



بهبود یادگیری مشارکتی در سیستمهای چندعاملی

Cooperative learning improvement in multi-agent systems

داریوش حسن پور آده

۹۳۰۸۱۶۴

استاد راهنما: دکتر پالهنک

خلاصه

اهمیت مشارکت و کار جمعی را نمی‌توان نادیده گرفت و می‌توان گفت راز بزرگ موفقیت انسان و حیوانات همین اجتماعی بودن است. محققان هوش مصنوعی که همیشه دنبال رسیدن به یک هوشمندی در ماشین بوده‌اند، در این راه از انسان و حیوانات الگوبرداری‌های زیادی داشته‌اند که یکی از این الگوبرداری‌ها شاخه سیستم‌های چندعاملی را در هوش مصنوعی به وجود آورده است. مثل جوامع انسانی و حیوانی که گاهی مشارکتی و گاهی رقابتی به زندگی می‌پردازند، سیستم‌های چندعاملی نیز در زیرشاخه‌های مشارکتی و رقابتی مورد پژوهش قرار می‌گیرند. در این پژوهش که با هدف بهبود یادگیری مشارکتی در سیستم‌های چندعاملی انجام خواهد شد تکیه بر روی مشارکت و مهمتر از آن یادگیری مشارکتی در سیستم‌های چندعاملی است. باوجود عمر کمی که سیستم‌های چندعاملی و مخصوصاً یادگیری مشارکتی دارد پژوهش‌های فراوانی در این زمینه صورت گرفته که می‌توان شروع آن را با [۱، ۲] دانست که با هدف افزایش سرعت یادگیری در یادگیری تقویتی ارائه شد. بعد از آن‌هم در تکمیل این کار پژوهش‌هایی صورت گرفته که معمولاً تقلید از یادگیری انسان است. ارائه ایده‌هایی چون تقلید، پند دهی، خبرگی و تخته‌سیاه که مفید هم بوده‌اند. نکته مهمی که در کارهای گذشته وجود دارد تخصیص یک وزن برای تمام جدول است در صورتی که ممکن است قسمتی

از جدولی که وزن بالایی می‌گیرد به درستی مقدار نگرفته باشد و باعث یادگیری نادرست شود. برای حل این مشکل می‌توان برای هر سلول وزن در نظر گرفت که به دلیل مشکل بودن محاسبه وزن زمان‌گیر خواهد شد. راه‌حل دیگر یک تقسیم‌بندی بهینه از جدول است. هدف نهایی این پژوهش نیز رسیدن به یک تقسیم‌بندی بهینه و روشی برای محاسبه وزن هر بخش جدول است. در پایان آزمایشهایی انجام خواهد شد که نشان‌دهنده نکات مثبت و منفی کار باشد. آزمایش‌ها در دو فاز انجام خواهد گرفت. فاز اول که مقایسه بین این روش و روشهای مشابه است و فاز دوم آزمایشهای انفرادی، که نشان‌دهنده مثبت و منفی بودن کار است.

کلمات کلیدی: ۱ - سیستم‌های چندعاملی ۲ - یادگیری مشارکتی ۳ - یادگیری تقویتی

موضوع کلی و زمینه اصلی تحقیق

یادگیری یکی از بزرگترین دغدغه‌های رشته‌های مختلفی چون روانشناسی، هوش مصنوعی و ... است که در هوش مصنوعی منجر به شاخه یادگیری ماشین شده است. میشل در [۳] تعریفی برای یادگیری آورده است.

تعریف ۱: یادگیری یعنی اینکه یک عامل رفتارش را بر اساس تجربیات گذشته‌اش، عوض کند. معمولاً در تحقیقات یادگیری ماشین از یادگیری انسان و حیوانات الگوبرداری میشود که این تقلیدها نتایج خوبی به دنبال داشته است؛ یکی از بزرگترین الگوهایی که می‌توان از زندگی انسان و حیوانات برداشت زندگی گروهی و جمعی آنهاست که گاه به مشارکت و گاه به رقابت می‌انجامد. در این پژوهش تمرکز بر روی مشارکت در زندگی جمعی است که ما را به مفهوم یادگیری مشارکتی میرساند. در سالهای اخیر تحقیقات زیادی در زمینه یادگیری مشارکتی انجام گرفته است در [۴] تعریفی برای سیستم چندعاملی مشارکتی ارائه شده است.

تعریف ۲: سیستمی که در آن چند عامل برای رسیدن به یک هدف و یا انجام یک وظیفه‌ی مشترک باهم همکاری میکنند، سیستم چندعاملی مشارکتی نامیده میشود.

همچنین اثبات شده که یادگیری مشارکتی از سرعت و قدرت بالاتری نسبت به یادگیری انفرادی برخوردار است. معمولاً در یادگیری مشارکتی از یادگیری تقویتی و معروف‌ترین عضو این خانواده یعنی یادگیری Q استفاده میشود و نتایج نیز قابل توجه بوده و نشان می‌دهد که این ترکیب می‌تواند یک روش یادگیری قدرتمند باشد. می‌توان بزرگترین دغدغه‌های این شاخه را به صورت سؤالاتی مطرح نمود.

۱. چه زمان باید اطلاعات عامل‌ها منتقل شود؟

۲. اطلاعات باید به کدام عامل فرستاده شود و یا از کدام عامل دریافت شود؟

۳. چه میزان اطلاعات باید ارسال شود؟

۴. چگونه این اطلاعات با اطلاعات خود عامل ترکیب شود؟

رسیدن به پاسخ سؤالات بالا می‌تواند یادگیری را تا حد زیادی بهینه نماید. در اکثر کارهایی که انجام شده به موضوع نحوه ترکیب بحث شده و معمولاً کل جدول ارسال می‌شود.

مروری بر کارها و نتایج گذشته

با وجود اینکه یادگیری مشارکتی زمینه نسبتاً جدیدی در یادگیری ماشین محسوب میشود کارهای فراوانی در این زمینه انجام شده است و همانگونه که گفته شد معمولاً الگوبرداری از رفتار انسانها یا حیوانات بوده است. می توان گفت اولین تلاش در زمینه یادگیری مشارکتی برمیگردد به [۱، ۲] که مکانیزمی برای افزایش سرعت یادگیری تقویتی با مشارکت بین عاملها ارائه نمودند. بعد از آن تان در [۵] سه روش انتقال برای یادگیری مشارکتی مطرح نموده و به مقایسه عملکرد چند عامل هنگام استفاده از یادگیری مشارکتی و بدون استفاده از یادگیری مشارکتی پرداخته که نشان میدهد در صورتیکه مشارکت به درستی پیاده سازی شود می تواند برای کلیه عاملها مفید باشد. روشی که در این مقاله استفاده شده بر اساس میانگین گیری از جداول Q عاملها است که نام روش را نیز SA^۱ نهاده است. بعد از آن برنجی و همکاران در [۶] به مزایای روش یادگیری مشارکتی پرداخته است.

برنجی در مقاله بعد خود نیز مفهوم یادگیری مشترک را مطرح نمودند که در آن سیاست بین عاملها مشترک بوده و عاملها می توانند آن را بروز رسانی نمایند [۷] سپس توافن در [۸] به تقلید که یکی از ابزارهای یادگیری انسان است اشاره نموده است. بعد از آن یادگیری جمعی که آن نیز برگرفته از انسان است توسط گارلند در [۹] مطرح شد که با الهام از ایده شناخت توزیع شده در علوم اجتماعی شکل گرفته است. سپس ایده پند دادن عاملهای باتجربه و پند گرفتن عاملهای تازه کار نونس در [۱۰] آمد؛ و بعد احمدآبادی یادگیری بر مبنای خبرگی را در [۱۱] ارائه نمودند، بعد از آن نیز یادگیری تخته سیاه در [۱۲، ۱۳] ارائه شده است. در سال ۱۳۹۲ خانم پاکیزه با در نظر گرفتن چندین معیار (مانند میزان پاداش جمع شده، میزان جریمه های اخذ شده و غیره توسط عامل) برای عوامل در مرحله ادغام دانش این عوامل، نشان دادند که این معیارها می توانند به صورت ضرایب ترکیب موثر برای ترکیب دانش عاملها استفاده شوند [۱۴]. چند سال بعد در سال ۱۳۹۵ آقای میرزایی با معرفی معیار کوتاه ترین مسیر تجربه شده یادگیری مشترک را تسریع بخشیدند و نشان دادند که این معیار کوتاه ترین مسیر تجربه شده مزایای بیشتری نسبت به معیارهای خانم پاکیزه دارد و می تواند نتایج بهتری بدست بدهد [۱۵].

ضرورت انجام، دیدگاه و اهداف تحقیق پیشنهادی

معمولاً یادگیری تقویتی را با یادگیری Q می شناسیم که مهمترین عضو خانواده یادگیری تقویتی است. اصلی ترین رکن در یادگیری Q جدولی با همین نام است که یادگیر باید بتواند در طول یادگیری به درستی این جدول را تکمیل نماید و معمولاً روشهایی که بر مبنای یادگیری تقویتی ارائه شده با هدف تسریع در بروز رسانی این جدول بوده است. تقریباً می توان گفت دلیل به وجود آمدن یادگیری تقویتی نیز همین بود و با این ایده که عاملها بتوانند با مشارکت زودتر جدول خود را کامل کنند شروع به کار نمود اما چالشهایی وجود داشته که باعث تولید ایده های فراوانی در این زمینه شده است.

روش هایی که تا به حال ارائه شده اند برای تمام جدول ضریبی تخصیص می دهند. در نهایت مجموع این

¹Simple Averaging

جداول با ضرایبی که جمع یک دارند محاسبه می‌شود. معمولاً روش‌های ارائه شده برای تولید همین ضرایب بوده است؛ اما اینکه برای تمام جدول یک مقدار در نظر گرفته شود می‌تواند گاهی نتیجه منفی داشته و مقادیر جدول را نادرست نماید. همچنین در نظر گرفتن و محاسبه‌ی معیاری برای تک‌تک موقعیت/عمل‌ها می‌تواند هزینه‌ی زمانی و مکانی زیادی به سیستم یادگیرنده تحمیل کند.

تحقیقات خانم پاکیزه و آقای میرزایی [۱۴، ۱۵] یک مبادله‌ای بین محاسبه‌ی ضرایب تاثیر کلی و جزئی در نظر نگرفته‌اند، بدین معنی که یا از اطلاعات کلی عامل‌ها برای ارائه‌ی یک معیاری استفاده می‌کنند یا به محاسبه‌ی معیاری برای تک‌تک حالت/عمل‌ها می‌پردازند، همچنین در هنگام ترکیب دانش خاصیت غیرافزایشی بودن دانش‌های عوامل را که در ماهیت مساله وجود در نظر نگرفته‌اند [۱۶]. همچنین تاثیر استفاده از دیگر روش‌های اکتشاف چون اپسیلون-حریصانه^۱ بررسی نشده است. همچنین ایراد دیگری که می‌توان به روش کوتاه‌ترین مسیر تجربه شده گرفت این است که ممکن است موقعیتی در گلوگاهی از امتیازهای منفی قرار بگیرد ولی یک واحد با موقعیت هدف فاصله دارد، حال در هنگام ترکیب دانش‌ها کوتاه‌ترین مسیر این موقعیت بیشترین تاثیر را برای جلب تمایل دیگر عوامل به گذر از این موقعیت خطرناک بکند در حالی که ممکن است احتمال موفقیت در هنگام گذر از این موقعیت ۲۰٪ یا کمتر باشد؛ لذا کوتاه‌ترین مسیر تجربه شده در این مثال لزوماً بهترین معیار برای ضریب ترکیب دانش عوامل نیست.

آنچه که در این پژوهش در پی آن هستیم محاسبه‌ی نرمی از معیارها می‌باشد که تعادلی بین کلی و جزئی نگری در هنگام ارزیابی و همچنین تاثیر دیگر روش‌های انتخاب عمل چون روش اپسیلون-حریصانه در افزایش بهره‌وری یادگیری مشارکتی در عوامل را بررسی خواهیم کرد. در صورت ایجاب وقت به مرحله‌ی ترکیب دانش‌های عوامل از دیدگاه ماهیت غیرافزایشی اطلاعات دانش‌ها نگاه خواهیم کرد و در صورت امکان روشی برای در نظر گرفتن این خاصیت مساله ارائه خواهیم داد.

نحوه ارزیابی دستاوردهای تحقیق

برای آزمایش روشهایی که در یادگیری مشارکتی ارائه می‌شود محیط‌های زیادی وجود دارد که دو مورد از مهمترین آن‌ها پلکان «مارپیچ» و «صید و صیاد» است که هر یک بر معیارهایی تأکید دارند. پلکان مارپیچ محیطی ساده و ایستا است و صید و صیاد یک محیط پیچیده و پویا است. خود آزمایش‌ها نیز به دو دسته مقایسه با روشهای قبلی و آزمونهای انفرادی تقسیم خواهد شد. تا هم بتوان به مقایسه با روشهای قبلی پرداخت و هم نکات مثبت کار را نمایش داد.

¹ ϵ -greedy

جدول زمانبندی تحقیق

فعالیت	زمان شروع	زمان خاتمه
مطالعه مقالات و شناخت کامل موضوع	–	اواسط آبان ۹۵
پیاده‌سازی	اواسط آبان ۹۵	اواخر آذر ۹۵
تجزیه تحلیل و آزمایش	اواخر آذر ۹۵	اواخر دی ۹۵
نگارش پایان‌نامه و آمادگی جهت دفاع	اواخر دی ۹۵	اواخر اسفند ۹۵
دفاع	فروردین ۹۶	

فهرست مراجع اصلی

- [1] S. D. Whitehead, "A complexity analysis of cooperative mechanisms in reinforcement learning," in *AAAI*, pp. 607–613, 1991.
- [2] S. D. Whitehead and D. H. Ballard, *A study of cooperative mechanisms for faster reinforcement learning*. University of Rochester, Department of Computer Science Rochester, NY, 1991.
- [3] T. M. Mitchell, "Machine learning. 1997," *Burr Ridge, IL: McGraw Hill*, vol. 45, p. 37, 1997.
- [4] L. Panait and S. Luke, "Cooperative multi-agent learning: The state of the art," *Autonomous agents and multi-agent systems*, vol. 11, no. 3, pp. 387–434, 2005.
- [5] M. Tan, "Multi-agent reinforcement learning: Independent vs. cooperative agents," in *Proceedings of the tenth international conference on machine learning*, pp. 330–337, 1993.
- [6] H. R. Berenji and D. Vengerov, "Cooperation and coordination between fuzzy reinforcement learning agents in continuous state partially observable markov decision processes," in *Fuzzy Systems Conference Proceedings, 1999. FUZZ-IEEE'99. 1999 IEEE International*, vol. 2, pp. 621–627, IEEE, 1999.
- [7] H. R. Berenji and D. Vengerov, "Advantages of cooperation between reinforcement learning agents in difficult stochastic problems," in *Fuzzy Systems, 2000. FUZZ IEEE 2000. The Ninth IEEE International Conference on*, vol. 2, pp. 871–876, IEEE, 2000.
- [8] C. Trevarthen, "Learning about ourselves, from children: Why a growing human brain needs interesting companions?," *Annual Report-Hokkaido University Research and Clinical Center for Child Development*, pp. 9–44, 2004.
- [9] A. Garland and R. Alterman, "Multiagent learning through collective memory," in *Adaptation, Coevolution and Learning in Multiagent Systems: Papers from the 1996 AAAI Spring Symposium*, pp. 33–38, 1996.
- [10] L. Nunes and E. Oliveira, "Cooperative learning using advice exchange," in *Adaptive agents and multi-agent systems*, pp. 33–48, Springer, 2003.
- [11] M. N. Ahmadabadi, M. Asadpur, S. H. Khodanbakhsh, and E. Nakano, "Expertness measuring in cooperative learning," in *Intelligent Robots and Systems, 2000.(IROS 2000). Proceedings. 2000 IEEE/RSJ International Conference on*, vol. 3, pp. 2261–2267, IEEE, 2000.

- [12] N. Carver and V. Lesser, “Evolution of blackboard control architectures,” *Expert systems with applications*, vol. 7, no. 1, pp. 1–30, 1994.
- [13] J. W. McManus and W. L. Bynum, “Design and analysis techniques for concurrent blackboard systems,” *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 26, no. 6, pp. 669–680, 1996.
- [14] E. Pakizeh, M. Palhang, and M. M. Pedram, “Multi-criteria expertness based cooperative q-learning,” *Applied intelligence*, vol. 39, no. 1, pp. 28–40, 2013.
- [15] M. ali mirzaei badizi, “Speed-up cooperative learning in multi-agent systems using shortest experimented path,” Master’s thesis, Department of Electrical and Computer Engineering, Isfahan University of Technology, Isfahan University of Technology, Isfahan 84156-83111, Iran, 3 2015.
- [16] V. Torra, Y. Narukawa, and M. Sugeno, *Non-Additive Measures*. Springer, 2014.