Josiah Hanna

Research Statement

Computers can solve many tasks provided a programmer can specify step-by-step instructions for mapping the inputs of a well-defined function into their corresponding outputs. Even when step-by-step instructions cannot easily be specified, computers can use *supervised* machine learning algorithms to solve tasks given only the right output for example inputs. However, there still remain many tasks for which algorithmic solutions do not exist and, though success is well defined, the right input-output mapping to achieve it are unknown.

I envision a world in which intelligent software agents – computers, robots, IoT devices – have the capability to interact with complex tasks and discover the steps to solve the task for themselves. In this world, software agents will learn the right actions (outputs) for any state of their world (inputs) by testing actions and observing how these actions lead to success or failure. My research aims to develop the algorithms that make this future possible and to integrate these algorithms into real world systems. **A key capability of such learning agents is the ability to predict how actions influence the state of the world and how actions lead to success or failure.** My research focuses on accurate prediction as a necessary step towards software that can learn to accomplish complex tasks. Effective prediction will lead to faster learning, better decision-making, and increased applicability to real world problems.

Towards this goal, my PhD thesis work has identified several sub-problems and made contributions towards solving them. In particular, I have made contributions to how a learning agent can **model the world**, **take informative action** to learn about the effects of its behavior, perform **counterfactual reasoning**, and **identify uncertainty in its predictions**. In addition to my PhD work, I have identified additional research problems crucial for achieving my long term goal: **learning appropriate abstractions**, **providing guarantees on prediction quality**, and **integrating effective prediction into real world task learning**. These research areas will provide a rich set of problems throughout my career.

My thesis contributions have primarily been made in the field of *reinforcement learning* (RL). The goals of the RL community are well-aligned with my own research goal and I will continue to contribute to this community throughout my career. Achieving my long-term goal will also involve collaborating with and contributing to other areas of artificial intelligence, computer science, and application areas. During my PhD, I have performed additional research in multi-agent systems, robotics, and transportation. This research has given me experience collaborating outside my primary research area and applying algorithms to complex real world problems.

Thesis Research

World Modeling Agents that can accurately model the world – predict how their actions change the state of the world – can make accurate predictions about the effectiveness of different behaviors. However, learning models of the world from scratch is still a challenging proposition. The first contribution of my thesis showed how a learning agent could leverage physics simulators to improve its predictions about the outcomes of its actions. Simulators provide a rich source of prior knowledge about action outcomes, however, by themselves simulators tend to inaccurately model the effects of actions. I introduced an algorithm that takes small amounts of data from the real world and combines it with simulation to improve an autonomous agent's ability to predict the effects of its actions. The agent then uses these improved

predictions to learn a better behavior for solving a task in the real world. **This algorithm led to an over 40% increase in the walking speed of a Softbank Nao robot compared to a state-of-the-art, hand-designed walking controller.** While this contribution has immediate applicability to the growing robotics industry, the contributed algorithm is applicable to any learning problem where a simulator is available a priori. I am in the process of applying it to chemical process control problems in collaboration with researchers at Exxon Mobil. My future goal is to improve the ability of agents to develop models of the world with less prior knowledge especially for tasks where good simulators are unavailable.

Taking Informative Actions An agent that wants to learn the effects of different behaviors can be more efficient with an appropriate choice of exploratory behavior. In particular, in tasks where rare outcomes can strongly influence performance it is important that an agent experience such outcomes to understand how the outcomes influence the agent's expected performance on the task. For example, when a robot is learning to walk it needs to experience falling in order to learn how to react when it is close to falling. Prior to my thesis, a common assumption in the reinforcement learning community was that the optimal evaluation behavior was the behavior to be evaluated. In my thesis work, I introduced an algorithm that an autonomous agent can use to learn an exploratory behavior that leads to more effective evaluation of a given behavior [7]. **This algorithm led to 85% more accurate prediction on reinforcement learning benchmarks.** In more recent work, I showed that the same algorithm could be used by a learning agent to find informative behaviors to help it learn to solve a task faster [4].

Counterfactual Reasoning Learning to predict outcomes in sequential decision-making tasks is typically done by taking actions and then learning the outcome. Reasoning about the outcomes of observed actions is known as *factual* reasoning. An alternative to factual reasoning is *counterfactual* reasoning. Counterfactual prediction refers to "what would have happened if I had taken action B instead of action A." The application of counterfactual reasoning to sequential decision-making tasks is a substantially more difficult task than in single-decision problems because the chosen action affects the decisions available to the agent in the future

Many widely-used algorithms for counterfactual reasoning in sequential decision-making tasks require the decision rule used in selecting observed actions be known. This requirement is typically violated in many realistic settings; this observation was underscored during a research internship at Google where I could see that two algorithms I had previously introduced were inapplicable due to this violated assumption. Motivated by this current limitation, I have introduced an algorithm for estimating the decision rule that (paradoxically) **provides more accurate counterfactual predictions than if the decision rule were known** [2]. This new algorithm will make state-of-the-art counterfactual prediction methods applicable to settings where they were previous inapplicable and even improves the accuracy of these methods. In addition to more accurate counterfactual prediction, I introduced a follow-up algorithm that uses similar ideas to increase the speed at which an agent learns a task [5].

Identifying Uncertainty in Predictions In real world applications it is not only useful for an agent to be able to predict the effects of its actions but also to understand it's uncertainty about these outcomes. My final thesis contribution is a method that allows an agent to provide confidence bounds on its predictions of measures of task success on a given task. Existing approaches for this problem had been proposed in the context of web-marketing systems where the agent may have millions of interactions with the task that it can use to estimate a confidence bound [12]. This amount of task experience is exceedingly high for applications such as robotics or health care. My thesis showed how models of the environment's dynamics could be used to tighten lower bounds on the performance of an untested behavior – **in some cases reducing the requisite data-set size from thousands to hundreds of interactions with the environment** [6]. This contribution closed the gap between what had been previously possible and the level of data efficiency required for more data-scarce applications. I also put forward theoretical bounds

to better understand the problem settings where our introduced confidence interval method would be less reliable. This contribution also makes an important step towards deploying autonomous agents on real world problems; in applications where safety is critical it is often unacceptable to deploy an untested policy without guarantees on its performance.

Additional Application-Related Research

While a large part of my research has been theoretical in nature, I also am committed to understanding how to tailor algorithms to real world applications. This commitment has motivated research in other application areas during my PhD thesis.

Robot Soccer In robotics, I have worked on the UT Austin Villa robot soccer team which programs a team of humanoid robots to compete in soccer matches at the annual RoboCup competition. In this domain, I integrated a novel robot perception pipeline with kicking and "approach ball" skills [9]. This work contributed to a **second place finish** at the 2016 RoboCup Standard Platform League finals. I also developed a learning pipeline for a simulated robot to learn to kick the ball precisely to different locations. This work allowed the development of passing behaviors and contributed to **a 2015 championship** in the RoboCup 3D Simulated Soccer League [8].

Intelligent Transportation Systems I have also worked on multi-agent systems and learning problems related to transportation research. The main focus of this research has been the problem of micro-tolling – placing a toll on every link of a road network in order to encourage drivers to take routes that lower congestion in the network. This work led to a novel algorithm for micro-tolling that was shown in simulations to **increase social welfare in transportation networks by up to 33%** [11, 10]. Follow-up research showed that the algorithm was robust to only a subset of agents responding to tolls and identified which agents the system would benefit from the most from being compliant with tolls [3]. Adaptive tolling is representative of the type of real world tasks that my research will allow autonomous agents to learn to solve. Towards this application, my collaborators and I have introduced an algorithm allowing a learning agent to learn to set tolls to reduce congestion in a traffic network [1].

Future Work

In my thesis research I have advanced the capability of autonomous agents, but there is still much work to be done. In addition to the near-term follow-up on my thesis research, there are also larger problems that will define my research in the next decade. For example, real world tasks involve predicting the effects of hundreds or even thousands of consecutive decisions, the state of the world is often only partially observable, and observations of success or failure may be sparse. Furthermore, different real world tasks will require different forms of guarantees on the quality of agent predictions. Finally, solutions to real world tasks may require complex behaviors; the problem of learning such behaviors is a hard search problem even with the ability to accurately predict the performance of any behavior. In the face of these problems, state-of-the-art prediction methods currently fall flat.

I believe the best way to address these challenges is to attempt to apply existing methods to an out-of-reach problem and use lessons learned on that problem to suggest problems for more theoretical study. I will focus my future research on effective prediction in robot soccer. Robot soccer is a complex task in which a robot must make thousands of decisions based on limited information and only observes success or failure in the relatively rare event that a goal is scored. As a concrete goal for the next decade of my research, I will develop prediction algorithms that allow a soccer playing robot to assess how changes to its behavior will affect its chances of winning or losing a game. Success at this application will have immediate consequences for the growing autonomous vehicle and drone industries. At the same time, I will

remain open to applications outside of robotics – such as health care or education – that can help develop knowledge as to when certain methods succeed or fail.

Learning Abstractions Towards scaling prediction algorithms to robot soccer and other similarly challenging domains, a necessary first step is approaching prediction at the right level of abstraction. Many complex tasks become much simpler when viewed with the right abstractions or when decomposed into subtask hierarchies. For example, a soccer-playing robot attempting to predict the effect of it's individual motor commands on a game is facing an intractable task. However, attempting to predict the effect of kicking the ball towards the opponent's goal is a much more tractable task. While humans can easily identify how a task can be broken into subtasks or what aspects of the world are relevant to a given tasks, it is difficult for a computer to develop the same hierarchies and abstractions. In the concrete application of robot soccer, **I will develop algorithms that allow a robot to automatically identify abstract actions, decompose the task into subtasks, and identify important features of a task** (e.g., does it matter that the opponent is wearing red). These algorithms will make the task of predicting behavior in the robot soccer domain tractable for current state-of-the-art prediction methods.

Guarantees on Prediction Quality For practitioners to trust algorithms – particularly on high-risk tasks – it will be important to establish guarantees about algorithm performance. For example, an agent making health-care treatment decisions should be able to provide guarantees on the likelihood that it is over-estimating the benefit of a treatment. In my future research I aim to provide practitioners the answers to the questions, "what are the odds this algorithm is over-estimating the effectiveness of a behavior?" "what is the worst-case outcome when deploying this algorithm?" "Is this algorithm guaranteed to treat people it interacts with fairly?" An important part of developing better prediction algorithms is automating the answers to questions like these. My vision is that **practitioners deploying algorithms from my research will be able to understand any potential risks.** This line of research will connect me to the emerging field of safe beneficial AI; this community is becoming increasingly important as government and industry seek ways to ensure that AI systems benefit society.

From Prediction to Better Real World Decision-making The ability to predict behavior outcomes is essential for software agents learning to solve tasks in the real world. However, there are still research problems that must be solved for an agent to automatically go from accurate prediction to improved ability to solve a task. For example, even once a robot soccer player can predict the outcome of any strategy, it still has the challenge of efficiently exploring the space of strategies to find a winning one. Finally, algorithms must be applied to a variety of real world problems and, if they fail to apply, the reasons will motivate further research. **A key feature of my research will be connecting theory to practice**; I believe real world problems should motivate bigger theoretical questions that will motivate the development of widely applicable algorithms.

Machine learning and data science are increasingly being used to solve real world problems. However, there are still many obstacles to applying these techniques to tasks that require a sequence of decisions to be made – the most commonly applied techniques fail to consider that decisions made at one point in time will influence the decisions that have to be made in the future. A key capability for solving such tasks is the ability to **predict outcomes of behaviors on task performance and the state of the world.** My research will develop this capability and allow software agents to learn to complete complex tasks without explicitly programmed instructions. The types of complex prediction tasks that drive my research are ubiquitous in our world, from health care to robotics and from education to finance. My research will scale existing methods to these problems and also identify new problems to push basic research forward. I am excited to pursue this work in a collaborative and innovative environment.

References

[1] Haipeng Chen, Bo An, Guni Sharon, Josiah P. Hanna, Peter Stone, Chunyan Miao, and Yeng Chai Soh. Dyetc: Dynamic electronic toll collection for traffic congestion alleviation. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI-18)*, February 2018.

[2] Josiah Hanna, Scott Niekum, and Peter Stone. Importance sampling policy evaluation with an estimated behavior policy. *arXiv preprint arXiv:1806.01347*, 2018.

[3] Josiah Hanna, Guni Sharon, Steve Boyles, and Peter Stone. Targeting compliant agents for opt-in microtolling. In *In Proceedings of the 33rd AAAI Conference on Artificial Intelligence*, 2019.

[4] Josiah P. Hanna and Peter Stone. Towards a data efficient off-policy policy gradient. In *AAAI Spring Symposium on Data Efficient Reinforcement Learning*, March 2018.

[5] Josiah P. Hanna and Peter Stone. Correcting sampling error in the monte carlo policy gradient estimator. In *NIPS Deep Reinforcement Learning Workshop*. 2019.

[6] Josiah P. Hanna, Peter Stone, and Scott Niekum. Bootstrapping with models: Confidence intervals for off-policy evaluation. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, pages 538–546. International Foundation for Autonomous Agents and Multiagent Systems, 2017.

[7] Josiah P. Hanna, Philip S. Thomas, Peter Stone, and Scott Niekum. Data-efficient policy evaluation through behavior policy search. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, 2017.

[8] Patrick MacAlpine, Josiah Hanna, Jason Liang, and Peter Stone. UT Austin Villa: RoboCup 2015 3D simulation league competition and technical challenges champions. In Luis Almeida, Jianmin Ji, Gerald Steinbauer, and Sean Luke, editors, *RoboCup-2015: Robot Soccer World Cup XIX*, Lecture Notes in Artificial Intelligence. Springer Verlag, Berlin, 2016.

[9] Jacob Menashe, Josh Kelle, Katie Genter, Josiah Hanna, Elad Liebman, Sanmit Narvekar, Ruohan Zhang, and Peter Stone. Fast and precise black and white ball detection for robocup soccer. In *RoboCup-2017: Robot Soccer World Cup XXI*. July 2017.

[10] Guni Sharon, Josiah P. Hanna, Tarun Rambha, Michael W. Levin, Michael Albert, Stephen D. Boyles, and Peter Stone. Real-time adaptive tolling scheme for optimized social welfare in traffic networks. In *Proceedings of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2017)*, May 2017.

[11] Guni Sharon, Michael W. Levin, Josiah P. Hanna, Tarun Rambha, Stephen D. Boyles, and Peter Stone. Network-wide adaptive tolling for connected and automated vehicles. *Transportation Research Part C*, 84:142–157, September 2017.

[12] Philip S. Thomas, Georgios Theocharous, and Mohammad Ghavamzadeh. High confidence off-policy evaluation. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence (AAAI)*, 2015.