

یافتن استراتژی غالب در بازی‌های Blotto با استفاده از اتماتاهاي يادگير

فرناز ابطحی محمد رضا میبدی سعید شیری

دانشکده مهندسی کامپیوتر و فناوری اطلاعات

دانشگاه صنعتی امیرکبیر

تهران، ایران

shiry@aut.ac.ir mmeybodi@aut.ac.ir farnaz_abtahi@yahoo.com

چکیده: در این مقاله، رویکردی مبتنی بر اتماتاهاي يادگير برای یافتن استراتژی غالب در بازی‌های Blotto ارائه می‌گردد. این دسته از بازی‌های رقابتی که از اهمیت ویژه‌ای در تئوری بازی برخوردارند، از دو جهت قابل توجه می‌باشند. اولاً این بازی‌ها در دنیای واقعی برای مدل‌سازی فرآیندهایی به کار می‌روند که در آن‌ها به منظور غلبه بر حریف، نیاز به توزیع بهینه منابع محدود بین چندین کار وجود دارد. ثانیاً رویکردی که در این بازی‌ها برای یافتن استراتژی غالب مورد استفاده قرار می‌گیرد را می‌توان برای مدل کردن هر فرآیند چندعامله رقابتی دیگر که دارای استراتژی غالب برای هریک از عامل‌ها می‌باشد به کار برد. با دانستن استراتژی غالب می‌توان تضمین کرد که عامل، همواره سودی بیشتر یا مساوی با سایر عامل‌ها به دست خواهد آورد. در روش پیشنهادی در این مقاله، هریک از بازیکنان دارای یک اتماتای يادگير می‌باشد که از آن، برای يادگيری و تصمیم‌گیری در مورد انتخاب اعمال کمک می‌گیرد. آزمایشات انجام‌شده نشان می‌دهند که با استفاده از این روش، استراتژی بازیکنان مجذب به اتماتای يادگير به تدریج به استراتژی غالب همگرا شده و این بازیکنان قادر به یافتن بهترین حالت تقسیم منابع و برد در بازی خواهند بود.

کلمات کلیدی: بازی‌های Blotto، يادگيری تقویتی، اتماتاهاي يادگير، استراتژی غالب

Finding Dominant Strategies in Blotto Games utilizing Learning Automata

F. Abtahi M. R. Meybodi S. Shiry

Department of Computer Engineering and Information Technology

Amirkabir University of Technology

Tehran, Iran

Abstract: In this paper, a new approach based on Learning Automata is proposed for finding dominant strategies in Blotto Games. These games which are of a special importance in the Game Theory, are remarkable from two viewpoints; First, these games are used to model a class of real-world competitive problems in which optimal allocation of limited resources is needed to overcome the opponents. Second, the approach used for finding dominant strategies in Blotto Games is applicable in any Multi-Agent System that has dominant strategies for any of the agents. Knowing the dominant strategy, it could be guaranteed that the payoff achieved by the agent is greater or at least equal to all the opponents. In the suggested algorithm, each player of the Blotto Game has a Learning Automaton and receives help from it to select actions. Experiments show that by using this approach, the strategy of those players that are equipped with learning automata gradually converges to the dominant strategy. These players can eventually find the optimal resource allocation and are then able to dominate the opponents regardless of how they play.

Keywords: Blotto Games, Learning Automata, dominant strategy.

- ۱ مقدمه

در تئوری بازی، غلبه^۱ هنگامی روی می‌دهد که یک استراتژی برای یک بازیکن، بدون توجه به این‌که رقبا چگونه بازی کنند، از سایر استراتژی‌ها بهتر بوده و سود بیشتری را نصیب آن بازیکن نماید. این استراتژی، استراتژی غالب^۲ نامیده می‌شود [۳]. مفهوم غلبه و استراتژی غالب در بسیاری از سیستم‌های چندعامله رقابتی، بازی‌ها و مسائل دنیای واقعی به یافتن مناسب‌ترین راه حل مسئله کمک

¹ Dominance

² Dominant Strategy

می‌کند، زیرا در صورت وجود استراتژی غالب، بازیکن قادر است در تمام تعادل‌های *Nash*³ بازی، آن استراتژی را اجرا می‌نماید و اطمینان داشته باشد که همواره برنده خواهد بود و یا در بدترین شرایط نتیجه‌ای برابر با رقبای خود به دست خواهد آورد.

گروهی از بازی‌های رقابتی که در مبحث تئوری بازی مطرح بوده و با استفاده از آن‌ها به خوبی می‌توان مفهوم استراتژی غالب را مورد مطالعه قرار داد، بازی‌های *Blotto*⁴ می‌باشد [۱۱، ۹، ۲، ۱]. این بازی‌ها، دسته‌ای از بازی‌های ماتریسی⁵ جمع‌صفر⁶ را شامل می‌شوند که در آن، بازیکنان باید به طور همزمان منابع محدودی را بین چندین کار یا شیء توزیع نمایند و امتیاز هر بازیکن برابر مجموع امتیازی خواهد بود که از هر کار یا شیء به دست می‌آورد.

بازی‌های *Blotto* براساس یک شخصیت خیالی به نام *Colonel Blotto* نامگذاری شده‌اند که در جنگ وظیفه توزیع بهینه سربازان بین N میدان مبارزه را به‌عهده داشته است، با دانستن این‌که:

(۱) در هر میدان، هر کدام از طرفین جنگ که سرباز بیشتری به آن محل اختصاص داده است پیروز خواهد شد،

(۲) هیچ‌یک از دو طرف اطلاع ندارند که رقیب چه تعداد سرباز را به هر میدان اختصاص می‌دهد،

(۳) طرفی که در اکثر میدان‌های جنگ بر حريف پیروز شود، در کل برنده جنگ خواهد بود.

انواع مختلفی از بازی‌های *Blotto* وجود دارد. یکی از نمونه‌های ممکن که دارای دو بازیکن می‌باشد به این صورت است که هریک از بازیکن‌ها سه عدد را روی کاغذ می‌نویسنند، به‌طوری که مجموع این سه عدد برابر عدد معینی شود. این کار بدون اطلاع از اعدادی که طرف مقابل می‌نویسد صورت می‌گیرد. فرض می‌کنیم در این بازی، بازیکن‌ها مجاز به نوشتن جایگشت‌های مختلف سه عدد مشخص نمی‌باشند. مثلاً (۱، ۲، ۳) و (۱، ۳، ۲) دو انتخاب به حساب نمی‌آیند. برای محاسبه امتیاز هر بازیکن، در صورتی که بیش از یک عدد از سه عدد بازیکن، از یکی از سه عددی که حریف نوشته بزرگ‌تر باشد، یک امتیاز مثبت به بازیکن تعلق خواهد گرفت. در غیر این صورت اگر کمتر از دو عدد از اعداد بازیکن دارای این ویژگی باشد، بازیکن یک امتیاز منفی می‌گیرد. و در صورتی که تمام اعداد بازیکن با اعداد رقیب مساوی باشند، بازیکن هیچ امتیازی دریافت نمی‌کند. فرض کنیم مجموع سه عدد باید برابر ۶ شود. در این حالت با توجه به قانون بازی ماتریس نتیجه برای بازیکن اول به صورت زیر خواهد بود:

| | (۱، ۱، ۴) | (۱، ۲، ۳) | (۲، ۲، ۲) |
|-----------|-----------|-----------|-----------|
| (۱، ۱، ۴) | ۰ | -۱ | -۱ |
| (۱، ۲، ۳) | ۱ | ۰ | -۱ |
| (۲، ۲، ۲) | ۱ | ۱ | ۰ |

جدول ۱- ماتریس نتیجه برای بازیکن اول در یک نمونه بازی *Blotto*

در ماتریس بالا، سطرها اعمال بازیکن اول و ستون‌ها اعمال بازیکن دوم را نشان می‌دهند. از آن جایی که مجموع امتیازات دو بازیکن برابر با صفر است، ماتریس نتیجه بازیکن دوم نیز مشابه همین ماتریس می‌باشد، تنها با این تفاوت که علامت امتیازات برای بازیکن دوم معکوس خواهد شد. با توجه به ماتریس نتایج، بهترین استراتژی (تعادل *Nash*) در اینجا (۲، ۲، ۲) می‌باشد و انتخاب آن، همواره منجر به دریافت امتیاز بیشتر یا مساوی با حریف می‌گردد.

از بازی‌های *Blotto* برای مدل‌سازی برخی از مسائل دنیای واقعی استفاده شده است. برای مثال در [۱۱]، این بازی برای مدل کردن انتخابات سال ۲۰۰۰ آمریکا که یکی از نزدیک‌ترین رقابت‌ها در تاریخ انتخابات ریاست جمهوری این کشور بوده به کار رفته است. همچنین در [۱۲]، از این بازی برای در مبارزه با تروریسم استفاده شده و این کار با محاسبه احتمال موفقیت تروریست در حمله خود، از طریق مدل‌سازی حمله و دفاع توسط بازی *Blotto* صورت گرفته است.

در بازی بالا و یا در هر فرآیند رقابتی دیگر، بسیار مطلوب است که بتوان استراتژی‌های غالب را مشخص کرد. بدین منظور می‌توان از تکنیک‌های یادگیری استفاده نمود. در این‌گونه بازی‌ها، بازیکنان هم‌زمان در حال یادگیری می‌باشند و سود هر عامل، علاوه بر اعمال خود او به اعمال سایر عامل‌ها نیز وابسته است. . به همین دلیل، محیط بازی غیرقطعی بوده و مناسب‌ترین روش برای یادگیری، استفاده از تکنیک‌های مبتنی بر یادگیری تقویتی می‌باشد. در بازی‌های ماتریسی، یادگیری از طریق تکرار بازی به دفعات متعدد صورت می‌گیرد. در هر تکرار، هریک از بازیکنان عملی را انتخاب کرده و انجام می‌دهد. سپس با توجه به عمل خود و اعمال سایر بازیکنان، امتیازی به بازیکن

³ Nash Equilibrium

⁴ Blotto Games

⁵ Matrix Games

⁶ Zero-Sum

تعلق می‌گیرد. با توجه به این امتیاز، بازیکن یاد می‌گیرد که عمل انجام شده تا چه حد مناسب بوده است و نتیجه این یادگیری را در دفعات بعدی اجرای بازی در هنگام انتخاب عمل به کار می‌برد.

در این مقاله، رویکردنی مبتنی بر اتماتای یادگیر⁷ برای پیاده‌سازی بازی‌های *Blotto* ارائه می‌گردد. اتماتای یادگیر که بر مبنای یادگیری تقویتی عمل می‌کند، ابزاری مناسب برای کمک به فرآیند یادگیری عامل‌ها در سیستم‌های چندعامله از جمله سیستم‌های رقابتی است. در رویکرد پیشنهادی، هریک از بازیکنان در بازی *Blotto* در بخش یادگیری خود دارای یک اتماتای یادگیر می‌باشد. اتماتای یادگیر به بازیکن کمک می‌کند تا در هر مرحله، با توجه به نتایجی که تا به حال به دست آمده، بهترین عمل را انتخاب نماید و سپس براساس پاداش یا جریمه‌ای که در اثر انجام این عمل دریافت می‌کند، احتمال انجام آن را در مراحل بعدی بازی به ترتیب افزایش یا کاهش دهد. آزمایشات انجام‌شده نشان می‌دهند که با استفاده از این روش، استراتژی بازیکنان مجهز به اتماتای یادگیر به تدریج به استراتژی غالب همگرا می‌گردد. استراتژی غالب که در بازی‌های *Blotto* به معنای بهترین حالت تخصیص منابع محدود به چندین کار می‌باشد، همواره بهترین نتیجه را برای بازیکن در پی خواهد داشت.

ادامه مقاله بین صورت سازماندهی شده است. در بخش ۲ مروری بر اتماتای یادگیر و الگوریتم یادگیری مورد استفاده در آن خواهیم داشت. بخش ۳ شامل آزمایشات انجام‌شده برای بررسی عملکرد روش پیشنهادی می‌باشد. در انتهای، در بخش ۴ نتایج حاصل از به کارگیری روش ارائه‌شده برای حل مسئله تخصیص بهینه منابع در بازی‌های *Blotto* را بر می‌شماریم.

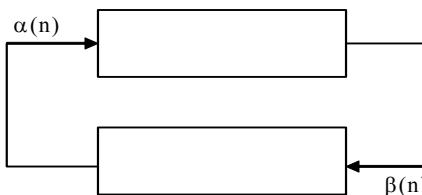
۲- اتماتای یادگیر

اتماتای یادگیر [۸,۷,۵,۴]، ماشینی است که می‌تواند تعدادی متنه‌ای عمل را انجام دهد. هر عمل انتخاب شده توسط یک محیط احتمالی ارزیابی می‌شود، نتیجه ارزیابی در قالب سیگنالی مثبت یا منفی به اتماتا داده می‌شود و اتماتا از این پاسخ در انتخاب عمل بعدی تأثیر می‌گیرد. هدف نهایی این است که اتماتا یادگیرد تا از بین اعمال خود، بهترین عمل را انتخاب کند. بهترین عمل، عملی است که احتمال دریافت پاداش از محیط را به حداقل برساند. کارکرد اتماتای یادگیر در تعامل با محیط، در شکل ۱ مشاهده می‌شود.

محیط را می‌توان توسط سه تایی $E \equiv \{\alpha, \beta, c\}$ نشان داد که در آن $\{\alpha_1, \alpha_2, \dots, \alpha_r\} \equiv \alpha$ مجموعه ورودی‌ها، $\{\beta_1, \beta_2, \dots, \beta_m\} \equiv \beta$ مجموعه خروجی‌ها و $\{c_1, c_2, \dots, c_r\} \equiv c$ مجموعه احتمال‌های جریمه می‌باشد. هرگاه $\beta_2 = 0$ به عنوان پاداش در نظر گرفته می‌شود. در محیط از نوع P است. در چنین محیطی $\beta_1 = 1$ به عنوان جریمه و $\beta_2 = 0$ به عنوان پاداش در نظر گرفته می‌شود. در محیط از نوع Q ، $\beta(n)$ می‌تواند به طور گسسته یک مقدار از مقادیر محدود در فاصله $[0, 1]$ را اختیار کند و در محیط از نوع S ، $\beta(n)$ متغیر تصادفی در فاصله $[0, 1]$ است. c_i احتمال این‌که عمل i نتیجه نامطلوب داشته باشد می‌باشد. در

محیط ایستا، مقادیر c_i بدون تغییر می‌مانند، حال آن‌که در محیط غیرایستا این مقادیر در طی زمان تغییر می‌کنند.

اتماتای یادگیر با ساختار ثابت توسط پنج تایی $\{\alpha, \beta, p, T, \phi\}$ نشان داده می‌شود که در آن $\{\alpha_1, \alpha_2, \dots, \alpha_r\} \equiv \alpha$ مجموعه عمل‌های اتماتا، $\{\beta_1, \beta_2, \dots, \beta_r\} \equiv \beta$ مجموعه ورودی‌های اتماتا، $\{p_1, p_2, \dots, p_r\} \equiv p$ بردار احتمال انتخاب هریک از عمل‌ها و اتماتا در لحظه n ، $T: \phi \times \beta \rightarrow \phi$ تابع تولید وضعیت جدید اتماتا و $\phi: G \rightarrow \alpha$ تابع خروجی می‌باشد که وضعیت کنونی اتماتا را به خروجی بعدی می‌نگارد.



شکل ۱- ارتباط بین اتماتای یادگیر و محیط

اتماتای یادگیر با ساختار متغیر را می‌توان توسط چهارتایی $\{\alpha, \beta, p, T\}$ نشان داد که $\{\alpha_1, \alpha_2, \dots, \alpha_r\} \equiv \alpha$ مجموعه اعمال اتماتا، $\{\beta_1, \beta_2, \dots, \beta_r\} \equiv \beta$ مجموعه ورودی‌های اتماتا، $\{p_1, p_2, \dots, p_r\} \equiv p$ بردار احتمال انتخاب هریک از عمل‌ها و $p(n+1) = T[\alpha(n), \beta(n), p(n)]$ الگوریتم یادگیری می‌باشد. الگوریتم زیر یک نمونه از الگوریتم‌های یادگیری خطی است. فرض می‌کنیم عمل i در مرحله n انتخاب شود.

⁷ Learning Automata (LA)

- پاسخ مطلوب از محیط

$$\begin{aligned} p_i(n+1) &= p_i(n) + a[1 - p_i(n)] \\ p_j(n+1) &= (1-a)p_j(n) \quad \forall j \neq i \end{aligned} \quad (1)$$

- پاسخ نامطلوب از محیط

$$\begin{aligned} p_i(n+1) &= (1-b)p_i(n) \\ p_j(n+1) &= (b/r-1) + (1-b)p_j(n) \quad \forall j \neq i \end{aligned} \quad (2)$$

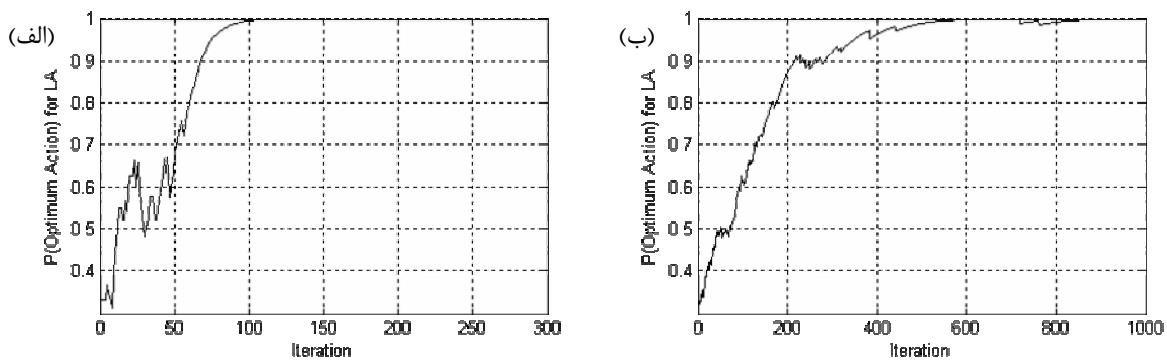
در روابط (1) و (2)، a پارامتر پاداش و b پارامتر جریمه می‌باشند. با توجه به مقادیر a و b سه حالت را می‌توان در نظر گرفت: اگر a و b با هم برابر باشند، الگوریتم را L_{RP} خیلی کوچکتر باشد، الگوریتم را L_{REP} و اگر b مساوی صفر باشد آن را L_{RI} می‌نامیم.

۳- یافتن استراتژی غالب در بازی Blotto با استفاده از اتوماتای یادگیری

هدف از این بخش یادگیری استراتژی غالب در بازی Blotto است، به طوری که بازیکن با اجرای آن، تحت هر شرایطی و بدون توجه به روش بازی حریف در بازی پیروز شود. برای پیاده‌سازی بازی Blotto از نمونه‌ای که در مقدمه توضیح داده شد، یعنی بازی اعداد استفاده کرده‌ایم. در رویکردی که در این بخش برای حل این بازی ارائه می‌دهیم، بازیکنان از اتوماتای یادگیری در فرآیند یادگیری و تصمیم‌گیری خود برای انتخاب بهترین عمل کمک می‌گیرند.

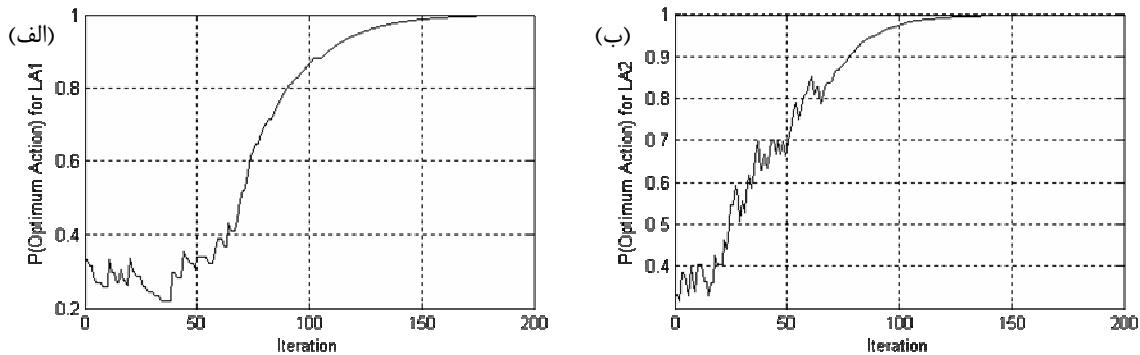
در این روش، هر بازیکن دارای یک اتوماتای یادگیر با ساختار متغیر می‌باشد. همان‌طور که در بخش ۲ گفته شد، اتوماتای یادگیر با ساختار متغیر با چهارتایی $\{\alpha, \beta, p, T\}$ نشان داده می‌شود. برای مدل‌کردن این بازی با اتوماتای یادگیر، α برابر با مجموعه اعمال هر عامل، یعنی $\{(1,1), (1,2), (1,3), (2,2), (2,3)\}$ قرار داده می‌شود. β برابر با امتیازی درنظر گرفته می‌شود که بازیکن در هر مرحله دریافت می‌کند. p بردار احتمال اعمال اتوماتا است که در ابتداء مقادیر آن به ازاء تمام اعمال، یکسان درنظر گرفته می‌شود. T نیز نشان‌دهنده الگوریتم یادگیری می‌باشد که در پیاده‌سازی این بازی، از الگوریتم یادگیری خطی مطابق با روابط (1) و (2)، و با شمای L_{RI} استفاده شده است.

بازی به صورت تکراری و با توجه به قانون بازی و ماتریس نتیجه در دفعات متعدد اجرا می‌گردد. در هر بار اجرای بازی، ابتداء اتوماتای یادگیر مربوط به هر بازیکن، مستقلًا یک عمل از مجموعه اعمال خود را با توجه به بردار احتمال انتخاب می‌کند. پس از اجرای اعمال توسط بازیکنان، با توجه به ماتریس نتایج امتیازی به هریک تعلق می‌گیرد. با توجه به این امتیاز و با استفاده از الگوریتم یادگیری اتوماتای یادگیر، بردار احتمال هر دو اتوماتا به روز می‌شود. این روند مرتبًا تکرار می‌شود تا جایی که بردار احتمال اتوماتا به پایداری برسد. برای انجام آزمایش دو حالت را در نظر می‌گیریم. ابتداء حالتی را بررسی می‌کیم که یکی از بازیکنان، دارای اتوماتای یادگیر است و بازیکن دیگر اعمال خود را به طور تصادفی انتخاب می‌نماید. همان‌طور که در نمودارهای شکل ۲ مشاهده می‌شود، احتمال انتخاب عمل بهینه توسط اتوماتا به سرعت به یک می‌رسد و استراتژی بازیکن به استراتژی غالب همگرا می‌گردد. مشاهده می‌کنیم که سرعت همگرایی و همواری نمودار به پارامتر پاداش، یعنی a وابسته است.

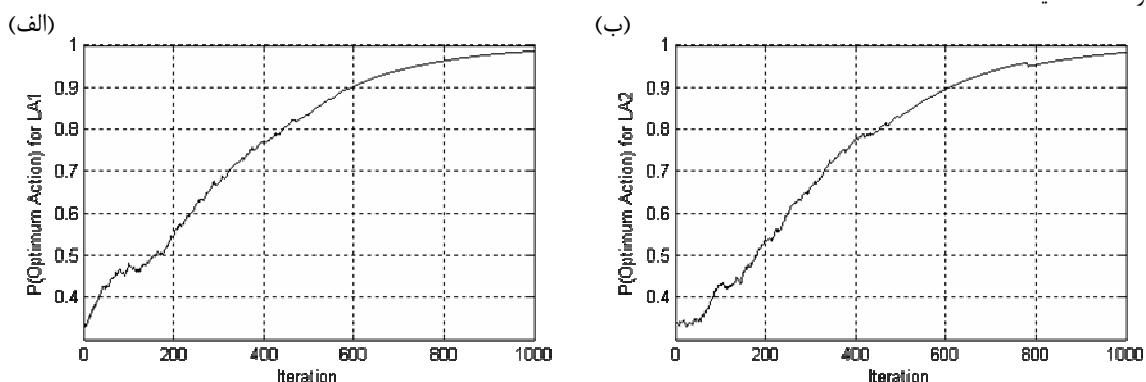


شکل ۲- تغییرات احتمال انتخاب عمل بهینه توسط بازیکن دارای اتوماتای یادگیر با نرخ یادگیری (الف) ۰/۱ و (ب) ۰/۰، در حالتی که فقط یکی از بازیکنان از اتوماتای یادگیر استفاده نماید.

در حالت دوم، در ساختار هر دو بازیکن از اتوماتای یادگیر استفاده می‌کنیم. در این وضعیت به دلیل این که بازیکنان به طور همزمان در حال یادگیری هستند، همگرایی با سرعت کمتری صورت می‌گیرد؛ زیرا در هر بار تکرار بازی، هر بازیکن سعی می‌کند بهترین رفتاری که تا آن لحظه یاد گرفته است را از خود نشان دهد و انتخاب عملی با سود بیشتر را برای حریف مشکل می‌سازد. این مسئله در نمودارهای شکل ۳ و ۴ قابل مشاهده می‌باشد. در این حالت نیز با کاهش پارامتر پاداش، همگرایی رفتار بازیکنان، کندتر شده ولی نوسانات کمتری از خود نشان می‌دهد.



شکل ۳- تغییرات احتمال انتخاب عمل بهینه توسط (الف) بازیکن ۱ و (ب) بازیکن ۲ با نرخ یادگیری ۰/۰۱، در حالتی که هر دو بازیکن از اتوماتای یادگیر استفاده نمایند.



شکل ۴- تغییرات احتمال انتخاب عمل بهینه توسط (الف) بازیکن ۱ و (ب) بازیکن ۲ با نرخ یادگیری ۰/۰۱، در حالتی که هر دو بازیکن از اتوماتای یادگیر استفاده نمایند.

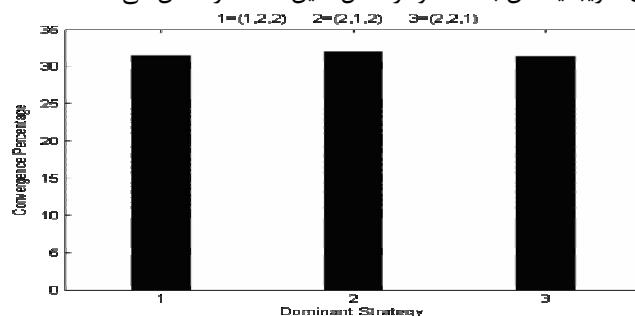
بازی مورد استفاده در آزمایش بالا، فقط دارای یک استراتژی غالب و تعادل *Nash* بوده و تعداد اعمال در آن محدود به سه عمل می‌باشد. هنگامی که مسئله بزرگتر بوده و یا قانون بازی متفاوت باشد، مجموعه اعمال نیز بزرگتر خواهد بود و ممکن است چندین استراتژی غالب برای بازیکن وجود داشته باشد. برای بررسی تأثیر افزایش اندازه مجموعه اعمال و همین‌طور افزایش تعداد استراتژی‌های غالب، نوع دیگری از بازی را مطرح می‌نماییم. در این بازی نیز که بسیار شبیه به بازی قبل می‌باشد، دو بازیکن سه عدد را روی کاغذ نویسنده، به طوری که مجموع آن‌ها برابر با عدد معینی شود. این کار بدون اطلاع از انتخاب حریف صورت می‌گیرد. تفاوت این بازی با بازی قبل در این است که اولاً بازیکن‌ها مجاز به نوشتن جایگشت‌های مختلف سه عدد می‌باشند؛ به این معنا که مثلًاً (۳،۱،۱) و (۰،۳،۱) دو انتخاب متفاوت به شمار می‌روند. ثانیاً در بازی جدید، هریک از سه عدد با عددی در همان جایگاه مقایسه می‌شود؛ یعنی عدد اول بازیکن با عدد اول حریف، عدد دوم او با عدد دوم حریف و عدد سوم او با عدد سوم حریف مقایسه شده و در این مقایسه‌ها اگر هریک از اعداد بازیکن از اعداد حریف بزرگتر باشد، به ازاء آن عدد یک امتیاز مثبت و اگر کوچکتر باشد یک امتیاز منفی به بازیکن تعلق می‌گیرد. در صورت تساوی، بازیکن امتیازی برای آن عدد دریافت نمی‌کند. درنهایت امتیازی که به بازیکن داده می‌شود، برابر با مجموع امتیازی خواهد بود که از هریک از سه عدد به دست می‌آورد.

اگر مجموع سه عدد می‌باشد برابر با ۵ شود، ماتریس نتیجه برای بازیکن اول به صورت زیر خواهد بود. همان‌طور که در بازی قبل گفته شد، به دلیل جمع صفر بودن بازی‌های *Blotto*، ماتریس نتیجه بازیکن دوم مشابه همین ماتریس می‌باشد، اما علامت امتیازات معکوس می‌گردد.

| | (۱,۱,۳) | (۱,۲,۲) | (۱,۳,۱) | (۲,۱,۲) | (۲,۲,۱) | (۳,۱,۱) |
|---------|---------|---------|---------|---------|---------|---------|
| (۱,۱,۳) | + | + | + | + | -1 | + |
| (۱,۲,۲) | + | + | + | + | + | 1 |
| (۱,۳,۱) | + | + | + | -1 | + | + |
| (۲,۱,۲) | + | + | 1 | + | + | + |
| (۲,۲,۱) | 1 | + | + | + | + | + |
| (۳,۱,۱) | + | -1 | 0 | 0 | 0 | 0 |

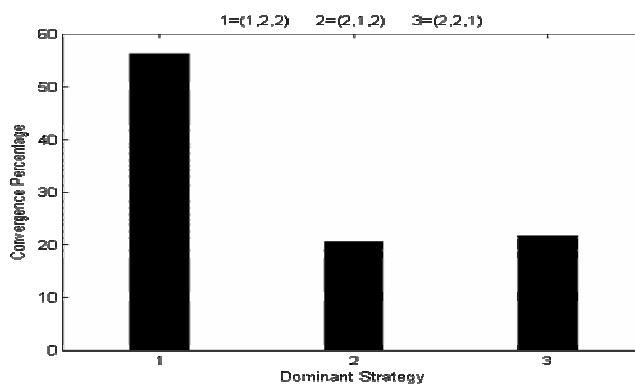
جدول ۲- ماتریس نتیجه برای بازیکن اول در یک نمونه بازی Blotto

همان‌طور که از جدول بالا می‌توان نتیجه گرفت، هر سه عمل $(1,1,3)$ ، $(2,1,2)$ و $(2,2,1)$ ، تعادل *Nash* و به عبارتی استراتژی غالب برای این بازی می‌باشند. برای انجام آزمایش، حالتی را در نظر می‌گیریم که هر دو بازیکن مجهز به اتماتاتی یادگیر بوده و در حال یادگیری هستند. زیرا در صورتی که بازیکن در این حالت بتواند بهترین استراتژی‌ها را بیابد، در وضعیتی که حریف قدرت یادگیری ندارد نیز قادر به این کار خواهد بود. اگر احتمال انتخاب اعمال در ابتدای بازی یکسان در نظر گرفته شوند، منطقی است که در دفعات متعدد اجراء، احتمال همگرایی به هریک از سه استراتژی تقریباً یکسان باشد. نمودار شکل ۵ این مسئله را نشان می‌دهد.



شکل ۵- درصد دفعات همگرایی به هریک از سه استراتژی غالب در ۱۰۰۰۰ بار اجرای الگوریتم با احتمال اولیه برابر برای تمام اعمال

اگر بخواهیم استراتژی بازیکن لزوماً به یکی از این سه استراتژی غالب همگرا گردد، می‌توانیم احتمال اولیه استراتژی مورد نظر را به میزان کمی افزایش دهیم. در این صورت احتمال همگرایی به آن استراتژی بیشتر خواهد شد. نمودار شکل ۶، وضعیتی را نشان می‌دهد که احتمال اولیه عمل $(1,2,2)$ نسبت به سایر اعمال، به میزان جزئی افزایش یافته است. مشاهده می‌شود که درصد دفعات همگرایی به عمل $(1,2,2)$ نسبت به $(2,1,2)$ و $(2,2,1)$ در مقایسه با حالت قبل بیشتر می‌گردد.



شکل ۶- درصد دفعات همگرایی به هریک از سه استراتژی غالب در ۱۰۰۰۰ بار اجرای الگوریتم پس از افزایش جزئی احتمال اولیه عمل $(1,2,2)$

۴- نتیجه‌گیری

در این مقاله، روشی مبتنی بر اتماتاتی یادگیر برای یافتن استراتژی غالب در بازی‌های *Blotto* ارائه گردید. با توجه به نتایج آزمایشات انجام شده، استفاده از اتماتاتی یادگیر در ساختار بازیکنان منجر به یادگیری استراتژی غالب توسط آنها و همگرایی سریع رفتار

آن‌ها به رفتار بهینه می‌گردد. این بازی را می‌توان مدلی از بسیاری از فرآیندهای دنیای واقعی که در آن‌ها نیاز به تقسیم بهینه منابع محدود بین چندین کار یا محل وجود دارد درنظر گرفت. به‌این ترتیب روش ارائه شده، در بسیاری از کاربردهای بزرگ و برای حل مسائل مطرح در سیستم‌های چندعامله رقابتی که دارای استراتژی غالب می‌باشند قابل استفاده بوده و دستیابی به حداکثر سود ممکن را تضمین می‌نماید.

مراجع

1. Roberson, B., "The Colonel Blotto Game", *Journal of Economic Theory*, vol. 29, no. 1, pp. 1-24, Springer Berlin, 2006.
2. Golman, R., Page, S. E., "General Blotto: Games of Allocative Strategic Mismatch", *Center for Complex Systems, University of Michigan*, 2006.
3. Tuyls, K., Nowe, A., "Evolutionary Game Theory and Multi-Agent Reinforcement Learning", *The Knowledge Engineering Review*, vol. 20, pp. 63-90, 2005.
4. Nowe, A., Verbeeck, K., Peeters, M., "Learning Automata as a Basis for Multi-Agent Reinforcement Learning", *Proceedings of First International Workshop on Learning and Adaptation in Multi-Agent Systems (LAMAS), Utrecht, the Netherlands*, 2005, pp. 71-85, 2006.
5. Shirazi, M.R., Meybodi, M.R. "Application of Learning Automata to Cooperation in Multi-Agent Systems", *Proceedings of First International Conference on Information and Knowledge Technology (IKT2003)*, pp. 338-349, 2003.
6. Kapetanakis, S., Kudenko, D., "Reinforcement Learning of Coordination in Cooperative Multi-Agent Systems", *Proceedings of Eighteenth National Conference on Artificial Intelligence*, pp. 326-331, 2002.
7. Shirazi, M. R., Meybodi, M. R., "Solving Iterated Prisoner's Dilemma Using Learning Automata", *Proceedings of First Iranian Conference on Mechatronics Engineering, ICME2003*, pp. 349-362, Qazvin, Iran, 2003.
8. Verbeek, K., Nowe, A., Peeters, M., Tuyls, K. "Multi-Agent Reinforcement Learning in Stochastic Single and Multi-Stage Games", *Lecture Notes in Computer Science*, Springer Berlin, vol. 3394, pp. 275-294, 2005.
9. Banerjee, B., Peng, J., "Convergence of No-Regret Learning in Multi-Agent Systems", *Proceedings of the First International Workshop on Learning and Adaptation in Multi-Agent Systems (LAMAS), Utrecht, the Netherlands*, 2005.
10. Bowling, M., Veloso, M. M., "Existence of Multi-Agent Equilibria with Limited Agents", *Technical Report CMU-CS-02104, Computer Science Department, Carnegie Mellon University*, 2002.
11. Merolla, J., Munger, M., Tofias, M., "Lotto, Blotto, or Frontrunner: The 2000 U. S. Presidential Election and the Nature of Mistakes", *Duke University*, 2003.
12. Powers, M. R., Shen, Z., "Colonel Blotto in the War on Terror: Implications for Event Frequency", *American Risk and Insurance Association (ARIA) Annual Meeting, Washington D.C.*, August 2006.