



بهبود یادگیری مشارکتی در سیستمهای چندعاملی

Cooperative learning improvement in  
multi-agent systems

داریوش حسن پور آده

۹۳۰۸۱۶۴

استاد راهنما: دکتر پالهنک

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.

پژوهش‌های زیادی در ترکیب یادگیری مشارکتی با یادگیری تقویتی صورت گرفته است، یادگیری تقویتی معمولاً میزان پاداش تجمعی تجربه شده برای تمامی عمل-حالت‌ها در یک جدولی جدول ذخیره می‌کند و در نهایت از مقادیر داخل این جدول برای استخراج سیاستی که توصیف‌کننده‌ی نحوه‌ی عمل کرد عامل برای تعامل با محیط استفاده می‌شود. از جمله کارهایی که در گذشته برای ترکیب یادگیری مشارکتی و یادگیری تقویتی انجام شده است ارائه ایده‌هایی چون تقلید، پند دهی، خبرگی و تخته‌سیاه می‌باشد که مفید هم بوده‌اند. نکته مهمی که در کارهای گذشته وجود دارد تخصیص یک وزن برای تمام جدول یادگیری تقویتی 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 اثر 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: در تعریف یادگیری گفته میشود که یک برنامه کامپیوتری از تجربه  $E$  با توجه به وظیفه  $T$  و معیار کارایی  $P$  یاد می‌گیرد، اگر کارایی آن بر روی  $T$  که با  $P$  اندازه‌گیری میشود با تجربه  $E$  افزایش یابد [۳].

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 [۵]. معمولاً در یادگیری مشارکتی از یادگیری تقویتی و معروف‌ترین عضو این خانواده یعنی یادگیری  $Q$  استفاده می‌شود و نتایج نیز قابل‌توجه بوده و نشان می‌دهد که این ترکیب می‌تواند یک روش یادگیری قدرتمند باشد. می‌توان بزرگترین دغدغه‌های این شاخه را به‌صورت 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 توجه شده است.

## 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. می‌توان 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 [۶] 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 به‌درستی پیاده‌سازی شود می‌تواند برای کلیه عاملها مفید باشد. روشی که در این مقاله استفاده شده بر اساس میانگین‌گیری از جداول  $Q$  عاملها است که نام روش را نیز SA<sup>۱</sup> نهاده است. بعد از آن برنجی و همکاران در [۷] 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 می‌توانند آن را بروز رسانی نمایند [۸] سپس تووافن در [۹] به تقلید که یکی از

<sup>1</sup>Simple Averaging

ابزارهای یادگیری انسان است اشاره نموده است. بعد از آن یادگیری جمعی که آن نیز برگرفته از انسان است توسط گارلند در [۱۰] 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 [۱۱] آمد؛ و بعد احمدآبادی یادگیری بر مبنای خبرگی را در [۱۲] ارائه نمودند، بعد از آن نیز یادگیری تخته‌سیاه در [۱۳، ۱۴] ارائه شده است. در سال ۱۳۹۲ (۲۰۱۳ م.) پاکیزه و همکاران با در نظر گرفتن چندین معیار (مانند میزان پاداش جمع شده، میزان جریمه‌های اخذ شده و غیره توسط عامل) برای عوامل در مرحله ادغام دانش این عوامل، نشان دادند که این معیارها می‌توانند به صورت ضرایب ترکیب موثر برای ترکیب دانش عامل‌ها استفاده شوند [۱۵]. چند سال بعد در سال ۱۳۹۵ (۲۰۱۶ م.) میرزایی با معرفی معیار کوتاه‌ترین مسیر تجربه شده یادگیری مشترک را تسریع بخشیدند و نشان دادند که این معیار کوتاه‌ترین مسیر تجربه شده مزایای بیشتری نسبت به معیارهای پاکیزه دارد و می‌تواند نتایج بهتری بدست بدهد [۱۶].

### ضرورت انجام، دیدگاه و اهداف تحقیق پیشنهادی

معمولاً یادگیری تقویتی را با یادگیری Q می‌شناسیم که مهمترین عضو خانواده یادگیری تقویتی است. اصلی‌ترین 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 می‌توان 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 ایده‌های فراوانی در این زمینه شده است.

روش‌هایی که تا به حال ارائه شده‌اند برای تمام جدول ضریبی تخصیص می‌دهند. در نهایت مجموع این جداول با ضرایبی که جمع یک دارند محاسبه می‌شود. معمولاً روش‌های ارائه شده برای تولید همین ضرایب بوده است؛ اما به نظر می‌رسد 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 پاکیزه و میرزایی [۱۵، ۱۶] 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 دارد را در نظر نگرفته‌اند [۱۷]. همچنین تاثیر استفاده از دیگر روش‌های اکتشاف چون اپسیلون-حریصانه<sup>۱</sup> بررسی نشده است.

در این پژوهش به بررسی امکان ارائه‌ی محاسباتی نرم برای معیارها خواهیم پرداخت، با این هدف که سعی شود تعادلی بین کلی و جزئی نگری به عملکرد عامل‌ها در هنگام ادغام دانش‌های آن‌ها برقرار شود. همچنین تاثیر دیگر روش‌های انتخاب عمل را در ترکیب با معیارهای ارائه شده را مورد بررسی قرار می‌دهیم

<sup>۱</sup>ε-greedy

و دستاوردهای این پژوهش را با در نظر گرفتن ماهیت غیرافزایشی بودن ذات مساله ارائه خواهیم داد.

### نحوه ارزیابی دستاوردهای تحقیق

برای آزمایش روشهایی که در یادگیری مشارکتی ارائه می شود 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.

### جدول زمانبندی تحقیق

فعالیت	زمان شروع	زمان خاتمه
مطالعه مقالات و شناخت کامل موضوع	-	اواسط آبان ۹۵
پیاده سازی	اواسط آبان ۹۵	اواخر آذر ۹۵
تجزیه تحلیل و آزمایش	اواخر آذر ۹۵	اواخر دی ۹۵
نگارش پایان نامه 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] R. E. Slavin, "Cooperative learning," *Review of educational research*, vol. 50, no. 2, pp. 315–342, 1980.
- [6] M. Tan, "Multi-agent reinforcement learning: Independent vs. cooperative agents," in *Proceedings of the tenth international conference on machine learning*, pp. 330–337, 1993.
- [7] 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.
- [8] 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.
  - [9] 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.
  - [10] 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.
  - [11] L. Nunes and E. Oliveira, “Cooperative learning using advice exchange,” in *Adaptive agents and multi-agent systems*, pp. 33–48, Springer, 2003.
  - [12] 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.
  - [13] N. Carver and V. Lesser, “Evolution of blackboard control architectures,” *Expert systems with applications*, vol. 7, no. 1, pp. 1–30, 1994.
  - [14] 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.
  - [15] 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.
  - [16] 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.
  - [17] V. Torra, Y. Narukawa, and M. Sugeno, *Non-Additive Measures*. Springer, 2014.