Conversation Derailment Forecasting with Graph Convolutional Networks

Item Type	Conference Paper
Author	Enas Altarawneh
Author	Ameeta Agrawal
Author	Michael Jenkin
Author	Manos Papagelis
Editor	Yi-ling Chung
Editor	Paul R\textbackslash"ottger
Editor	Debora Nozza
Editor	Zeerak Talat
Editor	Aida Mostafazadeh Davani
Abstract	Online conversations are particularly susceptible to derailment, which can manifest itself in the form of toxic communication patterns like disrespectful comments or verbal abuse. Forecasting conversation derailment predicts signs of derailment in advance enabling proactive moderation of conversations. Current state-of-the-art approaches to address this problem rely on sequence models that treat dialogues as text streams. We propose a novel model based on a graph convolutional neural network that considers dialogue user dynamics and the influence of public perception on conversation utterances. Through empirical evaluation, we show that our model effectively captures conversation dynamics and outperforms the state-of-the-art models on the CGA and CMV benchmark datasets by 1.5\textbackslash% and 1.7\textbackslash%, respectively.
Date	2023-07
Library Catalog	ACLWeb
URL	https://aclanthology.org/2023.woah-1.16
Accessed	1/22/2024, 3:00:18 PM
Place	Toronto, Canada
Publisher	Association for Computational Linguistics
Pages	160–169
Proceedings Title	The 7th Workshop on Online Abuse and Harms (WOAH)
Conference Name	WOAH 2023
DOI	10.18653/v1/2023.woah-1.16
Date Added	1/22/2024, 3:00:18 PM
Modified	1/22/2024, 3:00:18 PM

Attachments

- Altarawneh et al_2023_Conversation Derailment Forecasting with Graph Convolutional Networks.pdf

Detoxifying Online Discourse: A Guided Response Generation Approach for Reducing Toxicity in User-Generated Text

Item Type	Conference Paper
Author	Ritwik Bose
Author	Ian Perera
Author	Bonnie Dorr
Editor	Kushal Chawla
Editor	Weiyan Shi

Abstract The expression of opinions, stances, and moral foundations on social media often coincide with toxic, divisive, or inflammatory language that can make constructive discourse across communities difficult. Natural language generation methods could provide a means to reframe or reword such expressions in a way that fosters more civil discourse, yet current Large Language Model (LLM) methods tend towards language that is too generic or formal to seem authentic for social media discussions. We present preliminary work on training LLMs to maintain authenticity while presenting a community's ideas and values in a constructive, non-toxic manner.

Date 2023-07
Short Title Detoxifying Online Discourse
Library Catalog ACLWeb
URL https://aclanthology.org/2023.sicon-1.2
Accessed 1/22/2024, 3:16:09 PM
Extra 0 citations (Semantic Scholar/DOI) [2024-01-22]
Place Toronto, Canada
Publisher Association for Computational Linguistics
Pages 9–14
Proceedings Title Proceedings of the First Workshop on Social Influence in Conversations (SICon 2023)
Conference Name SICon 2023
DOI 10.18653/v1/2023.sicon-1.2
Date Added 1/22/2024, 3:16:09 PM
Modified 1/22/2024, 3:16:13 PM

Attachments

- Bose et al_2023_Detoxifying Online Discourse.pdf

Causal Inference and Natural Language Processing

Item Type Book Section
Author Wenqing Chen
Author Zhixuan Chu
Editor Sheng Li
Editor Zhixuan Chu
Abstract This chapter explores the intersection of two research fields: causal inference and natural language processing (NLP). We aim to answer two fundamental questions: (1) how can NLP aid in causal inference when working with textual data, and (2) how can causal inference theory enhance the robustness and interpretability of NLP models? We present the latest developments and challenges in each area. Firstly, we discuss the difficulties associated with performing causal inference with textual data, which stems from the unstructured and high-dimensional nature of the text. We demonstrate how NLP models can extract high-level semantic variables and how textual data can assume various roles in the causal graph based on Pearl's causal theory. Secondly, while NLP models have achieved remarkable success across different tasks, we highlight concerns about their reliability and robustness. NLP models are prone to learning spurious correlations, which are non-causal but correlated relationships. Thirdly, we provide an extensive overview of causality-driven models for NLP, examining various methods of integrating causality, including intervention-level and counterfactual-level debiasing techniques. Finally, we explore how causal interpretations can improve the interpretability of deep neural models in NLP, enabling a more profound understanding of the models.
Date 2023
Language en
Library Catalog Springer Link
URL https://doi.org/10.1007/978-3-031-35051-1_9

Accessed	1/22/2024, 1:08:04 PM
Extra	DOI: 10.1007/978-3-031-35051-1_9
Place	Cham
Publisher	Springer International Publishing
ISBN	978-3-031-35051-1
Pages	189-206
Book Title	Machine Learning for Causal Inference
Date Added	1/22/2024, 1:08:04 PM
Modified	1/22/2024, 1:08:04 PM

Tags:

Causal inference, Causal interpretations, Debiasing, Natural language processing

---

Measuring Conversational Uptake: A Case Study on Student-Teacher Interactions

Item Type	Preprint
Author	Dorottya Demszky
Author	Jing Liu
Author	Zid Mancenido
Author	Julie Cohen
Author	Heather Hill
Author	Dan Jurafsky
Author	Tatsunori Hashimoto
Abstract	In conversation, uptake happens when a speaker builds on the contribution of their interlocutor by, for example, acknowledging, repeating or reformulating what they have said. In education, teachers' uptake of student contributions has been linked to higher student achievement. Yet measuring and improving teachers' uptake at scale is challenging, as existing methods require expensive annotation by experts. We propose a framework for computationally measuring uptake, by (1) releasing a dataset of student-teacher exchanges extracted from US math classroom transcripts annotated for uptake by experts; (2) formalizing uptake as pointwise Jensen-Shannon Divergence (pJSD), estimated via next utterance classification; (3) conducting a linguistically-motivated comparison of different unsupervised measures and (4) correlating these measures with educational outcomes. We find that although repetition captures a significant part of uptake, pJSD outperforms repetition-based baselines, as it is capable of identifying a wider range of uptake phenomena like question answering and reformulation. We apply our uptake measure to three different educational datasets with outcome indicators. Unlike baseline measures, pJSD correlates significantly with instruction quality in all three, providing evidence for its generalizability and for its potential to serve as an automated professional development tool for teachers.
Date	2021-06-07
Short Title	Measuring Conversational Uptake
Library Catalog	arXiv.org
URL	http://arxiv.org/abs/2106.03873
Accessed	1/22/2024, 4:57:31 PM
Extra	31 citations (Semantic Scholar/arXiv) [2024-01-22] arXiv:2106.03873 [cs]
Repository	arXiv
Archive ID	arXiv:2106.03873
Date Added	1/22/2024, 4:57:31 PM
Modified	1/22/2024, 4:57:35 PM

Tags:

Computer Science - Computation and Language

Notes:

Comment: ACL 2021

Attachments

- arXiv.org Snapshot
- Demszky et al_2021_Measuring Conversational Uptake.pdf

Goal Awareness for Conversational AI: Proactivity, Non-collaborativity, and Beyond

Item Type	Conference Paper
Author	Yang Deng
Author	Wenqiang Lei
Author	Minlie Huang
Author	Tat-Seng Chua
Date	2023
Language	en
Short Title	Goal Awareness for Conversational AI
Library Catalog	DOI.org (Crossref)
URL	https://aclanthology.org/2023.acl-tutorials.1
Accessed	1/22/2024, 1:14:58 PM
Extra	6 citations (Semantic Scholar/DOI) [2024-01-22]
Place	Toronto, Canada
Publisher	Association for Computational Linguistics
Pages	1-10
Proceedings Title	Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 6: Tutorial Abstracts)
Conference Name	Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 6: Tutorial Abstracts)
DOI	10.18653/v1/2023.acl-tutorials.1
Date Added	1/22/2024, 1:14:58 PM
Modified	1/22/2024, 1:15:15 PM

Tags:

survey, tutorial

Attachments

- ACL2023-Tutorial-ConvAI.pdf

Causal Inference in Natural Language Processing: Estimation, Prediction, Interpretation and Beyond

Item Type	Journal Article
Author	Amir Feder

Author Katherine A. Keith
Author Emaad Manzoor
Author Reid Pryzant
Author Dhanya Sridhar
Author Zach Wood-Doughty
Author Jacob Eisenstein
Author Justin Grimmer
Author Roi Reichart
Author Margaret E. Roberts
Author Brandon M. Stewart
Author Victor Veitch
Author Diyi Yang

Abstract A fundamental goal of scientific research is to learn about causal relationships. However, despite its critical role in the life and social sciences, causality has not had the same importance in Natural Language Processing (NLP), which has traditionally placed more emphasis on predictive tasks. This distinction is beginning to fade, with an emerging area of interdisciplinary research at the convergence of causal inference and language processing. Still, research on causality in NLP remains scattered across domains without unified definitions, benchmark datasets and clear articulations of the challenges and opportunities in the application of causal inference to the textual domain, with its unique properties. In this survey, we consolidate research across academic areas and situate it in the broader NLP landscape. We introduce the statistical challenge of estimating causal effects with text, encompassing settings where text is used as an outcome, treatment, or to address confounding. In addition, we explore potential uses of causal inference to improve the robustness, fairness, and interpretability of NLP models. We thus provide a unified overview of causal inference for the NLP community.1

Date 2022-10-18
Short Title Causal Inference in Natural Language Processing
Library Catalog Silverchair
URL https://doi.org/10.1162/tacl_a_00511
Accessed 1/22/2024, 12:51:48 PM
Extra 118 citations (Semantic Scholar/DOI) [2024-01-22]
Volume 10
Pages 1138-1158
Publication Transactions of the Association for Computational Linguistics
DOI 10.1162/tacl_a_00511
Journal Abbr Transactions of the Association for Computational Linguistics
ISSN 2307-387X
Date Added 1/22/2024, 12:51:48 PM
Modified 1/22/2024, 12:51:52 PM

Attachments

- Feder et al_2022_Causal Inference in Natural Language Processing.pdf

Contents

- Snapshot

What Makes a Good Counter-Stereotype? Evaluating Strategies for Automated Responses to Stereotypical Text

Item Type	Conference Paper
Author	Kathleen Fraser
Author	Svetlana Kiritchenko
Author	Isar Nejadgholi
Author	Anna Kerkhof
Editor	Kushal Chawla
Editor	Weiyan Shi
Abstract	When harmful social stereotypes are expressed on a public platform, they must be addressed in a way that educates and informs both the original poster and other readers, without causing offence or perpetuating new stereotypes. In this paper, we synthesize findings from psychology and computer science to propose a set of potential counter-stereotype strategies. We then automatically generate such counter-stereotypes using ChatGPT, and analyze their correctness and expected effectiveness at reducing stereotypical associations. We identify the strategies of denouncing stereotypes, warning of consequences, and using an empathetic tone as three promising strategies to be further tested.
Date	2023-07
Short Title	What Makes a Good Counter-Stereotype?
Library Catalog	ACLWeb
URL	https://aclanthology.org/2023.sicon-1.4
Accessed	1/22/2024, 3:18:37 PM
Extra	1 citations (Semantic Scholar/DOI) [2024-01-22]
Place	Toronto, Canada

Publisher	Association for Computational Linguistics
Pages	25–38
Proceedings Title	Proceedings of the First Workshop on Social Influence in Conversations (SICon 2023)
Conference Name	SICon 2023
DOI	10.18653/v1/2023.sicon-1.4
Date Added	1/22/2024, 3:18:37 PM
Modified	1/22/2024, 3:18:41 PM

Attachments

- Fraser et al_2023_What Makes a Good Counter-Stereotype.pdf

A Survey of Challenges and Methods in the Computational Modeling of Multi-Party Dialog

Item Type	Conference Paper
Author	Ananya Ganesh
Author	Martha Palmer
Author	Katharina Kann
Editor	Yun-Nung Chen
Editor	Abhinav Rastogi
Abstract	Advances in conversational AI systems, powered in particular by large language models, have facilitated rapid progress in understanding and generating dialog. Typically, task-oriented or open-domain dialog systems have been designed to work with two-party dialog, i.e., the exchange of utterances between a single user and a dialog system. However, modern dialog systems may be deployed in scenarios such as classrooms or meetings where conversational analysis of multiple speakers is required. This survey will present research around computational modeling of "multi-party dialog", outlining differences from two-party dialog, challenges and issues in working with multi-party dialog, and methods for representing multi-party dialog. We also provide an overview of dialog datasets created for the study of multi-party dialog, as well as tasks that are of interest in this domain.
Date	2023-07
Library Catalog	ACLWeb
URL	https://aclanthology.org/2023.nlp4convai-1.12
Accessed	1/22/2024, 3:08:10 PM
Extra	0 citations (Semantic Scholar/DOI) [2024-01-22]
Place	Toronto, Canada
Publisher	Association for Computational Linguistics
Pages	140–154
Proceedings Title	Proceedings of the 5th Workshop on NLP for Conversational AI (NLP4ConvAI 2023)
Conference Name	NLP4ConvAI 2023
DOI	10.18653/v1/2023.nlp4convai-1.12
Date Added	1/22/2024, 3:08:10 PM
Modified	1/22/2024, 3:08:17 PM

Tags:

survey

Attachments

- Ganesh et al_2023_A Survey of Challenges and Methods in the Computational Modeling of Multi-Party.pdf

Measuring Lexico-Semantic Alignment in Debates with Contextualized Word Representations

Item Type	Conference Paper
Author	Aina Garí Soler
Author	Matthieu Labeau
Author	Chloé Clavel
Editor	Kushal Chawla
Editor	Weiyan Shi
Abstract	Dialog participants sometimes align their linguistic styles, e.g., they use the same words and syntactic constructions as their interlocutors. We propose to investigate the notion of lexico-semantic alignment: to what extent do speakers convey the same meaning when they use the same words? We design measures of lexico-semantic alignment relying on contextualized word representations. We show that they reflect interesting semantic differences between the two sides of a debate and that they can assist in the task of debate's winner prediction.
Date	2023-07
Library Catalog	ACLWeb
URL	https://aclanthology.org/2023.sicon-1.6
Accessed	1/22/2024, 3:19:12 PM
Extra	0 citations (Semantic Scholar/DOI) [2024-01-22]
Place	Toronto, Canada
Publisher	Association for Computational Linguistics
Pages	50–63
Proceedings Title	Proceedings of the First Workshop on Social Influence in Conversations (SICon 2023)
Conference Name	SICon 2023
DOI	10.18653/v1/2023.sicon-1.6
Date Added	1/22/2024, 3:19:12 PM
Modified	1/22/2024, 3:19:17 PM

Attachments

- Garí Soler et al_2023_Measuring Lexico-Semantic Alignment in Debates with Contextualized Word.pdf

Large Language Models respond to Influence like Humans

Item Type	Conference Paper
Author	Lewis Griffin
Author	Bennett Kleinberg
Author	Maximilian Mozes
Author	Kimberly Mai
Author	Maria Do Mar Vau
Author	Matthew Caldwell
Author	Augustine Mavor-Parker
Editor	Kushal Chawla
Editor	Weiyan Shi
Abstract	Two studies tested the hypothesis that a Large Language Model (LLM) can be used to model psychological change following exposure to influential input. The first study tested a generic mode of influence - the Illusory Truth Effect (ITE) - where earlier exposure to a statement boosts a later truthfulness test rating. Analysis of newly collected data from human and LLM-simulated subjects (1000 of each) showed the same pattern of effects in both populations; although with greater per statement

variability for the LLM. The second study concerns a specific mode of influence – populist framing of news to increase its persuasion and political mobilization. Newly collected data from simulated subjects was compared to previously published data from a 15 country experiment on 7286 human participants. Several effects from the human study were replicated by the simulated study, including ones that surprised the authors of the human study by contradicting their theoretical expectations; but some significant relationships found in human data were not present in the LLM data. Together the two studies support the view that LLMs have potential to act as models of the effect of influence.

Date 2023-07
Library Catalog ACLWeb
URL https://aclanthology.org/2023.sicon-1.3
Accessed 1/22/2024, 3:16:50 PM
Extra 2 citations (Semantic Scholar/DOI) [2024-01-22]
Place Toronto, Canada
Publisher Association for Computational Linguistics
Pages 15–24
Proceedings Title Proceedings of the First Workshop on Social Influence in Conversations (SICon 2023)
Conference Name SICon 2023
DOI 10.18653/v1/2023.sicon-1.3
Date Added 1/22/2024, 3:16:50 PM
Modified 1/22/2024, 3:16:53 PM

Attachments

- Griffin et al_2023_Large Language Models respond to Influence like Humans.pdf

---

MADNet: Maximizing Addressee Deduction Expectation for Multi-Party Conversation Generation

Item Type Conference Paper
Author Jia-Chen Gu
Author Chao-Hong Tan
Author Caiyuan Chu
Author Zhen-Hua Ling
Author Chongyang Tao
Author Quan Liu
Author Cong Liu
Editor Houda Bouamor
Editor Juan Pino
Editor Kalika Bali

Abstract Modeling multi-party conversations (MPCs) with graph neural networks has been proven effective at capturing complicated and graphical information flows. However, existing methods rely heavily on the necessary addressee labels and can only be applied to an ideal setting where each utterance must be tagged with an "@" or other equivalent addressee label. To study the scarcity of addressee labels which is a common issue in MPCs, we propose MADNet that maximizes addressee deduction expectation in heterogeneous graph neural networks for MPC generation. Given an MPC with a few addressee labels missing, existing methods fail to build a consecutively connected conversation graph, but only a few separate conversation fragments instead. To ensure message passing between these conversation fragments, four additional types of latent edges are designed to complete a fully-connected graph. Besides, to optimize the edge-type-dependent message passing for those utterances without addressee labels, an Expectation-Maximization-based method that iteratively generates silver addressee labels (E step), and optimizes the quality of generated responses (M step), is designed. Experimental results on

two Ubuntu IRC channel benchmarks show that MADNet outperforms various baseline models on the task of MPC generation, especially under the more common and challenging setting where part of addressee labels are missing.

Date 2023-12
Short Title MADNet
Library Catalog ACLWeb
URL https://aclanthology.org/2023.emnlp-main.476
Accessed 1/22/2024, 10:20:16 AM
Place Singapore
Publisher Association for Computational Linguistics
Pages 7681–7692
Proceedings Title Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing
Conference Name EMNLP 2023
DOI 10.18653/v1/2023.emnlp-main.476
Date Added 1/22/2024, 10:20:16 AM
Modified 1/22/2024, 10:20:16 AM

Attachments

- Gu et al_2023_MADNet.pdf

GIFT: Graph-Induced Fine-Tuning for Multi-Party Conversation Understanding

Item Type Conference Paper
Author Jia-Chen Gu
Author Zhenhua Ling
Author Quan Liu
Author Cong Liu
Author Guoping Hu
Editor Anna Rogers
Editor Jordan Boyd-Graber
Editor Naoaki Okazaki
Abstract Addressing the issues of who saying what to whom in multi-party conversations (MPCs) has recently attracted a lot of research attention. However, existing methods on MPC understanding typically embed interlocutors and utterances into sequential information flows, or utilize only the superficial of inherent graph structures in MPCs. To this end, we present a plug-and-play and lightweight method named graph-induced fine-tuning (GIFT) which can adapt various Transformer-based pre-trained language models (PLMs) for universal MPC understanding. In detail, the full and equivalent connections among utterances in regular Transformer ignore the sparse but distinctive dependency of an utterance on another in MPCs. To distinguish different relationships between utterances, four types of edges are designed to integrate graph-induced signals into attention mechanisms to refine PLMs originally designed for processing sequential texts. We evaluate GIFT by implementing it into three PLMs, and test the performance on three downstream tasks including addressee recognition, speaker identification and response selection. Experimental results show that GIFT can significantly improve the performance of three PLMs on three downstream tasks and two benchmarks with only 4 additional parameters per encoding layer, achieving new state-of-the-art performance on MPC understanding.
Date 2023-07
Short Title GIFT
Library Catalog ACLWeb
URL https://aclanthology.org/2023.acl-long.651
Accessed 1/22/2024, 1:27:07 PM

Place Toronto, Canada
Publisher Association for Computational Linguistics
Pages 11645–11658
Proceedings Title Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)
Conference Name ACL 2023
DOI 10.18653/v1/2023.acl-long.651
Date Added 1/22/2024, 1:27:07 PM
Modified 1/22/2024, 1:27:07 PM

Attachments

- Gu et al_2023_GIFT.pdf

“Mistakes Help Us Grow”: Facilitating and Evaluating Growth Mindset Supportive Language in Classrooms

Item Type Conference Paper
Author Kunal Handa
Author Margarett Clapper
Author Jessica Boyle
Author Rose E. Wang
Author Diyi Yang
Author David Yeager
Author Dorottya Demszky
Abstract Teachers’ growth mindset supportive language (GMSL)—rhetoric emphasizing that one's skills can be improved over time—has been shown to significantly reduce disparities in academic achievement and enhance students' learning outcomes. Although teachers espouse growth mindset principles, most find it difficult to adopt GMSL in their practice due the lack of effective coaching in this area. We explore whether large language models (LLMs) can provide automated, personalized coaching to support teachers' use of GMSL. We establish an effective coaching tool to reframe unsupportive utterances to GMSL by developing (i) a parallel dataset containing GMSL-trained teacher reframings of unsupportive statements with an accompanying annotation guide, (ii) a GMSL prompt framework to revise teachers’ unsupportive language, and (iii) an evaluation framework grounded in psychological theory for evaluating GMSL with the help of students and teachers. We conduct a large-scale evaluation involving 174 teachers and 1,006 students, finding that both teachers and students perceive GMSL-trained teacher and model reframings as more effective in fostering a growth mindset and promoting challenge-seeking behavior, among other benefits. We also find that model-generated reframings outperform those from the GMSL-trained teachers. These results show promise for harnessing LLMs to provide automated GMSL feedback for teachers and, more broadly, LLMs’ potentiality for supporting students’ learning in the classroom. Our findings also demonstrate the benefit of large-scale human evaluations when applying LLMs in educational domains.
Date 2023/12/01
Language en
Short Title “Mistakes Help Us Grow”
Library Catalog openreview.net
URL https://openreview.net/forum?id=SUAeMJKg6b&referrer=%5Bthe%20profile%20of%20Diyi%20Yang%5D(%2Fprofile%3Fid%3D~Diyi_Yang2)
Accessed 1/22/2024, 2:10:51 PM
Conference Name The 2023 Conference on Empirical Methods in Natural Language Processing
Date Added 1/22/2024, 2:10:51 PM

Modified 1/22/2024, 2:10:51 PM

Attachments

- Handa et al_2023_“Mistakes Help Us Grow”.pdf

My side, your side and the evidence: Discovering aligned actor groups and the narratives they weave

Item Type Conference Paper
Author Pavan Holur
Author David Chong
Author Timothy Tangherlini
Author Vwani Roychowdhury
Editor Anna Rogers
Editor Jordan Boyd-Graber
Editor Naoaki Okazaki
Abstract News reports about emerging issues often include several conflicting story lines. Individual stories can be conceptualized as samples from an underlying mixture of competing narratives. The automated identification of these distinct narratives from unstructured text is a fundamental yet difficult task in Computational Linguistics since narratives are often intertwined and only implicitly conveyed in text. In this paper, we consider a more feasible proxy task: Identify the distinct sets of aligned story actors responsible for sustaining the issue-specific narratives. Discovering aligned actors, and the groups these alignments create, brings us closer to estimating the narrative that each group represents. With the help of Large Language Models (LLM), we address this task by: (i) Introducing a corpus of text segments rich in narrative content associated with six different current issues; (ii) Introducing a novel two-step graph-based framework that (a) identifies alignments between actors (INCANT) and (b) extracts aligned actor groups using the network structure (TAMPA). Amazon Mechanical Turk evaluations demonstrate the effectiveness of our framework. Across domains, alignment relationships from INCANT are accurate (macro F1 \textgreater= 0.75) and actor groups from TAMPA are preferred over 2 non-trivial baseline models (ACC \textgreater= 0.75).
Date 2023-07
Short Title My side, your side and the evidence
Library Catalog ACLWeb
URL https://aclanthology.org/2023.acl-long.497
Accessed 1/22/2024, 1:31:57 PM
Extra 0 citations (Semantic Scholar/DOI) [2024-01-22]
Place Toronto, Canada
Publisher Association for Computational Linguistics
Pages 8938–8952
Proceedings Title Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)
Conference Name ACL 2023
DOI 10.18653/v1/2023.acl-long.497
Date Added 1/22/2024, 1:31:57 PM
Modified 1/22/2024, 1:32:02 PM

Attachments

- Holur et al_2023_My side, your side and the evidence.pdf

Zero-Shot Goal-Directed Dialogue via RL on Imagined Conversations

Item Type	Preprint
Author	Joey Hong
Author	Sergey Levine
Author	Anca Dragan
Abstract	Large language models (LLMs) have emerged as powerful and general solutions to many natural language tasks. However, many of the most important applications of language generation are interactive, where an agent has to talk to a person to reach a desired outcome. For example, a teacher might try to understand their student's current comprehension level to tailor their instruction accordingly, and a travel agent might ask questions of their customer to understand their preferences in order to recommend activities they might enjoy. LLMs trained with supervised fine-tuning or "single-step" RL, as with standard RLHF, might struggle which tasks that require such goal-directed behavior, since they are not trained to optimize for overall conversational outcomes after multiple turns of interaction. In this work, we explore a new method for adapting LLMs with RL for such goal-directed dialogue. Our key insight is that, though LLMs might not effectively solve goal-directed dialogue tasks out of the box, they can provide useful data for solving such tasks by simulating suboptimal but human-like behaviors. Given a textual description of a goal-directed dialogue task, we leverage LLMs to sample diverse synthetic rollouts of hypothetical in-domain human-human interactions. Our algorithm then utilizes this dataset with offline reinforcement learning to train an interactive conversational agent that can optimize goal-directed objectives over multiple turns. In effect, the LLM produces examples of possible interactions, and RL then processes these examples to learn to perform more optimal interactions. Empirically, we show that our proposed approach achieves state-of-the-art performance in various goal-directed dialogue tasks that include teaching and preference elicitation.
Date	2023-11-09
Library Catalog	arXiv.org
URL	http://arxiv.org/abs/2311.05584
Accessed	1/19/2024, 9:46:01 PM
Extra	4 citations (Semantic Scholar/arXiv) [2024-01-19] 4 citations (Semantic Scholar/DOI) [2024-01-19] arXiv:2311.05584 [cs]
DOI	10.48550/arXiv.2311.05584
Repository	arXiv
Archive ID	arXiv:2311.05584
Date Added	1/19/2024, 9:46:01 PM
Modified	1/19/2024, 9:46:11 PM

Tags:

Computer Science - Computation and Language, Computer Science - Machine Learning, Computer Science - Artificial Intelligence

Notes:

Comment: 25 pages, 6 figures

Attachments

- arXiv.org Snapshot
- Hong et al_2023_Zero-Shot Goal-Directed Dialogue via RL on Imagined Conversations.pdf

Contents

---

zhijing-jin/Causality4NLP_Papers

Item Type	Software
Programmer	Zhijing Jin
Abstract	A reading list for papers on causality for natural language processing (NLP)
Date	2024-01-21T14:49:41Z
Library Catalog	GitHub
URL	https://github.com/zhijing-jin/Causality4NLP_Papers
Accessed	1/22/2024, 1:09:05 PM
Extra	original-date: 2021-02-24T15:18:01Z
Date Added	1/22/2024, 1:09:05 PM
Modified	1/22/2024, 1:09:15 PM

Tags:

survey

Your spouse needs professional help: Determining the Contextual Appropriateness of Messages through Modeling Social Relationships

Item Type	Conference Paper
Author	David Jurgens
Author	Agrima Seth
Author	Jackson Sargent

Author Athena Aghighi
Author Michael Geraci
Editor Anna Rogers
Editor Jordan Boyd-Graber
Editor Naoaki Okazaki
Abstract Understanding interpersonal communication requires, in part, understanding the social context and norms in which a message is said. However, current methods for identifying offensive content in such communication largely operate independent of context, with only a few approaches considering community norms or prior conversation as context. Here, we introduce a new approach to identifying inappropriate communication by explicitly modeling the social relationship between the individuals. We introduce a new dataset of contextually-situated judgments of appropriateness and show that large language models can readily incorporate relationship information to accurately identify appropriateness in a given context. Using data from online conversations and movie dialogues, we provide insight into how the relationships themselves function as implicit norms and quantify the degree to which context-sensitivity is needed in different conversation settings. Further, we also demonstrate that contextual-appropriateness judgments are predictive of other social factors expressed in language such as condescension and politeness.
Date 2023-07
Short Title Your spouse needs professional help
Library Catalog ACLWeb
URL https://aclanthology.org/2023.acl-long.616
Accessed 1/22/2024, 1:32:24 PM
Place Toronto, Canada
Publisher Association for Computational Linguistics
Pages 10994–11013
Proceedings Title Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)
Conference Name ACL 2023
DOI 10.18653/v1/2023.acl-long.616
Date Added 1/22/2024, 1:32:24 PM
Modified 1/22/2024, 1:32:24 PM

Attachments

- Jurgens et al_2023_Your spouse needs professional help.pdf

From Multilingual Complexity to Emotional Clarity: Leveraging Commonsense to Unveil Emotions in Code-Mixed Dialogues

Item Type Conference Paper
Author Shivani Kumar
Author Ramaneswaran S
Author Md Shad Akhtar
Author Tanmoy Chakraborty
Abstract Understanding emotions during conversation is a fundamental aspect of human communication, driving NLP research for Emotion Recognition in Conversation (ERC). While considerable research has focused on discerning emotions of individual speakers in monolingual dialogues, understanding the emotional dynamics in code-mixed conversations has received relatively less attention. This motivates our undertaking of ERC for code-mixed conversations in this study. Recognizing that emotional intelligence encompasses a comprehension of worldly knowledge, we propose an innovative approach that integrates commonsense information with dialogue context to facilitate a deeper understanding of emotions. To

	achieve this, we devise an efficient pipeline that extracts relevant commonsense from existing knowledge graphs based on the code-mixed input. Subsequently, we develop an advanced fusion technique that seamlessly combines the acquired commonsense information with the dialogue representation obtained from a dedicated dialogue understanding module. Our comprehensive experimentation showcases the substantial performance improvement obtained through the systematic incorporation of commonsense in ERC. Both quantitative assessments and qualitative analyses further corroborate the validity of our hypothesis, reaffirming the pivotal role of commonsense integration in enhancing ERC.
Date	2023/12/01
Language	en
Short Title	From Multilingual Complexity to Emotional Clarity
Library Catalog	openreview.net
URL	https://openreview.net/forum?id=PWWg9q3S0C&referrer=%5Bthe%20profile%20of%20Shivani%20Kumar%5D(%2Fprofile%3Fid%3D~Shivani_Kumar1)
Accessed	1/22/2024, 10:30:46 AM
Conference Name	The 2023 Conference on Empirical Methods in Natural Language Processing
Date Added	1/22/2024, 10:30:46 AM
Modified	1/22/2024, 10:30:46 AM

Attachments

- Kumar et al_2023_From Multilingual Complexity to Emotional Clarity.pdf

---

What Boosts Fake News Dissemination on Social Media? A Causal Inference View

Item Type	Conference Paper
Author	Yichuan Li
Author	Kyumin Lee
Author	Nima Kordzadeh
Author	Ruocheng Guo
Editor	Hisashi Kashima
Editor	Tsuyoshi Ide
Editor	Wen-Chih Peng
Abstract	There has been an upward trend of fake news propagation on social media. To solve the fake news propagation problem, it is crucial to understand which media posts (e.g., tweets) cause fake news to disseminate widely, and further what lexicons inside a tweet play essential roles for the propagation. However, only modeling the correlation between social media posts and dissemination will find a spurious relationship between them, provide imprecise dissemination prediction, and incorrect important lexicons identification because it did not eliminate the effect of the confounder variable. Additionally, existing causal inference models cannot handle numerical and textual covariates simultaneously. Thus, we propose a novel causal inference model that combines the textual and numerical covariates through soft-prompt learning, and removes irrelevant information from the covariates by conditional treatment generation toward learning effective confounder representation. Then, the model identifies critical lexicons through a post-hoc explanation method. Our model achieves the best performance against baseline methods on two fake news benchmark datasets in terms of dissemination prediction and important lexicon identification related to the dissemination. The code is available at https://github.com/bigheiniu/CausalFakeNews.
Date	2023
Language	en
Short Title	What Boosts Fake News Dissemination on Social Media?
Library Catalog	Springer Link
Extra	0 citations (Semantic Scholar/DOI) [2024-01-22]
Place	Cham

Publisher Springer Nature Switzerland
ISBN 978-3-031-33383-5
Pages 234-246
Series Lecture Notes in Computer Science
Proceedings Title Advances in Knowledge Discovery and Data Mining
DOI 10.1007/978-3-031-33383-5_19
Date Added 1/22/2024, 1:05:53 PM
Modified 1/22/2024, 1:05:58 PM

Tags:

Causal inference on text, Fake news propagation

Understanding Client Reactions in Online Mental Health Counseling

Item Type Conference Paper
Author Anqi Li
Author Lizhi Ma
Author Yaling Mei
Author Hongliang He
Author Shuai Zhang
Author Huachuan Qiu
Author Zhenzhong Lan
Editor Anna Rogers
Editor Jordan Boyd-Graber
Editor Naoaki Okazaki
Abstract Communication success relies heavily on reading participants' reactions. Such feedback is especially important for mental health counselors, who must carefully consider the client's progress and adjust their approach accordingly. However, previous NLP research on counseling has mainly focused on studying counselors' intervention strategies rather than their clients' reactions to the intervention. This work aims to fill this gap by developing a theoretically grounded annotation framework that encompasses counselors' strategies and client reaction behaviors. The framework has been tested against a large-scale, high-quality text-based counseling dataset we collected over the past two years from an online welfare counseling platform. Our study show how clients react to counselors' strategies, how such reactions affect the final counseling outcomes, and how counselors can adjust their strategies in response to these reactions. We also demonstrate that this study can help counselors automatically predict their clients' states.
Date 2023-07
Library Catalog ACLWeb
URL https://aclanthology.org/2023.acl-long.577
Accessed 1/22/2024, 1:33:08 PM
Place Toronto, Canada
Publisher Association for Computational Linguistics
Pages 10358–10376
Proceedings Title Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)
Conference Name ACL 2023
DOI 10.18653/v1/2023.acl-long.577
Date Added 1/22/2024, 1:33:08 PM
Modified 1/22/2024, 1:33:08 PM

Attachments

- Li et al_2023_Understanding Client Reactions in Online Mental Health Counseling.pdf

---

Status, identity, and language: A study of issue discussions in GitHub

Item Type	Journal Article
Author	Jingxian Liao
Author	Guowei Yang
Author	David Kavaler
Author	Vladimir Filkov
Author	Prem Devanbu
Abstract	Successful open source software (OSS) projects comprise freely observable, task-oriented social networks with hundreds or thousands of participants and large amounts of (textual and technical) discussion. The sheer volume of interactions and participants makes it challenging for participants to find relevant tasks, discussions and people. Tagging (e.g., @AmySmith) is a socio-technical practice that enables more focused discussion. By tagging important and relevant people, discussions can be advanced more effectively. However, for all but a few insiders, it can be difficult to identify important and/or relevant people. In this paper we study tagging in OSS projects from a socio-linguistics perspective. First we argue that textual content per se reveals a great deal about the status and identity of who is speaking and who is being addressed. Next, we suggest that this phenomenon can be usefully modeled using modern deep-learning methods. Finally, we illustrate the value of these approaches with tools that could assist people to find the important and relevant people for a discussion.
Date	Jun 14, 2019
Language	en
Short Title	Status, identity, and language
Library Catalog	PLoS Journals
URL	https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0215059
Accessed	10/20/2021, 9:07:53 AM
Extra	6 citations (Semantic Scholar/DOI) [2022-07-07] Publisher: Public Library of Science
Volume	14
Pages	e0215059
Publication	PLOS ONE
DOI	10.1371/journal.pone.0215059
Issue	6
Journal Abbr	PLOS ONE
ISSN	1932-6203
Date Added	10/20/2021, 9:07:53 AM
Modified	7/7/2022, 12:22:58 PM

Tags:

Computer software, Entropy, Language, Open source software, Programming languages, Psycholinguistics, Semantics, Sociolinguistics

Attachments

- Liao et al_2019_Status, identity, and language.pdf
- Snapshot

e-THERAPIST: I suggest you to cultivate a mindset of positivity and nurture uplifting thoughts

Item Type Conference Paper
Author Kshitij Mishra
Author Priyanshu Priya
Author Manisha Burja
Author Asif Ekbal
Abstract The shortage of therapists for mental health patients emphasizes the importance of globally accessible dialogue systems alleviating their issues. To have effective interpersonal psychotherapy, these systems must exhibit politeness and empathy when needed. However, these factors may vary as per the user's gender, age, persona, and sentiment. Hence, in order to establish trust and provide a personalized cordial experience, it is essential that generated responses should be tailored to individual profiles and attributes. Focusing on this objective, we propose e-THERAPIST, a novel polite interpersonal psychotherapy dialogue system to address issues like depression, anxiety, schizophrenia, etc. We begin by curating a unique conversational dataset for psychotherapy, called PsyCon. It is annotated at two levels: (i) dialogue-level - including user's profile information (gender, age, persona) and therapist's psychotherapeutic approach; and (ii) utterance-level - encompassing user's sentiment and therapist's politeness, and interpersonal behaviour. Then, we devise a novel reward model to adapt correct polite interpersonal behaviour and use it to train e-THERAPIST on PsyCon employing NLPO loss. Our extensive empirical analysis validates the effectiveness of each component of the proposed e-THERAPIST demonstrating its potential impact in psychotherapy settings.
Date 2023/12/01
Language en
Short Title e-THERAPIST
Library Catalog openreview.net
URL https://openreview.net/forum?id=7UVOFuNk27&referrer=%5Bthe%20profile%20of%20Kshitij%20Mishra%5D(%2Fprofile%3Fid%3D~Kshitij_Mishra1)
Accessed 1/22/2024, 10:31:29 AM
Conference Name The 2023 Conference on Empirical Methods in Natural Language Processing
Date Added 1/22/2024, 10:31:29 AM
Modified 1/22/2024, 10:31:29 AM

Attachments

- Mishra et al_2023_e-THERAPIST.pdf

Causal Effects of Linguistic Properties

Item Type Conference Paper
Author Reid Pryzant
Author Dallas Card
Author Dan Jurafsky
Author Victor Veitch
Author Dhanya Sridhar
Editor Kristina Toutanova
Editor Anna Rumshisky
Editor Luke Zettlemoyer
Editor Dilek Hakkani-Tur
Editor Iz Beltagy
Editor Steven Bethard

Editor Ryan Cotterell
Editor Tanmoy Chakraborty
Editor Yichao Zhou
Abstract We consider the problem of using observational data to estimate the causal effects of linguistic properties. For example, does writing a complaint politely lead to a faster response time? How much will a positive product review increase sales? This paper addresses two technical challenges related to the problem before developing a practical method. First, we formalize the causal quantity of interest as the effect of a writer's intent, and establish the assumptions necessary to identify this from observational data. Second, in practice, we only have access to noisy proxies for the linguistic properties of interest —e.g., predictions from classifiers and lexicons. We propose an estimator for this setting and prove that its bias is bounded when we perform an adjustment for the text. Based on these results, we introduce TextCause, an algorithm for estimating causal effects of linguistic properties. The method leverages (1) distant supervision to improve the quality of noisy proxies, and (2) a pre-trained language model (BERT) to adjust for the text. We show that the proposed method outperforms related approaches when estimating the effect of Amazon review sentiment on semi-simulated sales figures. Finally, we present an applied case study investigating the effects of complaint politeness on bureaucratic response times.
Date 2021-06
Library Catalog ACLWeb
URL https://aclanthology.org/2021.naacl-main.323
Accessed 1/22/2024, 12:49:30 PM
Extra 28 citations (Semantic Scholar/DOI) [2024-01-22]
Place Online
Publisher Association for Computational Linguistics
Pages 4095–4109
Proceedings Title Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies
Conference Name NAACL-HLT 2021
DOI 10.18653/v1/2021.naacl-main.323
Date Added 1/22/2024, 12:49:30 PM
Modified 1/22/2024, 12:49:34 PM

Attachments

- Pryzant et al_2021_Causal Effects of Linguistic Properties.pdf

---

Cognitive Reframing of Negative Thoughts through Human-Language Model Interaction

Item Type Preprint
Author Ashish Sharma
Author Kevin Rushton
Author Inna Wanyin Lin
Author David Wadden
Author Khendra G. Lucas
Author Adam S. Miner
Author Theresa Nguyen
Author Tim Althoff
Abstract A proven therapeutic technique to overcome negative thoughts is to replace them with a more hopeful "reframed thought." Although therapy can help people practice and learn this Cognitive Reframing of Negative Thoughts, clinician shortages and mental health stigma commonly limit people's access to therapy. In this paper, we conduct a human-centered study of how language models may assist people in reframing negative thoughts. Based on psychology literature, we define a framework of seven linguistic

attributes that can be used to reframe a thought. We develop automated metrics to measure these attributes and validate them with expert judgements from mental health practitioners. We collect a dataset of 600 situations, thoughts and reframes from practitioners and use it to train a retrieval-enhanced in-context learning model that effectively generates reframed thoughts and controls their linguistic attributes. To investigate what constitutes a "high-quality" reframe, we conduct an IRB-approved randomized field study on a large mental health website with over 2,000 participants. Amongst other findings, we show that people prefer highly empathic or specific reframes, as opposed to reframes that are overly positive. Our findings provide key implications for the use of LMs to assist people in overcoming negative thoughts.

Date 2023-05-03
Library Catalog arXiv.org
URL http://arxiv.org/abs/2305.02466
Accessed 1/22/2024, 4:50:10 PM
Extra 7 citations (Semantic Scholar/arXiv) [2024-01-22] arXiv:2305.02466 [cs]
Repository arXiv
Archive ID arXiv:2305.02466
Date Added 1/22/2024, 4:50:10 PM
Modified 1/22/2024, 4:50:14 PM

Tags:

Computer Science - Computation and Language, Computer Science - Human-Computer Interaction, Computer Science - Social and Information Networks

Notes:

Comment: Accepted for publication at ACL 2023

Attachments

- arXiv.org Snapshot
- Sharma et al_2023_Cognitive Reframing of Negative Thoughts through Human-Language Model.pdf

---

Towards Zero-Shot Multilingual Transfer for Code-Switched Responses

Item Type Conference Paper
Author Ting-Wei Wu
Author Changsheng Zhao
Author Ernie Chang
Author Yangyang Shi
Author Pierce Chuang
Author Vikas Chandra
Author Biing Juang
Editor Anna Rogers
Editor Jordan Boyd-Graber
Editor Naoaki Okazaki
Abstract Recent task-oriented dialog systems have had great success in building English-based personal assistants, but extending these systems to a global audience is challenging due to the need for annotated data in the target language. An alternative approach is to leverage existing data in a high-resource language to enable cross-lingual transfer in low-resource language models. However, this type of transfer has not been widely explored in natural language response generation. In this research, we

investigate the use of state-of-the-art multilingual models such as mBART and T5 to facilitate zero-shot and few-shot transfer of code-switched responses. We propose a new adapter-based framework that allows for efficient transfer by learning task-specific representations and encapsulating source and target language representations. Our framework is able to successfully transfer language knowledge even when the target language corpus is limited. We present both quantitative and qualitative analyses to evaluate the effectiveness of our approach.

Date 2023-07
Library Catalog ACLWeb
URL https://aclanthology.org/2023.acl-long.417
Accessed 1/22/2024, 1:42:44 PM
Extra 0 citations (Semantic Scholar/DOI) [2024-01-22]
Place Toronto, Canada
Publisher Association for Computational Linguistics
Pages 7551–7563
Proceedings Title Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)
Conference Name ACL 2023
DOI 10.18653/v1/2023.acl-long.417
Date Added 1/22/2024, 1:42:44 PM
Modified 1/22/2024, 1:42:49 PM

Attachments

- Wu et al_2023_Towards Zero-Shot Multilingual Transfer for Code-Switched Responses.pdf

---

Conversation Modeling to Predict Derailment

Item Type Journal Article
Author Jiaqing Yuan
Author Munindar P. Singh
Abstract Conversations among online users sometimes derail, i.e., break down into personal attacks. Derailment interferes with the healthy growth of communities in cyberspace. The ability to predict whether an ongoing conversation will derail could provide valuable advance, even real-time, insight to both interlocutors and moderators. Prior approaches predict conversation derailment retrospectively without the ability to forestall the derailment proactively. Some existing works attempt to make dynamic predictions as the conversation develops, but fail to incorporate multisource information, such as conversational structure and distance to derailment. We propose a hierarchical transformer-based framework that combines utterance-level and conversation-level information to capture fine-grained contextual semantics. We propose a domain-adaptive pretraining objective to unite conversational structure information and a multitask learning scheme to leverage the distance from each utterance to derailment. An evaluation of our framework on two conversation derailment datasets shows an improvement in F1 score for the prediction of derailment. These results demonstrate the effectiveness of incorporating multisource information for predicting the derailment of a conversation.
Date 2023-06-02
Language en
Library Catalog ojs.aaai.org
URL https://ojs.aaai.org/index.php/ICWSM/article/view/22200
Accessed 1/22/2024, 3:29:55 PM
Rights Copyright (c) 2023 Association for the Advancement of Artificial Intelligence
Volume 17
Pages 926-935

Publication	Proceedings of the International AAAI Conference on Web and Social Media
DOI	10.1609/icwsm.v17i1.22200
ISSN	2334-0770
Date Added	1/22/2024, 3:29:55 PM
Modified	1/22/2024, 3:29:55 PM

Tags:

Web and Social Media

Attachments

- Yuan_Singh_2023_Conversation Modeling to Predict Derailment.pdf

---

Quantifying the Causal Effects of Conversational Tendencies

Item Type	Journal Article
Author	Justine Zhang
Author	Sendhil Mullainathan
Author	Cristian Danescu-Niculescu-Mizil
Abstract	Understanding what leads to effective conversations can aid the design of better computer-mediated communication platforms. In particular, prior observational work has sought to identify behaviors of individuals that correlate to their conversational efficiency. However, translating such correlations to causal interpretations is a necessary step in using them in a prescriptive fashion to guide better designs and policies. In this work, we formally describe the problem of drawing causal links between conversational behaviors and outcomes. We focus on the task of determining a particular type of policy for a text-based crisis counseling platform: how best to allocate counselors based on their behavioral tendencies exhibited in their past conversations. We apply arguments derived from causal inference to underline key challenges that arise in conversational settings where randomized trials are hard to implement. Finally, we show how to circumvent these inference challenges in our particular domain, and illustrate the potential benefits of an allocation policy informed by the resulting prescriptive information.
Date	October 15, 2020
Library Catalog	ACM Digital Library
URL	https://dl.acm.org/doi/10.1145/3415202
Accessed	1/22/2024, 12:53:55 PM
Extra	24 citations (Semantic Scholar/DOI) [2024-01-22]
Volume	4
Pages	131:1–131:24
Publication	Proceedings of the ACM on Human-Computer Interaction
DOI	10.1145/3415202
Issue	CSCW2
Journal Abbr	Proc. ACM Hum.-Comput. Interact.
Date Added	1/22/2024, 12:53:55 PM
Modified	1/22/2024, 12:54:01 PM

Tags:

causal inference, conversations, counseling

Attachments

- Zhang et al_2020_Quantifying the Causal Effects of Conversational Tendencies.pdf

Contents