiteit Groningen
David Caswell, Reynolds Journalism Institute
Nathanael Chambers, US Naval Academy
Yi-Pei Chen, The University of Tokyo
Jonathan Chen, UC Berkeley
Danilo Croce, University of Roma, Tor Vergata
Akash Dudhwal, The English and Foreign Languages University
Kevin Duh, Johns Hopkins University
Francis Ferraro, University of Maryland, Baltimore County
Mark Finlayson, FIU
Preston Fu, UC Berkeley
Guillermo Garrido, Google
Guillaume Grosjean, Réseau de Transport d'Électricité (RTE)
Sadid A. Hasan, Microsoft
Ruifang He, Tianjin University
Elvis Hsieh, UC Berkeley
I-Hung Hsu, USC Information Sciences Institute
Po Hu, Central China Normal University
Lifu Huang, Virginia Tech
Ruihong Huang, Texas A&M University
Junbo Huang, University of Hamburg
Yangfeng Ji, University of Virginia
Xiaomeng Jin, University of Illinois, Urbana Champaign
Sanjeev Kumar Karn, Siemens
Tracy Holloway King, Adobe Inc.
Udo Kruschwitz, University of Regensburg
Pasquale Lisena, EURECOM
Jian Liu, Beijing Jiaotong University
Stephanie M. Lukin, U.S. Army Research Laboratory
Cynthia Matuszek, UMBC
Margot Mieskes, University of Applied Sciences, Darmstadt
Ben Miller, Emory University
Seyedabolghasem Mirroshandel, Stony Brook University
Teruko Mitamura, Carnegie Mellon University
Roser Morante, UNED
Laurens Müter, Netherlands Police, Utrecht, the Netherlands
Minh Van Nguyen, University of Oregon

iv

Jessica Ouyang, University of Texas at Dallas
Gabriele Pergola, University of Warwick
Jerome Picault, RTE
Thierry Poibeau, LATTICE (CNRS & ENS/PSL)
Youssra Rebboud, EURECOM, sophia antiopolis
Georg Rehm, DFKI
German Rigau, UPV/EHU
Irene Russo, ILC CNR
Ori Shapira, OriginAI
Étienne Simon, University of Oslo
Reece Suchocki, University of Colorado Boulder
Douglas Summers-Stay, U.S. Army Research Laboratory
Sara Tonelli, FBK
Samia Touileb, University of Bergen
Raphael Troncy, EURECOM
Natalia Vanetik, Shamoon College of Engineering
Erik Velldal, University of Oslo
Clare Voss, Army Research Laboratory
Sijia Wang, Virginia Tech
Qi Yan, UBC
An-Zi Yen, National Yang Ming Chiao Tung University
Pengfei Yu, Department of Computer Science, University of Illinois at Urbana-Champaign
Ke Zhang, Dataminr, inc
Hanwen Zheng, Virginia Tech
Lilja Øvrelid, Dept of Informatics, University of Oslo
Gözde Şahin, Koç University

Keynote Talk
Event Detection: Something Olde and Something New

Heng Ji
University of Illinois at Urbana Champaign (UIUC)
2024-11-15 09:15:00 – Room: **Room 1**

Abstract: When the first event detection shared task was launched at the Message Understanding Conference (MUC-5) in 1993, building a system to detect just one event type in one language took a month of development. Thirty years later, we've seen remarkable progress: modern event detection systems can identify over 3,000 event types across hundreds of languages and multiple data modalities—including text, speech, images, and video. These systems now offer much higher accuracy, coverage, and portability, all at a fraction of the original cost, pushing the boundaries of what's possible in areas like disaster management, business analytics, decision support, and beyond. While these advancements are impressive, the exciting potential lies ahead.

In this talk, I will not only revisit the key techniques that have driven event detection forward but also explore what the next decades could hold in the era of large language models (LLMs). I will propose several PhD dissertation-worthy research directions that could shape the future of event detection, such as never-ending event knowledge base construction, event-based LLM knowledge updating, corpus-level complex event detection and analysis, multimodal event detection and knowledge integration, and emerging applications in situation understanding, hypothesis generation, simulation, and predictive analytics.

Bio: Heng Ji is a professor at Siebel School of Computing and Data Science, and an affiliated faculty member at Electrical and Computer Engineering Department, Coordinated Science Laboratory, and Carl R. Woese Institute for Genomic Biology of University of Illinois Urbana-Champaign. She is an Amazon Scholar. She is the Founding Director of Amazon-Illinois Center on AI for Interactive Conversational Experiences (AICE). She received her B.A. and M. A. in Computational Linguistics from Tsinghua University, and her M.S. and Ph.D. in Computer Science from New York University. Her research interests focus on Natural Language Processing, especially on Multimedia Multilingual Information Extraction, Knowledge-enhanced Large Language Models and Vision-Language Models, and AI for Science. The awards she received include Outstanding Paper Award at ACL2024, two Outstanding Paper Awards at NAACL2024, "Young Scientist" by the World Laureates Association in 2023 and 2024, "Young Scientist" and a member of the Global Future Council on the Future of Computing by the World Economic Forum in 2016 and 2017, "Women Leaders of Conversational AI" (Class of 2023) by Project Voice, "AI's 10 to Watch" Award by IEEE Intelligent Systems in 2013, NSF CAREER award in 2009, "Best of ICDM2013" paper award, "Best of SDM2013" paper award, ACL2020 Best Demo Paper Award, NAACL2021 Best Demo Paper Award, Google Research Award in 2009 and 2014, IBM Watson Faculty Award in 2012 and 2014 and Bosch Research Award in2014-2018. She was invited to testify to the U.S. House Cybersecurity, Data Analytics, & IT Committee as an AI expert in 2023. She was invited by the Secretary of the U.S. Air Force and AFRL to join Air Force Data Analytics Expert Panel to inform the Air Force Strategy 2030, and invited to speak at the Federal Information Integrity R&D Interagency Working Group (IIRD IWG) briefing in 2023. She is the lead of many multi-institution projects and tasks, including the U.S. ARL projects on information fusion and knowledge networks construction, DARPA ECOLE MIRACLE team, DARPA KAIROS RESIN team and DARPA DEFT Tinker Bell team. She has coordinated the NIST TAC Knowledge Base Population task 2010-2020. She served as the associate editor for IEEE/ACM Transaction on Audio, Speech, and Language Processing, and the Program Committee Co-Chair of many conferences including NAACL-HLT2018 and AACL-IJCNLP2022. She was elected as the North American Chapter of the Association for Computational Linguistics (NAACL) secretary 2020-2023. Her research has been widely supported by the U.S. government agencies (DARPA, NSF, DoE, ARL, IARPA, AFRL, DHS) and industry (Amazon, Google, Bosch, IBM, Disney).

<div align="center">

Keynote Talk

Detecting pivotal shifts during complex wildfires: hazards, incident response, community impacts and community response

</div>

<div align="center">

Lise St. Dennis
University of Colorado Boulder
2024-11-15 14:15:00 – Room: **Room 2**

</div>

Abstract: The impacts of wildfires are intensifying in the United States due to a combination of a warming climate, policies and practices contributing to a long-term buildup of fuels on the landscape and housing practices that continue to put more people in harm's way. Teasing apart and making sense of these contributing factors poses numerous challenges for researchers, land managers, and incident management organizations struggling to assess both our current situation and viable paths forward. Recent advances in natural language and data synthesis techniques are a critical component of making sense of this critical moment in time. In this talk I will present current research as part of the CU Earth Lab Wildfire Research Team and collaboration with the USDA Forest Service Wildfire Risk Management Science Team, Human Dimensions Lab to make progress in some of the most critical areas.

Bio: Lise is a research scientist at Earth Lab responsible for the Global Social Sensing Project, a research initiative to develop datasets Earth scale related to the societal impact and societal disruption of natural hazard events for use in Earth Sciences, natural hazards research, and for real-time response applications. Lise has a multidisciplinary background in computer science, data science, human-centered design, crisis informatics, emergency response, and natural hazards research. She holds degrees in Computer Science, Human-centered computing and brings over a decade of industry software design and development experience to her academic research role. Her doctoral research focused on the challenges of integrating social media into formal emergency response and as part of this research participated in over forty emergency response activations with the Virtual Operational Support Team (VOST) community. Her current research extends that knowledge into the Earth Analytics domain, finding new methods for capturing information related to emergency response and societal impacts at scale for use in Earth sciences and natural hazards research. Related work includes development of the ICS-209-PLUS dataset: a geospatial, research grade dataset combining daily snapshots of incident response, hazard characteristics and societal impacts. Design of a neural net classifier for organizing social media for emergency response and natural hazards research. In her free time, Lise loves using her art and design skills, spending time with family and hiking in Boulder County open space with her dogs. She is still an active member of the VOST community and co-lead for the PNWVOST Team.

Table of Contents

Program

09:15 - 09:00   *Opening Remarks*

09:15 - 10:00   *Keynote by Heng Ji: Event Detection: Something Olde and Something New*

10:00 - 10:30   *Session 1*

DEGREEˆ2: Efficient Extraction of Multiple Events Using Language Models
Philip Blair and Kfir Bar

An Incremental Clustering Baseline for Event Detection on Twitter
Marjolaine Ray, Qi Wang, Frédérique Mélanie-Becquet, Thierry Poibeau and Béatrice Mazoyer

10:30 - 11:00   *Coffee Break*

11:00 - 12:20   *Session 2*

BERTrend: Neural Topic Modeling for Emerging Trends Detection
Allaa Boutaleb, Jerome Picault and Guillaume Grosjean

MUMOSA, Interactive Dashboard for MUlti-MOdal Situation Awareness
Stephanie M. Lukin, Shawn Bowser, Reece Suchocki, Douglas Summers-Stay, Francis Ferraro, Cynthia Matuszek and Clare Voss

A Comprehensive Survey on Document-Level Information Extraction
Hanwen Zheng, Sijia Wang and Lifu Huang

Generative Approaches to Event Extraction: Survey and Outlook
Étienne Simon, Helene Olsen, Huiling You, Samia Touileb, Lilja Øvrelid and Erik Velldal

12:20 - 14:15   *Lunch Break*

14:15 - 15:00   *Keynote by Lise Ann St. Denis: Detecting pivotal shifts during complex wildfires: hazards, incident response, community impacts and community response*

Friday, November 15, 2024 (continued)

15:00 - 15:30   *Session 3*

                         When and Where Did it Happen? An Encoder-Decoder Model to Identify Scenario Context
Enrique Simon, Robert Vacareanu, Salena Torres Ashton, Adarsh Pyarelal, Clayton T. Morrison and Mihai Surdeanu

                         Reasoning and Tools for Human-Level Forecasting
Elvis Hsieh, Preston Fu and Jonathan Chen

16:00 - 15:30   *Break*

16:00 - 16:20   *Session 4*

                         Grounding Partially-Defined Events in Multimodal Data
Kate Sanders, Reno Kriz, David Etter, Hannah Recknor, Alexander Martin, Cameron Carpenter, Jingyang Lin and Benjamin Van Durme

16:20 - 17:00   *Panel*

17:00 - 17:15   *Concluding Remarks*