Large Language Models (LLMs) and Causality Extraction from Text Tutorial at FLAIRS-38

Wlodek Zadrozny
Computer Science Department
UNC Charlotte, Charlotte NC 28203
wzadrozn@charlotte.edu

Abstract

This tutorial explores the application of Large Language Models (LLMs), such as BERT, LLAMA, and GPT-3.5/4, to the extraction of causality from text documents, including identifying causes, effects, and actions in diverse texts, such as business, medical, and newswire domains. We also address challenges related to data availability and quality, such as varying definitions of causality. Causality extraction plays a crucial role in natural language understanding, particularly for building structured representations of medical and technical texts and for multimodal question answering. Participants will gain access to example code and links to related repositories. Beyond causality extraction, the session will connect these tasks to broader themes, such as the mathematics of hallucinations in generative models and best practices for effective prompting. Designed for participants with some familiarity with machine learning or natural language processing (NLP), and ideally LLMs, the tutorial should be both accessible and highly relevant.

Overview

Large Language Models (LLMs) have been the dominant research theme in natural language processing (NLP) since 2018, when BERT was made available by Google. Since then, we have seen great improvements in several versions of GPT, and research using publicly available Large Language Models such as LLAMA, Gemma, Falcon, Claude Mixtral and others. Causality extraction involves two key tasks: identifying phrases and clauses that describe causes, actions, and effects, and situating these elements within broader contexts. Examples include improving a patient's condition or explaining the political implications of an event and its causes. We note that human understanding is often based on "mechanical" action-effect or cause-effect models, and the desired outcome of causality extraction is often a cause-effect graph or hypergraph. Due to the generative nature of LLMs, they often hallucinate the locations of relevant action, cause, or effect passages. Additionally, their relatively shallow contextual understanding can hinder their performance in generating accurate interpretations, particularly for complex domains like political news. This tutorial will highlight these challenges and explore strategies to enhance the quality of causality extraction using LLMs. Topics will include prompt engineering, and other control strategies and architecture, as well as recent theoretical insights into the inevitability of hallucinations in generative models. Examples and case studies will be provided, along with links to code and other resources. The tutorial will present an overview of the field, showcasing recent findings up to April 2025. The material is designed for participants with some familiarity with machine learning or natural language processing (NLP), and ideally LLMs.

Outline of Tutorial

Part 1 (45 minutes): Introduction

a. What is causality extraction, and why it matters for NLP? (We will start with very brief introduction to the concepts of *causality* and *causality extraction* (see Notes below)).

b. What can LLMs do well and where they struggle? (This is changing on a weekly basis).

c. Strengths and limitations of LLMs in causality extraction. — Demonstrations: Examples and results

Part 2 (45 minutes): In-Depth Exploration

a. Challenges in causality extraction datasets: Data scarcity and quality issues

b. Application case studies: Enhancing extraction and controlling model behavior

c. Open problems: Mapping text to causal diagrams/graphs and addressing modalities

Acknowledgment. Parts of this tutorial are based on the author's collaborative work with his students and collaborators: Drs. Seetha Gopalakrishnan, Hossein Hematialam, and Luciana Garbayo (MD). In particular, the following works were used in its preparation: (Gopalakrishnan et al. 2023;

2024; Zadrozny et al. 2024; Hematialam and Zadrozny 2021; Gopalakrishnan et al. 2024).

Brief Biography

Dr. Wlodek Zadrozny joined the University of North Carolina at Charlotte in 2013 following a career at the IBM T.J. Watson Research Center. His contributions include work on the Watson project, the Jeopardy! playing machine, and were recognized by the 2013 AAAI Feigenbaum Prize. His research focuses on natural language understanding, particularly causality extraction from text using deep learning methods. His interests include applying mathematical techniques to classification and deep learning. Since 2018, Dr. Zadrozny has taught graduate-level deep learning courses annually at UNC Charlotte and twice at Duke University. He has published over 100 refereed papers and holds 60 granted patents. Dr. Zadrozny earned his PhD in Mathematics (with distinction) from the Polish Academy of Science.

Full text availability. For an extended version of this tutorial see `https://www.researchgate.net/publication/390545370_Large_Language_Models_LLMs_and_Causality_Extraction_from_Text_A_Survey`

Notes

(1) *Causation* and *causality* is of natural interest to everyone, and of scientific interest for each scientific domain from philosophy, to physics, to artificial intelligence and to medicine (Hitchcock and Woodward 2022; Hendricks and Symons 2022; Gallow 2022; Sawesi, Rashrash, and Dammann 2022). Aristotle (Shields 2022) introduced the four types of causes and argued that all four are necessary and sufficient for explanations. More recently theories of probabilistic probabilistic causation (Hitchcock 2022) have been developed. Perhaps the most prominent one presented in the works of J.Pearl, his students and collaborators (Pearl and Verma 1995; Pearl 2019; 2009; Bareinboim et al. 2022).

(2) Even though there are multiple well-established theories of causality, they seem have only limited impact on the subfield of NLP called "causality extraction." In NLP the starting point lies in commonsense views of cause and effect (Roemmele, Bejan, and Gordon 2011), summarized common dictionary definitions, e.g. in Merriam-Webster dictionary as "the direct relationship between an action or event and its consequence or result"and causality as "the relation between a cause and its effect or between regularly correlated events or phenomena." This common-sense intuition is operationalized as instructions for data annotators, e.g. in the "Five Tests for Causality," proposed in (Tan et al. 2022):

1. Why: The example is not causal if the reader is unable to construct a “Why” question regarding the Effect.
2. Temporal order: The example is not causal if the Cause does not precede the Effect in time.
3. Counterfactual: The example is not causal if the Effect is equally likely to occur or not occur without the Cause.
4. Ontological asymmetry: The example is not causal if the reader can readily swap the Cause and Effect claims in place.
5. Linguistic: The example is likely to be causal if it can be rephrased into “X causes Y” or “Due to X, Y.”

The properties 1-5 do not naturally map into the Aristotle types of causation, nor the Pearl's approach, nor other taxonomies of causal relations, e.g the seven views of causality in medical literature (Sawesi, Rashrash, and Dammann 2022). Therefore, for the purpose of this exposition, we see "causality extraction" as a separate domain of research.

References

Bareinboim, E.; Correa, J. D.; Ibeling, D.; and Icard, T. F. 2022. On pearl's hierarchy and the foundations of causal inference. *Probabilistic and Causal Inference*.

Gallow, J. D. 2022. The Metaphysics of Causation. `https://plato.stanford.edu/archives/fall2022/entries/causation-metaphysics/`.

Gopalakrishnan, S.; Chen, V. Z.; Dou, W.; Hahn-Powell, G.; Nedunuri, S.; and Zadrozny, W. 2023. Text to causal knowledge graph: A framework to synthesize knowledge from unstructured business texts into causal graphs. *Information* 14(7):367.

Gopalakrishnan, S.; Chen, V. Z.; Dou, W.; and Zadrozny, W. 2024. On the relation between k–l divergence and transfer learning performance on causality extraction tasks. *Natural Language Processing Journal* 6:100055.

Hematialam, H., and Zadrozny, W. 2021. Identifying condition-action statements in medical guidelines: Three studies using machine learning and domain adaptation.

Hendricks, V. F., and Symons, J. 2022. Logic and artificial intelligence. `https://plato.stanford.edu/archives/fall2022/entries/logic-ai/#cause`.

Hitchcock, C., and Woodward, J. 2022. Causal models. `https://plato.stanford.edu/archives/fall2022/entries/causal-models/`.

Hitchcock, C. 2022. Probabilistic causation. `https://plato.stanford.edu/archives/fall2022/entries/causation-probabilistic/`.

Pearl, J., and Verma, T. S. 1995. A theory of inferred causation. In *Studies in Logic and the Foundations of Mathematics*, volume 134. Elsevier. 789–811.

Pearl, J. 2009. *Models, reasoning and inference, 2nd Edition*. Cambridge, UK: CambridgeUniversityPress.

Pearl, J. 2019. The seven tools of causal inference, with reflections on machine learning. *Communications of the ACM* 62(3):54–60.

Roemmele, M.; Bejan, C. A.; and Gordon, A. S. 2011. Choice of plausible alternatives: An evaluation of commonsense causal reasoning. In *Proceedings of the AAAI Spring Symposium: Logical Formalizations of Commonsense Reasoning*, 90–95.

Sawesi, S.; Rashrash, M.; and Dammann, O. 2022. The representation of causality and causation with ontologies: A systematic literature review. *Online Journal of Public Health Informatics* 14(1):e4.

Shields, C. 2022. Aristotle. `https://plato.stanford.edu/archives/fall2022/entries/aristotle/#FouCauAccExpAde`.

Tan, F. A.; Hürriyetoğlu, A.; Caselli, T.; Oostdijk, N.; Nomoto, T.; Hettiarachchi, H.; Ameer, I.; Uca, O.; Liza, F. F.; and Hu, T. 2022. The causal news corpus: Annotating causal relations in event sentences from news. *arXiv preprint arXiv:2204.11714*.

Zadrozny, W.; Dou, W.; Chen, V. Z.; and Gopalakrishnan, S. 2024. Computer implemented method and system for integrative causal modeling and transfer. US Patent App. 18/525,940.