My Desktop
Prepare & Submit Proposals
Prepare Proposals in FastLane
New! Prepare Proposals (Limited proposal types)
Proposal Status
Awards & Reporting
Notifications & Requests
Project Reports
Submit Images/Videos
Award Functions
Manage Financials
Program Income Reporting
Grantee Cash Management Section Contacts
Administration
Lookup NSF ID

Preview of Award 1637937 - Annual Project Report

Cover |
Accomplishments |
Products |
Participants/Organizations |
Impacts |
Changes/Problems
| Special Requirements

Cover

Federal Agency and Organization Element to Which Report is Submitted:	4900
Federal Grant or Other Identifying Number Assigned by Agency:	1637937
Project Title:	NRI: Collaborative Research: A Framework for Hierarchical, Probabilistic Planning and Learning
PD/PI Name:	Marie E desJardins, Principal Investigator
Recipient Organization:	University of Maryland Baltimore County
Project/Grant Period:	09/01/2016 - 08/31/2019
Reporting Period:	09/01/2017 - 08/31/2018
Submitting Official (if other than PD\PI):	Marie E desJardins Principal Investigator
Submission Date:	08/29/2018
Signature of Submitting Official (signature shall be submitted in accordance with agency specific instructions)	Marie E desJardins

Accomplishments

*** What are the major goals of the project?**

The major goal of this project is to create a framework that will allow autonomous agents to create plans and execute them in rich, probabilistic, and partially observable environments, using hierarchical abstractions of the task environment. Agents will learn from low-level sensor perceptions to identify useful, repeatable patterns of behaviors, called subtasks. By learning from either experience or demonstrations, agents will compose various state-space abstractions into hierarchies, which are capable of solving complex, multi-stage tasks that would be intractable for a planner operating over primitive actions in the grounded state space. Subtasks are encoded in transferable structures, which can be reused to find solutions to novel, related tasks. Additionally, agents will interact with humans by interpreting natural language commands. Commands will serve as directions, suggested actions or abstract subtasks, both guiding learning and enabling the revision of plans based on feedback. Agents can be instructed on how to create plans and can be interrupted during their execution, affording a high

degree of human interaction. Thus, a simulated agent or robot using this framework will be able to cooperate alongside humans in unstructured, stochastic environments, while learning new behaviors and executing plans at varying levels of abstraction.

The framework under development depends on a new formalism, the abstract Markov decision process (AMDP), which encapsulates the state abstraction and termination conditions of a subtask into a self-contained planning problem. An AMDP is derived from an object-oriented Markov decision process, an extension of traditional Markov decision processes that factors the representation into objects and their attributes. Using this representation facilitates abstraction, since both state aggregation and non-uniform transformations can be defined readily: state aggregation collapses states by ignoring classes of objects and ranges of attributes, and non-uniform transformations map states into new composite objects at higher abstraction levels. AMDPs can also represent and evaluate propositional functions to assess complex relations or other state properties. By defining an AMDP with subtasks as actions, an AMDP hierarchy includes subtask AMDP as nodes in the hierarchy, actions of the subtasks as their children, and primitive actions as the leaves of the hierarchy. Taken together, these properties make AMDPs a well suited representation of subtasks for hierarchical planning problems.

AMDPs are Markovian with respect to the transitions of their state-action space, and are semi-Markov decision processes with respect to the “real” or ground MDP from which they are abstracted (actions in the AMDP are typically a sequence actions in the ground MDP). In essence, we aim to develop a process for autonomous bottom-up learning of structures that facilitate top-down reasoning.

*** What was accomplished under these goals (you must provide information for at least one of the 4 categories below)?**

Major Activities:

Decision-making agents face immensely challenging planning problems when operating in large environments to solve complex tasks. A hierarchy of AMDPs provides a framework for decomposing such problems into distinct, related subtasks or subgoals. AMDP hierarchies (Gopalan et al. 2017) grant considerable speedup over related recursively and hierarchically optimal methods such as MAXQ and options. Each AMDP acts as a subgoal, and each is itself a planning problem with a local model and state space abstracted from a ground MDP. Agents are able to plan more efficiently by using a reduced state space at the appropriate level of abstraction; however, our early work required their subtask models to be specified by a human expert (Gopalan et al. 2017)

In this project year, the UMBC team has focused on (1) learning the hierarchical ordering and internal models of abstract subtasks, (2) developing representations for learning task knowledge, and (3) interpreting natural language commands. Our primary joint activity with the Brown team has been in the form of a subtask exploring (4) improving how option models are computed in MDP and AMDP frameworks. Dr. desJardins is also serving on the dissertation committee for Nakul Gopalan, whose PhD research is studying how natural language instructions can be grounded to plans in the conetxt of AMDP learning.

For (1), we developed a framework uniting our technique for learning hierarchical structure from data (H-AMDP) with our new algorithm for model-based reinforcement learning, which we call Planning over AMDPs while Learning Models (PALM). H-AMDP constructs a hierarchy of AMDPs from data (solution trajectories, such as expert demonstrations). H-AMDP creates and links subtasks in a hierarchy by converting all nodes to AMDPs from a learned task hierarchy, the output of any existing task hierarchy learning algorithm, such as HierGen (Mehta, 2011). Then, an agent uses the AMDP hierarchy together with PALM to learn the models (transition probabilities and rewards) of all abstract tasks simultaneously, while deployed and exploring a (simulated) environment. PALM alleviates a great knowledge burden that was previously placed on human designers (namely, that the complete dynamics of all abstracts tasks had to be engineered and specified a priori by an expert). Thus, both the hierarchical ordering of AMDP subtasks and their internal workings can be learned entirely from data. Our results demonstrate that H-AMDP produces hierarchies that surpass hand-crafted ones in both efficiency and compactness, and PALM outperforms the existing method for model-based hierarchical reinforcement learning, R-MAXQ (Jong and Stone, 2008),

retaining the benefits of planning with AMDP hierarchies while removing the onerous knowledge requirements.

For (2), we investigated approximate state abstractions, which treat nearly-identical situations as equivalent, and derived theoretical guarantees of the quality of behaviors derived from four types of approximate abstractions. Additionally, we empirically demonstrated that approximate abstractions lead to reduction in task complexity and bounded loss of optimality of behavior in a variety of environments.

For (3), by grounding commands to all the tasks or subtasks available in a hierarchical planning framework, we arrived at a model capable of interpreting language at multiple levels of specificity ranging from coarse to more granular. We showed that the accuracy of the grounding procedure is improved when simultaneously inferring the degree of abstraction in language used to communicate the task. Leveraging hierarchy also improved efficiency: our approach enabled a robot to respond to a command within one second on 90% of our tasks, while baselines take over twenty seconds on half the tasks. Finally, we demonstrated that a real, physical robot can ground commands at multiple levels of abstraction allowing it to efficiently plan different subtasks within the same planning hierarchy.

For (4), we are investigating whether we can improve how option models are computed, in terms of both (a) learning options and their models and (b) using options to plan (as part of a hierarchy or on their own). The main insight we are exploiting to improve over current option models is that the option model should not depend on the exact number of lower-level actions taken in an execution of the option. Instead, we offer a variant of options that retains a *rough estimate* of the number of lower-level actions taken on a per-state basis. This value is most critical in figuring out how much to discount future plans. We have demonstrated:

1. A sample bound for learning options using this new model. (How many samples (s, o, s') are needed to determine *roughly* how many lower-level actions will be taken when o is executed in s?)
2. A bound on the value function when using the new, learned, option model, compared to using the usual option models.
3. Results from experiments in a variety of Taxi instances that showcase the potential of the method, enabling faster learning, with lower variance, under the new option model.

Specific Objectives:

Significant Results:

Key outcomes or Other achievements:

*** What opportunities for training and professional development has the project provided?**
Nothing to report.

*** How have the results been disseminated to communities of interest?**

We have published and presented papers at the 2016 ICML Workshop on Abstraction in Reinforcement Learning, the 2017 ICAPS Workshop on Integrated Execution (IntEx), and the 2017 International Conference on Automated Planning and Scheduling. We also presented an extended abstract as a poster presentation at the 2017 Conference on Reinforcement Learning and Decision Making. Dr. desJardins presented the research in guest seminars at SRI International (February 2017) and the University of Maryland, College Park (March 2017). Our current work on H-AMDP and PALM will be submitted to AAAI in September 2018. We plan to submit a joint paper with the Brown team on the option learning work during the upcoming project year, and have also been working on a joint journal paper descrxibing the overall AMDP framework.

*** What do you plan to do during the next reporting period to accomplish the goals?**

We are ramping up our use of task learning, both using deep learning methods and also using other more theoretically grounded statistical approaches. To understand the connection between our models and those in the literature, we are preparing a survey article on state and action abstraction.

We are also further developing our work on H-AMDP to incorporate a novel hierarchy structure learning algorithm specific to AMDPs. Currently, H-AMDP can be applied to any general algorithm takes solution trajectories as input and yields a MAXQ-style task hierarchy (which H-AMDP converts into an AMDP hierarchy). We propose a variant that goes directly to an AMDP hierarchy from sampled source data. Such an approach will loosen the assumptions of input data, permitting learning from partial and failure trajectories (for instance, demonstrations of what not to do, to learn from negative reinforcement). A secondary component will be the ability to score and rank different hierarchies proposed for the same domain, combine and collapse related hierarchies into a more general structure that facilitates transfer of experience learned in one domain to another. Similarly, to evaluate the relative benefit of two candidate AMDP hierarchies, we are researching methods to score them by quality and effectiveness. Ideally, an AMDP hierarchy would provide an acceptable tradeoff between compressing the information in each AMDP, simplifying the planning problem at each subtask while minimizing the error produced in the ground domain. Developing a scoring function that expresses this tradeoff will enable the agent to treat hierarchy learning hierarchies as a heuristic search problem. Error could be measured by calculating the policy generated by the hierarchy to the optimal ground MDP, but in practice, we would be required to rely on heuristics since this computation is intractable. Therefore, we are exploring various heuristic strategies to measure error for each node in the hierarchy. The ideal (but again intractable) method for measuring compression would be to calculate the performance of the hierarchy in a large sample of domains drawn from a target distribution, so again, we are working to develop heuristics for estimating compression. Additional directions include minimizing entropy of the hierarchy, measuring the amount of redundant work across subtasks, maximizing the portability of each subtask, and balancing tree height.

For PALM, we intend to extend model learning to be more general by incorporating parameterized reward models, where subtasks may share the same transition probabilities, but differ in reward based on some parameter such as a goal. Similarly, we plan to include cross-policy learning, such that when a subtask is executed, the resulting sampled transition updates not only the current task but is also propagated to non-descendant tasks in the hierarchy sharing the same executed subtask. In other words, one update can simultaneously be used to learn the models for other subtasks off-policy, whenever a sampled transition is valid to update its (abstract) model. One related extension we are continuing to investigate is object-parameterized option (OPO) generators, abstract structures that can dynamically identify subtasks based on the objects present in a domain. An OPO generator maintains a type signature that specifies the object classes and counts that must be satisfied for it to create a grounded option. The objects serve as task descriptors characterizing the particular instance of a subgoal that the OPO solves. Once grounded, OPOs are options in the traditional sense and can be used in planning. OPOs are promising because all of their components can be either expert-specified or learned from data in the form of solution trajectories for a given subtask. OPOs can be extended to AMDPs by defining a state abstraction function that masks object attributes not required by the parameterized policy and initiation/goal/failure functions.

Our next major goal is to apply AMDPs in partially observable settings, for both a complex video game environment, in a manipulator robot that can adhere to natural language commands.

Products

Books

Book Chapters

Inventions

Journals or Juried Conference Papers

View all journal publications currently available in the NSF Public Access Repository for this award.

The results in the NSF Public Access Repository will include a comprehensive listing of all journal publications recorded to date that are associated with this award.

Gopalan, Nakul and desJardins, Marie and Littman, Michael L. and MacGlashan, J. and Squire, S. and Tellex, Stefanie and Winder, John and Wong, Lawson L.. (2017). Planning with Abstract Markov Decision Processes. *27th International*

Conference on Automated Planning and Scheduling. . Status = Deposited in NSF-PAR Federal Government's License = Acknowledged. (Completed by Desjardins, null on 08/28/2017) Full text Citation details

Licenses

Other Conference Presentations / Papers

Shawn Squire*, John Winder, Matthew Landen, Stephanie Milani, and Marie desJardins (2017). *R-AMDP: Model-Based Learning for Abstract Markov Decision Process Hierarchies*. Third Multidisciplinary Conference on Reinforcement Learning and Decision Making (RLDM). . Status = PUBLISHED; Acknowledgement of Federal Support = Yes

John Winder, Shawn Squire, Matthew Landen, Stephanie Milani, and Marie desJardins (2017). *Towards Planning With Hierarchies of Learned Markov Decision Processes*. ICAPS workshop on Integrated Execution (IntEx). . Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Other Products

Other Publications

Nakul Gopalan, Marie desJardins, Michael L. Littman, James MacGlashan, Shawn Squire, Stefanie Tellex, John Winder, and Lawson L. S. Wong (2016). *Planning with Abstract Markov Decision Processes*. ICML Workshop on Abstraction in Reinforcement Learning. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Patents

Technologies or Techniques

Thesis/Dissertations

Websites

Participants/Organizations

Research Experience for Undergraduates (REU) funding

Form of REU funding support:	REU supplement
How many REU applications were received during this reporting period?	4
How many REU applicants were selected and agreed to participate during this reporting period?	4
REU Comments:	

What individuals have worked on the project?

Name	Most Senior Project Role	Nearest Person Month Worked
desJardins, Marie	PD/PI	2
Squire, Shawn	Graduate Student (research assistant)	6
Winder, John	Graduate Student (research assistant)	8
Brown, Elwin	Undergraduate Student	2
Carver, Noah	Undergraduate Student	2
Cocca, Caroline	Undergraduate Student	2
Fox, Charles	Undergraduate Student	2

Name	Most Senior Project Role	Nearest Person Month Worked
Landen, Matthew	Undergraduate Student	2
McNamara, Keith	Undergraduate Student	2
Milani, Stephanie	Undergraduate Student	2
Parr, Shane	Undergraduate Student	3
Symonette, Danilo	Undergraduate Student	2
Tinoco, Beatriz	Undergraduate Student	1
Oh, Ere	High School Student	2

Full details of individuals who have worked on the project:

Marie E desJardins
Email: mariedj@cs.umbc.edu
Most Senior Project Role: PD/PI
Nearest Person Month Worked: 2

Contribution to the Project: UMBC project lead and mentor to all participating students

Funding Support: N/A

International Collaboration: No
International Travel: Yes, Argentina - 0 years, 0 months, 6 days

Shawn Squire
Email: ssquire1@umbc.edu
Most Senior Project Role: Graduate Student (research assistant)
Nearest Person Month Worked: 6

Contribution to the Project: Led subproject to explore replanning and exploration/exploitation tradeoffs, contributed to AMDP learning activities, and supervised students developing new AMDP testbed domains

Funding Support: N/A

International Collaboration: No
International Travel: No

John Winder
Email: jwinder1@umbc.edu
Most Senior Project Role: Graduate Student (research assistant)
Nearest Person Month Worked: 8

Contribution to the Project: Led AMDP learning activity and mentored several undergraduate students

Funding Support: N/A

International Collaboration: No
International Travel: Yes, Argentina - 0 years, 0 months, 6 days

Elwin Brown
Email: elwin1@umbc.edu
Most Senior Project Role: Undergraduate Student
Nearest Person Month Worked: 2

Contribution to the Project: Elwin has taken the lead on the natural language processing aspects of the project, and has implemented basic statistical methods for summarizing student interactions in a series of text messages.

Funding Support: N/A

International Collaboration: No
International Travel: No

Noah Carver
Email: ncarver1@umbc.edu
Most Senior Project Role: Undergraduate Student
Nearest Person Month Worked: 2

Contribution to the Project: Noah is coming up to speed on the project and working on the AMDP learning research.

Funding Support: N/A

International Collaboration: No
International Travel: No

Caroline Cocca
Email: ccocca1@umbc.edu
Most Senior Project Role: Undergraduate Student
Nearest Person Month Worked: 2

Contribution to the Project: Caroline is taking a leadership role on the teamwork monitoring project, in particular in the area of developing machine learning approaches for identifying different patterns in student teamwork behaviors.

Funding Support: N/A

International Collaboration: No
International Travel: No

Charles Fox
Email: charfox1@umbc.edu
Most Senior Project Role: Undergraduate Student
Nearest Person Month Worked: 2

Contribution to the Project: Charles is working with Caroline Cocca on machine learning models and contributing to the software development.

Funding Support: N/A

International Collaboration: No
International Travel: No

Matthew Landen
Email: mlanden@umbc.edu
Most Senior Project Role: Undergraduate Student
Nearest Person Month Worked: 2

Contribution to the Project: Key contributor to AMDP learning activity, led the development of hierarchical AMDP structure discovery. Matthew was an undergraduate at UMBC, graduated in May 2017 with a 4.0 GPA as one of the department's Outstanding Students, and is currently pursuing a Ph.D. at Georgia Tech. Matthew was a coauthor on our ICAPS 2017 and RLDM 2017 papers, and will also be a coauthor on a submission to AAAI 2019. Matthew has a disability and is wheelchair-bound.

Funding Support: N/A

International Collaboration: No
International Travel: No

Keith McNamara
Email: keithmc@umbc.edu
Most Senior Project Role: Undergraduate Student
Nearest Person Month Worked: 2

Contribution to the Project: Contributed to the AMDP learning activity and took the lead on implementing a POMDP testbed domain. Keith will be a coauthor on a paper we plan to submit to AAAI 2008. Keith is an African American undergraduate who graduated in May 2018 and will enter the Ph.D. program at the University of Florida in Fall 2017.

Funding Support: N/A

International Collaboration: No
International Travel: No

Stephanie Milani
Email: stemila1@umbc.edu
Most Senior Project Role: Undergraduate Student
Nearest Person Month Worked: 2

Contribution to the Project: Worked on AMCP learning methods and developing novel test domains

Funding Support: N/A

International Collaboration: No
International Travel: No

Shane Parr
Email: sparr@umass.edu
Most Senior Project Role: Undergraduate Student
Nearest Person Month Worked: 3

Contribution to the Project: Shane started working in the MAPLE Lab in Summer 2016 as a recent high school graduate, just before the NRI award was made. He initially worked on developing alternative domains and running experiments. In Summer 2017 and Summer 2018, he returned as an undergraduate researcher on the NRI REU. He worked on methods for replanning and exploration vs. exploitation. He is currently an undergraduate CS major at UMass.

Funding Support: N/A

International Collaboration: No
International Travel: No

Danilo Symonette
Email: danilo2@umbc.edu
Most Senior Project Role: Undergraduate Student
Nearest Person Month Worked: 2

Contribution to the Project: Danilo has been acting as the lead system architect for the project. He has developed new software engineering skills and is helping to design the overall system.

Funding Support: N/A

International Collaboration: No
International Travel: No

Beatriz Tinoco
Email: btinoco1@umbc.edu
Most Senior Project Role: Undergraduate Student
Nearest Person Month Worked: 1

Contribution to the Project: Worked on multi-passenger taxi domain and developing a multi-time model for AMDP hierarchies.

Funding Support: N /A

International Collaboration: No
International Travel: No

Ere Oh
Email: eresoh1@gmail.com
Most Senior Project Role: High School Student
Nearest Person Month Worked: 2

Contribution to the Project: Developed novel application domains and helped to run experiments and collect data.

Funding Support: N/A

International Collaboration: No
International Travel: No

What other organizations have been involved as partners?

Name	Type of Partner Organization	Location
Brown University	Academic Institution	Providence RI

Full details of organizations that have been involved as partners:

Brown University

Organization Type: Academic Institution
Organization Location: Providence RI

Partner's Contribution to the Project:
Collaborative Research

More Detail on Partner and Contribution: This work is a collaborative project with Dr. Michael Littman and Dr. Stefanie Tellex at Brown University.

What other collaborators or contacts have been involved?

For the REU on teamwork monitoring, we have recently established a connection with Kobi Gal at Ben-Gurion University of the Negev.

Impacts

What is the impact on the development of the principal discipline(s) of the project?

This work connects the planning, robotics, and reinforcement learning communities, tackling challenges from all three areas that have not previously been explored collectively. In planning and robotics research, it is often assumed that models are provided correct and complete, making it a challenge to plan in large, uncertain domains that do not already have expert-defined dynamics. In RL, the model-free approach holds that rewards or value are sufficient to guide behavior, an assumption that can break down when rewards are given sparsely or in distal states that require a precise sequence of actions to complete. Thus, this work occupies a space in research that adds learning to planning and vice versa to reach a more powerful, general form of reasoning.

Our work draws upon and bridges more traditional symbolic reasoning with a statistical machine learning approach that learns from data. This approach will lead to increasingly more powerful, general reasoning systems that can operate at varying levels of abstraction, going from input at a perceptual and linguistic level to the execution of complex plans in rich environments.

What is the impact on other disciplines?
Nothing to report.

What is the impact on the development of human resources?
Nothing to report.

What is the impact on physical resources that form infrastructure?
Nothing to report.

What is the impact on institutional resources that form infrastructure?
Nothing to report.

What is the impact on information resources that form infrastructure?
Nothing to report.

What is the impact on technology transfer?
Nothing to report.

What is the impact on society beyond science and technology?

The broad impact of our research includes methods to construct hierarchies of tasks from demonstrations, potentially providing a more general audience the capability of producing autonomous agents for their particular field. Additionally, these hierarchies frequently map to human expectations of subtasks, allowing for rich interaction between the computer and human.

Changes/Problems

Changes in approach and reason for change
Nothing to report.

Actual or Anticipated problems or delays and actions or plans to resolve them
Nothing to report.

Changes that have a significant impact on expenditures
Nothing to report.

Significant changes in use or care of human subjects

Nothing to report.

Significant changes in use or care of vertebrate animals
Nothing to report.

Significant changes in use or care of biohazards
Nothing to report.

Special Requirements

Responses to any special reporting requirements specified in the award terms and conditions, as well as any award specific reporting requirements.
Nothing to report.