

# همکاری در سیستم‌های چند عامله با استفاده از اتوماتاهای یادگیر

محمد رضا خجسته و محمد رضا میبدی

شناختی خود و نیز براساس تأثیر مستقیم بر ورودی‌های قابل دریافت هم‌دیگر از طریق فعالیتهای ارتباطی با هم‌دیگر هماهنگ می‌کنند. دیگر عاملهای موجود در محیط نیز که اهدافی متقابل با هدف درازمدت این تیم دارند، حریفان یا دشمنان این تیم محسوب می‌شوند.

بعنوان بستری برای تست و پیاده‌سازی سیستم‌های چند عامله می‌توان به فوتیال روباتها یا روبوکاپ<sup>۱</sup> اشاره کرد. محیط فوتیال روباتها مثالی از یک محیط پیچیده است که در آن چند عامل باید جهت رسیدن به اهداف تیمی، با هم همکاری کنند [۳] تا [۶]. فرآیندهای رفتاری و تصمیم‌گیری می‌توانند از ساده‌ترین رفتارها، همانند حرکت مستقیم به طرف توپ تا پیچیده‌ترین استدلال‌ها که استراتژی‌های خود و تیم مقابل را روشن می‌سازند، تشکیل شوند.

به دلیل وجود پیچیدگی‌های موجود در بسترها می‌تواند بستر شبیه‌سازی فوتیال روباتها و روبارویی عاملهای موجود در چنین محیط‌هایی با حالات بسیار زیاد، متنوع و متغیر، ناگزیر به استفاده از روش‌های یادگیری ماشین می‌باشیم. تأکید در این مقاله بر روی سیستم‌هایی متشکل از چند عامل خودمختار است که می‌توانند در محیط‌های زمان واقعی، نیازی<sup>۲</sup>، نیاز به همکاری<sup>۳</sup> و دارای دشمن با اهداف متقابل<sup>۴</sup> عمل کنند [۲].

اتوماتاهای یادگیر بعنوان مدلی برای یادگیری، در محیطی تصادفی عمل نموده و قادر هستند که براساس ورودی‌های دریافت شده از محیط، احتمال انجام عملیات خود را به روز درآورند تا بتوانند از این طریق کارآیی خود را بهبود بخشنند. یکی از اهداف این مقاله بررسی کارآیی اتوماتای یادگیر در همکاری بین عاملهایی که انتخاب شبیه‌سازی فوتیال روباتها می‌باشد. با استفاده از بستر تست شبیه‌سازی فوتیال روباتها به بررسی کارآیی اتوماتای یادگیر در همکاری بین عاملهایی این تیم در پرداخته شده است. با پیاده‌سازی تیمهای ۲ نفره، ۵ نفره و ۱۱ نفره از عاملهایی که هر کدام از آنها به یک اتوماتای یادگیر مجهز شده است و مقایسه آنها با یک تیم بدون یادگیری و یا تیمهای یادگیر دیگر، کارآیی اتوماتای یادگیر در یادگیری یک کار تیمی جهت دست یافتن به یک هدف مشترک مورد ارزیابی قرار می‌گیرد.

بدلیل وجود تعداد حالات بسیار زیاد در دامنه‌های چند عامله پیچیده، داشتن روشی برای عمومی‌سازی حالات محیطی امری ضروری است انتخاب مناسب چنین روشی، در تعیین حالات و اعمال عامل نقشی تعیین کننده دارد. بهمین دلیل در این مقاله به معروفی تکنیک جدیدی بنام "تکنیک بهترین گوشه در مربع حالت" نیز می‌پردازیم. این روش فضای حالات پیوسته و بسیار وسیع عامل را به فضای حالات گستته و محدود نگاشت می‌کند، ضمن آنکه سعی می‌شود که مواردی مانند فاصله و زاویه که مفاهیمی پیوسته هستند در تعیین حالت عامل منظور شود. ایده نگاشت و عمومی‌سازی فضای دامنه فعالیت عاملهای موجود در یک تیم به یک شبکه از مربعهای حالت برای اولین بار در این مقاله مطرح گردیده است. در کارهای گذشته از جمله [۲]، از مفاهیمی پیوسته مانند فاصله، زاویه و

2. Robocup

3. Noisy

4. Collaborative

5. Adversarial

چکیده: عاملهای موجودات نرم‌افزاری هستند که بطور پیوسته و خود مختار در یک محیط که برای آن طراحی شده‌اند کار می‌کنند. برای عاملهای نیازهایی از قبیل واکنشی بودن نسبت به کنش‌های محیطی، خود مختاری در انتخاب مسیر و ادامه آن، قابلیت تطبیق و یادگیری و ... ضروری به نظر می‌رسد. امروزه مطالعه سیستم‌های مبتنی بر عاملهای به یک موضوع مهم، آکادمیک تبدیل شده است که کاربردهای تجاری و صنعتی فراوانی را نیز در بر دارد. در سیستم‌های چند عامله، چندین عامل هشتمند با قابلیت برقراری ارتباط با یکدیگر، جهت رسیدن به مجموعه‌ای از اهداف، با هم همکاری می‌کنند. بدلیل پیچیدگی‌های موجود در محیط‌های چند عامله پویا و متغیر نیاز به روش‌های یادگیری ماشین در چنین محیط‌هایی احساس می‌شود. اتوماتای یادگیر یک مدل انتزاعی است که تعداد محدودی عمل را می‌تواند انجام دهد. هر عمل انتخاب شده توسط محیطی احتمالی ارزیابی می‌گردد و پاسخی به اتوماتای یادگیر داده می‌شود. اتوماتای یادگیر از این پاسخ استفاده نموده و عمل خود برای مرحله بعد را انتخاب می‌کند. در این مقاله با استفاده از بستر تست شبیه‌سازی فوتیال روباتها به بررسی کارآیی اتوماتای یادگیر در همکاری بین عاملهایی عضو یک تیم پرداخته شده است. بدلیل وجود تعداد حالات بسیار زیاد در دامنه‌های چند عامله پیچیده، داشتن روشی برای عمومی‌سازی حالات محیطی، امری ضروری است که انتخاب مناسب چنین روشی، در تعیین حالات و اعمال عامل نقشی تعیین کننده دارد. در این مقاله همچنین به معرفی و پیاده‌سازی تکنیک "بهترین گوشه در مربع حالت" پرداخته شده است. با استفاده از این روش فضای حالات پیوسته و بسیار وسیع عامل به فضای حالات گستته و محدود نگاشته می‌شود. کارآیی این تکنیک در عمومی‌سازی حالات محیطی در یک دامنه چند عامله همکاری گوا مورد بررسی قرار گرفته است.

**کلید واژه:** اتوماتای یادگیر<sup>۱</sup>، عامل، سیستم‌های چند عامله، فوتیال روباتها، همکاری.

## ۱- مقدمه

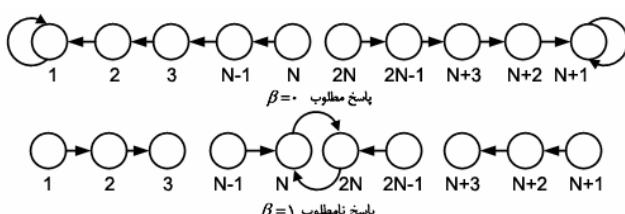
عامل، موجود خودمختاری است که ویژگی‌هایی از قبیل اجتماعی بودن، واکنشی بودن و پیش فعال بودن را دارا می‌باشد. عاملهای در محیطی زندگی می‌کنند که می‌تواند باز یا بسته باشد و نیز ممکن است که این محیط عاملهایی دیگر را در خود جای داده باشد. هر چند وضعیت‌هایی وجود دارد که یک عامل می‌تواند خودش به تنهایی عمل کند ولی معمول بر آن است که یک عامل با دیگر عامل‌ها ارتباط متقابل داشته باشد. عاملهای در جهت دست یافتن به اهداف خود یا جامعه‌ای که در آن زندگی می‌کنند با یکدیگر همکاری می‌کند. وقتی که یک گروه از عاملهای در یک سیستم چند عامله در یک هدف دراز مدت سهیم باشند، آنها تشکیل یک تیم را می‌دهند.

اعضای یک تیم رفاقت خود را براساس سازگار کردن فرآیندهای این مقاله در تاریخ ۳ شهریور ماه ۱۳۸۱ دریافت و در تاریخ ۶ مرداد ۱۳۸۲ بازنگری شد.

محمد رضا خجسته، آزمایشگاه محاسبات نرم، دانشکده مهندسی کامپیوتر، دانشگاه صنعتی امیرکبیر، تهران، ایران (email: khojasteh@ce.aut.ac.ir).

محمد رضا میبدی، آزمایشگاه محاسبات نرم، دانشکده مهندسی کامپیوتر، دانشگاه صنعتی امیرکبیر، تهران، ایران (email: meybodi@ce.aut.ac.ir).

1. Learning Automaton

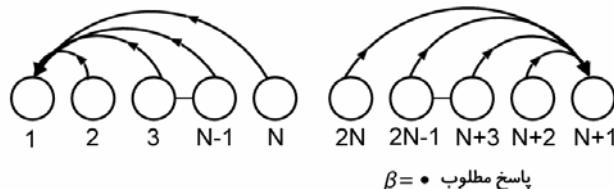
شکل ۲: نمودار تغییر وضعیت اتماتای  $L_{2N,2}$ .

مجموعه پاسخ

مجموعه ورودی



شکل ۱: ارتباط بین آتماتای یادگیر و محیط.



شکل ۴: نمودار تغییر وضعیت اتماتای Krinsky.

ارتباط بین آتماتای یادگیر و محیط را نشان می‌دهد.

## ۲-۲ اتماتاهای یادگیر با ساختار ثابت

اتماتاهای یادگیر با ساختار ثابت توسط ۵ تابی  $LA = \{\alpha, \beta, F, G, \phi\}$  نشان داده می‌شود که  $\alpha = \{\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_r\}$  مجموعه اعمال اتماتا،  $\beta = \{\beta_1, \beta_2, \beta_3, \dots, \beta_r\}$  مجموعه ورودی‌های اتماتا،  $F: \phi \times \beta \rightarrow \phi$  تابعی که براساس پاسخ محیط، وضعیت جدید را می‌یابد،  $G: \phi \rightarrow \alpha$  تابع خروجی که وضعیت کنونی را به خروجی بعدی می‌نگارد و  $\{\phi, \phi_1, \phi_2, \dots, \phi_k\} = \phi(n)$  مجموعه وضعیت‌های داخلی اتماتا می‌باشد.

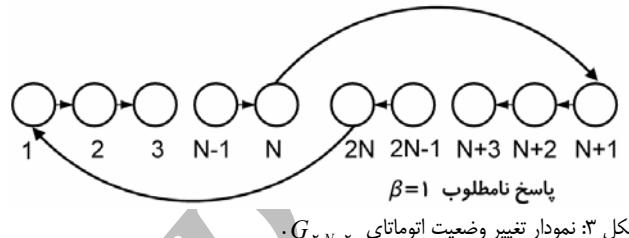
در ادامه به چند نمونه از اتماتاهای یادگیر با ساختار ثابت که در این مقاله از آنها استفاده شده است اشاره شده است.

- اتماتای  $L_{2N,2}$ : این اتماتا تعداد پاداش‌ها و جریمه‌های دریافت شده برای هر عمل را نگهداری کرده و تنها زمانی که تعداد جریمه‌ها بیشتر از پاداش‌ها می‌گردد، عمل دیگر را انتخاب می‌کند. نمودار تغییر وضعیت این اتماتای مطابق شکل ۲ می‌باشد.

- اتماتای  $G_{2N,2}$ : در این اتماتا برخلاف اتماتای  $L_{2N,2}$ ، عمل  $\alpha_2$  حداقل  $N$  بار انجام گیرد (پس از گرفتن  $N$  جریمه) تا اینکه عمل  $\alpha_1$  دوباره انتخاب شود. گراف تغییر وضعیت این اتماتا برای پاسخ نامطلوب مانند اتماتای  $L_{2N,2}$  بوده و برای پاسخ نامطلوب مطابق شکل ۳ می‌باشد.

- اتماتای Krinsky: این اتماتا زمانی که پاسخ محیط نامطلوب است، مانند  $L_{2N,2}$  رفتار می‌کند. اما برای پاسخ مطلوب هر وضعیت  $\phi_i$  ( $i = 1, 2, \dots, N$ ) به وضعیت  $\phi_j$  و هر وضعیت  $\phi_{N+i}$  ( $i = N+1, N+2, \dots, 2N$ ) به وضعیت  $\phi_{N+1}$  می‌رسد. بنابراین همیشه  $N$  پاسخ نامطلوب متولی لازم است تا اتماتا عمل خود را عوض کند. نمودار تغییر وضعیت این اتماتا برای پاسخ نامطلوب مانند اتماتای  $L_{2N,2}$  بوده و برای پاسخ مطلوب مطابق شکل ۴ می‌باشد.

- اتماتای Krylov: در این اتماتا زمانی که پاسخ محیط مطلوب است، تغییر وضعیت مانند اتماتای  $L_{2N,2}$  می‌باشد. اما زمانی که پاسخ محیط نامطلوب باشد، هر وضعیت  $(N+1, N+2, \dots, 2N)$  با احتمال  $1/N$  به وضعیت  $\phi_{i+1}$  و با احتمال  $1/N$  به وضعیت  $\phi_i$  مطابق شکل ۵ منتقل می‌شود.

شکل ۳: نمودار تغییر وضعیت اتماتای  $G_{2N,2}$ .

نظایر آن برای عمومی‌سازی فضای حالات عامل استفاده شده است. ادامه مقاله بصورت زیر سازماندهی شده است. در بخش ۲ به اختصار به شرح اتماتاهای یادگیر می‌پردازیم. در بخش ۳ "انکنیک بهترین گوشه در مربع حالت" برای عمومی‌سازی فضای حالات محیطی معروفی می‌گردد. در بخش ۴ به بحث و بررسی موضوع همکاری بین اعضای یک تیم چند عامله می‌پردازیم. در بخش ۵ نتایج آزمایشات ارایه می‌شود. بخش نهایی مقاله نتیجه‌گیری است.

## ۲- اتماتاهای یادگیر

اتماتاهای یادگیر مدل‌های انتزاعی هستند که در محیطی تصادفی عمل نموده و قادر هستند که براساس ورودی‌های دریافت شده از محیط، احتمال انجام عملیات خود را به روز درآورده تا بتوانند از این طریق کارآیی خود را بهبود بخشنند. یک اتماتای یادگیر تعداد محدودی عمل را می‌تواند انجام دهد. هر عمل انتخاب شده توسط محیطی احتمالی ارزیابی می‌گردد و پاسخی به اتماتای یادگیر داده می‌شود. اتماتاهای یادگیر از این پاسخ استفاده نموده و عمل خود برای مرحله بعد انتخاب می‌کند [۷] و [۸].

اتماتاهای یادگیر به دو گروه تقسیم می‌گردند:

- الف- اتماتای یادگیر با ساختار ثابت<sup>۱</sup>
- ب- اتماتای یادگیر با ساختار متغیر.<sup>۲</sup>

## ۱- محیط

محیط را می‌توان توسط سه تابی  $E \equiv \{\alpha, \beta, c\}$  تعریف نمود که  $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_r\}$  مجموعه ورودی‌ها،  $\beta = \{\beta_1, \beta_2, \dots, \beta_r\}$  مجموعه خروجی‌ها و  $c = \{c_1, c_2, \dots, c_r\}$  مجموعه احتمال‌های جریمه شدن می‌باشد.

هرگاه  $\beta_i$  دو مقداری باشد، محیط از نوع  $P$  می‌باشد. در چنین محیطی  $= \beta_i$  به عنوان جریمه و  $= \beta_i$  به عنوان پاداش در نظر گرفته می‌شود.  $c_i$  احتمال این است که عمل  $\alpha_i$  تغییر نامطلوب داشته باشد. در محیط پایدار<sup>۳</sup> مقادیر  $c_i$  بدون تغییر باقی می‌مانند، حال آن که در محیط ناپایدار<sup>۴</sup> این مقادیر در طی زمان تغییر می‌کنند. شکل ۱

1. Fixed Structure
2. Variable Structure
3. Stationary
4. Non-Stationary

|          |          |         |          |          |
|----------|----------|---------|----------|----------|
| $X - 22$ | $X - 12$ | $X - 2$ | $X + 8$  | $X + 18$ |
| $X - 21$ | $X - 11$ | $X - 1$ | $X + 9$  | $X + 19$ |
| $X - 20$ | $X - 10$ | $X$     | $X + 10$ | $X + 20$ |
| $X - 19$ | $X - 9$  | $X + 1$ | $X + 11$ | $X + 21$ |
| $X - 18$ | $X - 8$  | $X + 2$ | $X + 12$ | $X + 22$ |

شکل ۲۴ مربع اطراف مربع دربردارنده عامل مورد نظر در مربع  $X$ .

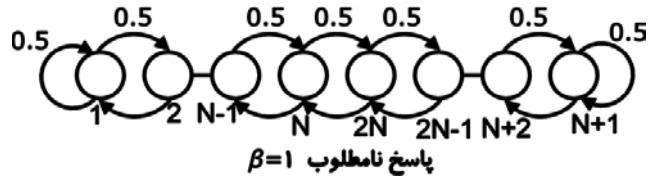
باشد، هدف پیدا کردن تابع  $V \rightarrow S$  می‌باشد. بدین صورت، پس از داشتن تابع  $f$ ، عامل می‌تواند از  $V$  برای یادگیری عمل مناسب خود در محیط استفاده نماید. علاوه بر این تابع، یک تابع تقسیم وظایف  $P: S \rightarrow M$  نیز مورد نیاز است. تابع  $P$  مجموعه حالات دامنه را بین عاملهای موجود در محیط تقسیم می‌کند. این تابع فضای حالت را به  $[M]$  بخش مجزا تقسیم می‌کند و هر بخش به حداقل یک عامل (جهت یادگیری و عمل در آن بخش) سپرده می‌شود.

با توجه به توضیحات فوق و با فرض این که مجموعه اعمال عامل  $A$  باشد، عامل در هر کدام از  $|V|$  حالت، دارای  $|A|$  عمل ممکن خواهد بود و بدین ترتیب، مجموعه مورد یادگیری عامل حداکثر شامل  $|A| \times |V|$  عضو خواهد بود. با انتخاب مناسب مجموعه‌های  $V$  و  $A$ ، امکان یادگیری مناسب با مطالعه‌ای محدود در دامنه‌ای پیچیده و همزمان فراهم می‌شود. مجموعه‌های  $V$  و  $A$  باید به گونه‌ای انتخاب شوند که تا حد امکان در برگیرنده کلیه حالات و اعمال باشد و نگاشتهای خوبی از مجموعه‌های حالات و اعمال ممکن در دامنه محیط در بردارنده عامل محسوب شود.

روش پیشنهادی برای عمومی‌سازی محیط بدین صورت است که فضای مستطیل شکل زمین فوتیال که محیط دامنه چند عامله محسوب می‌شود به ۱۵۰ مربع یکسان با ضلع ۷ متر تقسیم می‌شود. از این طریق محیط پیوسته پرامون عامل به محیطی گسته تبدیل می‌گردد. در هر لحظه از زمان بازی، هر عامل درون یکی از این مربع‌ها قرار دارد. با توجه به این نکته که هر عامل دارای دیدی محدود می‌باشد برای عاملی که در مربع حالت  $X$  قرار دارد، ۲۴ مربع اطراف آن را عنوان شاعع دید آن عامل در نظر می‌گیریم. به شکل ۲۴ توجه کنید.

در این شکل، مربعهای  $X+10, X+11, X+1, X-1, X-10, X-9, X+1, X+9$  را مربع‌های بلا فاصله اطراف مربع  $X$  می‌نامیم. شماره‌گذاری مربع‌ها از چپ به راست و بصورت ستونی و با شروع از شماره صفر برای اولین مربع (متناهی‌الیه چپ و بالا) تا شماره ۱۴۹ برای آخرین مربع (متناهی‌الیه راست و پایین) انجام گرفته است. هر سه تن دارای ۱۰ مربع و هر سطر دارای ۱۵ مربع می‌باشد. در شکل‌های ۷ و ۸، مربع‌های ۸ جهت اطراف مربع دربردارنده عامل (با اختیار جهت شمال به سوی دروازه حریف) نشان داده شده است. در روش پیشنهادی برای هر کدام از ۸ مربع اطراف عامل یعنی برای هر یک از ۸ جهت (شمال غربی، شمال، شمال شرقی، شرق، جنوب شرقی، جنوب، جنوب غربی و غرب) اطراف آن، یک کمیت عددی در نظر گرفته شده است.

به عنوان مثال، اگر عامل واقع در مربع حالت  $X$ ، یکی از عاملهای حریف را در درون یکی از مربع‌های شمال غرب خود (مربعهای  $X+9, X+8, X+18$  و  $X+19$ ) بینند، یک مقدار عددی منفی به کمیت عددی متناظر با جهت شمال غرب خود اضافه می‌کند. این وضعیت برای عاملهای هم تیمی برعکس است. به عنوان مثال، اگر عامل صاحب توب واقع در مربع حالت  $X$  (یکی از عاملهای خودی را در درون یکی از



شکل ۵: نمودار تغییر وضعیت اتوماتی Krylov

### ۳-۲ اتوماتی یادگیر با ساختار متغیر

اتوماتی یادگیر با ساختار متغیر توسط ۴ تائی  $\{\alpha, \beta, p, T\}$  نشان داده می‌شود که در آن  $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_r\}$  مجموعه عمل‌های اتومات،  $\beta = \{\beta_1, \beta_2, \dots, \beta_m\}$  مجموعه ورودی‌های اتومات،  $p = \{p_1, p_2, \dots, p_r\}$  بردار احتمال انتخاب هر یک از عملها و  $T(n+1) = T[\alpha(n), \beta(n), p(n)]$  الگوریتم یادگیری می‌باشد. در این نوع از اتومات‌ها، اگر عمل  $i$  در مرحله  $n$  ام انتخاب شود و پاسخ مطلوب از محیط دریافت نماید، احتمال  $p_i(n)$  کاهش یافته و سایر احتمالها افزایش می‌یابند. در هر حال، تغییرات به گونه‌ای صورت می‌گیرد تا حاصل جمع  $p_i(n)$  ها همواره مساوی یک باقی بماند. الگوریتم زیر یک نمونه از الگوریتم‌های یادگیری خطی در اتوماتی با ساختار متغیر است

الف- پاسخ مطلوب

$$p_i(n+1) = p_i(n) + a [1 - p_i(n)]$$

$$\forall j \quad j \neq i \quad p_j(n+1) = (1-a)p_j(n)$$

ب- پاسخ نامطلوب

$$p_i(n+1) = (1-b)p_i(n)$$

$$p_j(n+1) = \frac{b}{r-1} + (1-b)p_j(n) \quad \forall j \quad j \neq i$$

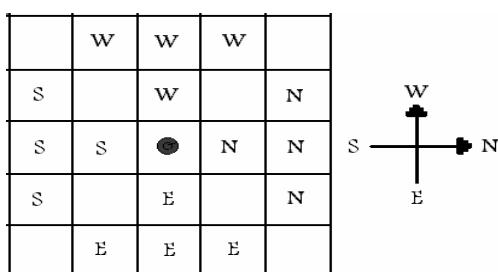
در روابط فوق،  $a$  پارامتر پاداش و  $b$  پارامتر جریمه می‌باشد. با توجه به مقادیر  $a$  و  $b$  سه حالت زیر را می‌توان در نظر گرفت. زمانی که  $a$  و  $b$  با هم برابر باشند، الگوریتم را  $L_{RP}$  می‌نامیم. وقتی که  $a$  از  $b$  از  $L_{REP}$  می‌نامیم و هنگامی که  $b$  خیلی کوچکتر باشد، الگوریتم را  $L_{RI}$  می‌نامیم. حافظه و زمان مورد نیاز برای پیاده‌سازی اتومات‌ها یادگیری با ساختار ثابت ( $O(1)$ ) و برای اتومات‌ها یادگیری با ساختار متغیر ( $O(m)$  می‌باشد که  $m$  تعداد اعمال اتومات است. برای مطالعه بیشتر درباره اتومات‌ها یادگیر می‌توان به مراجع [۷] تا [۱۱] مراجعه نمود.

### ۳- تکنیک بهترین گوشه در مربع حالت

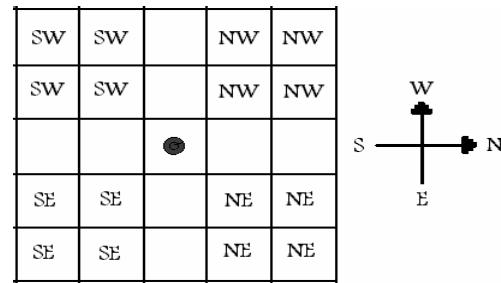
همان گونه که قبلاً اشاره شد، تعداد حالات در دامنه فوتیال روبوتیک شبیه‌سازی شده، بسیار زیاد است و لذا امکان در نظر گرفتن کلیه این حالات برای یک عامل عملاً غیرممکن است. بهمین دلیل ایجاد یک روش عمومی‌سازی مناسب از حالات محیطی امری ضروری است.

با فرض آنکه مجموعه حالات دامنه  $S$  و مجموعه حالات نگاشت  $V$

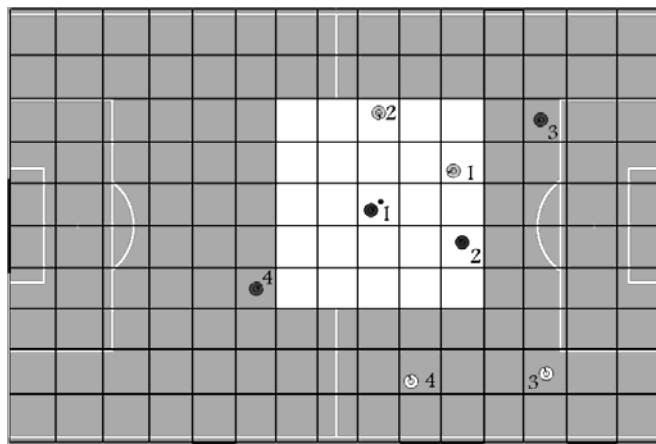
1. Linear Reward Penalty
2. Linear Reward Epsilon Penalty
3. Linear Reward Inaction



شکل ۶: مربعهای شمال، شرق، غرب، و جنوب مربع دربردارنده عامل با توجه به موقعیت عامل مورد نظر و چهار جهت اصلی.



شکل ۷: مربعهای شمال غرب، شمال شرق، جنوب غرب، و جنوب شرق مربع دربردارنده عامل با توجه به موقعیت عامل مورد نظر و چهار جهت.



شکل ۸: وضعیت بازیکنان هم تیمی و حریف با توجه به فضای محلی عامل صاحب توب در یک بازی ۴ در مقابل ۴.

می‌دهد. یعنی، عامل صاحب توب در هر سیکل، با توجه به اعداد بدست آمده برای ۸ گوشه مربع حالات در بردارنده خود، "بهترین گوشه در مربع حالت" حالت خود را محاسبه می‌کند. درباره چگونگی انتخاب ۲۴ مربع اطراف هر عامل که فضای محلی یا مهم عامل را تشکیل می‌دهد می‌توان به شکل ۹ مراجعه کرد. در این شکل، فضای محلی و مهم برای یک عامل در روش بهترین گوشه در مربع حالت، در مقایسه با کل فضای بازی نشان داده شده است. این فضای بجای کل فضای بازی برای تصمیم‌گیری محلی یک عامل مورد استفاده قرار می‌گیرد.

شکل ۹ یک بازی ۴ در مقابل ۴ را به تصویر کشیده است. در این شکل، برای بازیکن صاحب توب، بازیکن ۲ هم تیمی و بازیکن ۱ و ۲ از تیم حریف صاحب اهمیت هستند (به این دلیل که درون فضای محلی وی قرار دارند) و سایر بازیکنان نقشی در تصمیم‌گیری بازیکن صاحب توب ندارند.

با توجه به این که حالت هر عامل در محیط در هر لحظه تا حد زیادی به فاصله و زاویه دیگر عامل‌ها نسبت به آن عامل بستگی دارد نگاشت بایستی بصورتی انجام گیرد که حالت‌های یکسان به گوشه‌ای واحد و حالت‌های متفاوت به گوشه‌های متفاوت (در مربع حالت عامل) نگاشت شود.

سناریوی کلی برای هر عامل در شبیه‌سازیهای انجام گرفته بدین صورت است که اگر عامل صاحب توب، بازیکنی غیر از خود را در مسیر به سمت دروازه حریف ببیند، با تعیین حالت خود و انتخاب عمل بهینه در آن حالت (با استفاده از اتوماتای یادگیر وابسته به آن حالت)، سعی می‌کند حرکتی را انجام دهد که در جهت هدف تیمی که همانا برد است باشد.

اگر بازیکنی بازیکن تیم خود را صاحب توب ببیند، موقعیت خود و دیگر بازیکنان موجود در حوزه دید خود را به بازیکن صاحب توب اعلام می‌کند تا بازیکن صاحب توب، حتی المقدور، بهترین تخمین را از حالتی که در آن قرار دارد داشته باشد. این امر با توجه به این نکته که دید هر بازیکن

مربعهای جنوب غرب خود (مربعهای X-۲۱ و یا X-۱۱ و یا X-۱۲) ببیند، یک مقدار عددی مثبت به کمیت عددی متناظر با جهت جنوب غرب خود اضافه می‌کند. لازم به ذکر است که اندازه این افزایش (منفی و یا مثبت) در ضریب اطمینان هر جهت با معکوس فاصله عامل حریف و یا عامل هم تیمی از عامل مورد نظر (و در واقع صاحب توب) در آن جهت، متناسب است. در واقع، عامل‌های نزدیک‌تر (و درون فاصله محلی تا ۲۴ مربع حالت پیرامون عامل صاحب توب) تاثیر بیشتر و عامل‌های دورتر (و درون فاصله محلی تا ۲۴ مربع حالت پیرامون عامل صاحب توب) تاثیر کمتری را بر روی ضرایب اطمینان عامل صاحب توب خواهند داشت و عامل‌های خارج از فاصله محلی (خارج از ۲۴ مربع حالت پیرامون عامل)، تاثیری در تغییر ضرایب اطمینان عامل صاحب توب ندارند.

هر عامل با محاسبه موقعیت تمامی عامل‌های هم تیمی و حریف پیرامون خود (درون ۲۴ مربع)، ۸ عدد در اختیار خواهد داشت. با متناظر کردن این اعداد با ۸ جهت پیرامون عامل صاحب توب، عامل صاحب توب دارای ۸ حالت و ۸ عمل (که متناظر با ارسال توب به سمت مرکز ۸ مربع بالاً فاصله پیرامون عامل می‌باشند) خواهد بود. لازم ذکر است که تقسیم زمین به مربعهای ۷×۷ متر مربعی براساس تجربه کسب شده در کار با محیط شبیه‌ساز و میزان جابجایی توب در اثر هر ضربه حاصل شده است و عملاً وابسته به دامنه انتخابی است. بدین ترتیب، فضای حالت پیرامون عامل به ۸ حالت کاهش می‌یابد و مشکل نگاشت حالات متفاوت به یک حالت بخصوص تا حد زیادی از بین می‌رود.

ممکن است در یک لحظه چندین عامل هم تیمی و یا حریف در درون ۲۴ مربع حالت پیرامون مربع حالت عامل وجود داشته باشند. حال اگر عامل یا عامل‌هایی درون همان مربعی قرار داشته باشند که عامل صاحب توب در آن قرار دارد، عامل صاحب توب بر حسب آنکه عامل‌های فوق در کدام یک از ۸ جهت خود قرار دارند کمیت‌های عددی خود را تعییر

جدول ۱: نتایج میانگین گلهای زده در ۵۰ بازی بین تیمهای یادگیر ۲ نفره با تیم بدون یادگیر ۲ نفره.

| ۶    | ۵    | ۴    | ۳    | ۲   | ۱   |                        |
|------|------|------|------|-----|-----|------------------------|
| ۵-۱۶ | ۵-۱۲ | ۴-۹  | ۴-۷  | ۳-۳ | ۲-۱ | $L_{\gamma N, \gamma}$ |
| ۵-۶  | ۳-۱۳ | ۲-۱۰ | ۲-۸  | ۱-۵ | ۱-۲ | $G_{\gamma N, \gamma}$ |
| ۴-۱۷ | ۳-۱۵ | ۳-۱۱ | ۳-۷  | ۲-۴ | ۱-۲ | Krinsky                |
| ۳-۱۲ | ۳-۱۱ | ۳-۹  | ۲-۶  | ۱-۴ | ۱-۲ | Krylov                 |
| ۴-۱۷ | ۳-۱۵ | ۳-۱۳ | ۲-۱۰ | ۲-۵ | ۱-۳ | Q                      |

همکاری مورد ارزیابی قرار دهیم.

#### ۱-۴ شبیه‌سازیها برای تیم ۲ نفره

چندین سری شبیه‌سازی انجام گرفته است. در اولین سری شبیه‌سازیها، به پیاده‌سازی تیمهای ۲ نفره از عاملها می‌پردازیم. این سری از شبیه‌سازیها با دو روش برای تعیین حالت هر عامل در محیط خود انجام می‌گیرد. یکی از این دو روش یک عمومی‌سازی ساده و دیگر روش، تکنیک "بهترین گوشش در مربع حالت" که قبلاً در این مقاله به آن اشاره شده است می‌باشد. در روش عمومی‌سازی ساده، کلیه حالات محیط به ۴ حالت برای بازیکن دارای توب و ۴ حالت برای بازیکن بدون توب خالصه می‌شود [۱۳]. برای هریک از این چهار حالت یک اتوماتای یادگیر با ساختار ثابت و با عمق حافظه ۳ درنظر گرفته می‌شود. هر اتوماتا، دارای یکی از دو عمل پاس به هم تیمی و یا شوت به طرف دروازه حریف می‌باشد.

نتایج آزمایشها اولیه [۱۳] نشان دادند که تیم دارای اتوماتای یادگیری در مقایسه با یک تیم بدون یادگیری به سرعت یاد می‌گیرد که در چه حالاتی، باستی چه عملی را انجام دهد و به همین دلیل می‌تواند براحتی بر حریف خود غلبه کند. به دلیل این که یادگیری همزمان با بازی انجام می‌گیرد تیم اتوماتای یادگیر می‌تواند در حین بازی خود را با نحوه بازی تیم حریف تا حد زیادی تطبیق دهد. تعداد ۵۰ بازی را بین تیمهای اتوماتای با ساختار ثابت و تیمی مبتنی بر یادگیری Q با تیم بدون یادگیری انجام دادیم که نتایج این بازیها در جدول ۱ آمده است. در این جدول، عدد سمت راست نشان‌دهنده تعداد گلهای زده توسط تیم یادگیر و عدد سمت چپ نشان‌دهنده تعداد گلهای زده توسط تیم بدون یادگیری می‌باشد. همانگونه که مشاهده می‌شود کلیه بازیها به سود تیمهای یادگیر به پایان رسیده است. لازم بذکر است که در جدول ۱، ستون ۱ به معنی نتیجه تجمعی بازی از سیکل ۰ تا سیکل ۰، ستون ۹۹۹ به معنی نتیجه تجمعی بازی از سیکل ۱۰۰۰ تا سیکل ۱۹۹۹، ...، و ستون ۶ به معنی نتیجه تجمعی بازی از سیکل ۵۰۰۰ تا آخر بازی (سیکل ۵۹۹۹) می‌باشد. همان‌گونه که این جدول نشان می‌دهد، تفاوت زیادی بین اتوماتاهای یادگیر مختلف در سری شبیه‌سازیها انجام شده مشاهده نمی‌شود، هرچند که روش‌های یادگیری Q و اتوماتای Krinsky اندکی بهتر از دیگر اتوماتاهای عمل کرده است.

همان‌گونه که جدول ۱ نشان می‌دهد، تیم اتوماتای یادگیر موفق شده است تدریجاً در حین بازی، انجام عمل صحیح پاس و شوت را در ۴ حالت تعريف شده برای آن (در این سری از شبیه‌سازیها) فرا بگیرد. به همین دلیل اکثر گلهای دریافتی تیم اتوماتای یادگیر در نیمه اول (۳۰۰۰ سیکل اول) و اکثر گلهای زده شده توسط تیم اتوماتای یادگیر در نیمه دوم بازی (۳۰۰۰ سیکل دوم) به ثمر می‌رسد. باید خاطرنشان کرد که روش یادگیری Q که در دامنه روبوکاب استفاده شده است دارای محدودیت‌هایی است. به عنوان مثال، در این دامنه، هر عامل دارای

محدود است و هر بازیکن کنار و پشت خود را نمی‌بیند، برای بازیکن صاحب توب حیاتی است و لذا در این مورد (و تنها در این مورد) از امکان شنوایی (و آنهم بصورت محدود) استفاده شده است. در واقع در چنین حالتی، بازیکن بدون توب، مدل دنیای خود را به بازیکن صاحب توب هم تیمی خود اعلام می‌کند و وی را در تعیین حالت و به تبع آن، تعیین عمل مناسب در آن حالت، بازیکن بدون توب نیز، به مرکز مربع بالاfrac{اصله} (از مربع حالت در بردارنده خود) می‌رود که به منظور دریافت توب (در صورت لزوم) از بازیکن هم تیمی صاحب توب، مناسب تشخیص می‌دهد. بدین نحو، در شبیه‌سازیها انجام گرفته، بازیکن بدون توب نیز می‌تواند دارای حالات و اعمال خاص (حرکت به سمت یکی از ۸ جهت اطراف خود به منظور دریافت توب) باشد. یعنی بازیکن می‌تواند "حرکت بدون توب" انجام دهد. باید توجه داشت که عمل مناسب انتخابی عامل صاحب توب در هر حالت، الزاماً به معنی پاس و یا شوت نمی‌باشد و در بعضی موارد، ممکن است دریبل و یا نگهدارشتن موقعت توب به منظور یافتن روزنای برای انجام عمل مناسب باشد.

همچنین در روش پیشنهادی، بازیکن خود را ملزم به حرکت رو به جلو (در همه حالات) نمی‌داند و در صورت لزوم از مفاہیم اوت کدن توب، پاس به عقب، و حتی بازکردن بازی و ارسال توب به فضای خالی (با احتمال دریافت مناسب توب برای بازیکن هم تیمی و یا خود در چند سیکل بعد) نیز استفاده می‌کند [۱۳].

لازم به ذکر است که یادگیری در کلیه شبیه‌سازیها انجام گرفته کاملاً چندعامله و توزیع شده است و برخلاف روش‌های موجود، بازیکن صاحب توب (بدون توب)، قصد خود را از ارسال (دریافت) توب به (از) بازیکن بدون توب (صاحب توب) اعلام نمی‌دارد و بدین ترتیب، هر عامل در انتخاب عمل بهینه خود کاملاً خودمنختار است و در حین انجام بازی بدون ایجاد ارتباط با دیگر بازیکنان سعی می‌کند که با آنها همکاری کنند. شبیه‌سازیها انجام شده برای تیمهای ۲ نفره، ۵ نفره و ۱۱ نفره [۱۳] که بعداً در این مقاله ارایه خواهد گردید نشان دهنده این قابلیت می‌باشد.

#### ۴- همکاری بین اعضای یک تیم چند عامله با استفاده از اتوماتای یادگیر

تا حال از روش‌های مختلفی از جمله یادگیری Q، الگوریتم‌های ژنتیک، درختهای تصمیم‌گیری و یادگیریهای رفتاری [۲] برای یادگیری عامل‌های فوتیالیست استفاده شده است. در این بخش از مقاله، توانایی اتوماتای یادگیر به منظور ایجاد همکاری بین عامل‌های موجود در یک تیم روبوکاب برای رسیدن به یک هدف مشخص تیمی که همانا برد باشد را مورد بررسی قرار می‌دهیم. در آزمایشها انجام شده در این مقاله، چند بازیکن فوتیال که هر کدام مجهز به یک اتوماتای یادگیر می‌باشند را در مقابل چند بازیکن بدون توانایی یادگیری و یا دارای روش‌های دیگر یادگیری قرار داده‌ایم تا این طریق مدل اتوماتای یادگیر را در ایجاد

جدول ۲: نسبت گل زده به گل خورده در روش‌های مختلف یادگیری با استفاده از تکنیک بهترین گوشه در مربع حالت و در بازیهای ۲ در مقابل ۲ (در حین آموزش)

| نسبت گل زده به گل خورده در حین آموزش | نام روش  |
|--------------------------------------|--|
| ۱/۲۶                                 | اتوماتای $L_{2N,2}$ با عمق حافظه ۳ (۱۰ بازی)     |
| ۱/۱۷                                 | اتوماتای $L_{2N,2}$ با عمق حافظه ۲ (۱۰ بازی)     |
| ۱/۲۵                                 | اتوماتای $G_{2N,2}$ با عمق حافظه ۲ (۱۰ بازی)     |
| ۱/۳۰                                 | اتوماتای Krylov با عمق حافظه ۲ (۱۰ بازی)         |
| ۰/۸۲                                 | اتوماتای Krinsky با عمق حافظه ۲ (۱۰ بازی)        |
| ۱/۱۸                                 | اتوماتای $L_{RP}$ (۳۰ بازی)                      |
| ۱/۰۵                                 | اتوماتای $L_{RP}$ (۱۵ بازی)                      |
| ۱/۰۷                                 | اتوماتای $L_{REP}$ با نسبت پارامتر ۰/۱ (۱۵ بازی) |
| ۱/۱۲                                 | اتوماتای $L_{REP}$ با نسبت پارامتر ۰/۰ (۱۵ بازی) |
| ۱/۰۲                                 | یادگیری Q (۱۰ بازی)                              |

می‌باشد. لهذا برای عمومی‌سازی بهتر حالات محیطی از تکنیک "بهترین گوشه در مربع حالت" استفاده خواهیم کرد. همچنین تعداد اعمال هر عامل که به پاس دادن به هم تیمی و یا شوت به دروازه حریف خلاصه می‌شد نیز افزایش داده شد. کلیه شبیه‌سازیهایی که در ادامه این مقاله آمده است با توجه به دو مورد فوق انجام گرفته است و شبیه‌سازیهای ۲ بازیکن در مقابل ۲ بازیکن مجدداً انجام گرفت.

برای آزمایش‌های جدید دو تیم با نامهای "یادگیری کامل" و "نگاشت ثابت" ساخته شدند. در تیم "یادگیری کامل" در هر مربع حالت از فضای محیط عامل، ۸ اتوماتا تعییه کرده‌ایم. در این تیم یک عامل، اتوماتا و یا اتوماتاهایی که به آنها تجهیز شده است با حرکت در محیط با خود را نمی‌کند. تیم "نگاشت ثابت" یک تیم بهینه شده با دست است. این دو با تیم‌های دیگر مقایسه گردیده است [۱۳].

در یک سری از آزمایش‌های تیم یادگیر  $L_{RP}$  با استفاده از تکنیک "بهترین گوشه در مربع حالت" را در چند بازی آموزشی در مقابل تیم  $L_{2N,2}$  ۲ نفره که از عمومی‌سازی ساده حالات محیطی بخشن قبل استفاده می‌کرد، قراردادیم. مجموعاً ۳۰ بازی بین تیم  $L_{RP}$  با تیم  $L_{2N,2}$  برگزار کردیم که نتیجه ۶ بازی اول آنها در زیر آمده است. در مجموع ۳۰ بازی، ۲۸ بار تیم  $L_{RP}$  برنده بازی بود و یک بار نیز بازی به تساوی کشیده شد [۱۳].

می‌توان از تعداد گلهای زده و خورده تیم‌های یادگیر مختلف زمانی که در مقابل تیم بدون یادگیری بازی می‌کنند استفاده کرد و تاثیر روش‌های یادگیری متفاوت را در کارایی تیم زمانی که از تکنیک بهترین گوشه در مربع حالت استفاده می‌شود را مورد بررسی قرار داد. نتایج حاصله که در جدول ۲ آمده است نشان می‌دهد که عامل در تعیین حالت و انتخاب عمل مناسب در آن حالت بهتر عمل می‌نماید. همچنین می‌توان این نتیجه را گرفت که روش بهترین گوشه در مربع حالت، عمومی‌سازی مناسبتری از حالات محیط و نگاشت حالات یکسان به یک حالت واحد را انجام می‌دهد.

#### ۲-۴ شبیه‌سازیها برای تیم ۵ نفره

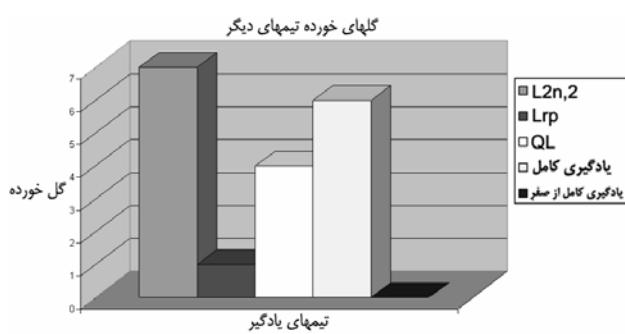
قبل از این که به ارایه نتایج شبیه‌سازیها برای تیم ۵ نفره پردازیم بینیم در چه مواقعی اتوماتای یادگیری که عامل به آن مجهرز است پاداش یا جریمه دریافت می‌کند. چگونگی دادن پاداش و یا جریمه به یک عمل به صورت زیر انجام می‌گیرد: اگر یک بازیکن عملی را انجام دهد در

دید محدود است و لذا نمی‌تواند از بقیه عاملها، محیط و همین طور تاثیر (بخصوص درازمدت) اعمال خود بر محیط و دیگر عاملها دیدی کامل داشته باشد. از این رو برای پیاده‌سازی یادگیری  $Q$ ، شبیه‌سازی ساده‌ای از روش یادگیری  $Q$  براساس روش بکار برده شده در [۲] (روش TPOT\_RL) انجام شده است و تهها تاثیرات کوتاه مدت اعمال انجام گرفته توسعه عامل، در تغییر مقادیر  $Q$  نقش دارند.

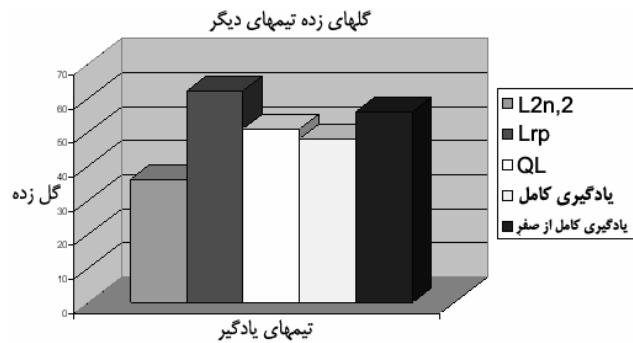
درباره شبیه‌سازیها دو نکته قابل ذکر است. نکته اول این که در کلیه شبیه‌سازیهای انجام شده، از عامل ثالثی به منظور مشاهده روند کار عامل‌ها استفاده نشده است بلکه از خود عامل جهت قضایت تاثیر عمل خود استفاده کرده‌ایم که این خود می‌تواند دلیلی بر تاکید بر یادگیری در یک سیستم چندعامله در شبیه‌سازیهای انجام شده در این مقاله باشد. نکته دوم این است که عامل نمی‌تواند همه تغییرات در محیط خود را مشاهده کند و همواره حالات مخفی و پوشیده در محیط وجود دارد و به همین دلیل در اکثر موارد، اولین تاثیر قابل مشاهده عمل برای عامل انجام‌دهنده آن عمل، تخمین مناسبی از تاثیر درازمدت عمل انجام شده توسط عامل می‌باشد. قضایت براساس اولین تاثیر قابل مشاهده توسعه (دید محدود) خود بازیکن، در کلیه شبیه‌سازیهای انجام شده در این مقاله رعایت شده است.

شبیه‌سازیهای دیگر نشان دادند که تیم‌های مجهز به اتوماتای یادگیر ۲ نفره بر تیم‌های بدون یادگیری با تعداد نفرات بیشتر (۳ یا ۴ نفره) نیز غلبه نمودند. نکته قابل توجه دیگر که از آزمایشها حاصل شد این است که یادگیری چندعامله تا حدی وابسته به حریف است. یعنی مقدار از قبل تعیین شده‌ای برای همگرایی مقادیر حافظه بازیکنان وجود ندارد. این بدان معنی است که بازیکنان در بعضی موارد که از نظر ما منطقاً باید پاس بدنه‌ند، یادگرفته‌اند که بهتر است به سمت دروازه حریف شوت کنند و این یکی از محسان یادگیری (بخصوص یادگیری همزمان با بازی) می‌باشد.

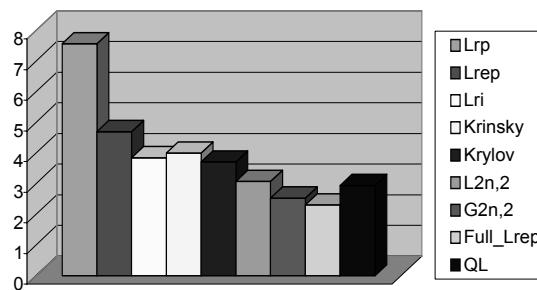
آزمایش‌هایی که تا بحال انجام گرفته است دارای دو ایراد می‌باشد. اولین ایراد این است که امکان نگاشت حالات بسیار متفاوتی از دنیای پیرامون عامل صاحب توب به یک حالت واحد وجود دارد و لذا ممکن است عملی که عامل بعنوان عمل صحیح در آن حالت انتخاب کرده است، بهترین عمل ممکن برای آن عامل نباشد. شبیه‌سازی‌ها نشان دادند که عامل در شناخت حالت خود دچار مشکل می‌باشد. برای حل این مشکل می‌توان از روش "عمومی‌سازی" بهتری از حالات محیطی برای عامل استفاده کرد. ایراد دیگر تعداد کم اعمال تعریف شده برای هر عامل



شکل ۱۱: گلهای خورده تیمهای یادگیر ۵ نفره در ۱۰ شبیه‌سازی (در حین آموزش).



شکل ۱۰: گلهای زده تیمهای یادگیر ۵ نفره در ۱۰ شبیه‌سازی (در حین آموزش).



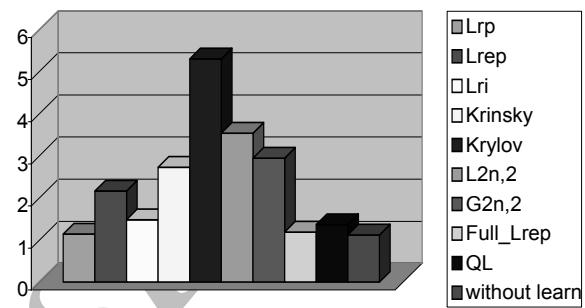
شکل ۱۳: مقایسه نسبت میانگین گلهای زده به خورده در هر بازی پس از آموزش (آزمایشی) در بازیهای تیم بدون یادگیری با تیمهای دیگر.

بازیهای ۲ نفره در هیچ دیداری بازنشده نبودند. نتایج آزمایشها در اشکال ۱۰ و ۱۱ آمده است.

### ۳-۴ شبیه‌سازیها برای تیم ۱۱ نفره

شبیه‌سازیها ارایه شده در این قسمت به بررسی یادگیری در نحوه همکاری در بین عاملهای موجود در یک تیم ۱۱ نفره و مقایسه آن با دیگر تیمهای می‌پردازد. با توجه به این که در بازیهای ۱۱ نفر در مقابل ۱۱ نفر، ممکن است تعداد گلهای رد و بدل شده خیلی زیاد نباشد به منظور بررسی نتایج شبیه‌سازیها در این سری (و سریهای بعد)، از معیارهای دیگری هم بمنظور نشان دادن کارایی استفاده می‌شود. پس از بررسی معیارهای مطرح در همکاری بین بازیکنان یک تیم فوتbal، معیارهای زیر بنظر مناسب می‌باشد.

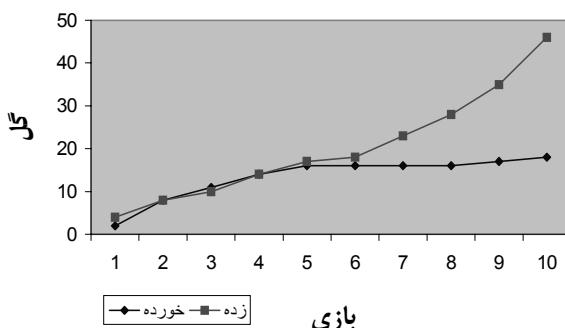
- درصد مالکیت توب توسط تیم خودی در مقایسه با مورد مشابه در تیم حریف در حین بازی.
  - درصد گردش توب در  $1/3$  زمین خودی،  $1/3$  میانی زمین، و  $1/3$  زمین حریف در حین بازی.
  - حداقل زمان در اختیار داشتن توب بصورت ممتد توسط تیم که بر حسب سیکل سنجیده می‌شود.
  - حداقل تعداد رد و بدلهای متولی توب بدون برخورد با حریف.
  - میانگین درصد خطای (اعمال) بازیکنان تیم خودی در حین بازی.
- برای سازماندهی ۱۱ بازیکن درون زمین برای هر تیم آرایش تیمی ۴-۳-۳-۳ انتخاب گردیده است. همانند شبیه‌سازیها قبلی، در این سری از آزمایشها نیز برای مقایسه روند یادگیری در تیمهای یادگیر، یک تیم بدون یادگیری ایجاد کردیم. این تیم بدون یادگیری، به غیر از مساله یادگیری در همه موارد از جمله شکل تیمی مشابه تیمهای یادگیر است. در شبیه‌سازیها انجام شده تیمهای اتوماتی یادگیر موفق شدند پس از انجام تعدادی محدود بازی آموزشی بر حریف بدون یادگیری غلبه کنند. آنها همچنین توانستند تیم "نگاشت ثابت" را شکست دهند. شکل‌های ۱۲



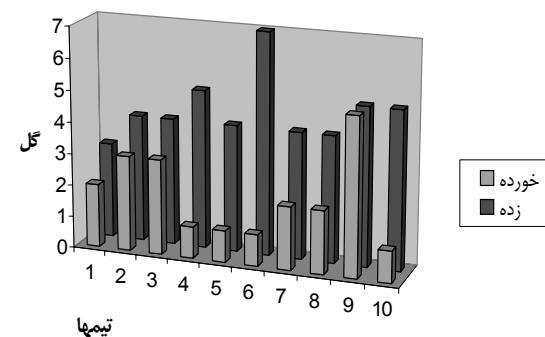
شکل ۱۲: مقایسه نسبت میانگین گلهای زده به خورده در هر بازی آموزشی در بازیهای تیم بدون یادگیری با تیمهای دیگر.

صورتی پاداش می‌گیرد که تیم صاحب توب بعد از خاتمه این عمل هنوز تیم اوست و توب به سود تیم او جلوتر رفته است (بطور مثال تیم او ۵ گل زده است) و در غیر این صورت بازیکن جرمیه می‌شود. همچنین اگر انجام دهنده آخرین عمل، خود من بوده‌ام (یا اگر هنوز نتیجه آخرین عمل خود را ندیده‌ام) و تیم صاحب توب، تیم حریف است و اگر فاصله موقعیت قبلی توب تا موقعیت فعلی توب خیلی زیاد نیست (بطور مثال به تیم حریف ۵ نزدیک) بازیکن جرمیه می‌شود و در غیر این صورت به او پاداش داده می‌شود. نتیجه عمل که پاداش یا جرمیه می‌باشد توسط اتوماتی یادگیر استفاده می‌شود تا حالت داخلی خود را به روز نماید. چگونگی به روز آمدن حالت اتوماتا بستگی به نوع اتوماتای یادگیر دارد. برای تیمهای ۱۱ نفره که در ادامه این قسمت به آن پرداخته می‌شود روش دادن پاداش یا جزا نیز به همین ترتیب خواهد بود.

در ادامه به بررسی شبیه‌سازیها برای تیمهای ۵ نفره می‌پردازیم. مهمترین تغییری که در این شبیه‌سازیها نسبت به شبیه‌سازیها قبل از انجام گرفته است دادن شکل تیمی به تیم و استفاده از بازیکنان تیم در پستهای تخصصی بوده است. دادن شکل تیمی با توجه به افزایش نفرات تیم جهت برقراری نظم و همکاری هر چه بهتر در بین عاملهای بمنظور ضروری می‌رسید. علت انتخاب ۵ نفر برای یک تیم، شبیه‌سازی مواردی چون مسابقات رویکاپ با روابطهایی با اندازه‌های متعدد و بزرگ و نیز فوتbal داخل سالنی بوده است. در شبیه‌سازیها ۵ نفره، تیمهای یادگیر ۵ نماینده تیمهای یادگیر با شکل تیمی  $L_{2N,2}$  بعنوان نماینده تیمهای یادگیر بازیکنان تیمی  $L_{RP}$  بعنوان نماینده تیمهای یادگیر با ساختار متغیر،  $Q$  بعنوان نماینده‌ای از تیمهای دارای روش یادگیری به غیر از اتوماتای یادگیر، و "تیم یادگیری کامل" [۱۳] استفاده کردیم. در شبیه‌سازیها این قسمت، از ۱ دروازه‌بان، ۲ بازیکن کناری چپ و راست، یک دفاع عقب و یک بازیکن حمله استفاده شد است. به هر کدام از این بازیکنان فضای مشخصی از زمین که در واقع محدوده‌ای از مربعهای حالت می‌باشد تخصیص داده شده است. در بازیهای انجام شده، تیمهای یادگیر، همانند



شکل ۱۵: مقایسه گلهای زده و خورده در طی ۱۰ بازی آموزشی متولی با پارامترهای تصادفی اضافه شده زمانیکه تیم اتوماتای  $L_{R_{EP}}$  با تیم بدون یادگیری بازی می‌کند.



شکل ۱۶: مقایسه گلهای خورده (میله های جلو) و گلهای زده (میله های عقب) تیمهای مختلف در مقابل تیم دستنویس (ثابت)، از چپ به راست: بدون یادگیری، Q، Krinsky، Krylov،  $L_{RN,2}$ ،  $L_{REP}$ ،  $L_{RI}$ ،  $L_{RP}$ ،  $L_{R_{EP}}$ ،  $G_{RN,2}$ ،  $G_{REP}$ ،  $G_{RI}$ ،  $G_{RP}$ .

تیم 2001 Saloo توسط نودا نوشته شده است. نودا در شرح تیم خود تقلید را عنوان اولین قدم برای سازگاری یک عامل با عامل‌های دیگر در محیطی چندعامله می‌داند [۱۶]. او از یک شبکه عصبی بازگشتی برای یادگیری و برنامه‌ریزی استفاده نموده است. این شبکه عصبی دارای ۲ وظیفه است: پیشگویی محیط و شناخت بازی. شبکه عصبی استفاده شده می‌تواند عنوان یک تمیزدهنده بین انواع بازیها آموزش داده شده مورد استفاده قرار گیرد. وی نتیجه‌گیری کرده است که این شبکه به همراه معماری وی می‌تواند انواع گوناگونی از بازیها را مجددًا تولید کند [۱۶]. در آزمایش‌های این قسمت از تیم اتوماتای  $L_{R_{EP}}$  و تیم یادگیری کامل  $L_{R_{EP}}$  استفاده شده است. این انتخاب بدین دلیل بوده است که این تیمهای در شبیه‌سازی‌های قبلی [۱۲] تا [۱۵] نتایج خوبی را تولید کرده است. در اولین سری از شبیه‌سازی‌ها به ارزیابی تیم اتوماتای یادگیر در شرایطی که نویز در محیط وجود داشته باشد می‌پردازیم. پارامترهای متعددی در کارگزار شبیه‌سازی فوتbal وجود دارند که می‌توان با تغییر آنها، شرایط حاکم بر زمین بازی، حرکت بازیکنان، حرکت توپ و ... را تغییر داد و یا در آنها اختلال ایجاد کرد.

در اولین نمونه از شبیه‌سازی‌های، اثر پارامترهای rand که معرف مقدار نویز در موارد مختلف هستند مورد بررسی قرار داده شده است. بدینصورت که پارامتر player\_rand را از مقدار  $0/0.1$  تا  $0/0.5$ ، پارامتر ball\_rand را از  $0/0.1$  تا  $0/0.5$  و بالاخره پارامتر kick\_rand را از  $0/0.1$  تا  $0/0.5$  تغییر دادیم. پارامتر اول باعث ایجاد اختلال (نویز) توسط کارگزار در حرکت بازیکن می‌شود و پارامترهای دوم و سوم، همین نقش را در مرور حرکت توپ در زمین و زدن ضربه به توپ بر عهده دارند. ۱۰ بازی متولی بین تیم  $L_{R_{EP}}$  (با یادگیری از صفر) و تیم بدون یادگیری برگزار نمودیم. هدف از این شبیه‌سازی این بود که بیننیم آیا روش‌های یادگیری پیشنهادی در این مقاله می‌توانند در این گونه شرایط نیز موثر واقع شوند. شکل ۱۵ نتایج را (تصویر تجمعی) نشان می‌دهد. در این شکل، کارآبی روش یادگیری پیشنهادی (اتوماتای یادگیر  $L_{R_{EP}}$ ) به صورت روش مشاهده می‌شود. همانگونه که شکل نشان می‌دهد، با افزایش تعداد بازیها، تیم اتوماتای یادگیر موفق می‌شود که رفته با شرایط محیطی سازگار شود و فاصله خود را با تیم بدون یادگیری بیشتر و بیشتر نماید. حتی در شرایط با نویز بالا نیز تیم اتوماتای یادگیر می‌تواند خود را با شرایط محیطی سازگار کند. گرچه بدليل تاکید بر خصوصیت خودمنظری عامل‌های تیمهای پیشنهادی از حداقل ارتباطات بین بازیکنان در حین بازی سود می‌برد ولی بدليل اینکه کanal ارتباطی بین بازیکنان شلوغ و فاقد اطمینان است، برای همین حجم اندک ارسال (و نه تبادل) اطلاعات اتکا به ارتباطات بین عامل‌ها در چنین دامنه‌هایی قابل اطمینان نیست. تیمهای اتوماتای

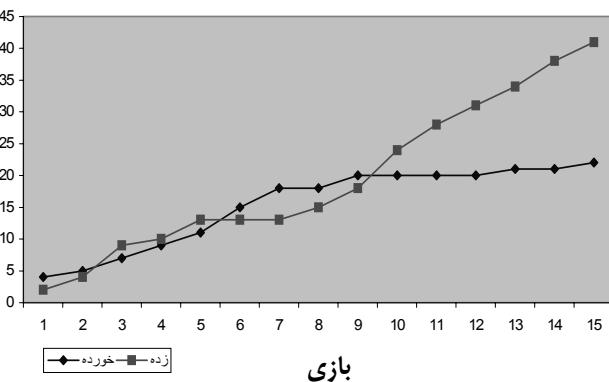
و ۱۳ نتایج کلی شبیه‌سازیها برای بازیهای انجام شده بین تیمهای یادگیر و تیم بدون یادگیری را در دوسری بازیهای آموزشی (۱۵ بازی) و بازیهای آزمایشی (۳ بازی پس از بازیهای آموزشی) خلاصه می‌کنند. همانگونه که مشاهده می‌شود با تعدادی محدود بازیهای آموزشی تیمهای اتوماتای یادگیر با ساختار ثابت به برتری نسبی می‌رسند. در مورد بازیهای آزمایشی (پس از آموزش و تنها با استخراج مقادیر یادگرفته شده) به نظر می‌رسد که برتری نسبی (در غلبه بر تیم بدون یادگیری) با تیمهای اتوماتای یادگیر با ساختار متغیر می‌باشد. جزییات این آزمایشها در [۱۳] آمده است. شکل ۱۴ گلهای زده و خورده تیمهای یادگیر را در مقابل تیم "نگاشت ثابت" خلاصه می‌کند.

## ۵- تستهای ارزیابی برای تیمهای مبتنی بر اتوماتای یادگیر

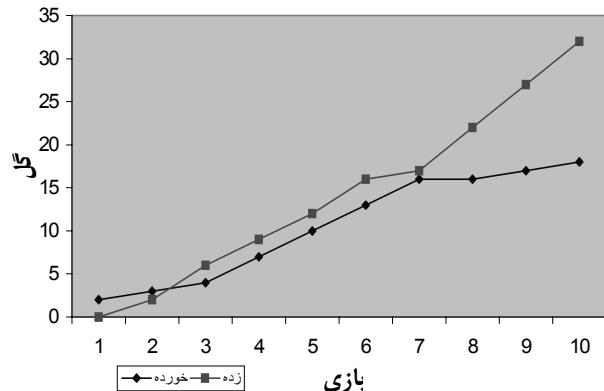
در بخش‌های قبلی، عملکرد اتوماتاهای یادگیر را در انجام یک کار گروهی در بین عاملهای عضو یک تیم مورد بررسی قرار دادیم. تمام شبیه‌سازیهای انجام شده در موارد فوق، در شرایط طبیعی دامنه برگزار شدند. در این بخش به ارزیابی تیمهای یادگیر از طریق بازی با بعضی از تیمهای شرکت‌کننده در مسابقات روبوکاپ جهانی می‌پردازیم.

گرچه تیم اتوماتای یادگیر بدليل پیاده‌سازی نکردن مواردی مانند استراتژی حمله، استراتژی دفاع، مدل کردن حریف، تغییر استراتژی بازی در زمانهای خاص، مربی و ... یک تیم مسابقه محسوب نمی‌شود ولی بازی با تیمهای شناخته شده که در مسابقات روبوکاپ شرکت می‌کنند، می‌تواند ما را در ارزیابی مدل اتوماتاهای یادگیر در یادگیری عامل‌ها یاری کند. لازم به ذکر است که تیم اتوماتای یادگیر یک تیم تحقیقاتی می‌باشد و صرفاً برای مطالعه و بررسی توانایهای اتوماتاهای یادگیر در همکاری بین عامل‌ها در انجام یک کار گروهی استفاده شده است.

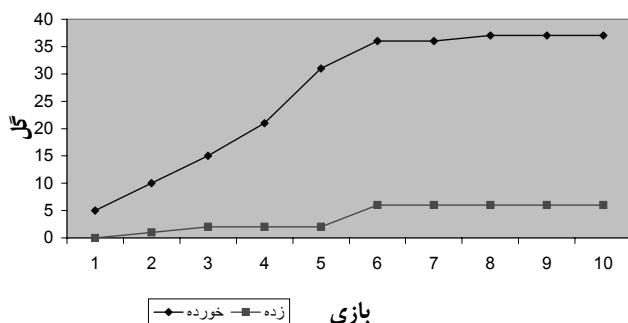
یادآوری می‌شود که کد پایه که برای پیاده‌سازی تیمهای اتواتهای یادگیر مورد استفاده قرار گرفته است، کد CMUnited98 [۲] می‌باشد. لازم به ذکر است که این تیم در مقایسه با تیمهای جدیدتر از مهارت‌های Sharif Arvand 2000، FuzzyFoo 2000، 2001، Yberoos 2000 و Saloo 2001 که از نظر فردی و گروهی بالایی بر خوردار نیست. تیمهای Sharif Arvand 2000، FuzzyFoo 2000، 2001، Yberoos 2000 و Saloo 2001 که از نظر می‌توانند عنوان حریف بازیها انتخاب شوند. از بین چهار تیم فوق، تیم Saloo 2001 بدليل نزدیکتر بودن مهارت‌های فردی با تیم خودی مناسب‌ترین تیم برای مسابقه به منظور ارزیابی روش اتوماتای یادگیر تشخیص داده شد و بهمین دلیل بازیها با این تیم انجام گرفت [۱۳]. برای مشاهده نتایج بازیها با دیگر تیمهای فوق می‌توان به [۱۳] مراجعه کرد.



شکل ۱۷: مقایسه گلهای زده و خورده در طی ۱۰ بازی آموزشی متوالی با برداشتمند بازیکن جناح چپ از تیم یادگیر زمانیکه تیم اتوماتای  $L_{REP}$  با تیم بدون یادگیری بازی می‌کند.



شکل ۱۶: مقایسه گلهای زده و خورده در طی ۱۰ بازی آموزشی متوالی بدون امکان ارتباط مابین بازیکنان بین تیم اتوماتای  $L_{REP}$  زمانیکه با تیم بدون یادگیری بازی می‌کند.

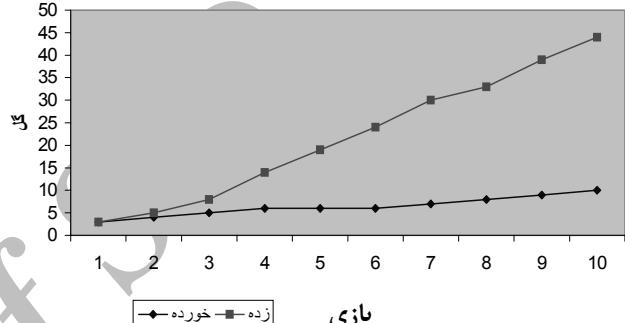


شکل ۱۹: مقایسه گلهای زده و خورده در طی ۱۰ بازی آموزشی متوالی بین تیم اتوماتای  $L_{REP}$  با تیم بدون یادگیری با افزودن بردار باد (۰ و ۵۰) به زمین بازی (جهت باد، عمود بر عرض زمین بازی و تیم یادگیر در خلاف جهت باد).

در شبیه‌سازیهای سری بعد، تیم اتوماتای  $L_{REP}$  را (با یادگیری از صفر) در چند بازی در مقابل تیم Saloo قرار دادیم. همانگونه که نتایج نشان می‌دهند [۱۳]، تیم Saloo در تمام بازیهای اولیه برندۀ بازی بوده است و با نفاذیت گل نسبتاً بالایی (میانگین گل زده  $5/7$  گل در مقابل گل خورده) در هر بازی اولیه در جدول ۳ آمده است. همانگونه که آمار بازی بدست آمده از ۷ بازی اولیه در جدول ۳ آمده است. همانگونه که آمار بازی نشان می‌دهد، تیم Saloo برتری مطلقی بر تیم اتوماتای یادگیر (در شروع یادگیری خود) دارد.

در ادامه آزمایشها تعداد بازیهای بین تیمها را افزایش دادیم. در طی ۱۵۰ بازی آموزشی (۲۵ ساعت بازی آموزشی متوالی) بین تیم اتوماتای یادگیر با تیم Saloo (که نتایج ۷ بازی اول آنها در جدول ۳ آمده است)، تیم اتوماتای یادگیر رفته بهتری ارائه داد و نتایج بازیها به مساوی و در نهایت به برد پیوسته تیم اتوماتای یادگیر (در بازیهای آخر) انجامید. آمار ۷ بازی پس از آموزش ۲۵ ساعتی انجام شده بین تیم اتوماتای یادگیر و تیم Saloo در جدول ۴ آمده است. در این ۷ بازی، تیم اتوماتای یادگیر میانگین گل زده  $3/6$  گل در مقابل گل خورده  $1/0$  در هر بازی را داشته است. همانگونه که از روی آمار بازی نیز مشخص است، تیم Saloo هر چند در تعدادی از معیارها، دارای نتایج بهتری از تیم اتوماتای یادگیر می‌باشد، اما برتری مطلق خود را بر تیم اتوماتای یادگیر آن گونه که در شروع بازی وجود داشت را از دست داده است.

یک دیگر از معیارهایی که برای ارزیابی کارآیی همکاری در بین عاملهای تعریف شده است، معیار درصد اعمال صحیح انجام گرفته و همچنین خطای یک بازیکن در طول بازی است. در این قسمت و به منظور مقایسه، میانگین درصد خطای بازیکنان تیم یادگیر را در ۷ بازی



شکل ۱۸: مقایسه گلهای زده و خورده در طی ۱۰ بازی آموزشی متوالی بین تیم اتوماتای  $L_{REP}$  با تیم بدون یادگیری با افزودن بردار باد (۰ و ۵۰) به زمین بازی (جهت باد، عمود بر طول زمین بازی).

یادگیر حتی در چنین دامنه‌هایی می‌توانند موفق باشند. بدین منظور، ۱۰ بازی متوالی آموزشی بین تیم اتوماتای یادگیر  $L_{REP}$  (با یادگیری از صفر) و تیم بدون یادگیری صورت دادیم که نتایج این ۱۰ بازی در شکل ۱۶ نشان داده شده است. کارآیی روش یادگیری پیشنهادی (اتوماتای یادگیر  $L_{REP}$ ) در این شکل دیده می‌شود. مشاهده می‌شود که با گذشت زمان و افزایش تعداد بازیها، تیم یادگیر تدریجاً خود را با شرایط نبود ارتباطات مطمئن تطبیق می‌دهد و در جهت نیل به هدف پیش می‌رود.

در آزمایشها دیگری ارزیابی عملکرد تیم در صورت موافق شدن با خرابی و اشکال در عملکرد چند عامل مورد بررسی قرار گرفت. ۳ بازیکن از جناح چپ تیم اتوماتای یادگیر را برداشتیم. با توجه به آرایش ۴-۳-۳-۴، این بازیکنان عبارت بودند از بازیکن شماره ۲ از خط دفاعی تیم (دفاع چپ)، بازیکن شماره ۶ از خط وسط تیم (هافبک چپ)، و بازیکن شماره ۱۰ از خط حمله تیم (فوروارد چپ). تیم ۸ نفری را در ۱۵ بازی متوالی آموزشی در برابر تیم بدون یادگیری که از ۱۱ بازیکن استفاده می‌کرد قرار دادیم. نتایج این بازیها در شکل ۱۷ آمده است. همانطور که نتایج نشان می‌دهد، تیم اتوماتای یادگیر ۸ نفره موفق شده است که در حین آموزش، بر نبود بازیکنان جناح چپ خود فائق آید و تدریجاً بازیها را به سود خود به پایان برساند. این امر حاکی از مناسب بودن روش یادگیری مورد استفاده می‌باشد.

در آزمایشها دیگری، به بررسی اثر جریان باد در زمین بازی بر روی یادگیری عامل‌ها پرداختیم. نتایج حاصله در شکل‌های ۱۷ و ۱۸ داده شده است. همانطور که مشاهده می‌شود، تیم اتوماتای یادگیر پیشنهادی موفق شده است که در حین آموزش، بر جریان باد در زمین فائق آید و تدریجاً بازیها را به سود خود به پایان برساند.

جدول ۴: میانگین آمار برای ۷ بازی آخر (پس از آموزش ۲۵ ساعته) بین تیم آتوماتای  $L_{REP}$  با تیم SALOO.

|       |   |
|-------|---|
| ۴۵    | درصد مالکیت توب توسط تیم حریف (Saloo)                                   |
| ۵۵    | درصد مالکیت توب توسط تیم خودی ( $L_{REP}$ )                             |
| ۲۴/۵  | درصد گردش توب در ۱/۳ زمین حریف (Saloo)                                  |
| ۴۲    | درصد گردش توب در ۱/۳ میانی زمین   |
| ۳۳/۵  | درصد گردش توب در ۱/۳ زمین خودی ( $L_{REP}$ )                            |
| ۱۱۲/۷ | ماکریزم زمان در اختیار داشتن توب بصورت ممتد توسط تیم حریف (Saloo)       |
| ۱۳۴/۲ | ماکریزم زمان در اختیار داشتن توب بصورت ممتد توسط تیم خودی ( $L_{REP}$ ) |
| ۸/۸   | ماکریزم تعداد رد و بدلهای متوالی توب توسط تیم حریف (Saloo)              |
| ۱۲    | ماکریزم تعداد رد و بدلهای متوالی توب توسط تیم خودی ( $L_{REP}$ )        |

جدول ۵: مقایسه میانگین درصد اعمال (پاس) صحیح هر بازیکن در دو روش درخت تصمیم (تیم CMUNITED98) و آتوماتای یادگیر (تیم پیشنهادی).

| درصد درستی اعمال (ازبایی پاس) | درخت تصمیم گیری               |
|-------------------------------|-------------------------------|
| ۶۵                            | آتوماتای یادگیر ( $L_{REP}$ ) |
| ۷۶/۴                          |                               |

زیادی در بالا بردن سرعت همگرایی مؤثر است.

## ۱-۵ آرایشهای تیمی دیگر

تا بحال، تمام شبیه‌سازیها براساس تیمهایی با آرایش ۳-۳-۴ انجام داده شد. در این قسمت به تاثیر آرایشهای تیمی دیگر بر روی همکاری عامل‌ها می‌پردازیم. بدین لحاظ غیر از آرایش ۳-۳-۴، تیمهایی مشابه ( $L_{REP}$ ) با آرایشهای ۴-۴-۲، ۱-۴-۶، ۳-۵-۲، ۳-۵-۲، و ۳-۴-۳ ایجاد نمودیم و یک سری بازیهای دوره‌ای بین آنها برگزار نمودیم. شکل ۲۰ اطلاعات بدست آمده از شبیه‌سازیها را نشان می‌دهد. همانگونه که نتایج نشان می‌دهد، داشتن یک شکل تیمی مناسب، در نتیجه‌گیری تیمی بسیار موثر است. با توجه به آمار بدست آمده، بیشترین تعداد گل مربوط به شکل تیمی ۳-۵-۲ و کمترین تعداد گل خورده مربوط به شکل تیمی ۴-۴-۲ است. ضمن آنکه کمترین تعداد گل زده (و یکی از بیشترین تعداد گل خورده) مربوط به شکل تیمی ۳-۳-۴ است از نتایج آزمایشها همچنین می‌توان نتیجه گرفت که در مقابل تیمهای دفاعی از یک شکل تیمی استفاده کنیم که بیشترین گل زده را دارد و یا در مقابل تیمهای حمله‌ای از شکل تیمی استفاده کنیم که آمار کمترین تعداد گل دریافتی را دارد. همچنین می‌توان نتیجه گرفت که در صورتی که نتیجه به سود تیم ماست و زمان زیادی به انتهای بازی نمانده است، شکل تیمی دفاعی (با کمترین گل خورده) را انتخاب کنیم و در صورتی که نتیجه به ضرر تیم ماست و زمان کمی تا به انتهای بازی باقی است، شکل تیمی حمله‌ای (با بیشترین گل زده) را انتخاب نماییم. در حالت عادی بازی می‌توان یک شکل تیمی که بین دو حالت فوق‌الذکر باشد استفاده کرد (مانند شکل تیمی ۳-۴-۳ که در شبیه‌سازیها این مقاله استفاده شده است).

## ۶- نتیجه‌گیری

تحقیق ارائه شده در این مقاله، اولين تحقیق جدی در درباره استفاده از آتوماتاهای یادگیر در همکاری در سیستم‌های چند عامله در محیط

جدول ۳: میانگین آمار برای ۷ بازی اول بین تیم آتوماتای  $L_{REP}$  با تیم SALOO.

|      |   |
|------|---|
| ۵۲/۶ | درصد مالکیت توب توسط تیم حریف (Saloo)                                   |
| ۴۷/۴ | درصد مالکیت توب توسط تیم خودی ( $L_{REP}$ )                             |
| ۱۰/۵ | درصد گردش توب در ۱/۳ زمین حریف (Saloo)                                  |
| ۴۷   | درصد گردش توب در ۱/۳ میانی زمین   |
| ۴۲/۵ | درصد گردش توب در ۱/۳ زمین خودی ( $L_{REP}$ )                            |
| ۴۳/۵ | ماکریزم زمان در اختیار داشتن توب بصورت ممتد توسط تیم حریف (Saloo)       |
| ۱۱۲  | ماکریزم زمان در اختیار داشتن توب بصورت ممتد توسط تیم خودی ( $L_{REP}$ ) |
| ۱۴   | ماکریزم تعداد رد و بدلهای متوالی توب توسط تیم حریف (Saloo)              |
| ۸    | ماکریزم تعداد رد و بدلهای متوالی توب توسط تیم خودی ( $L_{REP}$ )        |

جدول ۵: مقایسه درصد خطای میانگین هر بازیکن در بازیکن تیم  $L_{REP}$  در ۷ بازی اول و ۷ بازی آخر (پس از آموزش ۲۵ ساعته) در مقابل تیم Saloo.

| درصد خطای | در ۷ بازی اول | در ۷ بازی آخر |
|-----------|---------------|---------------|
| ۴۰/۱      |               |               |
| ۲۴/۶      |               |               |

اول و نیز ۷ بازی آخر (پس از ۲۵ ساعت آموزش) بدست آورده‌یم که در جدول ۵ آمده است. از نتایج جدول ۵ مشخص است که درصد انجام اعمال صحیح هر بازیکن از تیم آتوماتای یادگیر، در اثر یادگیری که در حین بازی واقع شده است بهبود یافته است.

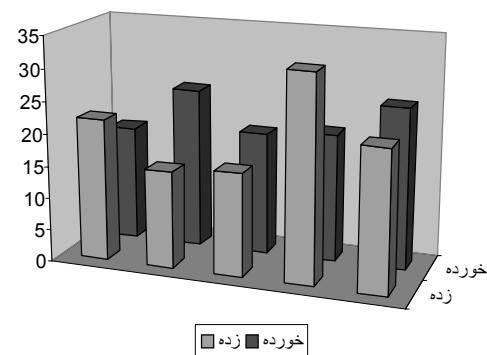
در تیم CMUnited98 CMUnited98 دو لایه برای رفتار چند عامله (ازبایی پاس) که بصورت غیرهمزن و با استفاده از درخت تصمیم آموزش داده می‌شود و رفتار تیمی (انتخاب پاس که بصورت همزمان و با استفاده از روشی بر مبنای یادگیری  $Q$  آموزش داده می‌شود) در نظر گرفته شده است [۲]. در روش پیشنهادی هر دو لایه فوق‌الذکر با هم ترکیب شده‌اند. بدین صورت که پس از تعدادی بازی آموزشی تیم را در مقابل تیم حریف در بازی آزمایشی قرار می‌دهیم، ولی چه در بازیهای آموزشی و چه در بازی‌های آزمایشی روش یادگیری یکسان و همان آتوماتاهای یادگیر است و ارزیابی یک عمل از انتخاب آن عمل جدا نشده است. به همین دلیل برای مقایسه کارآیی روش پیشنهادی با روش استون در تیم CMUnited98 CMUnited98 باستی درصد اعمال درست و غلط توسط عاملهای تیم خودی را با درصدهای پاس درست و غلط توسط عاملهای تیم CMUnited99 مقایسه کنیم. جدول ۶ درصد میانگین خطای روش پیشنهادی را با روش درخت تصمیم استون مقایسه می‌کند. توجه کنید که استون نتیجه ۶۵٪ میانگین اعمال صحیح را برای ارزیابی پاس (دومین لایه یادگیری روش لایه‌ای خود) بدست آورده است. همانگونه که این جدول نشان می‌دهد، روش پیشنهادی در این مقاله بر روش استون برتری دارد و بازیکنان پس از یادگیری کافی، درصد خطای کمتری نسبت به بازیکنان تیم CMUnited98 CMUnited98 مرتکب می‌شوند.

نتایج بازیهای تیم آتوماتای یادگیر با تیمهای دیگر ذکر شده در ابتدای این بخش در [۱۳] آمده است. نتایج شبیه‌سازیها برای تیمهایی که از آتوماتاهای یادگیر دیگری مانند الگوریتم‌های تخمین زن [۱۷] و یا از الگوریتم Pursuit پیوسته [۱۸] استفاده کرده‌اند نیز در [۱۳] آمده است. نتایج گزارش شده در [۱۳] نشان می‌دهند که این الگوریتمها به مقدار

- [11] M. R. Meybodi and S. Lakshmivarahan, "On a class of learning algorithms which have a symmetric behavior under success and failure," *Springer-Verlag Lecture Notes in Statistics*, pp. 145-155, 1984.
- [12] محمد رضا خجسته و محمد رضا میدی / تکنیک "بهترین گوشه در مربع حالت برای عمومی سازی حالات محیطی در یک دامنه چند عامله همکاری گرای"، مجموعه مقالات هشتمین کنفرانس سالانه انجمان کامپیوتر ایران، صفحات ۴۵۵-۴۶۴، دانشگاه فردوسی مشهد، مشهد، اسفند ۱۳۸۱.
- [13] محمد رضا خجسته، "همکاری در سیستمهای چند عامله با استفاده از اتوماتای یادگیری"، پایان نامه کارشناسی ارشد، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، بهار ۱۳۸۱.
- [14] محمد رضا خجسته و محمد رضا میدی، "ازیابی اتوماتای یادگیر در همکاری بین عاملها در یک سیستم چند عامله پیچیده"، مرکز تحقیقات انفورماتیک، آزمایشگاه محاسبات نرم / دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، بهار ۱۳۸۱.
- [15] محمد رضا خجسته و محمد رضا میدی، "اتوماتای یادگیر بعنوان مدلی برای همکاری در یک تیم از عاملها"، مجموعه مقالات هشتمین کنفرانس سالانه انجمان کامپیوتر ایران، صفحات ۱۲۶-۱۱۵، دانشگاه فردوسی مشهد، مشهد، اسفند ۱۳۸۱.
- [16] I. Noda, *Team Description: Saloo*, AIST & PREST, Japan, 2001.
- [17] M. A. L. Thathachar and P. S. Sastry, "A new approach to the design of reinforcement schemes for learning automata," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 15, no. 1, pp. 168-175, Jan./Feb. 1985.
- [18] B. J. Oomen and J. K. Lanctot, "Discretized pursuit learning automata," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 20, no. 4, pp. 931-938, Jul./Aug. 1990.

محمد رضا خجسته شرح حال ایشان در زمان انتشار نشریه در دسترس نبود.

محمد رضا میدی تحصیلات خود را در مقاطع کارشناسی و کارشناسی ارشد اقتصاد بتریپ در سالهای ۱۳۵۲ و ۱۳۵۶ از دانشگاه شهید بهشتی و در مقاطع کارشناسی ارشد و دکتری علوم کامپیوتر بتریپ در سالهای ۱۳۵۹ و ۱۳۶۲ از دانشگاه اوکلاهما امریکا به پایان رسانده است و هم اکنون استاد دانشکده مهندسی کامپیوتر دانشگاه صنعتی امیرکبیر می باشد. نامبرده قبل از پیوستش به دانشگاه صنعتی امیرکبیر در سالهای ۱۳۶۲ الی ۱۳۶۴ استادیار دانشگاه میشیگان غربی و در سالهای ۱۳۶۴ الی ۱۳۷۰ دانشیار دانشگاه اوهاوی در ایالات متحده امریکا بوده است. زمینه های تحقیقاتی مورد علاقه ایشان عبارتند از: الگوریتم های موازی، پردازش موازی، محاسبات نرم و کاربردهای آن، شبکه های کامپیوتری و مهندسی نرم افزار.



شکل ۲۰: مقایسه گل های زده و خورده توسط تیمهای یادگیر اتوماتای  $L_{REP}$  با شکلهای تیمی متفاوت در برابر همدیگر (از چپ به راست ۳-۶-۳، ۴-۳-۲، ۴-۳-۳، ۳-۴-۳، ۳-۵-۲).

شبیه ساز فوتbal رویاتها بشمار می رود. از طریق پیاده سازی تیمهای ۲ نفره، ۵ نفره و ۱۱ نفره از عامل هایی که هر کدام از آنها مجهز به یک اتوماتای یادگیر می باشد توانایی اتوماتاهای یادگیر در ایجاد همکاری در سیستمهای چند عامله مورد بررسی و ارزیابی قرار گرفت. یک روش جدید عمومی سازی حالات محیط به نام "بهترین گوشه در مربع حالت" نیز معرفی و پیاده سازی گردید.

## مراجع

- [1] G. Weiss, *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*, The MIT Press, London, 1999.
- [2] P. Stone, *Layered Learning in Multi Agent Systems*, Ph.D. Thesis, School of Computer Science, Carnegie Mellon University, Dec. 1998.
- [3] I. Noda, *Team GAMMA: Agent Programming on Gaea*, in H. Kitano, editor, RoboCup-97: Robot Soccer World Cup I, pp. 500-507, Springer Verlag, Berlin, 1998.
- [4] RoboCup web page, at URL <http://www.robocup.org>, 1997.
- [5] H. Kitano, editor, RoboCup-97: Robot Soccer World Cup I, Springer Verlag, Berlin, 1998.
- [6] D. Andre *et al.*, *Soccer Server Manual*, Version 4.0, Technical Report RoboCup 1998-001, RoboCup, 1998.
- [7] K. S. Narendra and M. A. L. Thathachar, *Learning Automata: An Introduction*, Prentice-Hall Inc., 1989.
- [8] P. Mars, J. R. Chen, and R. Nambir, *Learning Algorithms: Theory and Applications*, in *Signal Processing, Control and Communications*, CRC Press, Inc., pp. 5-24, 1996.
- [9] S. Lakshmivarahan, *Learning Algorithms: Theory and Applications*, New York, Springer-Verlag, 1981.
- [10] M. R. Meybodi and S. Lakshmivarahan, " $\epsilon$ -optimality of a general class of absorbing barrier learning algorithms", *Information Sciences*, vol. 28, pp. 1-20, 1982.