Towards Context-aware Knowledge Entailment from Health Conversations

Saeedeh Shekarpour,[1], Faisal Alshargi,[2], Mohammadjafar Shekarpour
[1] University of Dayton, Dayton, United States
[2] University of Leipzig, Leipzig, Germany
sshekarpour1@udayton.org, alshargi@informatik.uni-leipzig.de, mj.shekarpour@gmail.com

Abstract

Despite the competitive efforts of leading companies, cognitive technologies such as chatbot technologies still have limited cognitive capabilities. One of the major challenges that they face is knowledge entailment from the ongoing conversations with a user. Knowledge entailment implies entailing facts that indicate opinions, beliefs, expressions, requests, and feelings of a particular user about a particular target during conversations. The entailed pieces of knowledge will evolve the background knowledge graph of cognitive technologies and advance their contextual inference and reasoning capabilities. Although the Natural Language Processing (NLP) community deals with the Recognizing Textual Entailment (RTE) task, it is treated in a static manner where the predefined hypothesis is typically fed to the learning model, and then the model decides whether it is an entailment or not. However, since the discourse of conversations is dynamic and unpredictable, the traditional RTE approach does not suffice in the context of conversational agents. In this vision paper, we demonstrate our work in progress as to inject background knowledge into machine learning approaches where it entails facts using domain-specific ontologies and contextualized knowledge. Further, we propose investigating solutions for extending or transferring this approach to other domains. We frame our discussion in a case study related to mental health conversations.

Introduction

Knowledge entailment has applicability in various cognitive technologies such as conversational AI interfaces (chatbot technologies), which recently gained the competitive efforts of leading companies. The existing implementations of this technology have limited cognitive capabilities where they fail to perceive users' opinions, beliefs, expressions, requests, and feelings. Knowledge entailment (also known as knowledge perception) primarily from the text and secondarily from conversations between a bot (i.e., conversational agent) and a user is a major challenge. Knowledge entailment mainly implies entailing facts that demonstrate opinions, beliefs, expressions, requests, and feelings of a particular user about a particular target (object) from conversations. These entailed pieces of knowledge will evolve the background knowledge about the user. Richer background knowledge will extend the future contextual inference and reasoning capabilities of cognitive technologies built up on top. Although the Natural Language Processing (NLP) community deals with the Recognizing Textual Entailment (RTE) task, it is treated in a static manner where the predefined hypothesis is typically fed to the learning model, and then the model decides whether it is entailment or not. However, the knowledge entailment task, particularly in our context (conversations being dynamic, unpredictable, and complex), requires a model for the purposes of not only learning entailments but also for inferring all possible hypotheses (including entailing, contradicting, etc). Recently, the combination of knowledge representation and machine learning has been in the center of attention towards reaching an explainable, accountable, and fair AI which will exhibit more robust intelligence and reliable capabilities (Holzinger et al. 2017; Samek, Wiegand, and Müller 2017). Knowledge representation provides essential conceptualization (domain ontology), contextual entities, associated facts, and, more importantly, relations between entities and concepts. In this paper, we describe our work in progress as it proposes an ontology-based knowledge entailment approach over the discourse of conversations. We demonstrate our envisioned plan in an illustrative mode to display the open research areas required the future attention of the community. This paper is organized as follows: Section 2 discusses the limitations of the state-of-the-art. Section 3 presents the problem statement, followed by Section 4, which showcases a case study. Next, we demonstrate our ultimate envisioned plan. We close with the remarks related to the applicabilities of our knowledge entailment approach in chatbot technologies.

Premise	Hypothesis	E	C	N
An older man is drinking orange juice at a restaurant	1) A man is drinking juice.	✓	✗	✗
	2) Two women are at a restaurant drinking wine.	✗	✓	✗
	3) A man in a restaurant is waiting for his meal to arrive.	✗	✗	✓
Children going home from school	1) The children are at the library.	✗	✓	✗
	2) The school children head home.	✓	✗	✗
	3) The children are walking in the afternoon.	✗	✗	✓

Figure 1: Samples of data for RTE task, two given premises with the possible hypotheses. The RTE model determines whether a given hypothesis is an Entailment (E), or Contradiction (C) or Neural (N).

Limitations of the state-of-the-art

RTE (Dagan et al. 2010) is a task in NLP where it determines whether two given sentences (i) contradict each other, (ii) are semantically unrelated to each other, or (iii) one of them (premise) entails the other one (hypothesis). Figure 1 showcases multiple examples. For instance for the given premise `An older man is drinking orange juice at a restaurant`, three hypotheses are listed. The first premise `A man is drinking juice.` is an entailment (E) of the premises whereas the second hypothesis `Two women are at a restaurant drinking wine` is a contradiction C and the third one `A man in a restaurant is waiting for his meal to arrive.` is neutral (N). The work presented in (Bowman et al. 2015) was published in the Stanford Natural Language Inference (SNLI) corpus, which is far larger than all of the other existing resources of its type. It contains more than 500K pairs of sentences, which are annotated using the labels E (entailment), C (contradiction), and N (neutral). The RTE task is substantially important in information extraction, text summarization, classification, and machine translation. The NLP community deals with static RTE tasks where the hypothesis is fed to the learning model, and then the model decides it is an entailment or not, while in scenarios such as knowledge entailment from conversations, we have to develop a generative model where the model can dynamically generate hypotheses and there is no predefined hypothesis.

Problem Statement

In the era of contemporary conversational AI, the first major deficiency attributes to the lack of a convincing approach for knowledge entailment from conversations, e.g., whether or not a chatbot learns the user by entailing knowledge from ongoing conversations, and evolves its underlying knowledge for future conversation management. The second deficiency is that the available approaches are solely data-driven approaches (i.e., machine learning approaches) or rule-based approaches, and in both cases, the inference capabilities are limited to the underlying data and rules. Thus, they fail to overcome unpredicted situations. The machine learning approaches are solely data-driven. Advancing them with explicit knowledge will result in faster convergence on sparse data. Furthermore, it makes them explainable, compliant to the domain, and more robust against noise. Figure 2A shows the static RTE task which predicts (discriminates) the proper label (entailment, contradiction, and neutral) for the given input text (premise) and the given hypothesis. In this scenario the input hypothesis is supposed to be given by the user. In our envisioned model (Figure 2B) there is no need to worry about the hypothesis because they are automatically fed to a generative model using the existing facts from the background knowledge graph. We plan to extend a knowledge entailment approach (which is a neural network approach) fed with domain-specific ontologies to contextual as well as personalized knowledge graphs; it will not only be a data-driven approach but also a knowledge-driven approach. To present a clear and practical vision of our proposed scenario, we frame a case study on the health domain which entails knowledge from the conversations about mental health.

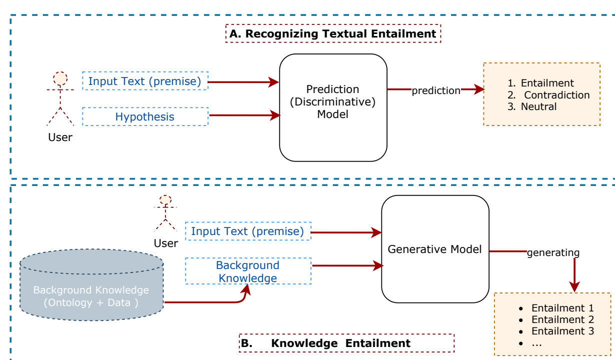

Figure 2: Part (A) is a textual entailment module, which is a discriminative model, and Part (B) is a knowledge entailment module, which is a generative model.

Case Study

Figure 3 demonstrates our expectations from a knowledge entailment approach over the conversations. There is a given excerpt from our underlying conversation dataset (will be introduced in the following). This excerpt shows semantics referring to insomnia which is a subject question in most of the questionnaires (such as PHQ-8 and PHQ-9) for depression disorder. However, entailing these semantics requires considering the indicators of insomnia, in addition to the contextual information from several lines in the course of the conversation. Considering a given question from PHQ-9 inquiring about the status of the patient's sleep, our expectation is that our envisioning approach can entail the piece of knowledge that "the patient has a sleep disorder often". In the following, we introduce the sources of knowledge which will be incorporated in our approach.

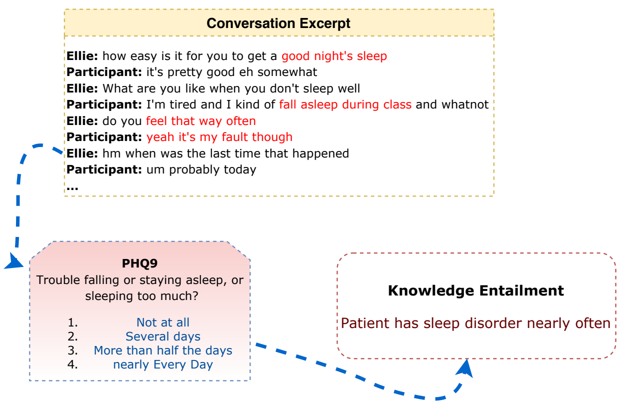

Figure 3: A conversation excerpt, with the entailed statement based on the background knowledge from PHQ9.

PHQ-9 Ontology and Lexicon: The Diagnostic and Statistical Manual of Mental Disorders (DSM) (Association et al. 2013) suggests that clinical depression can be diagnosed through the presence of a set of symptoms over a fixed period of time. The PHQ-9 (Löwe et al. 2004) is a nine-item depression scale that incorporates DSM-V. It can be utilized to screen, diagnose, and measure the severity of depression. We are building an ontology from PHQ-9 where it incorporates all the concepts, depression symptoms and relevant phrases.

Dataset: The Distress Analysis Interview Corpus Wizard-of-Oz (DAIC-WoZ) interview database (Gratch et al. 2014; DeVault et al. 2014) consists of clinical diagnostic interviews designed to support the diagnosis of psychological disorders such as anxiety, depression, and post-traumatic stress disorder. This corpus (DAIC) comprises recorded interviews between a patient (participant) and a computerized animated virtual interviewer "Ellie". It contains data from 189 interviews, including transcripts, audio, and video recordings, and PHQ depression questionnaire responses.

Personalized Healthcare Knowledge Graph (PHKG): The work presented in (Gyrard et al. 2018) introduces PHKG, which is described as a representation of all relevant medical knowledge and personal data for a patient. PHKG can support the development of innovative applications such as digital personalized coach applications that can keep patients informed, help to manage their chronic condition, and empower physicians to make effective decisions on health-related issues or receive timely alerts as needed through continuous monitoring. Typically, PHKG formalizes medical information in terms of relevant relationships between entities. For instance, a knowledge graph (KG) for asthma can describe causes, symptoms, and treatments for asthma, and a PHKG can be the subgraph containing just those causes, symptoms, and treatments that are applicable to a given patient. In our case study, PHKG is limited to the knowledge about the patient which is determined from conversations.

Envisioned Plan

Figure 4 schematically shows our envisioned plan as the given data in the background knowledge graphs (personalized health graph and contextualized graph). Then, our knowledge entailment approach will drive further knowledge from conversations. The third two validation and quality assurance strategies will be applied to determine whether entailed knowledge is valid or not. This step might rely on manual approaches such as crowd-sourcing or automatic approaches such as graph completion and reasoning to validate entailed knowledge. Finally, the newly entailed knowledge is added to PHKG and contextualized graphs. Having an iteration over this cycle or upcoming conversation will help to both entail further knowledge or augment our entailment approach.

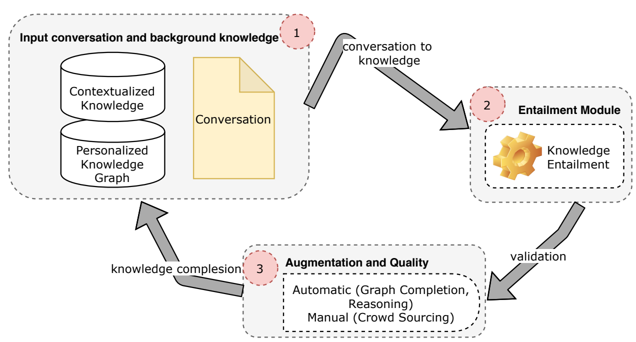

Figure 4: The process of entailing new knowledge from conversations, validating that, and adding to the personalized knowledge graph.

Our model will entail triples (subject-predicate-object) from conversations where the subject is the ongoing user, and predicates and objects represent opinions, beliefs, expressions, requests, and feelings belonging to the user. Figure 5 demonstrates the transformation of the input text into a graph that contains all the entailed facts about the patient. We assume that all the required relations (predicates) and possible objects are declared in the domain ontology. If we develop an attention model that is fed with the context (cognitive ontology, context, and personalized knowledge) along with the user utterance then possibly one or multiple relations and objects acquire higher weight (attention) – meaning they are entailed from the input utterance. Furthermore, the higher the volume of conversations with the user, the better the context-aware knowledge entailment. The key novel part of this work is combining domain knowledge representation and machine learning approaches to provide robust, explainable, and context-aware solutions.

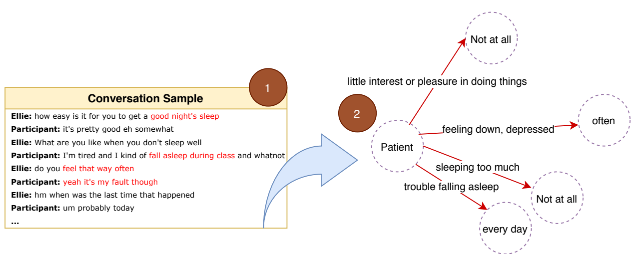

Figure 5: The process of transforming a given ongoing conversation to a graph containing entailed facts about the patient.

Applicability in Chatbot Technology

A chatbot is typically an Artificial Intelligence (AI)-based application designed to simulate a conversation with human users in a continuous and common sense manner (Lee, Oh, and Choi 2017). This assistance can reduce the cognitive load for the user, especially in high-pressure situations such as surgical operations, battlefields, and disaster preparedness and response. Furthermore, it is promising and effective in everyday life activities, such as retail, travel, news, and entertainment. However, despite the recent competitive efforts and investments of leading companies (e.g., Facebook (Messenger), Microsoft (Cortana), Apple (Siri), Google (Duplex), WeChat, and Slack), the existing implementations do not provide impressive cognitive capabilities. For example, state-of-the-art chatbots still struggle with simple conversational domains, such as task ordering (Microsoft challenge (Li et al. 2018, 2016), bAbI project of Facebook (Bordes, Boureau, and Weston 2016; Weston et al. 2015)), and is still far from complicated conversations in a variety of domains. Our proposed work, if successful, is a complementary step for future chatbot technologies.

References

Association, D.-. A. P.; et al. 2013. Diagnostic and statistical manual of mental disorders. *Arlington: American Psychiatric Publishing* .

Bordes, A.; Boureau, Y.-L.; and Weston, J. 2016. Learning end-to-end goal-oriented dialog. *arXiv preprint arXiv:1605.07683* .

Bowman, S. R.; Angeli, G.; Potts, C.; and Manning, C. D. 2015. A large annotated corpus for learning natural language inference. *arXiv preprint arXiv:1508.05326* .

Dagan, I.; Dolan, B.; Magnini, B.; and Roth, D. 2010. Recognizing textual entailment: Rational, evaluation and approaches–erratum. *Natural Language Engineering* 16(1): 105–105.

DeVault, D.; Artstein, R.; Benn, G.; Dey, T.; Fast, E.; Gainer, A.; Georgila, K.; Gratch, J.; Hartholt, A.; Lhommet, M.; et al. 2014. SimSensei Kiosk: A virtual human interviewer for healthcare decision support. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, 1061–1068.

Gratch, J.; Artstein, R.; Lucas, G. M.; Stratou, G.; Scherer, S.; Nazarian, A.; Wood, R.; Boberg, J.; DeVault, D.; Marsella, S.; et al. 2014. The distress analysis interview corpus of human and computer interviews. In *LREC*, 3123–3128. Citeseer.

Gyrard, A.; Gaur, M.; Shekarpour, S.; Thirunarayan, K.; and Sheth, A. 2018. Personalized Health Knowledge Graph. In *ISWC 2018 Contextualized Knowledge Graph Workshop*.

Holzinger, A.; Biemann, C.; Pattichis, C. S.; and Kell, D. B. 2017. What do we need to build explainable AI systems for the medical domain? *arXiv preprint arXiv:1712.09923* .

Lee, D.; Oh, K.-J.; and Choi, H.-J. 2017. The chatbot feels you-a counseling service using emotional response generation. In *2017 IEEE International Conference on Big Data and Smart Computing (BigComp)*, 437–440. IEEE.

Li, X.; Lipton, Z. C.; Dhingra, B.; Li, L.; Gao, J.; and Chen, Y.-N. 2016. A user simulator for task-completion dialogues. *arXiv preprint arXiv:1612.05688* .

Li, X.; Wang, Y.; Sun, S.; Panda, S.; Liu, J.; and Gao, J. 2018. Microsoft dialogue challenge: Building end-to-end task-completion dialogue systems. *arXiv preprint arXiv:1807.11125* .

Löwe, B.; Kroenke, K.; Herzog, W.; and Gräfe, K. 2004. Measuring depression outcome with a brief self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). *Journal of affective disorders* 81(1): 61–66.

Samek, W.; Wiegand, T.; and Müller, K.-R. 2017. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *arXiv preprint arXiv:1708.08296* .

Weston, J.; Bordes, A.; Chopra, S.; Rush, A. M.; van Merriënboer, B.; Joulin, A.; and Mikolov, T. 2015. Towards ai-complete question answering: A set of prerequisite toy tasks. *arXiv preprint arXiv:1502.05698* .