Causal Domain Adaptation for Information Extraction from Complex Conversations

Xue Li[1][0000−0001−5451−1867]

University of Amsterdam,
Amsterdam, Netherlands
x.li3@uva.nl

Abstract. *Complex conversations* can be seen everywhere on the web from email lists to discussion forms. Being able to more effectively extract entities and their relations from these conversations would be an important contribution to conversational content analysis. Despite the success of current information extraction systems, their use in complex conversations is challenging due to, among other reasons, the existence of *long-tail* entities that are unrepresented in standard training corpora (e.g. news). Moreover, in general the distribution of entities in the target domain is frequently different from that of the training domain, which requires the algorithms to be able to perform domain adaptation. In this research, we will focus on identifying domain shifts that might impact information extractions systems and we aim to propose a causal framework for domain adaptation in information extraction.

Keywords: Complex Conversations · Information Extraction · Domain Adaptation · Causality.

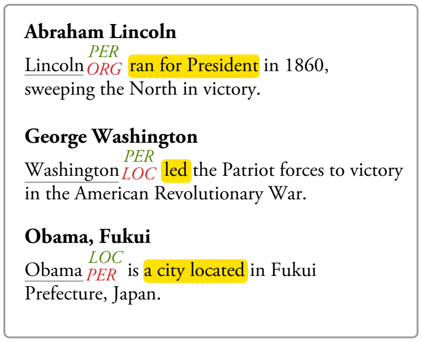
1 Introduction

Complex conversations are characterized by 1) their long-form nature; 2) extension across sessions; and 3) frequent reference to domain-specific background knowledge and material. Examples of a complex conversation include:

- a mailing list where participants discuss the development of a new standard for 5G;
- a long-running Slack chat on the development of an EU proposal with multiple partners;
- meeting minutes from a series of meetings discussing the development of new governmental policy.

Recent work on conversational AI and dialogue extraction have primarily focused on the development of interactive systems such as chatbots or assistants [11], while complex conversations remain an under-explored topic.

One way to analyze conversations is through the construction and use of Knowledge Graphs (KGs) [28]. KGs are graph-structured knowledge bases (KBs)



Abraham Lincoln

Lincoln ran for President in 1860, sweeping the North in victory.

George Washington

Washington led the Patriot forces to victory in the American Revolutionary War.

Obama, Fukui

Obama is a city located in Fukui Prefecture, Japan.

Fig. 1. Examples extracted from Wikipedia (title in bold) illustrate bias in NER. Entities of interest are underlined, green superscript indicates the correct category for the entity and red subscript indicates the prediction from the Stanford NER tagger [17]. The related contextual information to infer the correct types is highlighted in yellow. These failing results show that the model relies heavily on the representation learned from the training dataset.

that store factual information in form of relationships between entities [20]. To construct knowledge graphs from complex conversations, we need to extract the entities (nodes) with their types and their relations (edges). In particular, the process includes a set of core Natural Language Processing (NLP) tasks related to entities in Information Extraction (IE): Named Entity Recognition (NER) and Coreference Resolution (CR). More broadly information extraction is a key topic for web research [18] and in particular the semantic web [18].

A challenge with applying information extraction, especially in complex conversations, is the existence of long-tail entities [6,9,12,13,15,26]. This is exemplified in Figure 1, which shows for an NER task that given biases in the training dataset, current models can fail to to categorize entity mentions correctly even with related context.

Even with the advent of highly successful deep learning based information extraction models using large-scale pre-trained language models (e.g. Bert [8]), challenges in dealing with long-tail data still exist [1, 19]. This is because many machine learning models have the default assumption that the training and target data follow the same distribution. However, as mentioned, in real-world applications, this assumption is often not true, especially in-cases with rare, *long-tail* entities. This phenomenon is referred to as *domain shift* [24]. Effectively adapting to changes in the test distribution is referred to as *domain adaptation* [24]. For IE/NLP tasks, there are several types of domain adaption methods including sample-based, feature-based, and inference-based methods [14].

One problem of domain shift is that the data representation is domain-specific when transforming data from one domain to match another. For example, the same entity mention Lincoln is more likely to be an organization in the automotive domain than when it occurs in the American history domain. Therefore, developing approaches that represent data in a domain-invariant space is important avenue for investigation [14]. We would like to learn invariant predictors that perform consistently irrespective of domain.

To that end, causality presents an avenue worth exploring, as it is a principled way to reason about shifts [2,16,22,23]. Only very recently has this been explored in the context of "stress tests" for NLP models [27] as well as distantly supervised NER [31]. However, both these approaches require that the causal graph is known a priori, including the graphical structure of the domain shift. In a general setting we might not know, in advance, the true underlying causal graph, or how the domain affect the features or the label.

Our goal, therefore, is to investigate new methods that are able to learn predictors that transfer across domains and do not require a known causal graph. We will focus on domain adaption specifically for information extraction tasks.

2 State of the Art

2.1 Information Extraction on the Web

Web data normally contains a large amount of data without annotations in various domains. There are a number of approaches for IE on web [18]. These approaches can be categorized broadly into two main classes: (1) semi-supervised and unsupervised systems and (2) supervised systems.

For the semi-supervised setting, many works tackle this problem with distant-supervised data-centric methods that generate distant annotations for unlabelled data. One recent work that focuses on IE from conversational data on the web is ConvSearch [29]. This paper addresses two main challenges in conversational search of online shopping domain: the imperfect entity attributes with multi-turn utterances in conversations and the lack of in-domain annotations for training due to the long-tail entities. ConvSearch combines dialogues and search systems to tackle the first challenge, as well as a jump-start dialog generation method (M2M-UT) for generating utterance and building dialogues to solve the second challenge. M2M-UT builds dialog outlines for online shopping from e-commerce search behavior data, and fills them with the generated utterances. As a result they introduce a new dataset CSD-UT for the online shopping domain. Though training with the newly introduced distantly annotated dataset improved performance, the method is limited to the shopping domain. To use this method in different domains, new datasets need to be generated. Additionally, the method does not take potential biases in the data generating process into consideration.

For supervised systems, recent work [21] introduces Learning To Adapt with Word Embeddings (L2AWE), a model-centric method that exploits the distributional representation of named entities for adapting the entity types predicted by a NER system trained on source generic schema (pre-defined entity types in any NER system) to a given schema. L2AWE takes word embeddings obtained from word2vec or Bert and learns to map between the source schema and target schema. L2AWE is able to achieve fairly good results on the target datasets #Microposts2015, #Microposts2016 and WNUT-17, however there is no comparisons between the distributional changes of the source domain and the target domain.

There is also work on generating knowledge graphs from a different domain. [7] proposed an overall pipeline for generating knowledge graphs from scientific domains by integrating different NLP tasks into the framework. The work aims to utilize state-of-the-arts from different NLP tasks such as entity extraction and relation extraction for knowledge graph generation. However, academic data normally has more structured formats and formal language than conversational data. Using a pipeline with unified dictionaries makes it possible to use different models from NLP tasks to generate knowledge graphs. The work does not aim at improving individual model performance. Instead, we focus on improving robustness of each model for different domains.

2.2 Causal NLP

Until recently, there has been limited attention at the intersection of causal inference and NLP. A recent survey [10] provides a systematic review of the existing literature, classifying it in two general directions: 1) NLP helps causality (i.e. estimating causal effects from text); 2) Causality helps NLP (i.e. improving the robustness and interpretability of NLP through the use of causal reasoning). In our research, we focus on the latter direction, and in particular on IE tasks.

Causality for Improving Robustness in NLP Veitch et al. [27] describe one of the few methods that exploit the knowledge of the causal relationship between features, labels and domain variables to improve model robustness. In particular, they consider the problem of learning a predictor f that predicts a label Y from features X, assuming there is an additional variable Z that captures the domain information. They consider two settings: (1) the relationship between the features X and the label Y is causal, i.e. $X \rightarrow Y$, or (2) this relationship is anti-causal, i.e. $X \leftarrow Y$. For each of these two settings, they propose different regularization terms.

Veitch et al. [27] consider a general setting in which the label Y and the domain Z might be confounded, which means that there is no general predictor f that can transfer stably across domains since $P(Y|f(X))$ might change for different values of Z. Thus, as a second-best option, they propose to identify sets of features that are not caused by the domain Z. In contrast, our research focuses on the case in which these stable predictors exist. Furthermore, we will also consider the case in which the causal graph is unknown.

Moreover, in this work the tasks addressed are easier to manipulate than typical NER tasks. In particular, the tasks are predicting the usefulness and sentiment of online reviews, in which a synthetic confounder is generated by substituting some tokens that should have no effect on the prediction. Both tasks use binary variables for the domain and label, which simplifies the problem, while in IE tasks we have often several possible classes.

Causality for distant supervision in NER Another example of using causality in NLP is the work of Zhang et al. who apply it to the task of distantly supervised NER [31]. Their method D-DSNER [31] addresses the *intra-dictionary* *bias* and *inter-dictionary bias* for this task in which dictionaries are mention-type pairs pre-acquired from existing knowledge bases such as Wikipedia.

In the current NER model they investigate, positive and negative instances are generated by matching input texts with the dictionary. The problem they identify is that these distantly supervised models are highly dependent on the quality of the dictionary.

To address this problem, D-DSNER [31], they employ a causal graph to identify bias in the NER model stemming from the dictionary and subsequently propose corresponding adjustments (e.g. to the dictionary) based on the causal theory.

In summary, the current state-of-the-art methods for IE on the web still have limitations generalizing across different domains and identifying the biases in data generating processes. Causality provides a powerful tool for tackling domain shifts in the invariant-feature space, however, using causality to help with domain adaptation in information extraction is under-explored.

3 Problem Statement

Given the above, our main research question is:

> *To what extent can domain adaptation for information extraction in the context of complex conversations be improved through the use of causality?*

Specifically, we can break down this into the following sub-research questions:

- **RQ1** *What is the performance of current state-of-the-art methods of IE for complex conversations in different domains?*
  Here, we aim to characterize the performance of information extraction models when applied to complex conversations in the setting where there are domain shifts. These experimental results can also be seen as baseline results. Preliminary results are shown in Section 6.
- **RQ2** *What are the types of domain shifts in IE tasks and how can we systematically identify them?*
  This research question focuses on identifying various distribution shifts between training data (source domain) and test data (target domain) that occur frequently in IE tasks. We are interested in what are the characteristics of the benchmark datasets that state-of-the-art models leverage to make predictions, e.g. spurious correlations. For example, most models fail to use the context, but only rely on the mention representations, which introduces bias and lack of robustness [12]. In our work, we would like to identify distributional shifts systematically. In particular, we will start by focusing on NER.
- **RQ3** *How can we use causality to reason about domain shift, even without a known causal graph?*
  This is the most challenging part of the research. The goal is to learn an invariant predictor so that the performance on IE tasks is consistent between

domains. The first step to address this will be applying the causal invariant regularizer [27] in an NER task. In particular, we will modify the approach of Zhang et al. to see if the use of this regularizer improves performance. Subsequently, if this is effective, we will investigate performance for domain adaptation given the distribution shifts identified in sub-question 2.

The targeted contributions of this research can be summarized as:

- a characterization of the performance of the state-of-the-art IE models on complex conversations from different domains;
- identification of distribution shifts for IE tasks; and
- novel methods for domain adaptation in IE based on causality.

4 Research Methodology and Approach

With the aim of addressing the research questions in Section 3, we will use different methodologies in different phrases. We first carry out experimental research and use empirical analysis to understand the performance of the state-of-the-art models from benchmarks in IE for complex conversations in the target domain as the baseline. We then use exploratory research to understand the specific type of distribution shifts in IE tasks. With this better understanding, we can carry out theoretical research to adapt existing causal inference frameworks for reasoning about such shifts. Last, we will evaluate the proposed framework on complex conversations across different domains and compare the results with the baseline.

IE systems contain many different tasks such as NER, CR, etc. We propose to start our research with NER first, and then extend to other tasks. Our aim is to create a framework that attempts to identify distribution shifts automatically and learn invariant predictors across different domains based on ideas from causality. This framework should be able to adapt across different domains.

Experiments will be carried out with Pytorch [1] as the training framework for machine learning models. Several NLP toolkits will be used such as NLTK [2], spacy [3] and sikit-learn[4]. The Huggingface[5] package will also be used for comparing performance with state-of-the-art Bert-based models.

5 Evaluation Plan

As mentioned in Section 4, we first carry out our research on NER tasks. The benchmark dataset CoNLL-2003 [25] will be used as our source domain. Two other datasets are proposed as target domains to evaluate our proposed approach, specifically, email conversations (CEREC) [5] and social media (W-NUT) [6].

CoNLL-2003 contains 1393 news articles from the Reuters Corpus with 4 different types.

[1] https://pytorch.org/ [2] https://www.nltk.org/ [3] https://spacy.io/
[4] https://scikit-learn.org/stable/ [5] https://huggingface.co/

The email conversation dataset CEREC [5] contains 36,448 email messages with 4 types of entities. Email conversations are one type of complex conversation which is characterized by their long forms, diverse language variations, and a huge variety of surface forms for each entity.

The W-NUT dataset [6], focuses on emerging and rare entities. W-NUT dataset consists of annotated texts from YouTube comments, Twitter posts, and StackExchange contents, which is composed of 5,691 posts and 3,890 entity mentions. Although some of the data might not have the long-form back and forth conversational features, it is still valuable to evaluate the performance on long-tail entities. All contents are user-generated and across different domains.

State-of-the-art NER models (e.g. Bert) will be trained on the CoNLL-2003 dataset combined with small amounts of data from the CEREC and W-NUT datasets. The model will be tested on the rest of the CEREC and W-NUT datasets. This unregularized model is our baseline. A causally-regularized model will be trained and tested on the same split as above and compared.

The evaluation metrics that will be used are *Accuracy, Precision, Recall*, and *F-measure*. Since the distribution of the data over the entity types is unbalanced, we will also calculate the *macro-average* and *micro-average* [30]. The macro-averaged measure gives equal weight to each class, regardless of their frequencies. Micro-averaged weight each class with respect to its number of instances.

6 Results

In our K-CAP 2021 paper [15], we started investigating IE in complex conversations by focusing on the performance of the current state-of-the-art models for cross-document coreference resolution in emails. Coreference resolution is the task of finding all mentions in text that refer to the same real-world entities. We can see that the coreferent relations in ECB+ are more structural and formal. The entities are more likely to exist in Wikipedia or other knowledge bases. On the other hand, the entities in the email conversations are contextual and require additional background knowledge.

In our paper, we investigated the different performances of state-of-the-art models (e.g. [3]) on the hand annotated corpus from the CEREC email dataset [5] with respect to the benchmark ECB+ news dataset [4]. The CD-CR model in [3] contains three components: a span_embedder, a span_scorer and a pairwise_scorer. In the first step, it extracts all possible mentions from the text and encodes them with the span_embedder, then it prunes the mentions given the score generated by span_scorer. Next, the mentions are paired and the pairs of mentions are scored by pairwise_scorer in terms of likelihood of being coreferent. Finally, the coreference chain will be obtained by agglomerative clustering. The results in [15] show that the CD-CR model for the email test set has an F1 score of 27.4, which is a 7 points drop compared to an F1 score of 34.4 for the ECB+ dataset. This drop has shown that the CD-CR model cannot generalize easily to the email setting due to the language variation in the email conversations and the less-frequently-used-entities caused distribution shift from different domains.

Our first next step will be applying the causal invariant regularizer [27] to the D-DSNER approach [31]. If the approach is effective, we will then apply the regularized model to the CEREC and W-NUT datasets as described in our evaluation plan.

7 Conclusions and Future Work

Complex conversations are important source of data on the web. They show a large distribution shift from benchmark datasets due to language variation and long-tail entities. Thus, applying current state-of-the-art IE models is challenging because of they struggle with generalization on data from different domains. Causality is a principled way to help with domain adaptation. Our research will investigate new methods for using such a principled approach in the context of information extraction.

Acknowledgements I would like to thank my supervisors Prof. Dr. Paul Groth and Dr. Sara Magliacane for their supervision and feedback. This research is funded by the Dutch Research Council (NWO) through grant MVI.19.032

References

1. Bommasani, R., Hudson, D.A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M.S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A.S., Creel, K., Davis, J.Q., Demszky, D., Donahue, C., Doumbouya, M., Durmus, E., Ermon, S., Etchemendy, J., Ethayarajh, K., Fei-Fei, L., Finn, C., Gale, T., Gillespie, L., Goel, K., Goodman, N.D., Grossman, S., Guha, N., Hashimoto, T., Henderson, P., Hewitt, J., Ho, D.E., Hong, J., Hsu, K., Huang, J., Icard, T., Jain, S., Jurafsky, D., Kalluri, P., Karamcheti, S., Keeling, G., Khani, F., Khattab, O., Koh, P.W., Krass, M.S., Krishna, R., Kuditipudi, R., et al.: On the opportunities and risks of foundation models. CoRR **abs/2108.07258** (2021), https://arxiv.org/abs/2108.07258
2. Bühlmann, P.: Invariance, causality and robustness. Statistical Science **35**(3), 404–426 (2020)
3. Cattan, A., Eirew, A., Stanovsky, G., Joshi, M., Dagan, I.: Streamlining cross-document coreference resolution: Evaluation and modeling **abs/2009.11032** (2020)
4. Cybulska, A., Vossen, P.: Using a sledgehammer to crack a nut? lexical diversity and event coreference resolution. In: LREC 2014 (2014)
5. Dakle, P.P., Moldovan, D.: CEREC: A corpus for entity resolution in email conversations. In: Proceedings of the 28th International Conference on Computational Linguistics. pp. 339–349. International Committee on Computational Linguistics, Barcelona, Spain (Online) (Dec 2020), https://www.aclweb.org/anthology/2020.coling-main.30
6. Derczynski, L., Nichols, E., van Erp, M., Limsopatham, N.: Results of the WNUT2017 shared task on novel and emerging entity recognition. In: Proceedings of the 3rd Workshop on Noisy User-generated Text. pp. 140–147. Association for Computational Linguistics, Copenhagen, Denmark (Sep 2017). https://doi.org/10.18653/v1/W17-4418, https://aclanthology.org/W17-4418

7. Dessì, D., Osborne, F., Recupero, D.R., Buscaldi, D., Motta, E.: Generating knowledge graphs by employing natural language processing and machine learning techniques within the scholarly domain. CoRR **abs/2011.01103** (2020), https://arxiv.org/abs/2011.01103
8. Devlin, J., Chang, M., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. CoRR **abs/1810.04805** (2018), http://arxiv.org/abs/1810.04805
9. van Erp, M., Mendes, P., Paulheim, H., Ilievski, F., Plu, J., Rizzo, G., Waitelonis, J.: Evaluating entity linking: An analysis of current benchmark datasets and a roadmap for doing a better job. In: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16). pp. 4373–4379. European Language Resources Association (ELRA), Portorož, Slovenia (May 2016), https://aclanthology.org/L16-1693
10. Feder, A., Keith, K.A., Manzoor, E., Pryzant, R., Sridhar, D., Wood-Doughty, Z., Eisenstein, J., Grimmer, J., Reichart, R., Roberts, M.E., Stewart, B.M., Veitch, V., Yang, D.: Causal inference in natural language processing: Estimation, prediction, interpretation and beyond (2021)
11. Gao, J., Galley, M., Li, L.: Neural approaches to conversational AI. CoRR **abs/1809.08267** (2018), http://arxiv.org/abs/1809.08267
12. Ghaddar, A., Langlais, P., Rashid, A., Rezagholizadeh, M.: Context-aware Adversarial Training for Name Regularity Bias in Named Entity Recognition. Transactions of the Association for Computational Linguistics **9**, 586–604 (07 2021). https://doi.org/10.1162/tacl_a_00386, https://doi.org/10.1162/tacl_a_00386
13. Ilievski, F., Vossen, P., Schlobach, S.: Systematic study of long tail phenomena in entity linking. In: Proceedings of the 27th International Conference on Computational Linguistics. pp. 664–674. Association for Computational Linguistics, Santa Fe, New Mexico, USA (Aug 2018), https://aclanthology.org/C18-1056
14. Kouw, W.M., Loog, M.: A review of domain adaptation without target labels. IEEE Transactions on Pattern Analysis and Machine Intelligence **43**(3), 766–785 (2021). https://doi.org/10.1109/TPAMI.2019.2945942
15. Li, X., Magliacane, S., Groth, P.: The challenges of cross-document coreference resolution for email. In: Proceedings of the 11th on Knowledge Capture Conference. p. 273–276. K-CAP '21, Association for Computing Machinery, New York, NY, USA (2021). https://doi.org/10.1145/3460210.3493573, https://doi.org/10.1145/3460210.3493573
16. Magliacane, S., van Ommen, T., Claassen, T., Bongers, S., Versteeg, P., Mooij, J.M.: Domain adaptation by using causal inference to predict invariant conditional distributions. CoRR **abs/1707.06422** (2017), http://arxiv.org/abs/1707.06422
17. Manning, C.D., Surdeanu, M., Bauer, J., Finkel, J.R., Bethard, S., McClosky, D.: The stanford corenlp natural language processing toolkit. In: ACL (System Demonstrations). pp. 55–60. The Association for Computer Linguistics (2014), http://dblp.uni-trier.de/db/conf/acl/acl2014-d.htmlManningSBFBM14
18. Martínez-Rodríguez, J., Hogan, A., López-Arévalo, I.: Information extraction meets the semantic web: A survey. Semantic Web **11**(2), 255–335 (2020). https://doi.org/10.3233/SW-180333, https://doi.org/10.3233/SW-180333
19. McCoy, T., Pavlick, E., Linzen, T.: Right for the wrong reasons: Diagnosing syntactic heuristics in natural language inference. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. pp. 3428–3448. Association for Computational Linguistics, Florence, Italy (Jul 2019). https://doi.org/10.18653/v1/P19-1334, https://aclanthology.org/P19-1334

20. Nickel, M., Murphy, K., Tresp, V., Gabrilovich, E.: A review of relational machine learning for knowledge graphs. Proceedings of the IEEE **104**(1), 11–33 (Jan 2016). https://doi.org/10.1109/jproc.2015.2483592, http://dx.doi.org/10.1109/JPROC.2015.2483592
21. Nozza, D., Manchanda, P., Fersini, E., Palmonari, M., Messina, E.: Learningtoadapt with word embeddings: Domain adaptation of named entity recognition systems. Information Processing Management **58**(3), 102537 (2021). https://doi.org/https://doi.org/10.1016/j.ipm.2021.102537, https://www.sciencedirect.com/science/article/pii/S0306457321000455
22. Peters, J., Bühlmann, P., Meinshausen, N.: Causal inference by using invariant prediction: identification and confidence intervals. Journal of the Royal Statistical Society. Series B (Statistical Methodology) pp. 947–1012 (2016)
23. Peters, J., Janzing, D., Schölkopf, B.: Elements of causal inference: foundations and learning algorithms. The MIT Press (2017)
24. Ramponi, A., Plank, B.: Neural unsupervised domain adaptation in NLP—A survey. In: Proceedings of the 28th International Conference on Computational Linguistics. pp. 6838–6855. International Committee on Computational Linguistics, Barcelona, Spain (Online) (Dec 2020). https://doi.org/10.18653/v1/2020.coling-main.603, https://aclanthology.org/2020.coling-main.603
25. Tjong Kim Sang, E.F., De Meulder, F.: Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition. In: Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003. pp. 142–147 (2003), https://aclanthology.org/W03-0419
26. Tu, J., Lignos, C.: TMR: Evaluating NER recall on tough mentions. In: Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Student Research Workshop. pp. 155–163. Association for Computational Linguistics, Online (Apr 2021). https://doi.org/10.18653/v1/2021.eacl-srw.21, https://aclanthology.org/2021.eacl-srw.21
27. Veitch, V., D'Amour, A., Yadlowsky, S., Eisenstein, J.: Counterfactual invariance to spurious correlations: Why and how to pass stress tests (2021)
28. Weikum, G., Dong, L., Razniewski, S., Suchanek, F.M.: Machine knowledge: Creation and curation of comprehensive knowledge bases. CoRR **abs/2009.11564** (2020), https://arxiv.org/abs/2009.11564
29. Xiao, L., Ma, J., Dong, X.L., Martínez-Gómez, P., Zalmout, N., Chen, W., Zhao, T., He, H., Jin, Y.: End-to-end conversational search for online shopping with utterance transfer. In: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. pp. 3477–3486. Association for Computational Linguistics, Online and Punta Cana, Dominican Republic (Nov 2021). https://doi.org/10.18653/v1/2021.emnlp-main.280, https://aclanthology.org/2021.emnlp-main.280
30. Yang, Y., Liu, X.: A re-examination of text categorization methods. In: Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. p. 42–49. SIGIR '99, Association for Computing Machinery, New York, NY, USA (1999). https://doi.org/10.1145/312624.312647, https://doi.org/10.1145/312624.312647
31. Zhang, W., Lin, H., Han, X., Sun, L.: De-biasing distantly supervised named entity recognition via causal intervention (2021)