بسمه تعالی

دانشگاه تهران

پردیس دانشکده های فنی

دانشکده مهندسی برق و کامپیوتر

پیشنهاد و فرم حمایت از پایان‌نامه تحصیلات تکمیلی

کارشناسی ارشد ☒ **دکتری** ☐

شماره مرجع : *

*شماره مرجع، توسط معاونت پژوهشی پردیس دانشکده‌های فنّی هنگام صدور ابلاغ درج خواهدشد.

1- خلاصه اطلاعات پایان نامه

عنوان پایان نامه به زبان فارسی:
بررسی یادگیری سلسله‌مراتبی در قالب سیستم‌های یادگیری در حوزه یادگیری تقویتی

عنوان پایان نامه به زبان انگلیس

The Study of the Hierarchical Learning in the Context of the Learning Systems in Reinforcement Learning

نوع پایان نامه: **بنیادی**

پردیس/دانشکده: فنی **دانشکده/گروه:**مهندسی برق و کامپیوتر
مقطع تحصیلی : کارشناسی ارشد **رشته و گرایش تحصیلی: هوش مصنوعی و رباتیک**
تاریخ پیشنهاد:تاریخ تصویب:

2- اطلاعات اساتید راهنما و مشاورین

نوع مسئولیت	نام و نام‌خانوادگی	مرتبه علمی	محل خدمت	امضاء
استاد راهنما (مجری)	دکتر نیلی احمدآبادی	استاد	دانشگاه تهران – دانشکده فنی	
استاد راهنمای دوم(حسب نیاز)	دکتر وهابی	استادیار	مرکز پژوهش‌های ببنیادی	
استاد مشاور				
استاد مشاور دوم(برای دکتری)				

3- اطلاعات دانشجو

نام و نام‌خانوادگی:یاسمن رازقی **شماره دانشجویی:** 810194508 **رشته و گرایش تحصیلی: هوش مصنوعی و رباتیک دانشکده:** مهندسی برق و کامپیوتر**مقطع تحصیلی:**کارشناسی ارشد
پست الکترونیک:yasamanrazeghi7@gmail.com **تلفن ثابت:** 02122053854 **تلفن همراه:**09124331187

4- مشخصات موضوعی پایان نامه

تعریف مسأله، هدف و ضرورت اجرا (حداکثر سه صفحه)

یادگیری تقویتی[1]

یادگیری تقویتی مدلی است که در سالیان اخیر برای بررسی یادگیری موجودات زنده به کار گرفته شده‌است و با توجه به شواهد و نتایج بدست آمده موفقیت چشم‌گیری در این حوزه داشته‌است. این مدل توانسته است بسیاری از رفتارهای انسان و دیگر موجودات زنده را مانند ایجاد عادات رفتاری توجیه کرده و پاسخ مدلسازی قابل قبولی برای این رفتارها ارائه کند. از طرفی دیگر شواهد و اطلاعاتی که از مطالعات آسیب‌های مغزی، دستکاری‌های دارویی و ثبت‌های متفاوت مغزی به دست آمده‌اند میتوانند این چهارچوب را تائید کنند. هم‌چنین این مدل ها توانسته است سیستم‌های یادگیری مغز را که بر مبنای دوپامین[2] کار می‌کنند و یکی از پایه‌ای‌ترین نقش‌ها در فرایند یادگیری را دارد توجیه کند [3][2][1]. این موفقیت در توجیه شواهد یادگیری در انسان موجب تقویت این حوزه و جذب بسیاری به این حوزه شده‌است.

سیستم‌های یادگیری

نظرهای متفاوتی در رابطه با سیستم‌های متفاوت یادگیری در انسان ارائه شده‌است. بر طبق شواهد به نظر می‌رسد یکی از این تئوری‌ها تطابق بیشتری با داده‌های رفتاری و مغزی انسانی داشته و در حال حاضر تعداد بسیاری از دانشمندان این حوزه را به خود درگیر کرده‌است. براساس این تحقیقات به نظر می‌رسد که پستانداران از دو نوع سیستم برای یادگیری انتخاب درست استفاده می‌کنند: سیستم دارای مدل [3]و سیستم بدون مدل.[4] در سیستم اول که سیستم دارای مدل شناخته می‌شود عامل تلاش می‌کند که بر مبنای اطلاعاتی که از محیط [5]پیرامون خود به دست آورده است، محیط پیرامون را در غالب یک مدل ذهنی یاد بگیرد و با استفاده از این مدل ذهنی یک درخت تصمیم‌گیری تشکیل دهد و با محاسبه‌ی امید ریاضی مسیرهای متفاوت در این درخت بهترین مسیر را انتخاب کند. این روش برای رسیدن به هدف از نظر آماری بهینه است اما حجم محاسبات و در نتیجه تلاش ذهنی زیادی نیاز دارد. در سیستم دوم که سیتم بدون مدل است، عامل مدلی از محیط ندارد و با بازخورد از محیط انتخاب های خود را بهینه می کند. به این صورت که عامل ابتدا برای هر یک از

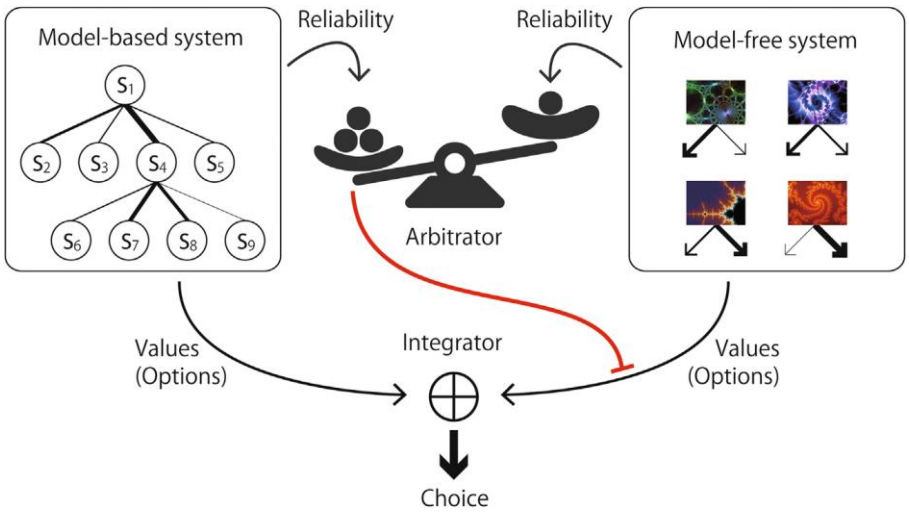
[1] Reinforcement Learning

[2] Dopamine
[3] Model Based
[4] Model Free
[5] Environment

انتخاب‌های خود بسته به شرایط یک ارزش ذهنی در نظر دارد، و بر اساس آن ارزش ذهنی اولیه انتخاب خود را انجام داده و از محیط پاداشی[6] دریافت می‌کند و از این پاداش یا بازخورد برای بهبود ارزش اعمال استفاده می‌کند. این پاداش می‌تواند با تصور پیشین عامل متفاوت باشد که ازین تناقض برای بهبود ارزش‌های ذهنی خود استفاده می‌کند تا اگر عامل در شرایط محیطی مشابه قرار گرفت تخمین‌های واقع‌بینانه‌تری از ارزش انتخاب‌های خود داشته باشد. این روش، روش بهینه ای نیست اما هزینه‌ی محاسباتی پایین‌تری دارد و بسته به محیط می‌تواند پاداش‌هایی به اندازه‌ی کافی خوب برای عامل داشته باشد. همچنین هزینه‌ی محاسباتی پایین این سیستم سرعت تصمیم گیری را نیز می‌تواند بسیار بالا برد. حضور این دو سیستم در کنار هم باعث بروز انعطاف‌پذیری[7] در یادگیری در شرایط یادگیری متفاوت و در محیط‌های گسترده‌ی متفاوت شود. شواهد مغزی بسیاری هم برای کد شدن نتایج این دو سیستم در مغز و نحوه‌ی تعامل و همکاری آن‌ها وجود دارد [4].

تعامل این دو سیستم با تعاریف متفاوتی بیان شده‌است، در یکی از این تعاریف دو سیستم هم زمان رقابت می کنند و یک سیستم میانجی با توجه به اطمینان‌پذیری[8] این دو سیستم از بین آنها انتخاب می کند. با این معنا که در لحظه هر سیستمی که با توجه به تصمیم‌های پیشین خود اطمینان‌پذیری بیش‌تری داشته باشد برای تصمیم گیری استفاده می‌شود. شماتیک این تعامل را می‌توان در شکل شماره‌ی یک دید [17].

در تعریفی دیگر این دو سیستم با هم همکاری می کنند. به این شکل که سیستم بدون مدل به‌صورت آنلاین سبب بروز جنبه‌های مختلف رفتار انسانی و سیستم دارای مدل به‌صورت آفلاین همواره در حال تصحیح رفتار سیستم بدون مدل است [18][19].



شکل شماره ی یک یکی از نظرهای تعامل دو سیستم را نشان می دهد

[6] Reward
[7] Flexibility
[8] Reliability

مطالعات در مورد نحوه‌ی کار این دو سیستم و بروز نشانه‌های مغزی و رفتاری این دو و هم‌چنین تعامل رقابتی و یا همکاری این دو سیستم در موجودات زنده هم‌چنان ادامه دارد

یادگیری سلسله مراتبی[9]

یادگیری تقویتی ابزاری قدرتمند در حوزه‌ی یادگیری ماشین و مدل‌سازی رفتار انسان است. یک عامل یادگیرنده‌ی تقویتی می‌تواند با زندگی بر پایه‌ی آزمون و خطا و تنها با استفاده از پاداش‌هایی که از محیط دریافت می‌کند اقدام به یادگیری و دست‌یابی به سیاست بهینه نماید. تعداد حالات تعداد نمونه‌های مورد نیاز عامل جهت یادگیری نحوه‌ی رفتار در کل حالات محیط به‌صورت نمایی افزایش میاید و این یعنی پیچیدگی محاسباتی تابعی نمایی از تعداد حالات است. از این مشکل تحت عنوان نفرین ابعاد[10] یاد می‌کنند در اغلب روش‌های یادگیری تقویتی با این مشکل روبه‌رو هستیم. این مشکل کارایی و کاربردی بودن روش‌های یادگیری تقویتی را در مسائل دنیای واقعی با چالش روبه‌رو کرده است از این رو راه‌کارهایی برای غلبه بر این چالش ارائه گشته که در حوزه‌ی یادگیری تقویتی سلسله مراتبی پوشش داده می‌شود.

امروزه شواهد زیادی در حوزه‌ی علوم اعصاب یافت شده است که سازوکار شبیه به یادگیری تقویتی و نیز یادگیری تقویتی سلسله‌مراتبی را در مغز نشان می‌دهد. [6][7]

یادگیری تقویتی سلسله مرابی در فرایند یادگیری به کلاسی از روش‌های یادگیری اشاره داده که در مقیاس بالاتر روش‌های یادگیری تقویتی را اعمال می‌نماید. در واقع ایده‌ی اصلی آن به ایده‌ی استفاده از زیررویه‌ها در زبان‌های برنامه‌نویسی بسیار نزدیک است. در زبان‌های برنامه‌نویسی اجرای یک رویه هم می‌تواند هم با اجرای زیررویه‌ها و هم با اجرای دستورهای بدوی انجام گیرد که که اجرای هر زیررویه نیز می‌تواند شامل اجرای دستورات بدوی و یا زیررویه‌های دیگر باشد. در یادگیری تقویتی سلسله مراتبی یک وظیفه می‌تواند به تعدادی زیروظیفه تقسیم شود. بنابراین در صورتی مع سیاست انجام زیروظیفه مشخص باشه، نیازی به یادگیری وظیفه‌ی اصلی نیست و می‌توان از دانش مربوط به زیروظیفه‌ها استفاده نمود. بدین ترتیب جهت انجام وظیفه چندین زیروظیفه به همراه تعدادی کنش‌های بدوی انجام می‌گیرد. و درنهایت یه ساختار سلسله‌مراتبی از وظایف ارائه می‌گردد.

یادگیری سلسله مراتبی به جهت محاسبات، سادگی برای ما فراهم می‌کند که از لحاظ کارایی بهتر عمل خواهد کرد و شاید در همه‌ی موارد ما را به تصمیم بهینه نرساند اما تصمیم‌هایی به اندازه‌ی کافی خوب را برای ما به همراه خواهد داشت. البته داشتن این مزیت نیازمند تعریف درست زیرمسئله‌هاست. روش‌هایی برای پیدا کردن درست زیرمسئله‌ها مثل روش‌های مبتنی بر گراف[8][9]، روش‌های مبتنی بر فرکانس ارائه شده است. [10][11]

[9] Hierarchical Reinforcement Learning
[10] Curse od Dimension

بیان سوال اصلی

همانطور که گفتیم در مدل کردن دنیای واقعی با استفاده از یادگیری تقویتی در برخی موارد دچار مشکل می‌شویم. این مشکل از آنجایی به دست می‌آید که مسائل دنیای واقعی عموما بزرگ‌اند و در یادگیری تقویتی دچار مشکل نفرین ابعاد می‌شویم. اگر یادگیری سلسله‌مراتبی ممکن باشد، سرعت یادگیری در مسائل بزرگ دنیای واقعی افزایش میابد. از طرفی گفتیم که دو ساختار مدل‌محور و مدل‌آزاد برای یادگیری تقویتی پیشنهاد شده‌است.

با توجه به موارد ذکرشده، سوال این پژوهش به این سمت می‌رود که اولا بررسی کند که هنگامی که یادگیری سلسله‌مراتبی به‌صورت شفاف ممکن باشد، رفتار یادگیری انسان چگونه از یادگیری مدل‌محور و مدل‌آزاد استفاده می‌کند؟

دوما در شرایطی که مدل سلسله‌مراتبی به‌صورت شفاف در مسئله وجود ندارد، دیده‌شده است که افراد زیراهدافی را انتخاب می‌کنند و مسئله‌ی کوتاه‌ترین مسیر را حل می‌کنند. اگر در یک مسئله‌ی یادگیری سلسه‌مراتبی شفاف وجود نداشته باشد، روش یادگیری چگونه است؟ آیا افراد وظیفه را به یادگیری چند وظیفه می‌شکنند؟ در صورتی که پاسخ مثبت است، یادگیری پس از شکستن چگونه از روش‌های یادگیری مدل‌محور و مدل‌آزاد در حل زیرمسائل استفاده می‌کند؟

روشها و فنون اجرایی طرح

چشم‌اندازهایی که برای انجام این پژوهش در نظر گرفته شده است شامل مطالعه‌ی آزمایش‌های رفتاری موجود و استفاده از آنها برای طراحی آزمایش، جمع‌آوری داده از افراد، مدل‌سازی تعامل دو سیستم، ارائه‌ی شواهد مغزی برای مدل‌سازی.

۱- طراحی آزمایش

در مرحله‌ی اول باید با مطالعه‌ی آزمایش‌های موجود در هر یک از حوزه‌های یادگیری سلسله‌مراتبی و یادگیری بر مبنای دو سیستم دارای مدل و بدون مدل، به طراحی آزمایشی بپردازیم که علاوه بر داشتن پارامتری از بروز رفتاری برای جدا کردن دو نوع سیستم یادگیری دارای مدل و بدون مدل، دارای ساختار سلسله‌مراتبی بوده تا رابطه‌ی این دو سیستم را بتواند بر اساس سوال پژوهش بدست آورد.

۲- جمع‌آوری داده

بعد از مرحله‌ی طراحی، در طول آزمایش اطلاعات رفتاری افراد ثبت شده تا در مرحله‌ی تحلیل مورد بررسی قرار گیرد.

۳- مدل‌سازی و بررسی داده‌های

مدل‌سازی قدم مهمی در انجام این پروژه است، با توجه به شواهد رفتاری و نیازهایی که آزمایش بر اساس آن‌ها طراحی شده‌بود به تحلیل داده‌ها و ارائه‌ی نتایج بپردازیم.

۴-تصویربرداری عملکردی تشدید مغناطیسی

در این مرحله تلاش خواهیم کرد تا شواهد مغزی‌ای برای ادعاهای و نتایج پژوهش نیز با گرفتن داده‌های تصویربرداری تشدید مغناطیسی در حین انجام آزمایش، ارائه دهیم.

پیشینه تحقیق (همراه با ذکر منابع اساسی)

سیستم‌های یادگیری

یکی از نتایج مهم تحقیقات انجام گرفته در چهارچوب یادگیری تقویتی حضور چندین روش یادگیری در عوامل یادگیری است. بسیاری از حضور دو سیستم یادگیرنده با نام‌های مختلف توضیح می‌دهند. بر اساس این چهارچوب «دو سیستم یادگیری» بسیاری از ابعاد متفاوت رفتار انسانی قابل توجیه می‌شود که توضیح یکی از موردقبول‌ترین این تئوری‌ها در قسمت مطالعات پیشین ارائه شده است. مواردی مثل نحوه‌ی ایجاد عادات و یا انواع بایاس‌های رفتاری توسط این دو سیستم قابلیت پشتیبانی دارد. یکی از تعاریف این دو سیستم، سیستم یادگیری بر مبنای مدل و بدون مدل هستند. پژوهش‌ها در چند سال اخیر به سمت نحوه‌ی کار این دو سیستم و شناسایی روش‌هایی برای تمایز این دو

سیستم ارائه شده است. در این میان تسک‌هایی متفاوتی که با تحلیل داده‌های رفتاری و مغزی به بررسی و مدل‌سازی و یافتن شواهد برای نحوه‌ی بروز کار این دو سیستم در عامل طراحی شده‌اند.[15][14][13]

در حوزه‌ی یادگیری ماشین فرایادگیری به معنای یادگیریِ یادگیری است. به صورت شهودی الگوریتم‌های فرایادگیری از تجربه‌ها استفاده می‌کنند تا تا جنبه‌های متفاوتی از الگوریتم یادگیری خود را بهبود بخشند. این یادگیری بهبود یافته از الگوریتم یادگیری اولیه عملکرد بهتری خواهد داشت.

این مفهوم اولین بار در سال ۱۹۷۹ بیان شد و به بررسی روند عامل‌های یادگیرنده‌ای که با استفاده از نوعی کنترل روی فرآیند یادگیری خود به بهبود این روند می‌پردازند، پرداخت. می‌توان فرایادگیری را اگاهی نسبت به فرآیند یادگیری در ناخودآگاه عامل مستقل از دانش شخصی عامل تعریف کرد. در واقع این مفهوم می‌تواند به خودکار کردن تصمیم‌های انسان و بهینه کردن این تصمیم‌ها در حین یادگیری بپردازد.

بعضی از فلاسفه اعتقاد دارند که روش های علمی در واقع یکی از حالت‌های پیاده‌سازی فرایادگیری هستند.

این حوزه همچنین دارای مدل‌سازی‌ها و فرمول‌بندی‌های ریاضی مخصوص به خود است که می‌تواند رده‌ی فرایادگیری، شروع فرآیند فرایادگیری، پروسه‌ی اضافه شدن تجربه، دامنه‌ها، پارامترهای فرایادگیری، دانش‌های قبلی عامل و الگورتیم‌های متفاوت یادگیری را مدل کند.

[1] Montague, P. Read, Peter Dayan, and Terrence J. Sejnowski. "A framework for mesencephalic dopamine systems based on predictive Hebbian learning." The Journal of neuroscience 16.5 (1996): 1936-1947.
[2] Fiorillo, Christopher D., Philippe N. Tobler, and Wolfram Schultz. "Discrete coding of reward probability and uncertainty by dopamine neurons." Science299.5614 (2003): 1898-1902.
[3]Daw, Nathaniel D., and Kenji Doya. "The computational neurobiology of learning and reward." Current opinion in neurobiology 16.2 (2006): 199-204.
[4] Dolan, Ray J., and Peter Dayan. "Goals and habits in the brain." Neuron 80.2 (2013): 312-325
[5] Norton, L. & Walters, D (2005). Encouraging meta-learning through personal development planning: first year students' perceptions of what makes a really good student. PRIME (Pedagogical Research In Maximising Education), in-house journal, Liverpool Hope University, 1 (1) 109-124.
[6] N. D. Daw, Y. Niv, and P. Dayan, "Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control," Nat. Neurosci., vol. 8, no. 12, pp. 1704–1711, 2005.
[7]M. M. Botvinick, Y. Niv, and A. C. Barto, "Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective," Cognition, vol. 113, no. 3, pp. 262–280, 2009.
[8]S. Mannor, I. Menache, A. Hoze, and U. Klein, "Dynamic abstraction in reinforcement learning via clustering," in Proceedings of the twenty-first international conference on Machine learning, 2004, p. 71.
[9]S. Mahadevan and M. Maggioni, "Proto-value Functions: A Laplacian Framework for Learning Representation and Control in Markov Decision Processes.," J. Mach. Learn. Res., vol. 8, no. 2169–2231, p. 16, 2007.
[10]M. Stolle and D. Precup, "Learning options in reinforcement learning," in SARA, 2002, pp. 212–223.
[11]C. Shi, R. Huang, and Z. Shi, "Automatic discovery of subgoals in reinforcement learning using unique-dreiction value," in Cognitive Informatics, 6th IEEE International Conference on, 2007, pp. 480–486.
[12] Dayan, Peter, and Yael Niv. "Reinforcement learning: the good, the bad and the ugly." Current opinion in neurobiology 18.2 (2008): 185-196.

[13] Dolan, Ray J., and Peter Dayan. "Goals and habits in the brain." Neuron 80.2 (2013): 312-325.
[14] Kahneman, Daniel. Thinking, fast and slow. Macmillan (2011).
[15] Dayan, Peter. "Rationalizable irrationalities of choice." Topics in cognitive science 6.2 (2014): 204-228.
[16]Daw, Nathaniel D., et al. "Model-based influences on humans' choices and striatal prediction errors." Neuron 69.6 (2011): 1204-1215.
[17]Lee, S.W., Shimojo, S., and O'Doherty, J.P. (2014).Neuron 81, this issue, 687–699.
[18]N.D. Daw, Y. Niv, P. Dayan "Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control" Nat Neurosci, 8 (2005), pp. 1704–1711
[19]S.J. Gershman, A.B. Markman, A.R. Otto "Retrospective revaluation in sequential decision making: a tale of two systems" J Exp Psychol Gen, 143 (2014), pp. 182–194

5- مصوبه شورای پژوهشی و تحصیلات تکمیلی‌دانشکده مهندسی برق و کامپیوتر

5-1-فرم پیشنهاد و حمایت از پایان‌نامه در تاریخ.....در شورای پژوهشی و تحصیلات تکمیلی دانشکده /گروه مطرح و نظرشورا به شرح زیر اعلام می‌شود:

☐ تصویب شد ☐ نیاز به اصلاح دارد ☐ به تصویب نرسید

5-2-عنوان طرح جامع تحقیقات استاد راهنما:

5-3- آیا پایان نامه پیشنهادی مرتبط با طرح جامع تحقیقات استاد راهنما/مشاور/گروه آموزشی/ دانشکده می باشد:

☐ بلی ☐ خیر

امضا استاد راهنما

امضاء رئیس / معاون پژوهشی و تحصیلات تکمیلی‌دانشکده مهندسی .....

شماره:

تاریخ:

معاون محترم آموزشی وتحصیلات تکمیلی پردیس دانشکده های فنی

با سلام و احترام,

فرم پیشنهاد و حمایت از پایان‌نامه کارشناسی ارشد / رساله دکتری آقای / خانم .....

با عنوان .....

به راهنمایی آقای / خانم دکتر .....

در شورای پژوهشی و تحصیلات تکمیلی دانشکده مهندسی ..... مورخ ..... به تصویب رسید.

خواهشمند است دستور فرمایید اقدامات مقتضی انجام شود.

امضاء رئیس / معاون پژوهشی و تحصیلات تکمیلی دانشکده مهندسی .....

شماره:

تاریخ:

معاون محترم پژوهشی پردیس دانشکده های فنی

با سلام و احترام ,

به پیوست فرم پیشنهاد و حمایت از پایان‌نامه تحصیلات تکمیلی با مشخصات مذکور که به تصویب شورای پژوهشی وتحصیلات تکمیلی دانشکده مهندسی ..... رسیده است، جهت دستور اقدام مقتضی تقدیم می‌شود.

امضاء معاون آموزشی و تحصیلات تکمیلی پردیس دانشکده های فنّی

رونوشت: معاون محترم پژوهشی و تحصیلات تکمیلی دانشکده مهندسی ..... : جهت اطلاع و پیگیری