

بسمه تعالی



دانشگاه تهران

پردیس دانشکده هایفنی

دانشکده مهندسی برق و کامپیوتر

پیشنهاد و فرم حمایت از پایان نامه تحصیلات تکمیلی

☐

دکتری

☒

کارشناسی ارشد

شماره مرجع : *

*شماره مرجع، توسط معاونت پژوهشی پردیس دانشکده‌های فنی هنگام صدور ابلاغ درج خواهد شد.

1- خلاصه اطلاعات پایان نامه

عنوان پایان نامه به زبان فارسی:	
عنوان پایان نامه به زبان انگلیسی	
نوع پایان نامه:	بنیادی <input checked="" type="checkbox"/> کاربردی <input type="checkbox"/> توسعه‌ای <input type="checkbox"/>
پروفسور/دانشگاه/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 نام خانوادگی	مرتبه علمی	محل خدمت	امضاء
استاد راهنما (مجری)	دکتر نیلی	استاد	دانشگاه تهران – دانشکده فنی	
استاد راهنمای دوم (حسب نیاز)				
استاد مشاور				
استاد مشاور دوم (برای دکتری)				

3- اطلاعات دانشجو

نام و نام خانوادگی: یاسمن رازی 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
فمراه: 09124331187

4- مشخصات موضوعی پایان نامه

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

یادگیری تقویتی¹

یادگیری تقویتی مدلی است که در سالیان اخیر برای بررسی یادگیری موجودات زنده به کار گرفته شده است و با توجه به شواهد و نتایج بدست آمده موفقیت چشم‌گیری در این حوزه داشته است. 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 تأیید کنند. هم‌چنین این مدل‌ها توانسته است سیستم‌های یادگیری مغز را که بر مبنای دوپامین² کار می‌کنند و یکی از پایه‌ای‌ترین نقش‌ها در فرایند یادگیری را دارد توجیه کند [1][2][3]. این موفقیت در توجیه شواهد یادگیری در انسان موجب تقویت این حوزه و جذب بسیاری به این حوزه شده است.

سیستم‌های یادگیری

نظریات متفاوتی در رابطه با سیستم‌های متفاوت یادگیری در انسان ارائه شده است. 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 کرده است. براساس این تحقیقات به نظر می‌رسد که پستانداران از دو نوع سیستم برای یادگیری انتخاب درست استفاده می‌کنند: سیستم دارای مدل³ و سیستم بدون مدل⁴. در سیستم اول که سیستم دارای مدل شناخته می‌شود عامل تلاش می‌کند که بر مبنای اطلاعاتی که از محیط⁵ 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 محاسبه 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 سیستم بدون مدل است، عامل مدلی از محیط ندارد و با بازخورد از محیط انتخاب‌های خود را بهینه می‌کند. به این صورت که عامل ابتدا برای هر یک از انتخاب‌های خود بسته به شرایط یک ارزش ذهنی در نظر دارد، و بر اساس آن ارزش ذهنی اولیه

¹ Reinforcement Learning

² Dopamine

³ Model Based

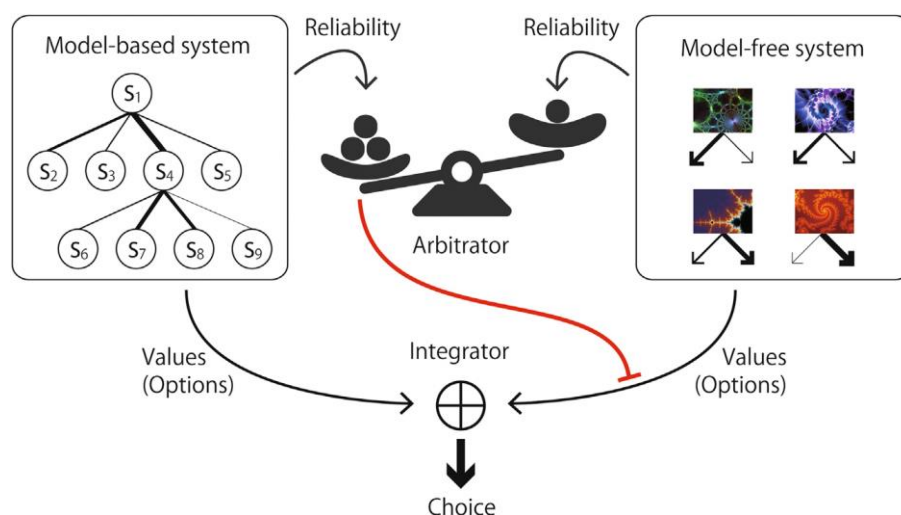
⁴ Model Free

⁵ Environment

انتخاب خود را انجام داده و از محیط پاداشی⁶ دریافت می‌کند و از این پاداش یا بازخورد برای بهبود ارزش اعمال استفاده می‌کند. این پاداش می‌تواند با تصور پیشین عامل متفاوت باشد که از این 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 بهینه‌ای 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 تصمیم‌گیری را نیز می‌تواند بسیار بالا برد. حضور این دو سیستم در کنار هم باعث بروز انعطاف‌پذیری⁷ 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 هم‌زمان رقابت می‌کنند و یک سیستم میانجی با توجه به اطمینان‌پذیری⁸ این دو سیستم از بین آنها انتخاب می‌کند. با این معنا که در لحظه هر سیستمی که با توجه به تصمیم‌های پیشین خود اطمینان‌پذیری بیشتری داشته باشد برای تصمیم‌گیری استفاده می‌شود. شماتیک این تعامل را می‌توان در شکل شماره‌ی یک دید [17].

در تعریفی دیگر این دو سیستم با هم همکاری می‌کنند. 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 از نظرهای تعامل دو سیستم را نشان می‌دهد

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 مراتبی⁹

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 می‌یابد و این یعنی پیچیدگی محاسباتی تابعی نمایی از تعداد حالات است. از این مشکل تحت عنوان نفرین ابعاد¹⁰ 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 سلسله‌مراتبی 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 باشد، 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 به 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 of Dimension

بیان سوال اصلی

سوال اصلی این پروژه را می‌توان به بررسی تعامل بین این دو سیستم یادگیری بر مبنای مدل و بدون مدل نسبت داد و اینکه آیا این تعامل می‌تواند ناشی از یک رفتار سلسله‌مراتبی در روند یادگیری باشد یا خیر؟ پژوهش‌های زیادی به روابط بین این دو سیستم یادگیری و یا بروز رفتار سلسله‌مراتبی در فرایند یادگیری انسانی و همچنین پیدا کردن شواهد مغز برای هر کدام یک از این سیستم‌ها انجام شده‌است. در این پروژه بر آنیم تا رابطه‌ی این رفتار یادگیری سلسله‌مراتبی را در قالب تعریف دو سیستم یادگیری دارای مدل و بدون مدل بررسی کنیم. و به این سوال که آیا تقسیم مسئله به ساختارهای کوچک‌تر در یادگیری سلسله‌مراتبی خود از این دو سیستم یادگیری دارای مدل و بدون مدل تبعیت می‌کنند یا خیر پاسخ دهیم.

این پژوهش از چندین جهت می‌تواند جالب باشد اول اینکه شناخت بهتر و درست‌تر ما نسبت به تعامل سیستم‌های یادگیرنده‌ی انسانی می‌تواند ما را به سمت استفاده‌ی بهتر از این سیستم‌ها سوق دهد. همچنین در مواردی که به دلایلی مثل بروز اعتیاد، بیماری‌های متفاوت و عوامل مداخله‌گر دیگر رفتار یکی از این سیستم‌ها دچار اختلال شود با دانستن نحوه‌ی تعامل شاید بتوان در روند درمانی این اختلالات دخالت کرد.

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. [13][14][15]

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 آگاهی 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 روش‌های 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 الگوریتم‌های 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." *Science* 299.5614 (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-direction 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 تکمیلی دانشکده مهندسی برق و کامپیوتر

5-1- فرم پیشنهاد و حمایت از پایان نامه در تاریخ..... در شورای پژوهشی و تحصیلات تکمیلی دانشکده /گروه		
مطرح و نظرشورا به شرح زیر اعلام می شود:		
<input type="checkbox"/> تصویب شد	<input type="checkbox"/> نیاز به اصلاح دارد	<input type="checkbox"/> 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:		
<input type="checkbox"/> خیر	<input type="checkbox"/> بلی	
امضا استاد راهنما		
امضاء رئیس / معاون پژوهشی و تحصیلات تکمیلی دانشکده مهندسی		

شماره:
تاریخ:
معاون محترم آموزشی و تحصیلات تکمیلی پردیس دانشکده های فنی
با سلام و احترام،
فرم پیشنهاد و حمایت از پایان نامه 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 پایان نامه تحصیلات تکمیلی با مشخصات مذکور که به تصویب شورای پژوهشی و تحصیلات تکمیلی دانشکده مهندسی رسیده است، جهت دستور اقدام مقتضی تقدیم می شود.

امضاء معاون آموزشی و تحصیلات تکمیلی پردیس دانشکده های فنی

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