بهار ۱۴۰۰، دوره ۶، شماره ۱، پیاپی ۲۰
DOR :20.1001.1.2538161.1400.6.1.1.9

تاریخ دریافت مقاله: ۱۳۹۹/۲/۱۵
تاریخ پذیرش مقاله: ۱۳۹۹/۸/۲۲

نوع مقاله: پژوهشی

مدل‌سازی محاسباتی نظریه ذهن: از مشاهده تا یادگیری

آشنا گرگان محمدی
دانشجوی کارشناسی ارشد، دانشکده ریاضی، آمار و علوم کامپیوتر، دانشگاه تهران، تهران، ایران
پست الکترونیکی: ashena.mohammadi@ut.ac.ir

محمد گنج‌تابش *
دانشیار، دانشکده ریاضی، آمار و علوم کامپیوتر، دانشگاه تهران، تهران، ایران
پست الکترونیکی: mgtabesh@ut.ac.ir

چکیده

نظریه ذهن به توانایی انسان در استنتاج حالت‌های ذهنی خود و دیگران گفته می‌شود. این توانایی در روابط اجتماعی و یادگیری فرد نقش اساسی ایفا می‌کند. مطالعه محاسباتی این توانایی می‌تواند پاسخی بر چالش‌های کنونی حوزه هوش مصنوعی ارائه کند. دیدگاه شبیه‌سازی-نظریه با بهره‌گیری از سازوکار نورون‌های آینه‌ای، در یادگیری فرد از طریق مشاهده رفتار دیگران اثرگذار است. بنابراین، تعبیه کردن نظریه ذهن در یک عامل هوشمند موجب می‌شود که آن عامل با مشاهده رفتارهای یک عامل خبره، به عملکردی مشابه او در تعامل با محیط اطراف خود دست یابد. در این مطالعه، عامل هوشمند را به مدلی از نورون‌های آینه‌ای مجهز می‌کنیم و آن را در محیط آونگ واژگون قرار می‌دهیم. خواهیم دید که این عامل هوشمند با مشاهده رفتار عامل خبره در محیط به درکی از اهداف و دانش عامل خبره دست می‌یابد که به کمک آن، آونگ واژگون را در حالت تعادل نگه می‌دارد.

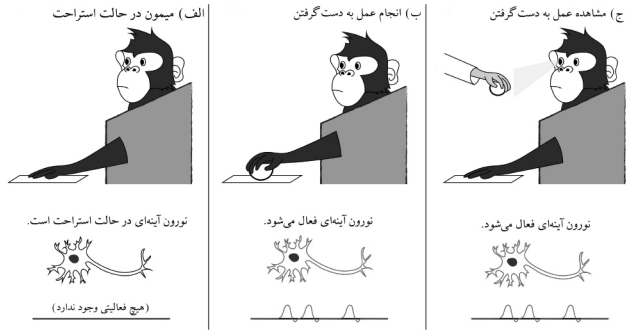
واژه‌های کلیدی: نظریه ذهن، نورون‌های آینه‌ای، شبکه‌های عصبی ضربه‌ای.

۱ مقدمه

ما انسان‌ها قادریم تا با استنتاج و تحلیل عقاید، اهداف، باورها و احساسات دیگران، تعامل مؤثری با آن‌ها و محیط اطراف خود داشته باشیم و بر مجموعه دانسته‌های خود بیفزاییم [۱، ۲]. این توانایی که به آن نظریه ذهن گویند، یکی از بحث‌برانگیزترین موضوعات فلسفی-روان‌شناختی است. در گذشته فلاسفه و روان‌شناسان بر این باور بودند که نظریه ذهن نوعی از توانایی است که طی آن، فرد حالت‌های ذهنی خود و دیگران را به صورت فرض‌هایی از نظریه‌های خام استنتاج می‌کند [۳، ۴]. این دیدگاه، به دیدگاه نظریه-نظریه[1] معروف است. در مقابل این دیدگاه، دیدگاه شبیه‌سازی-نظریه[2] قرار دارد که قدمت کمتری داشته و بیان می‌کند که افراد با اتخاذ زاویه دید شخص مقابل، حالت‌های ذهنی او را استنتاج می‌کنند [۳، ۴]. در دیدگاه نظریه-نظریه، نظریه ذهن معمولاً به صورت یک توانایی ذاتی در نظر گرفته می‌شود؛ حال آنکه در دیدگاه شبیه‌سازی-نظریه، باور غالب بر این است که این توانایی طی مراحل رشد، یاد گرفته می‌شود [۱]. با وجودی که بحث‌های فراوانی بر سر این دو دیدگاه وجود

* نویسنده مسئول

1- Theory-Theory
2- Simulation-Theory



شکل ۱: سازوکار یک نورون آینه‌ای [۷]. نورون آینه‌ای نه تنها در زمانی که میمون جسم را به دست می‌گیرد، بلکه هنگامی که فرد دیگری را در حال به دست گرفتن آن جسم مشاهده می‌کند، فعال می‌شود.

داشته و همچنان نیز این بحث‌ها ادامه دارد، مطالعات اخیر عصب‌شناختی پرده از برخی رازهای این توانایی گشوده‌اند.

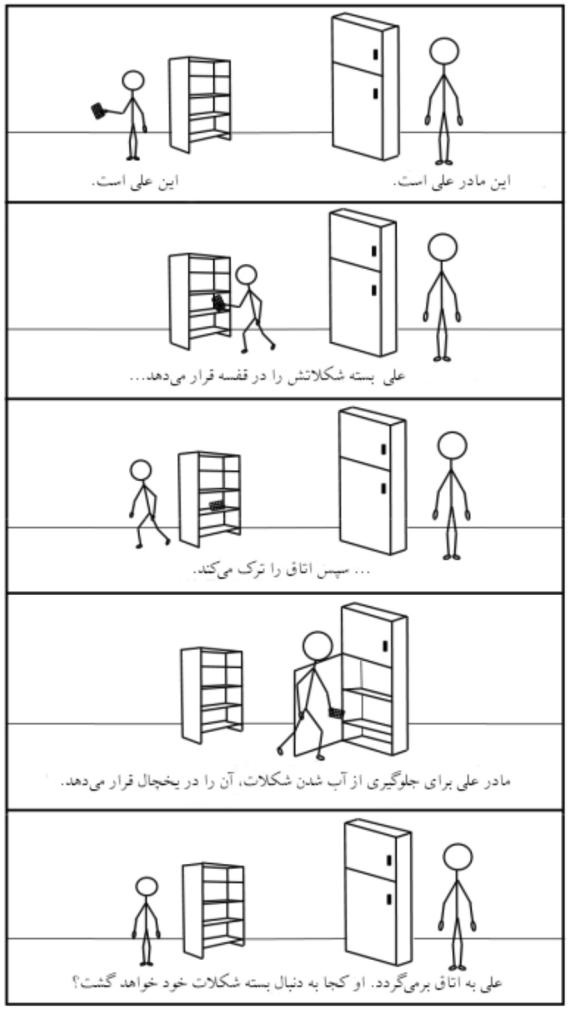
در سال ۱۹۷۸، مقاله‌ای با عنوان «آیا شامپانزه‌ها دارای نظریه ذهن هستند» منتشر شد [۵]. در این مقاله به بررسی نورون‌های خاصی در این حیوانات پرداخته شده بود. این نورون‌ها که در ناحیه F5 قشر پیش‌حرکتی واقع بودند، نه تنها هنگامی که شامپانزه جسمی مانند موز را به دست می‌گرفت فعال می‌شدند، بلکه وقتی انسان یا شامپانزه دیگری نیز همان جسم را به دست می‌گرفتند، فعالیت‌های مشابهی از خود نشان می‌دادند (شکل ۱). این نورون‌ها به واسطه عملکرد آینه‌وارشان به نورون‌های آینه‌ای مشهور شدند [۵، ۴]. در سال‌های پس از آن و با پیشرفت روش‌های تصویربرداری مغز، قسمت‌هایی در مغز انسان با سازوکاری مشابه یافت شد که به آن‌ها، سیستم نورون‌های آینه‌ای اطلاق می‌شود [۳، ۶]. این سیستم به عنوان یکی از زیرساخت‌های سطح پایین نظریه ذهن از دیدگاه شبیه‌سازی-نظریه به‌شمار می‌رود.

به منظور شناخت بیشتر این توانایی و همچنین، تعبیه آن در ماشین‌ها و سیستم‌های هوشمند، دانشمندان بر آن شدند که به کمک روش‌های محاسباتی نیز این توانایی را مورد بررسی قرار دهند. این مطالعات غالباً از دیدگاه نظریه-نظریه صورت گرفته‌اند. با این حال، مدل‌هایی شناختی از دیدگاه شبیه‌سازی-نظریه نیز به مسئله پرداخته‌اند که می‌توانند اساس مطالعات محاسباتی قرار گیرند. در این مطالعه سعی داریم به بررسی مدل‌سازی محاسباتی نظریه ذهن، به‌ویژه از دیدگاه شبیه‌سازی-نظریه، بپردازیم. سؤال اساسی در این مطالعه آن است که آیا می‌توان با تعبیه سیستم نورون‌های آینه‌ای در یک عامل هوشمند، آن را به سطوح پایین نظریه ذهن مجهز کرد؟ برای پاسخ به این سؤال، مسئله را به توانایی یادگیری از طریق مشاهده رفتار دیگران تقلیل دادیم؛ چراکه نظریه ذهن در فرایند یادگیری اثرگذار است. فرد با مشاهده رفتار دیگران در محیط، نحوه تعامل با محیط پیرامون را می‌آموزد. به عبارت دیگر، فرد با استنتاج دانش دیگران از محیط و اهداف آن‌ها می‌تواند نحوه تعامل با محیط اطراف خود را آموخته و با بهره‌گیری از نظریه ذهن، می‌تواند حالت‌های ذهنی خود و فرد مقابل را از هم تمیز دهد. مطالعات حاکی از آن است که سیستم نورون‌های آینه‌ای در فرایند آموزش افراد نقش دارند [۸]. بنابراین، به کمک شبیه‌سازی سیستم نورون‌های آینه‌ای، سعی در مدل‌سازی این توانایی در عامل هوشمند داریم. بدین منظور از شبکه‌های عصبی ضربه‌ای بهره خواهیم برد. عامل هوشمند را در محیط ساده آونگ واژگون قرار داده و با ارائه مدلی ساده، به حل مسئله مطرح شده می‌پردازیم.

در ادامه، ابتدا مروری بر مطالعات محاسباتی صورت گرفته روی نظریه ذهن و نیم‌نگاهی بر مدل‌های شناختی که نقشه راه ما در ارائه مدل محاسباتی نظریه ذهن شده‌اند، خواهیم داشت. سپس به بررسی شبکه‌های عصبی ضربه‌ای و روش شبیه‌سازی سازوکار نورون‌های آینه‌ای در مدل پیشنهادی خواهیم پرداخت. در انتها نیز، پس از ارائه آزمایش‌ها و نتایج صورت گرفته روی محیط آونگ واژگون، به بحث و بررسی مدل پیشنهادی و چالش‌های پیش رو برای تعمیم آن به محیط‌های پیچیده‌تر می‌پردازیم.

۲ مرور کارهای پیشین

مدل‌های مختلف نیازمند ابزارهای متفاوتی جهت مطالعه و اعتبارسنجی می‌باشند. در همین راستا، نظریه ذهن نیز می‌تواند به کمک آزمایش‌های روان‌شناختی مورد مطالعه و اعتبارسنجی قرار گیرد. یکی از مشهورترین این آزمایش‌ها، آزمایش باور غلط[1] است. باور غلط بیان می‌دارد که یک فرد ممکن است به بخشی از دانسته‌های فرد مقابل دسترسی نداشته باشد. آزمایش‌های باور غلط ماجرایی را به تصویر می‌کشند که در آن حداقل دو فرد وجود دارند و یکی از آن‌ها (فرد الف) به تمامی اطلاعات صحیح دسترسی دارد، درحالی‌که بخشی از اطلاعات از دید دیگری (فرد ب) پنهان است. سپس از فرد مورد آزمون درباره اندیشه‌های فرد الف و فرد ب سؤال می‌شود (شکل ۲). انتظار می‌رود که فرد مورد آزمون بتواند باور غلط فرد ب را تشخیص دهد [۱، ۹]. در مطالعات محاسباتی نظریه ذهن نیز معمولاً از این آزمایش برای اعتبارسنجی مدل‌ها استفاده شده و آن‌ها را به کمک فرآیندهای تصمیم‌گیری مارکوف (نیمه مشاهده‌پذیر)[2] مدل می‌کنند [۱۰، ۱۱، ۱۲]. غالب مدل‌های محاسباتی مطرح شده برای نظریه ذهن بر اساس آمار بیزی می‌باشند. به غیر از مدلی که به کمک شبکه‌های بیزی پویا[3] در سال ۲۰۰۵ ارائه شد و به منظور استنتاج وضعیت‌های ذهنی مشخصی (موافقت، مخالفت، تمرکز، علاقه‌مندی، تفکر، و عدم قطعیت) از روی حالات چهره و سر افراد توسعه یافته بود [۱۳]، بقیه مدل‌های بیزی به کمک آزمایش باور غلط مورد مطالعه و ارزیابی قرار گرفته‌اند. در سال ۲۰۰۷، کیلنر[4] سیستم نورون‌های آینه‌ای را از دیدگاه بیزی مورد بررسی قرار داد و مدلی از نظریه ذهن بر اساس دیدگاه شبیه‌سازی- نظریه ارائه کرد [۱۴]. در سال ۲۰۱۲، گوپنیک[5] و ولمن[6]



شکل ۲: نمونه‌ای از آزمایش باور غلط [۹]. در این آزمایش انتظار می‌رود فرد مورد آزمون، باور غلط علی را تشخیص داده و در نتیجه پاسخ دهد که علی در قفسه به دنبال بسته شکلات خود خواهد گشت.

مدل‌های سببی و روش‌های یادگیری بیزی تا آن سال را مورد بررسی قرار دادند [۱۵]. این مدل‌ها با بهره‌گیری از استنتاج بیزی، به استنتاج خواست‌ها، باورها و یا اهداف بر مبنای دیدگاه نظریه-نظریه می‌پردازند. در سال ۲۰۱۴ نیز، مدلی جامع‌تر تحت عنوان نظریه ذهن بیزی برای استنتاج خواست‌ها و باورها بر همین اساس ارائه شد [۱۰].

با پیشرفت شبکه‌های عصبی مصنوعی و الگوریتم‌های یادگیری، مطالعات محاسباتی نظریه ذهن نیز سمت و سوی دیگری به خود گرفت. در سال ۲۰۱۸، شرکت دیپ‌مایند[7] مدل نظریه ذهن ماشین[8] را،

1- False Belief Task
2- (Partially Observable) Markov Decision Processes
3- Dynamic Bayesian Networks
4- Kilner
5- Gopnik
6- Wellman
7- DeepMind
8- Machine ToM (ToMnet

که بر اساس شبکه‌های عصبی مصنوعی است، ارائه کرد [۱۶]. این مدل متشکل از سه زیرشبکه کوچک‌تر است که به کمک قوانین فرایادگیری[1] آموزش دیده است. این زیرشبکه‌ها به منظور توصیف عامل نمایش داده شده به مدل، استنتاج حالت ذهنی و پیش‌بینی تصمیمات آینده او ارائه شده‌اند. هدف این مدل، شبیه‌سازی نظریه ذهن به کمک فراشناخت[2] است. این مدل را می‌توان نقطه عطفی در مطالعات محاسباتی نظریه ذهن از دیدگاه هوش مصنوعی دانست؛ به‌طوری‌که پس از ارائه آن، به‌کارگیری نظریه ذهن در هوش مصنوعی و سیستم‌های هوشمند، چالش‌های آن، و روش‌های احتمالی در بین سال‌های ۲۰۱۸ تا ۲۰۲۱ به دقت مورد بحث و بررسی قرار گرفت [۱۷، ۱۱، ۲، ۹]. از نتایج این بررسی‌ها، می‌توان به بهره‌گیری از روش‌های یادگیری تقویتی معکوس[3] برای مدل‌سازی نظریه ذهن در هوش مصنوعی اشاره کرد [۱۲]. با وجود این پیشرفت‌ها در مدل‌های مطرح شده، همچنان جای خالی مدلی که شواهد زیستی به‌دست آمده را در خود جای دهد، حس می‌شد.

در سال ۲۰۲۰، شواهد به‌دست آمده از داده‌های مربوط به تصویربرداری از مغز انسان کنار هم قرار گرفتند و مدل نظریه ذهنِ مغز[4] برای روبات‌های انسان‌نما به کمک شبکه‌های عصبی مصنوعی ارائه شد [۱۸]. این مدل شبکه‌ای چندلایه است که تنها دو لایه آن قابل آموزش می‌باشند. هر دوی این لایه‌ها از شبکه‌های عصبی ضربه‌ای بهره گرفته و به منظور بازشناسی اشیاء در شبکه اصلی تعبیه شده‌اند.

در همین سال، شرکت دیپ‌مایند مدل موزیرو[5] را ارائه کرد [۱۹] که گرچه از آن به عنوان مدلی محاسباتی از نظریه ذهن یاد نشده است، ولی می‌توان آن را به تعبیری بهره‌مند از نظریه ذهن دانست. این مدل با ترکیب روش جستجوی درختی (مدل‌سازی فراشناخت) و یک مدل از پیش آموزش داده شده (مشابه آنچه ما عامل خبره نامیدیم)، تابع پاداش و سیاست حرکت (مشابه استنتاج اهداف، مقاصد، و دانش عامل خبره) را می‌آموزد. همچنان، سؤال‌هایی مطرح است که پاسخ به آن‌ها در راستای برچسب‌گذاری این مدل به عنوان مدل محاسباتی نظریه ذهن و یا عاملی هوشمند و بهره‌مند از نظریه ذهن، اجباری است. سؤال نخست آن است که این مدل تا چه حد به دانش عامل خبره دست می‌یابد و سیاست‌های خود را هم‌راستا با او تعیین می‌کند؟ سؤال دوم و مهم‌تر آن است که آیا این مدل از لحاظ زیستی و شواهد روان‌شناختی توجیه‌پذیر است؟ برای پاسخ به این سؤالات، اطلاعات بیشتری از مدل نیاز است تا بتوان به طور قطعی اظهار نظر کرد. با این حال، این مدل نمونه موفقی از یک عامل هوشمند در محیط‌های تعاملی است.

تمامی مدل‌های مطرح شده در بالا (به غیر از [۱۴]، [۱۸] و [۱۹]) بر اساس دیدگاه نظریه-نظریه می‌باشند. در [۱۸] نیز تأکید شده است که تلاشی برای توجیه هیچ کدام از دیدگاه‌های کلاسیک نظریه ذهن نشده است. با این حال، دیدگاه شبیه‌سازی-نظریه در مدل‌های شناختی همواره مورد توجه بوده است. این مدل‌ها غالباً بر پایه قوانین یادگیری تداعی‌گرا[6] مطرح شده‌اند [۲۰، ۶] که می‌توانند چراغ راه مدل‌سازی‌های محاسباتی باشند، همانطورکه ما را در این مسیر راهنمایی کردند. در ادامه، به بررسی مدل مطرح شده در این مطالعه می‌پردازیم.

۳ روش‌ها

در این پژوهش، با الهام از [۲۰]، از شبکه‌های عصبی ضربه‌ای با قانون یادگیری مبتنی بر زمان ضربه[7] بهره می‌بریم. عامل هوشمند با بهره‌گیری از این شبکه و مشاهده یک عامل خبره، تعامل با محیط پیرامون خود را می‌آموزد. در ادامه، پس از بررسی دقیق‌تر شبکه‌های

1- Metalearning
2- Metacognition
3- Inverse Reinforcement Learning
4- BrainToM
5- MuZero
6- Associative learning
7- Spike-Time Dependent Plasticity (STDP)

عصبی ضربه‌ای، به تشریح مدل پیشنهادی خواهیم پرداخت.

۱.۳ شبکه‌های عصبی ضربه‌ای

در این شبکه‌ها از سیگنال‌های دودویی، که به آن‌ها ضربه[1] گفته می‌شود، برای انتقال و پردازش اطلاعات استفاده می‌شود. ضربه‌ها در طول زمان کد می‌شوند و پیام مورد نظر بین جمعیت‌های نورونی شبکه به‌وسیله اتصالات سیناپسی انتقال می‌یابند. این اتصالات می‌توانند از الگوهای خاصی پیروی کنند؛ به طور مثال، ممکن است بین تمام نورون‌های دو جمعیت ارتباط وجود داشته باشد و یا ممکن است هر نورون پیش(پس)سیناپسی به تعداد مشخصی نورون پس(پیش)سیناپسی به صورت تصادفی مرتبط باشد. هر اتصال سیناپسی بین دو نورون، دارای وزنی است که نمایانگر قدرت اتصال بین دو نورون پیش و پس‌سیناپسی است [۲۱]. بنابراین، برای طراحی شبکه مورد نظر، لازم است نورون‌ها، الگوی ارتباطی بین آن‌ها، و سازوکار یادگیری برای تغییر وزن‌های سیناپسی مشخص شوند.

یکی از پرکاربردترین مدل‌های نورونی مورد استفاده در این شبکه‌ها، مدل نورونی تجمیع و آتش نشتی[2] است. این مدل نورونی، در عین سادگی، توانایی تولید الگوهای متداول ضربه در نورون‌های مغز انسان را دارد [۲۲]. بنابراین، برای طراحی و پیاده‌سازی شبکه عصبی عامل هوشمند، از این مدل نورونی بهره می‌گیریم. دینامیک نورونی در مدل نورونی تجمیع و آتش نشتی به صورت روابط ۱ است:

$$\tau.\frac{du}{dt} = -(u - u_{rest}) + R.I(t),$$
$$\text{if } u(t) = \theta \implies \text{Fire} + \text{Reset}, \qquad (۱)$$

که در آن، u اختلاف پتانسیل غشای نورون، u_{rest} اختلاف پتانسیل غشای نورون در حالت استراحت، R مقاومت غشای نورون و $I(t)$ جریان ورودی نورون در طول زمان است [۲۱].

قانون یادگیری اصلی در این شبکه‌ها نیز، یادگیری مبتنی بر زمان ضربه است. این قانون یادگیری در واقع شکل تغییر یافته قانون یادگیری تداعی‌گرای هب[3] است. در قانون یادگیری هب، تداوم برانگیخته شدن نورون پس‌سیناپسی به واسطه فعالیت همزمان نورون پیش‌سیناپسی، موجب تقویت اتصال بین آن‌ها می‌شود [۲۰]. در مقابل، در قانون یادگیری مبتنی بر زمان ضربه، اختلاف زمانی بین برانگیخته شدن دو نورون نقش مهمی در تقویت و یا تضعیف وزن‌های سیناپسی بین آن‌ها ایفا می‌کند. به عبارت دقیق‌تر، قانون یادگیری مبتنی بر زمان ضربه بیان می‌دارد که اگر نورون پیش‌سیناپسی در بازه زمانی مشخصی قبل از نورون پس‌سیناپسی فعال شود، در برانگیخته شدن آن تأثیرگذار بوده و تداوم این فعالیت موجب تقویت وزن سیناپسی بین آن‌ها می‌شود. همچنین، اگر نورون پس‌سیناپسی در بازه زمانی مشخصی قبل از نورون پیش‌سیناپسی فعال شود، بدان معنی است که فعالیت نورون پیش‌سیناپسی تأثیری در فعالیت نورون پس‌سیناپسی ندارد و لذا تداوم این الگو موجب تضعیف سیناپس بین این دو نورون می‌شود [۲۳]. بر همین اساس، تغییرات وزن‌های سیناپسی به کمک روابط ۲ صورت می‌گیرد:

$$\Delta w = \begin{cases} A_+(w).\exp(\frac{-|\Delta t|}{\tau_+}) & \text{if } t_{pre} \le t_{post}, \\ A_-(w).\exp(\frac{-|\Delta t|}{\tau_-}) & \text{if } t_{pre} > t_{post}. \end{cases} \qquad (۲)$$

که در آن، $|\Delta t| = |t_{post} - t_{pre}|$، $\tau_\pm$ ثابت‌های زمانی و $A_\pm$ بیانگر شدت تقویت و یا تضعیف وزن‌های سیناپسی می‌باشند [۲۱].

همچنین، شواهد زیستی نشان می‌دهند که دوپامین نقش فعالی در سیستم جزا و پاداش دارد و در فرایندهای درگیر در نظریه ذهن مؤثر است. به همین منظور، در طراحی شبکه، از تعمیمی از یادگیری مبتنی بر زمان ضربه، با عنوان قانون یادگیری مبتنی بر زمان ضربه پاداش‌محور[4] بهره می‌بریم که با افزودن دینامیک دوپامین به معادله ۲ به دست می‌آید [۲۴]. معادلات حاصل به صورت روابط ۳ خواهند بود:

1- Spike
2- Leaky Integrate-and-Fire neurons (LIF)
3- Hebbian associative learning
4- Reward-modulated STDP

$$\frac{dc}{dt} = -\frac{c}{\tau_c} + \mathrm{STDP}(\Delta t)\delta(t - t_{pre/post}),$$
$$\frac{dw}{dt} = cd, \qquad (۳)$$
$$\frac{dd}{dt} = -\frac{d}{\tau_d} + \mathrm{DA}(t),$$

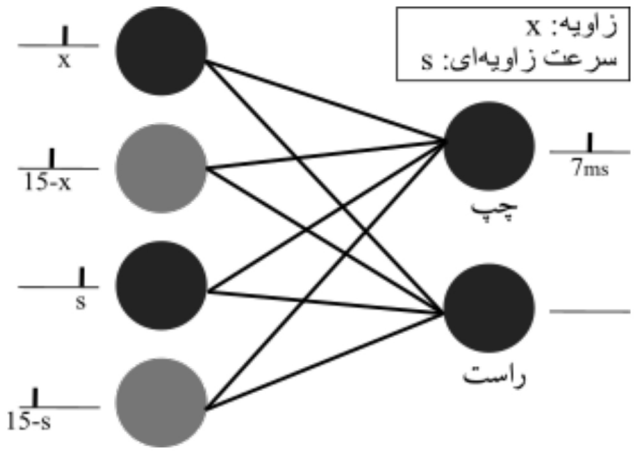
که در آن، c اثر بلندمدت تغییرات، d میزان دوپامین ترشح شده، و DA تابع دوپامین حاصل از فعالیت نورون‌های مسئول ترشح دوپامین می‌باشند [۲۱].

۲.۳ مدل پیشنهادی

همانطور که گفته شد، این مدل بر پایه دیدگاه شبیه‌سازی-نظریه و سازوکار نورون‌های آینه‌ای طراحی شده است. برای شبیه‌سازی سازوکار نورون‌های آینه‌ای، جمعیت نورونی خروجی را متناسب با فعالیت عامل خبره، وادار به ضربه زدن می‌کنیم. به عبارت دیگر، در یک لحظه، تصویر فعالیت انتخاب شده توسط عامل خبره به عامل هوشمند می‌رسد و در نتیجه، نورون‌های خروجی مانند یک آینه عمل کرده و نورون‌های مربوط به همان فعالیت، فعال می‌شوند.

شبکه‌ای که برای عامل هوشمند طراحی شده است، شبکه‌ای دو لایه و تنها شامل لایه ورودی و خروجی است (شکل ۳). ورودی شامل چهار نورون و خروجی شامل دو نورون است. فعال شدن هر یک از نورون‌های خروجی نمایانگر یکی از دو حرکت چپ و یا راست است. همچنین، درصورتی‌که هیچ نورونی فعال نشود، بدان معنی است که حرکت باید به صورت تصادفی انتخاب شود. در زمان یادگیری، به ازای هر فعالیت عامل خبره، نورون‌های متناسب با آن فعالیت (با توجه به مدل‌سازی نورون‌های آینه‌ای) فعال می‌شوند. توجه داشته باشید، از آنجایی که عامل هوشمند از پیش نمی‌داند که کدام فعالیت‌ها تصادفی و کدام فعالیت‌ها هوشمند هستند، به ازای حرکات تصادفی نیز نورون خروجی متناسب فعال خواهد شد. پس انتظار می‌رود که عامل هوشمند با توجه به رفتار عامل خبره، یاد بگیرد که چه زمان‌هایی لازم است به صورت تصادفی عمل کند.

برای آموزش شبکه از قانون یادگیری مبتنی بر زمان ضربه پاداش‌محور استفاده می‌کنیم. این قانون یادگیری بر اساس یادگیری تداعی‌گرا و قانون یادگیری تقویتی است. پاداش و جزا در عامل هوشمند بر اساس حالت ذهنی خود او و درک او از محیط تعریف می‌شود. با این حال، شبکه عصبی مربوط به عامل هوشمند باید توانایی استنتاج فعالیت‌هایی از عامل



شکل ۳: شبکه عصبی ضربه‌ای پیشنهادی. نورون‌های تیره رنگ در ورودی نورون‌های تحریکی و نورون‌های کمرنگ، نورون‌های مهاری هستند.

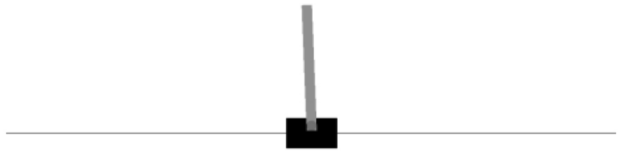
خبره که برای او (عامل خبره) منجر به پاداش یا جزا می‌شود را نیز داشته باشد. قانون یادگیری مبتنی بر زمان ضربه پاداش‌محور این امکان را به عامل هوشمند می‌دهد که علاوه بر لحاظ نمودن دانش خود، با توجه به آینه تصویر فعالیت عامل خبره، دانش خود را از نحوه تعامل با محیط تقویت کند. در ادامه، به بررسی دقیق‌تر این فرایند در محیط آونگ واژگون، به عنوان مثالی ساده از تعامل یک عامل هوشمند با محیط، خواهیم پرداخت.

۴ آزمایش‌ها و نتایج

برای بررسی مدل پیشنهادی، محیط آونگ واژگون (شکل ۴) را برگزیدیم. در این محیط، هدف برقراری تعادل در یک سیستم آونگ واژگون است. آونگ واژگون متشکل از یک گاری است که یک آونگ روی آن تعبیه شده است. برای برقراری تعادل در این سیستم، باید آونگ را تنها با حرکت دادن گاری به چپ و یا راست، تا حد امکان به صورت عمودی نگه داریم. همچنین، میزان جابه‌جایی افقی گاری نباید از یک حد بیشتر باشد. وضعیت آونگ در هر لحظه با چهار عدد حقیقی تعیین می‌شود: مکان گاری، سرعت گاری، زاویه آونگ، و سرعت زاویه‌ای آونگ. در زمانی‌که آونگ در حالت عمود قرار دارد، زاویه صفر و در صورت انحراف از این وضعیت، با توجه به جهت انحراف، مقداری مثبت و یا منفی خواهد بود.

عامل خبره انتخاب شده، به کمک الگوریتم تپه‌نوردی[1] در

1- Hill Climbing



شکل ۴: محیط آونگ واژگون. آونگ به صورت معلق روی گاری قرار دارد و هدف، حفظ تعادل آن به کمک حرکت دادن گاری به سمت چپ یا راست است.

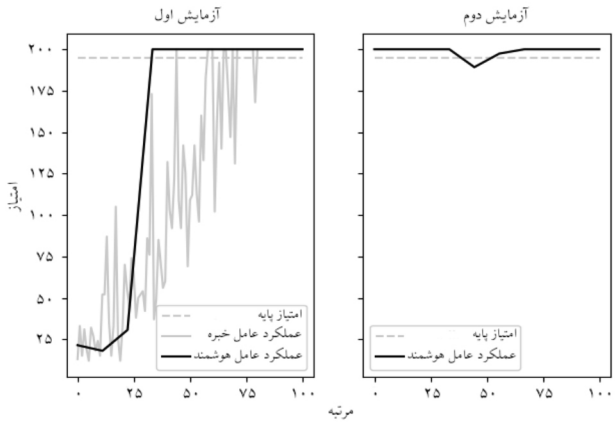
این محیط آموزش داده شده است. عامل هوشمند ۱۰۰ مرتبه عامل خبره را مشاهده می‌کند (هر مرتبه شامل لحظه شروع فعالیت عامل خبره در محیط تا لحظه شکست او در برقراری تعادل است). نتایج را در قالب دو آزمایش مختلف بررسی می‌کنیم: در آزمایش اول، در مرتبه‌های نخست، حرکات تصادفی به عامل خبره اضافه می‌شود که به مرور زمان، میزان انتخاب‌های تصادفی حرکات در او کاهش می‌یابد. بدین ترتیب، عامل خبره از دید عامل هوشمند، در حال یادگیری در محیط خواهد بود. در آزمایش دوم، عامل خبره همواره بهترین عملکرد خود را به نمایش می‌گذارد و تنها در زمان‌هایی که آونگ در حالت نسبتاً عمودی قرار دارد و حرکت به چپ و یا راست تعادل آونگ را برهم نمی‌زند، حرکتی را به صورت تصادفی انتخاب می‌کند.

از آنجایی که مکان و سرعت گاری با توجه به فعالیت عامل خبره قابل استنتاج است، تنها از داده‌های مرتبط با زاویه آونگ برای ورودی شبکه استفاده می‌کنیم. همچنین، با توجه به این‌که بازه اعداد ورودی پیوسته و ضربه‌ها سیگنال‌هایی دودویی و گسسته هستند، ابتدا اعداد را در بازه زمانی ۱۵ میلی‌ثانیه‌ای، که دقت مورد نیاز را فراهم می‌کند، گسسته‌سازی می‌کنیم. فرض کنید x عدد به‌دست آمده پس از گسسته‌سازی در بازه ۱۵ میلی‌ثانیه‌ای باشد. به ازای هر داده، دو نورون در نظر می‌گیریم: نورون اول در لحظه x و نورون دوم در لحظه $x - ۱۵$ فعال می‌شوند. انتظار می‌رود نورون اول به عنوان نورون تحریکی و نورون دوم به عنوان نورون مهاری عمل کنند.

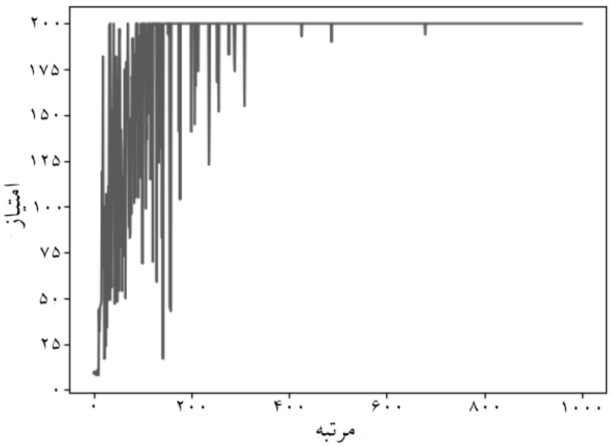
داده مربوط به فعالیت عامل خبره برای فرایند آینه، در میلی‌ثانیه هفتم به شبکه عصبی عامل هوشمند می‌رسد. به عبارت دیگر، نورون خروجی متناسب با فعالیت عامل خبره، در میلی‌ثانیه هفتم وادار به ضربه می‌شود. سیستم پاداش و جزا در عامل هوشمند بر اساس موفقیت یا عدم موفقیت عامل خبره تعریف می‌شود. مادامی‌که عامل خبره در حال تلاش برای برقراری تعادل است، میزان دوپامین ثابتی در شبکه ترشح می‌شود و از آنجایی که خروجی شبکه بر اساس رفتار عامل خبره است، به محض شکست او، ترشح دوپامین متوقف می‌شود. به بیان دیگر، عامل هوشمند تلاش‌های عامل خبره را مفید می‌داند و به محض این‌که شکست او را می‌بیند، متوجه می‌شود که رفتارهای عامل خبره نتیجه مطلوب نداشته و لذا دیگر دوپامینی ترشح نمی‌شود.

در آزمایش اول، که عامل خبره خود در حال یادگیری است، عامل هوشمند نیز در اوایل به صورت تصادفی عمل می‌کند و به مرور که انتخاب‌های تصادفی عامل خبره کاهش می‌یابد، عامل هوشمند به هدف عامل خبره آگاه می‌شود و رفتار او را در جهت تحقق هدف می‌آموزد. همچنین، در آزمایش دوم که عامل خبره همواره بهترین عملکرد را دارد، عامل هوشمند به خوبی هدف و رفتار مناسب را یاد می‌گیرد. به عنوان حالت کنترل نیز، حالتی را در نظر گرفتیم که عامل خبره کاملاً به صورت تصادفی عمل کند. همانطور که انتظار می‌رفت، عامل هوشمند در این حالت توانایی درک هدف را نداشته و او نیز به صورت تصادفی عمل می‌کند.

هر دو آزمایش دارای نوعی پیچیدگی هستند. در آزمایش اول، عامل هوشمند باید تشخیص دهد که کدامیک از فعالیت‌های عامل خبره در جهت دستیابی او به هدفش صورت گرفته است. با کاهش انتخاب‌های تصادفی عامل خبره، عامل هوشمند متوجه می‌شود که در چه شرایطی او با شکست مواجه می‌شود. همچنین عامل هوشمند می‌تواند رفتارهای هوشمندانه را از رفتارهای تصادفی که صرفاً به هدف شناخت محیط صورت گرفته‌اند، تفکیک کند. در این حالت، بعد از حدود ۴۰ تا ۵۰ مرتبه، عامل هوشمند به بهترین عملکرد خود دست می‌یابد. در آزمایش دوم، از آنجایی که عامل هوشمند همواره عامل خبره را در بهترین عملکرد ممکن می‌بیند، دیرتر می‌تواند به درک شرایط شکست برسد. در حقیقت، او متوجه می‌شود که این رفتارها مناسب هستند و باید هر رفتار دیگر را در خود مهار کند. با این حال، همچنان این احتمال وجود دارد که با توجه به برخی مشاهدات آتی، خود را در شرایطی گرفتار کند که کمتر وضعیتی مشابه آن را پیش‌تر مشاهده کرده باشد و لذا نتواند به خوبی تعادل را برقرار کند و امتیاز کمتری کسب کند. عملکرد عامل هوشمند در دو آزمایش، امتیاز مورد نیاز برای داشتن بهترین عملکرد، و عملکرد عامل خبره در آزمایش اول در شکل ۵ قابل مشاهده است.



شکل ۵: نمودار عملکرد در آزمایش‌های مختلف در هر مرتبه. نمودار خط ممتد، عملکرد عامل خبره در آزمایش اول و نمودار نقطه‌چین و خط-نقطه، به ترتیب، متوسط امتیاز کسب شده توسط عامل هوشمند در آزمایش اول و دوم پس از هر ۱۰ مرتبه را نشان می‌دهند. خط‌چین نشان‌دهنده آستانه امتیاز قابل قبول است.



شکل ۶: عملکرد عامل هوشمندی که با آزمون و خطا در محیط آونگ واژگون به کمک روش یادگیری Q ، نحوه تعامل با محیط را یاد می‌گیرد [۲۵].

۵ بحث و بررسی

یکی از مزیت‌های مدل پیشنهاد شده، تعداد محدود مرتبه‌ها جهت یادگیری می‌باشد. همانطور که دیدیم، این مدل قادر است در کمتر از ۱۰۰ مرتبه بهترین عملکرد را بیاموزد. این در حالی است که مدلی از شبکه‌های عصبی کلاسیک، با بهره‌گیری از روش یادگیری تقویتی Q و تنها بر اساس آزمون و خطای عامل هوشمند در محیط، به بیش از ۳۰۰ مرتبه آموزش نیاز دارد [۲۵]. عملکرد این عامل هوشمند در شکل ۶ قابل مشاهده است.

محیط آونگ واژگون محیطی ساده است و شاید بهترین مثال برای بررسی مدل محاسباتی نظریه ذهن نباشد. با این حال، تعمیم مدل به محیط‌های پیچیده‌تر تغییراتی بنیادین را می‌طلبد. طراحی سازوکار مهاری مناسب در زمان یادگیری، پردازش دنباله فعالیت‌های صورت گرفته در محیط، کدگذاری مناسب محرک‌های ورودی، بازتعریف سازوکار نورون‌های آینه‌ای به منظور پوشش فعالیت‌های پیچیده‌تر، و همچنین، تعریف مناسب سیستم پاداش و جزای انتزاعی برای عامل هوشمند از جمله مواردی هستند که باید در طراحی مدل نهایی لحاظ شوند. تعریف صحیح موارد ذکر شده موجب می‌شوند که نه تنها حساسیت مدل نسبت به تغییر محیط عمل کمتر باشد، بلکه بتواند با تعداد محدودی مشاهده، چگونگی تعامل با محیط را بیاموزد.

یکی از سؤالاتی که در اینجا ممکن است مطرح شود این است که آیا مدل ارائه شده همان روش یادگیری باناظر نیست؟ پاسخ منفی است، زیرا افعال درست و غلط در یادگیری باناظر مشخص است درحالی‌که در اینجا، هیچ دانشی از درست یا غلط بودن رفتار عامل خبره در دست نیست. سپس ممکن است مطرح شود که این روش دقیقاً همان روش یادگیری تقویتی است و مدل ارائه شده قابل رقابت با مدل‌های مشهور در ادبیات یادگیری تقویتی نخواهد بود. در پاسخ باید گفت که در مدل‌های مطرح شده بر اساس یادگیری تقویتی، عامل خبره‌ای وجود ندارد؛ بلکه عامل هوشمند خود با آزمون و خطا در محیط آموزش می‌بیند و لذا نیاز است سیستم پاداش و جزا مهندسی شود. یکی از مزیت‌های مدل‌سازی نظریه ذهن، از بین رفتن همین نیاز به مهندسی‌سازی تابع پاداش است.

۶ نتیجه‌گیری

نظریه ذهن و توانایی‌های مرتبط با آن، تأثیر بسزایی در یادگیری و تعامل ما انسان‌ها با محیط اطرافمان دارد. مدل‌سازی محاسباتی آن می‌تواند در جهت شناخت مجهولات و نیز توانمندسازی عامل‌های هوشمند به‌کار گرفته شود. تا کنون مدل‌های محاسباتی متنوعی برای پاسخ به نیازهای مطرح شده ارائه شده‌اند. با این حال، همچنان نیاز به بهبود و تغییر نگرش در توسعه مدل‌های محاسباتی نظریه ذهن حس می‌شود. به کمک شواهد موجود از سیستم نورون‌های آینه‌ای و در نتیجه آن، دیدگاه شبیه‌سازی-نظریه، می‌توان مدلی محاسباتی مبتنی بر شبکه‌های عصبی ضربه‌ای ارائه کرد. هدف از ارائه چنین مدلی، افزایش توجیه‌پذیری زیستی و روان‌شناختی، مد نظر قرار دادن دیدگاه نوین‌تر شبیه‌سازی-

نظریه در مطالعات محاسباتی، و نیز بهبود یادگیری و عملکرد سیستم‌های هوشمند در محیط‌های تعاملی است. در این پژوهش، با ارائه مدلی ساده از یک عامل هوشمند در یک محیط ساده تعاملی، نشان دادیم که طراحی چنین مدلی امکان‌پذیر است و می‌تواند تأثیر بسزایی در بهبود فرآیند آموزش عامل‌های هوشمند و عملکرد آن‌ها داشته باشد.

مراجع

[1] M. A. Sabbagh and L. C. Bowman, *Theory of Mind*, pp. 1–39. American Cancer Society, 2018.

[2] F. Cuzzolin, A. Morelli, B. Cîrstea, and B. J. Sahakian, "Knowing me, knowing you: theory of mind in ai," *Psychological Medicine*, vol. 50, no. 7, p. 1057–1061, 2020.

[3] K. Shanton and A. Goldman, "Simulation theory," *WIREs Cognitive Science*, vol. 1, no. 4, pp. 527–538, 2010.

[4] V. Gallese and A. Goldman, "Mirror neurons and the simulation theory of mind-reading," *Trends in Cognitive Sciences*, vol. 2, no. 12, pp. 493 – 501, 1998.

[5] D. Premack and G. Woodruff, "Does the chimpanzee have a theory of mind?," *Behavioral and Brain Sciences*, vol. 1, no. 4, p. 515–526, 1978.

[6] C. Catmur, C. Press, and C. Heyes, "Mirror neurons from associative learning," *The Wiley Handbook on the Cognitive Neuroscience of Learning*, pp. 515–537, 2016.

[7] "Mirror neurons in action.." `https://sitn.hms.harvard.edu/flash/2016/mirror-neurons-quarter-century-new-light-new-cracks/`.

[8] A. Eren, "Exploring the relationships among mirror neurons, theory of mind, and achievement goals: Towards a model of achievement goal contagion in educational settings," *Educational Research Review*, vol. 4, no. 3, pp. 233–247, 2009.

[9] T. Rusch, S. Steixner-Kumar, P. Doshi, M. Spezio, and J. Gläscher, "Theory of mind and decision science: Towards a typology of tasks and computational models," *Neuropsychologia*, vol. 146, p. 107488, 2020.

[10] C. L. Baker and J. B. Tenenbaum, "Chapter 7 - modeling human plan recognition using bayesian theory of mind," in *Plan, Activity, and Intent Recognition* (G. Sukthankar, C. Geib, H. H. Bui, D. V. Pynadath, and R. P. Goldman, eds.), pp. 177 – 204, Boston: Morgan Kaufmann, 2014.

[11] J. B. Hamrick, "Analogues of mental simulation and imagination in deep learning," *Current Opinion in Behavioral Sciences*, vol. 29, pp. 8 – 16, 2019. SI: 29: Artificial Intelligence (2019).

[12] J. Jara-Ettinger, "Theory of mind as inverse reinforcement learning," *Current Opinion in Behavioral Sciences*, vol. 29, pp. 105 – 110, 2019. SI: 29: Artificial Intelligence (2019).

[13] R. el Kaliouby and P. Robinson, "Generalization of a vision-based computational model of mind-reading," in *Affective Computing and Intelligent Interaction* (J. Tao, T. Tan, and R. W. Picard, eds.), (Berlin, Heidelberg), pp. 582–589, Springer Berlin Heidelberg, 2005.

[14] J. M. Kilner, K. J. Friston, and C. D. Frith, "The mirror-neuron system: a bayesian perspective," *Neuroreport*, vol. 18, no. 6, pp. 619–623, 2007.

[15] A. Gopnik and H. M. Wellman, "Reconstructing constructivism: Causal models, bayesian learning mechanisms, and the theory theory," *Psychological Bulletin*, vol. 138, pp. 1085 – 1108, 2012.

[16] N. Rabinowitz, F. Perbet, F. Song, C. Zhang, S. M. A. Eslami, and M. Botvinick, "Machine theory of mind," in *Proceedings of the 35th International Conference on Machine Learning* (J. Dy and A. Krause, eds.), vol. 80 of *Proceedings of Machine Learning Research*, (Stockholmsmässan, Stockholm Sweden), pp. 4218–4227, PMLR, 10–15 Jul 2018.

[17] A. F. T. Winfield, "Experiments in artificial theory of mind: From safety to story-telling," *Frontiers in Robotics and AI*, vol. 5, p. 75, 2018.

[18] Y. Zeng, Y. Zhao, T. Zhang, D. Zhao, F. Zhao, and E. Lu, "A brain-inspired model of theory of mind," *Frontiers in Neurorobotics*, vol. 14, p. 60, 2020.

[19] J. Schrittwieser, I. Antonoglou, T. Hubert, K. Simonyan, L. Sifre, S. Schmitt, A. Guez, E. Lockhart, D. Hassabis, T. Graepel, *et al.*, "Mastering atari, go, chess and shogi by planning with a learned model," *Nature*, vol. 588, no. 7839, pp. 604–609, 2020.

[20] C. Keysers and V. Gazzola, "Hebbian learning and predictive mirror neurons for actions, sensations and emotions," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 369, no. 1644, p. 20130175, 2014.

[21] W. Gerstner, W. M. Kistler, R. Naud, and L. Paninski, *Neuronal dynamics: From single neurons to networks and models of cognition*. Cambridge University Press, 2014.

[22] A. N. Burkitt, "A review of the integrate-and-fire neuron model: I. homogeneous synaptic input," *Biological cybernetics*, vol. 95, no. 1, pp. 1–19, 2006.

[23] J. Sjöström and W. Gerstner, "Spike-timing dependent plasticity," *Spike-timing dependent plasticity*, vol. 35, no. 0, pp. 0–0, 2010.

[24] E. M. Izhikevich, "Solving the Distal Reward Problem through Linkage of STDP and Dopamine Signaling," *Cerebral Cortex*, vol. 17, pp. 2443–2452, 01 2007.

[25] Y. H. Liu, *PyTorch 1.x Reinforcement Learning Cookbook: Over 60 recipes to design, develop, and deploy self-learning AI models using Python*. Packt Publishing, 2019.