Research Statement

He He

1 Overview

Modern intelligent systems work with real-time, dynamic data and centers around human users. Increasingly, we need service that goes beyond a passive model where the user feeds in input and requests an output. For example, a self-driving car must actively infer any possible change in the environment given the current observation, coordinate with other autonomous cars, and communicate with the passengers about their needs. A smart home system needs to decide when and where to collect the sensor data and learn about user habits and preferences over time. Such applications call for intelligent agents that are able to interact with the world (including humans), collect needed information, and make decisions with partial information.

My research is motivated by the insight that the input / data acquisition is an integrative part of a learning model and that a model is always situated in a dynamic world. I am interested in developing efficient models that not only makes predictions, but can also decide when and what information is needed for a particular prediction, learning through interaction with the environment, and communicate and collaborate with humans. To develop effective interactive learning paradigms, one must combine innovations from machine learning, natural language processing, human computer interaction, and social sciences. Therefore, my work spans from principled learning algorithms [1, 2] to live interactive platforms [3, 4].

2 Prior Work

2.1 Dynamic Information Acquisition

Information is crucial to making predictions and they usually come at some expense, ranging from computational cost such as hashing a string to monetary cost such as conducting a medical exam. Gathering all information is often not possible, nor necessary. For example, when doctors diagnose diseases, they start with the basic data (e.g. age, blood pressure, body temperature); only when these observations are not sufficient to make a decision do they move on to more costly tests such as MRI or X-ray. I formalized this problem as a sequential decision-making process: for each instance, the model uses a *policy* to sequentially select from a set of features based on past predictions until it is confident enough to finalize the prediction. The goal is to achieve a good trade-off between feature cost and prediction accuracy. Thus it learns which features are most cost-effective on an instance basis. I applied this framework to both multiclass classification [1] and structured prediction (dependency parsing [5]). The latter is one of the first few works that study feature computation cost in natural language processing, and our parser instantiates less than 30% feature templates without hurting the accuracy.

At a high level, the model learns to search a space—subset of features in the above setting. The initial success in feature selection motivated me to think about algorithms that learn to search the solution space of combinatorial optimization problems, e.g. integer linear programming (ILP). Can we exploit common problem structures across instances to learn promising search paths? Can we predict the future cost of a node without expanding it? To this end, I designed an algorithm that learns to prune and prioritize nodes on a branch-and-bound tree for solving ILPs. Given the same amount of budget (time or number of nodes explored), our solver consistently finds better solution compared to a commercial solver, Gurobi.

My approach is based on imitation learning: at training time, an *oracle* with knowledge of the ground truth can reverse-engineer the best features; thus we can learn a policy to mimic the oracle's behavior. One challenge in this setting is that the oracle's policy may not be in our learning space. Therefore, I proposed an adaptive learning target that interpolates the oracle's decision and the learner's decision to provide a learning curriculum, which enjoys the no-regret guarantee and better empirical performance than directly learning from the oracle.

2.2 Incremental Language Understanding

The dynamic information acquisition setting assumes that all information is available upon request, an assumption taken by most natural language processing (NLP) tasks. Moving beyond this static mode, I started to explore online settings where the data comes in a stream. Here the model must decide when to take actions given insufficient information. For example, on the Internet of Things, the central controller keeps receiving real-time data from multiple devices (e.g. lighting, heating, security systems), and must decide if more data should be collected before an action is triggered. One interesting NLP task in this setting is simultaneous interpretation, where the translator translates *while* the speaker is talking. The challenge in this setting is that the critical information needed for a prediction may not be available until some time in far future, making it hard to trade off accuracy for speed/time. For example, when translating from verb-final languages (e.g. Japanese) to verb-medial languages (e.g. English), a naive translator often has to wait until the end for the verb to produce a sensible translation. Our findings [6] show that compared to translators, simultaneous interpreters use a unique set of skills to balance speed and accuracy, e.g. break long sentences into smaller, independent units, summarizing content, reordering words according to the target language grammar. Inspired by these observations, my work proposed the first framework [7] that unifies machine translation and interpretation techniques in a reinforcement learning setting. In addition to sentence segmentation which is used by most prior work, we have shown that predicting future words [7] and reordering words [8] improve the speed-accuracy trade-off.

2.3 Collaboration and Communication with Humans

Humans play a central role in many AI applications. An intelligent agent need to interact not only with the environment, but also with the users to understand their needs. One desired component for such an agent is a dialogue system, which enables humans to communicate with the agent in natural language. To this end, my work [4] has studied a new symmetric collaborative dialogue setting, where two agents, each with private knowledge, must communicate to achieve a common goal, e.g. finding a common item in both agents' private knowledge bases. A typical dialogue in our setting consists of information exchange, question answering, coordination and inference. I designed a neural network model that bridges two main approaches to dialogue: semantic parsing that grounds utterances to logical forms and encoder-decoder models that embed utterances to a vector space. Our model have the best of both worlds by combining structured knowledge and unstructured conversation context on an embedded graph. Human evaluation has shown that our dialogue agent can successfully complete the task by collaborating with users in a humanlike way.

Communication and collaboration are two-way processes thus an effective agent must infer about its parter, e.g. what they want, what their next steps are. To tackle this problem, I focused on an online game setting,[1] where players compete with each other to answer trivia questions as fast as possible. The goal is to design an agent that can play with humans and adapt to different playing strategies (e.g. cautious, aggressive). I proposed a deep reinforcement learning model which learns a latent distribution of player strategies as it plays with multiple users repeatedly [9]. The model then adjust its strategy depending on who it is playing against. We have shown that by explicitly modeling different user behavior, the agent is able to exploit weak players more effectively.

3 Current and Future Work

Broadly, I am interested in developing robust, efficient, and natural interactive approaches, with natural language as the key communication channel.

My ongoing work focuses on **robust dialogue systems**. One common failure mode of current data-driven dialogue models is that once the conversation deviates from the training data (i.e. the model no longer fully understands the dialogue context), the model falls to some disastrous state (e.g. repeating itself,

[1] `http://protobowl.com`

generating generic responses). Such brittleness prevents us from deploying dialogue models and learning through user interaction. Currently, I am exploring two key ideas to tackle this problem. First, we often have strong prior knowledge about the user intention given a dialogue domain (e.g. booking tickets, chatting about movies), which allows us to build a high-precision parser that maps utterances to simple logical forms. Second, the dialogue agent has an operating zone where the user utterance can be parsed and responded to properly, and it should drive the conversation into this region. With these guidelines, our simple rule-based bot augmented with data-driven generation has achieved better performance on multiple tasks compared to more complex neural models. Looking forward, the end goal of this project is to build dialogue agents that improve themselves through chatting with people, much as modern search engines learn from user click feedback. This would require significant advancement in language generation, reinforcement learning, and human-in-the-loop learning, e.g. how to refine an utterance toward a certain goal, how to do safe exploration and solicit feedback in a conversation, and when to ask for human help to respond.

Modern intelligent agents need to interact with a population of users who have different preferences and habits. Expanding my work on opponent modeling [9], I am interested in developing **personalized, adaptive systems** that are automatically adjusted to different individuals and occasions. One example is *text style transfer*. People change their vocabulary and sentence structure depending on who they are addressing and where they are speaking, e.g. talking to their kid vs talking to their boss, tweeting vs writing an email. I have just started looking at how we can extract idiosyncratic patterns from text of a certain style and build models that apply these patterns to a given piece of text. When interacting with a population of users to learn strategies targeting at each individual, another challenge is to balance exploration and exploitation across multiple subjects. We do not want to repeatedly discover common knowledge about many individuals. One related application is *online education / tutoring systems*. While each student may prefer different amount of time on each topics, most of them would follow a common syllabus. To scale an interactive system to a large number of users with diverse needs, I will explore several research questions: How can we discover a shared "basis" of user behavior? How can we transfer knowledge among individuals and quickly adapt to new users or new scenarios? Is there an optimal strategy for selecting a human subject when exploring different actions?

The success of an interactive system hinges on the quality of feedback (reward signal) it receives. In many real world applications, the feedback is delayed and insufficient to point the model to the right direction. Considering a dialogue agent troubleshooting a network failure with a user, only at the end of the conversation does the agent receives a 1–5 rating from the user. The sparse reward, large space of natural language, plus non-stationary human users pose significant challenge to current learning methods—the model will need to interact with the world for a long time before it learns anything useful. To address these issues, I am interested in **eliciting and learning from rich supervision signals in language**. When humans communicates in natural language, oftentimes the feedback is implicit in their responses. For example, short, generic responses can indicate losing interest, such as "okay", "fair enough", and "I guess so". In more goal-oriented settings, they may also directly express negative feelings, e.g. "I don't want that.", "No, I was asking...". To take advantage of these signals, we need an empathetic component in the system that "senses" where things go wrong, thus alleviate the credit assignment problem. Another direction is to proactively ask users or a human teacher for stronger supervision. This can involve demonstration of a task, instruction in natural language, or paraphrasing / repeating a request. In this setting, we must also balance learning and the cost of richer supervision, which connects to my previous work on dynamic information acquisition. In addition to human feedback, we can also extract prior knowledge from documents accompanying a task, such as manuals, guides, and instructions.

References

[1] He He, Hal Daum'e III, and Jason Eisner. Imitation learning by coaching. In *Neural Information Processing Systems (NIPS)*, 2012.

[2] He He, Hal Daum'e III, and Jason Eisner. Learning to search in branch and bound algorithms. In *Neural Information Processing Systems (NIPS)*, 2014.

[3] Jordan Boyd-Graber, Brianna Satinoff, He He, and Hal Daumé III. Besting the quiz master: Crowdsourcing incremental classification games. In *Empirical Methods in Natural Language Processing*, 2012.

[4] He He, Anusha Balakrishnan, Mihail Eric, and Percy Liang. Learning symmetric collaborative dialogue agents with dynamic knowledge graph embeddings. In *Association for Computational Linguistics (ACL)*, 2017.

[5] He He, Hal Daum'e III, and Jason Eisner. Dynamic feature selection for dependency parsing. In *Empirical Methods in Natural Language Processing (EMNLP)*, 2013.

[6] He He, Jordan Boyd-Graber, and Hal Daum'e III. Interpretese vs. translationese: The uniqueness of human strategies in simultaneous interpretation. In *North American Chapter of the Association for Computational Linguistics (NAACL)*, 2016.

[7] Alvin C. Grissom II, He He, Jordan Boyd-Graber, John Morgan, and Hal Daumé III. Don't until the final verb wait: Reinforcement learning for simultaneous machine translation. In *Proceedings of Empirical Methods in Natural Language Processing (EMNLP)*, 2014.

[8] He He, Alvin Grissom II, Jordan Boyd-Graber, John Morgan, and Hal Daumé III. Syntax-based rewriting for simultaneous machine translation. In *Proceedings of Empirical Methods in Natural Language Processing (EMNLP)*, 2015.

[9] He He, Jordan Boyd-Graber, and Hal Daum'e III. Opponent modeling in deep reinforcement learning. In *Proceedings of the International Conference of Machine Learning*, 2016.