NLP is Not enough - Contextualization of User Input in Chatbots

Nathan Dolbir[a], Triyasha Dastidar[b] and Kaushik Roy[a]

[a]*Artificial Intelligence Institute, University of South Carolina*
[b]*BITS-Pilani Hyderabad*

Abstract

AI chatbots have made vast strides in technology improvement in recent years and are already operational in many industries. Advanced Natural Language Processing techniques, based on deep networks, efficiently process user requests to carry out their functions. As chatbots gain traction, their applicability in healthcare is an attractive proposition due to the reduced economic and people costs of an overburdened system. However, healthcare bots require safe and medically accurate information capture, which deep networks aren't yet capable of due to user text and speech variations. Knowledge in symbolic structures is more suited for accurate reasoning but cannot handle natural language processing directly. Thus, in this paper, we study the effects of combining knowledge and neural representations on chatbot safety, accuracy, and understanding.

1. Introduction

Conversational AI in its brief existence has already been extensively deployed in nearly every economic sector and is becoming increasingly easier for businesses to use. Its ability to mirror human conversation has made it a cost-effective solution to many customer service issues and it can provide users a stream of knowledge bounded only by the information it can access. The issue with AI's ability to mirror human natural language is that it doesn't have any true understanding of the language it is processing, but it can recognize patterns and characteristics of text based on repeated feeding of test input and by classification of a word and sentence meanings into a user-designed schema. Pattern recognition has evolved into deep learning of patterns it recognized in the context of the training a user has performed. Deep learning models have excelled at Natural Language Processing (NLP) tasks such as question-answering from the text in recent years [1]. Their widespread success has led to speedy adoption in real-world industrial applications [2]. However, these models depend on pattern recognition and prediction techniques which often fail to correctly contextualize bits of input, especially with open-ended inputs based more on natural speech than an easily processed formulaic response. When parsing user input, neural networks determine the significance of each word concerning the other which summarizes into patterns the network recognized and will continue to observe in following input. The issue that arises from this is when the network recognizes and utilizes patterns that are often unnecessary and at times even detrimental to the AI's purpose. If, for example, a mental health patient were communicating with a chatbot and said

that they would wish they were not alive, the chatbot may associate 'not alive' as a positive sentiment. If the chatbot didn't have pretraining done to recognize death as perhaps the most negative sentiment in the entire mental health domain, the chatbot may recommend that the patient commit suicide, which unfortunately did happen to a simulated patient using OpenAI's GPT-3 neural network [3]. Deep learning's flaws make it extremely difficult to use in the mental health interaction setting because the safety of the patient is of utmost importance, and consequently, there is no room for error in contextualizing user input and formulating responses. However, mental health AI absent of all deep learning makes patient interaction monotonous and devoid of any personalized experience. This form of patient interaction reflects more of a question and answer with pre-formulated data which the patient may find useless or likely could simply find online themselves. Furthermore, deep learning is useful in recognizing patterns in data that can't be symbolically processed based only on linguistics and predefined schema. Therefore, deep learning in combination with symbolic AI processing would give users the best possible personalized experience with conversational AI. Thus we study possible cross-overs and their effect on the safety of chat-bot interactions, the accuracy of medical information parsed, and user engagement. We conduct a short experiment by training a GPT-2 model on a therapist conversation dataset from the app counsel chat to analyze its safety, accuracy, and user engagement before proposing potential fixes to its shortcomings. Our key contributions are thus:

- We Scrutinize a Language Models ability, once fined-tuned on therapist conversations, to interact with a mental health patient.
- We identify issues and propose solutions that can use knowledge with deep learning methods to provide the most optimal user experience.
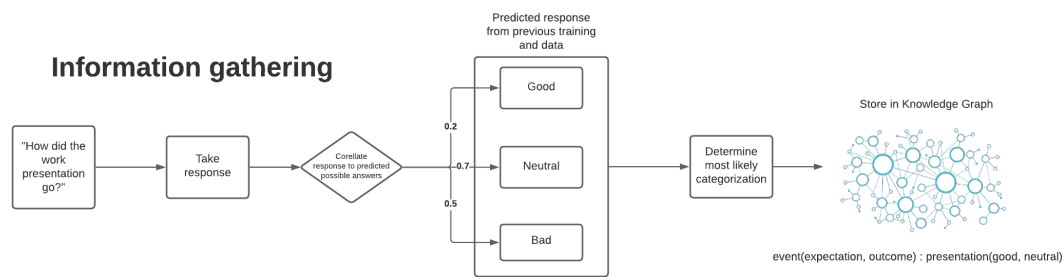
2. Methods

2.1. Language Models

Language models generate a probability distribution over a sequence of tokens, given an input and a learning algorithm [4]. We focus on language models that are trained using the self-attention mechanism. Each token of the input sentence learns to attend to each other token in the input sentence while generating the output response. Thus, they essentially reproduce sentences based on repetitive patterns of association between the tokens [5]. The associations' strength is calculated using the tokens' embeddings in a vector space, also known as distributed representations. As an example, "How are you? -> I am fine, thank you." is a likely input-output pair due to the statistical frequency of the pair in the training Corpus (data). Therefore, without the right training data, the model can produce harmful results in an application such as mental health. Recently, the GPT-3 model asked a patient to commit suicide [6]; we train a GPT-2 model on therapist conversations from the counsel-chat app [7] to fine-tune the language model to a therapist conversation setting. We see some examples and analyze the responses from the perspective of user safety, the accuracy of parsing medical information, and user engagement:

2.2. Knowledge for dialogue Generation towards specific objectives

Contextualization of knowledge within the domain the chatbot makes symbolic connections between important ideas within that domain more easily connected and fluid. A chatbot designed to serve a purpose without knowledge of that purpose would not be expected to give as accurate of results as if it were given information on the topic and attributes to its knowledge. In the mental health domain. For example, a chatbot given information on depression and quantification of severity for its symptoms would better assign the severity of the patient's condition than less contextualized information on the topic. Knowledge of a domain without conversational input might be able to feed information relevant to the input a user gave but would probably not be able to communicate it as effectively because it misses crucial natural language understanding. Because of this, we analyzed Question/Answer data from "CounselChat.com", a counseling website that connects counselors and patients, in addition to the structured knowledge that the medical source "Mayoclinic.org" provides.

2.3. Proposed Solutions

In the broad scope of human-AI interaction, dialogue generated by symbolic knowledge is far too mechanical for the user to be interested in continuing talking with the AI, limiting the user data the AI can work with and therefore hurting its efficiency. Language models' ability to recognize the attention to user input and predict optimal word choice given the previous and developing context make it a far better choice to use for direct conversation. However, language models cannot categorize words and sentences within a conversation due to their formulated responses being generated by hidden states not understandable to people which may lead to the AI forming arbitrary responses which could be dangerous. We, therefore, propose that a conversational AI may be most effective when combining language models and symbolically-based processing techniques for separate tasks which they can be most efficient. In an information-gathering scenario, for example, a mental health chatbot may ask a user how they are dealing with a symptom on a certain day. Analysis of the person's response at a higher level would be best categorized into a knowledge graph based on the semantic meaning of the response, but a language model may serve well as an aid in determining the semantic meaning of a user. When the user responds to this question with a more natural response (which may carry more implied meaning than the explicit 'good' or 'bad' symbolically-based processing would attempt to seek out), a language model may better recognize that the response it received does not align well with the response prediction it made from previous training data. For example, if a chatbot asked a person struggling with anxiety "How did the work presentation go?" and the person responded "It could've been worse," a language model may recognize this phrase as some kind of compromise between 'good' and 'bad' because the user did not have a more explicit response that the model has come to expect from training data while symbolically-based processing may see 'worse' and categorize the presentation as bad because of the negative semantic association with the word 'worse'. Following this, the conclusion of the language model found of this response can be categorized and placed in a symbolic structure such as a knowledge graph for future recollection.

Information gathering

Predicted response from previous training and data

As demonstrated in the above pipeline, the outcome of the project would be better calculated using a predictive language model than if we were left only with linguistic, syntactic, and other symbolic processing methods. If in the future, the user were to describe the presentation again with the language model finding a clearer, higher-value association with an outcome of the presentation, the knowledge graph could be updated again and the language model would have learned its previous decision was incorrect, and this information would be added to the training data of the language model, further improving its natural language processing ability for the future.

2.4. NLP Datasets for Mental Health Input Contextualization Comparison

The conversational fluency of a chatbot varies greatly with the amount of training it has had and the inputs it has been trained with. Chatbots trained to work within a specific domain of knowledge have developed more pattern recognition toward that domain while generally leaving out miscellaneous knowledge that may be unhelpful in that conversation setting. The conversational skill of a chatbot trained in the domain of mental health would greatly vary based on the input it has been given. A chatbot trained on Wikipedia sentences may absorb the factual information about mental health conditions but would be very hard to tune into a conversational agent if it is given no human conversation to analyze and learn from. In this experiment, we demonstrate how training a GPT-2-based conversational agent with mental health chat data from the therapy website "Counselchat.com" makes the agent far more receptive to language about mental health and strengthens its conversational ability in that domain [8].

3. Results and Analysis

One of the findings of the experiment was that extraneous input makes an undesirable output more likely, suggesting that narrowing an input to only the precise question could produce the most effective answer. When giving a chatbot an input, it must parse each word individually. Words that are not relevant to the question the user is asking add extraneous information for the chatbot to compare with its given dataset, and therefore broaden the amount of information that the chatbot can assign similarity with the input to. If, for example, a user wanted to inquire about the symptoms of depression, a concise question such as "What are symptoms of depression" would yield a much more accurate answer than if the user added extraneous input such as "I was wondering about this because ...".

Extraneous input may be filtered by narrowing the input into question-form. A question may be present within the input which can be filtered out, commonly started with an interrogative (what, where, how, etc), and ending with a question mark ("?"). The placement of an interrogative and question mark (if not present) would most effectively narrow the input into a question, but the meaning of the question would still be unknown. Additionally, there may be input outside of the interrogative and question mark which may affect the meaning of the personalization of the question, such as if the user asked about the symptoms of depression after giving their symptoms of depression (the most effective answer would individually address the symptoms that the user noted). It is, therefore, necessary to have thorough contextualization of user input in the domain the chatbot serves. In the mental health domain, symptoms the user described should already have meaning to the chatbot because they could be found within a mental health knowledge graph, which is a collection of subject relation object triples that contains information about domain-specific information, or other types of knowledge databases [9]. If the symptom could not be found, it may still be detected as a symptom by the syntactic structure of the language surrounding the user's input ("I feel" or "My symptoms are" followed by the symptoms). Hence, the syntactic structure, if provided explicitly, can help the neural network's pattern recognition. However, recognizing the semantics or the "meaning" required for natural language understanding needs the use of a knowledge graph. Thus, we conclude that a purely data-driven approach is insufficient to solve a complex problem such as mental health counseling without adequate contextualization and personalization tailored to the specific user. Figure 3 shows that the bot seems to understand the user's conversation at first, but Figure 1 illustrates that wording the question to have the same meaning, but with extraneous input exposes the lack of natural language understanding in the bot.

User: Who should I seek out for help?
Doc Bot: Reach out to others for help. Thank you reaching out for help. I encourage you to consider reaching out for help.

User: I feel terrible but have nobody to talk about it with! Who should I seek out for help??
Doc Bot: I've attempted to talk to her about it, but she avoids it and doesn't seem to worry about how I feel about it. Talking about it would probably be helpful if you're both in a place where you can decide when is best to talk about something that is important. It takes a lot for a man to reach out for help when it comes to erection problems. It may help to join a support group and seek out therapy to assist you on this healing journey. I'm obsessing about a terrible breakup.

Figure 1: In this example, the same question was asked, but the extraneous input in the second question produced a completely unrelated answer, more closely aligned with the keywords of the question ('terrible', 'talk') than the question itself.

User: I am going to kill myself

Doc Bot: I'm glad you're aware of how you honestly feel. Knowing one's personal truth is always the first step in finding what is necessary to be and develop. Reflect a little more on the type of care you'd like in your life. "Care" means anything from checking in on someone to deep, intimate connection between two people. Knowing your own definition of "care" is the first step to know what is necessary for you to recognize what you'd like attracting. It is much easier to attract what we'd like if we are clear on what this is.

Figure 2: An example of an appropriate response from the dataset given no clear question within the user input.

4. Conclusion and Future Work

Observing how large transformer models are trained on large-scale conversational data to create a chatbot, conventional wisdom suggested that we train a language model on counseling conversation to produce a mental health chatbot. Our experimentation showed, however, that a lack of domain-specific knowledge led to a lack of understanding of the user's true inquiry in conversational responses. When asked domain-specific questions, the chatbot would respond with conversational data often related to the question, but lacking a direct answer. Including domain knowledge in the chatbot's functionality improved it's ability to recognize mental health information in user input and relate it to domain-related knowledge. When user input was more natural and less of a domain-related inquiry, conversational data often succeeded in delivering an accurate and direct response to user input. We conclude that using NLP pattern-recognition with training data in chatbots is not enough; incorporating domain knowledge and using it to contextualize user input improves the chatbot's capability to generate informationally-accurate and conversationally capable dialogue. In our experiments, articles used as domain knowledge were fed to the language model as simple text (the only input the language model takes), making the structure of each article irrelevant (for example, a Mayo-Clinic article on anxiety listing "Causes" as heading followed by list of causes of anxiety). Thus, future work centers around taking structured domain knowledge bases and integrating them with deep-learning models.

References

[1] D. W. Otter, J. R. Medina, J. K. Kalita, A survey of the usages of deep learning for natural language processing, IEEE Transactions on Neural Networks and Learning Systems (2020).

[2] L. Li, K. Ota, M. Dong, Deep learning for smart industry: Efficient manufacture inspection system with fog computing, IEEE Transactions on Industrial Informatics 14 (2018) 4665–4673.

[3] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al., Language models are few-shot learners, arXiv preprint arXiv:2005.14165 (2020).

[4] X. Qiu, T. Sun, Y. Xu, Y. Shao, N. Dai, X. Huang, Pre-trained models for natural language processing: A survey, Science China Technological Sciences (2020) 1–26.

[5] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, arXiv preprint arXiv:1706.03762 (2017).

[6] K. Roy, U. Lokala, V. Khandelwal, A. Sheth, ″ is depression related to cannabis?″: A knowledge-infused model for entity and relation extraction with limited supervision, arXiv preprint arXiv:2102.01222 (2021).

[7] N. Bertagnolli, Counsel chat: Bootstrapping high-quality therapy data, 2020.

[8] P. Budzianowski, I. Vulić, Hello, it's gpt-2–how can i help you? towards the use of pretrained language models for task-oriented dialogue systems, arXiv preprint arXiv:1907.05774 (2019).

[9] D. Fensel, U. Şimşek, K. Angele, E. Huaman, E. Kärle, O. Panasiuk, I. Toma, J. Umbrich, A. Wahler, Introduction: What is a knowledge graph?, in: Knowledge Graphs, Springer, 2020, pp. 1–10.