Research Statement

Amogh Mannekote

March 10, 2022

My primary research interest lies in the area of natural-language based dialogue systems. I am particularly interested in problems pertaining to the design and development of dialogue systems that go beyond simple slot-filling. To this end, I am keen on studying problems falling into two broad areas: (1) problems surrounding the modeling of various "non-traditional" task-oriented dialogue paradigms, which go beyond simple slot-filling, and (2) human-centric issues surrounding the authoring of dialogue systems including, but not limited to the creation of dialogue datasets, data annotation, and bootstrapping of dialogue systems with small datasets.

Non-Traditional Task-Oriented Dialogue Paradigms

Since "non-traditional" is not a term-of-art, I will begin by defining it. I seek to use it as an umbrella term to refer to the paradigms of dialogue systems that attempt to go beyond the role of a passive *service provider* (e.g., booking flights, making reservations) towards more "active" roles such as those of a *collaborative partner*, a *tutor*, or an *instruction giver* [1]. Building automated dialogue systems for these paradigms of dialogue presents several unique challenges. My most recent research efforts to this end have been directed at addressing three such challenges that I have encountered in the area:

1. Lack of large open datasets. The lack of large, open dialogue datasets such as MultiWOZ for these specialized settings precludes the use of commonly used transfer-learning based approaches. To tackle this problem, I have been looking into data augmentation and few-shot learning techniques for Natural Language Understanding (NLU) and Response Generation (NLG). Current approaches treat the problem of data augmentation for NLU as a sequence generation problem - going from Machine Representation (MR) to text. However, they mostly fail to account for the dialogue context, the task context, and stylistic aspects of the utterances, all of which are key to achieving realistic examples of utterances.

2. Presence of highly domain-specific characteristics. The presence of highly domain-specific characteristics (e.g., style, vocabulary, responses rooted in deep domain expertise) makes learning ML models for these non-traditional dialogue paradigms more challenging. Unlike simple, everyday domains such as hotel reservations and taxi-ride booking, settings involving tutoring or collaboration require a significantly higher level of domain expertise to come up with an appropriate response. For example, in a tutoring context, a human tutor's next move is based on a combination of several factors such as task context, dialogue history, personal attributes, domain expertise, and pedagogical strategies.

3. Multimodal Task Streams. In contrast to the service provider role, which typically involves interfacing only with a *static* external knowledge base, collaborative and instructional dialogues usually involve a combination of static (e.g., a static document serving as a knowledge base) as well as *dynamic* task contexts. Due to the dynamically varying nature of the context, such data streams can be construed as a separate modality (e.g., code, music, UI elements). As part of the Earsketch-CAI project, I am investigating the effectiveness of models such as GraphCodeBERT [2] to incorporate the code context into improve natural language understanding task peformance.

Human-Centric Issues Surrounding the Authoring of Dialogue Systems

This line of research is motivated by my belief that the developer tools can have a real impact on the quality and realism of the datasets and models that get developed.

1. Democratization of Dialogue Systems Development. Pre-trained Large Language Models (PTLMs) such as GPT-2, BART, and T5 have ushered in an era of few-shot transferring of dialogue models to new domains with minimal training data. They also expose simpler, more decoupled interfaces for the dialogue system author in the form of schemas [4] and natural language prompts, which are more accessible to machine learning non-experts compared to approaches such as fine-tuning. In tandem with commercially available and open-source tools that enable dialogue system authoring, we are heading towards a situation akin to web development, where domain experts can bootstrap their own dialogue systems with minimal training data or ML/NLP expertise.

However, achieving this would require data collection efforts to ensure that the benchmark datasets are reflective of the descriptions generated by a non-expert crowd. One archetypical example of such an effort is the SGD-X dataset [3], which collected paraphrasings of natural language schema descriptions in the original SGD dataset [4] from crowdworkers to better reflect the distribution of descriptions written by laymen. My study *Schemas in the Wild* was geared towards this goal.

2. Automatically Collecting Labels from Human Activity. As part of a course project, and along the lines of Sood et al. [5], I am working on collecting eye-gaze data of humans reading transcripts of task-oriented dialogues from a publicly available task-oriented dialogue such as MultiWOZ in an attempt to build a model of how much attention do human readers pay to various sections/turns of a dialogue while understanding it. Such a model could then be potentially used to augment a neural attention model by treating the human gaze information as an added supervision signal.

More broadly, I believe the question of "How can we triangulate signals from external data sources such as UI events, eye-gaze, and task contexts to make the most of our dialogue datasets?" is interesting from both a HCI as well as an ML perspective.

Past Research Projects

1. Schemas in the Wild: Investigating Human Factors in the Development of Task-Oriented Dialogue Schemas

Schema-guided dialogue aims to transfer models of task-oriented dialogue to new target domains using only a structured representation of the target domain. This allows for the sharing of knowledge among different services as well as more efficient handling of unseen services. However, despite the recent rise in research interest in this paradigm, we know very little about the human factors surrounding the development of these dialogue schemas.

To address this research gap, we conducted a study with 22 participants, where we presented our participants with two real-world dialogue tasks and asked them to 1) write natural-language instructions for the tasks to human agents, and 2) develop dialogue schemas for the same tasks for automated dialogue systems. Through the findings from our study, we sought to guide the design of schema-authoring tools for dialogue system developers, inform future efforts to create benchmark datasets, and inspire further human studies to better understand the gap between human-human and human-machine communication in the realm of task-oriented dialogue.

2. Unsupervised Clustering of Discussion Forum Posts

In an attempt to perform unsupervised dialogue act induction of discussion forum posts taken from CS1/2 courses, I experimented with learning techniques that made use of supervisory signals both intrinsic as well as extrinsic to the dataset. These approaches included semi-supervised clustering, transfer-learning based approaches based on closely matching StackOverflow posts, and topic modeling using BERT (combined with HDBSCAN). Although we finally ended up choosing a supervised learning approach to robustly categorize the posts in the interest of the project's objectives, analyzing the failure cases of the various representation methods and unsupervised modeling techniques offered me a deeper, first-hand insight into many of the challenges involved in applying vanilla NLP methods designed for long-running text to shorter utterances.

3. Using Granular Annotations to Increase Out of Distribution Generalizability of Gradient-Descent based Machine Learning Models

With a small number of fine-grained human annotations, we showed that we can greatly increase the trustworthiness (and accuracy) of a neural network - for out-of-domain samples by penalizing high magnitudes of input gradients at regions of the input space that are known to be irrelevant (as part of the training cost function). The system we developed allowed end users and model developers to annotate a small number of examples with information about the (ir)relevant regions of the input space for a classification task using a web-based annotation tool. We demonstrated this approach over both image and text data.

4. Developing an Evaluation Framework for a Commercial Text2SQL System

This work was done as part of my role as a Data Scientist at Cuddle AI (now Crux Intelligence[1]). The platform's Natural Language Understanding module was responsible for parsing complex, business intelligence queries such as *Dollar Sales for East by state last months* into a structured representation that would then be converted to a SQL query. The hybrid ML/rule-based nature of the system made it effort-intensive to verify that improvements to one set of queries did not negatively impact the performance of a different type of query.

My work involved defining metrics at various levels of strictness and granularities to come up with a single measure of progress that enabled multiple teams working on the different pieces of the pipeline to have a shared, linear view of progress.

Open-Source Projects

1. *nonechucks* - A PyTorch library for dynamically handling data pipeline failure

nonechucks[2] is an open source library for PyTorch that allows developers to transparently handle invalid or unwanted samples in their machine learning data pipelines. It also extends the existing construct of PyTorch's Data Transforms to function as Data Filters. *nonechucks* was also featured in GitHub's worldwide list of trending Python repositories in the week following its release.

2. *Rasa-Frames*: A Frame-Based Dialogue Framework

Most slot-filling dialogue frameworks do not allow tracking multiple mentions of the same slot during a conversation, while it is quite natural to want to do so

[1] https://cruxintelligence.com/

[2] https://github.com/msamogh/nonechucks

(e.g., comparing two different hotel options). *Rasa Frames*[3] is an open-source dialogue systems framework inspired by the Microsoft FRAMES dataset that allows multiple dialogue "frames" to be handled in parallel in a single dialogue.

References

[1] Traum. David. "Computational Approaches to Dialogue". en. In: *The Routledge Handbook of Language and Dialogue*. Ed. by Weigand Edda. 1st ed. New York, NY : Routledge, [2017] — Series: Routledge Handbooks in Linguistics: Routledge, June 2017, pp. 143–161. ISBN: 978-1-315-75058-3. DOI: 10.4324/9781315750583-10. URL: https://www.taylorfrancis.com/books/9781317612599/chapters/10.4324/9781315750583-10 (visited on 10/13/2021).

[2] Daya Guo et al. "GraphCodeBERT: Pre-training Code Representations with Data Flow". In: *arXiv:2009.08366 [cs]* (Sept. 2021). arXiv: 2009.08366. URL: http://arxiv.org/abs/2009.08366 (visited on 03/10/2022).

[3] Harrison Lee et al. "SGD-X: A Benchmark for Robust Generalization in Schema-Guided Dialogue Systems". In: *arXiv:2110.06800 [cs]* (Oct. 2021). arXiv: 2110.06800. URL: http://arxiv.org/abs/2110.06800 (visited on 01/08/2022).

[4] Abhinav Rastogi et al. "Towards Scalable Multi-domain Conversational Agents: The Schema-Guided Dialogue Dataset". In: (2019). URL: http://arxiv.org/abs/1909.05855.

[5] Ekta Sood et al. "Improving Natural Language Processing Tasks with Human Gaze-Guided Neural Attention". In: *arXiv:2010.07891 [cs]* (Oct. 2020). arXiv: 2010.07891. URL: http://arxiv.org/abs/2010.07891 (visited on 03/10/2022).

[3] https://github.com/msamogh/rasa-frames