نشریه مهندسی برق و مهندسی کامپیوتر ایران، سال ۱، شماره ۲، پاییز و زمستان ۱۳۸۲

همکاری در سیستمهای چند عامله با استفاده از اتوماتاهای یادگیر

محمدرضا خجسته و محمدرضا میبدی

چکیده: عاملها موجودات نرم‌افزاری هستند که بطور پیوسته و خود مختار در یک محیط که برای آن طراحی شده‌اند کار می‌کنند. برای عاملها 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 عاملها به یک موضوع مهم آکادمیک تبدیل شده است که کاربردهای تجاری و صنعتی فراوانی را نیز در بر دارد. در سیستمهای چندعامله، چندین عامل هوشمند با قابلیت برقراری ارتباط با یکدیگر، جهت رسیدن به مجموعه‌ای از اهداف، با هم همکاری می‌کنند. بدلیل پیچیدگیهای موجود در محیطهای چندعامله 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 فوتبال روباتها به بررسی کارآیی اتوماتای یادگیر در همکاری بین عاملهای عضو یک تیم پرداخته شده است. بدلیل وجود تعداد حالات بسیار زیاد در دامنه‌های چندعامله پیچیده، داشتن روشی برای عمومی‌سازی حالات محیطی، امری ضروری است چرا که انتخاب مناسب چنین روشی، در تعیین حالات و اعمال عامل نقشی تعیین‌کننده 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 گسسته و محدود نگاشته می‌شود. کارآیی این تکنیک در عمومی‌سازی حالات محیطی در یک دامنه چندعامله همکاری‌گرا مورد بررسی قرار گرفته است.

کلید واژه: **اتوماتای یادگیر[1]، عامل، سیستمهای چندعامله، فوتبال روباتها، همکاری.**

۱- مقدمه

عامل، موجود خودمختاری است که ویژگیهایی از قبیل اجتماعی بودن، واکنشی بودن و پیش فعال بودن را دارا می‌باشد. عاملها 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. هرچند 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. عاملها در جهت دست یافتن به اهداف خود یا جامعه‌ای که در آن زندگی می‌کنند با یکدیگر همکاری می‌کند. وقتی که یک گروه از عاملها در یک سیستم چندعامله در یک هدف دراز مدت سهیم باشند، آنها تشکیل یک تیم را می‌دهند.

اعضای یک تیم رفتار خود را براساس سازگار کردن فرآیندهای 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 چندعامله می‌توان به فوتبال روباتها یا روبوکاپ[2] اشاره کرد. محیط فوتبال 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 تیم خود و تیم مقابل را روشن می‌سازند، تشکیل شوند.

به دلیل وجود پیچیدگیهای موجود در بسترهایی مانند بستر شبیه‌سازی فوتبال روباتها و رویارویی 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، نویزی[3]، نیاز به همکاری[4] و دارای دشمن با اهداف متقابل[5] 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 بخشند. یکی از اهداف این مقاله بررسی کارآیی اتوماتای یادگیر در همکاری بین عاملهای عضو یک تیم در یک بستر تست شبیه‌سازی فوتبال روباتها می‌باشد. با استفاده از بستر تست شبیه‌سازی فوتبال روباتها به بررسی کارآیی اتوماتای یادگیر در همکاری بین عاملهای عضو یک تیم پرداخته شده است. با پیاده‌سازی تیمهای ۲ نفره، ۵ نفره و ۱۱ نفره از عاملها که هر کدام ازآنها 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.

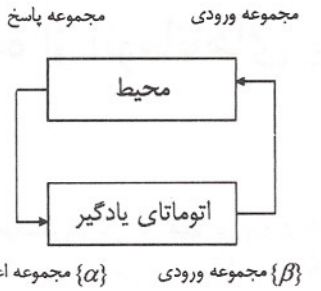
بدلیل وجود تعداد حالات بسیار زیاد در دامنه‌های چندعامله پیچیده، داشتن روشی برای عمومی‌سازی حالات محیطی امری ضروری است انتخاب مناسب چنین روشی، در تعیین حالات و اعمال عامل نقشی تعیین‌کننده دارد. بهمین دلیل در این مقاله به معرفی تکنیک جدیدی بنام "تکنیک بهترین گوشه در مربع حالت" نیز می‌پردازیم. این روش فضای حالات پیوسته و بسیار وسیع عامل را به فضای حالات گسسته 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 مفاهیم پیوسته مانند فاصله، زاویه و

این مقاله در تاریخ ۳ شهریور ماه ۱۳۸۱ دریافت و در تاریخ ۶ مرداد ۱۳۸۲ بازنگری شد.

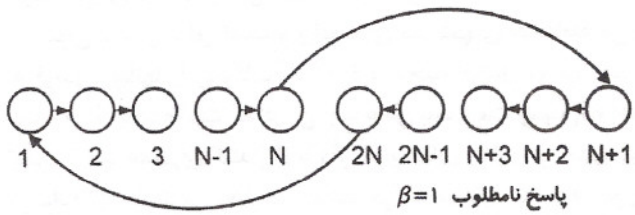
محمدرضا خجسته، آزمایشگاه محاسبات نرم، دانشکده مهندسی کامپیوتر، دانشگاه صنعتی امیرکبیر، تهران، ایران (email: khojasteh@ce.aut.ac.ir).

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

1. Learning Automaton
2. Robocup
3. Noisy
4. Collaborative
5. Adversarial



شکل ۱: ارتباط بین اتوماتای یادگیر و محیط.



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

نظایر آن برای عمومی‌سازی فضای حالات عامل استفاده شده است.

ادامه مقاله بصورت زیر سازماندهی شده است. در بخش ۲ به اختصار به شرح اتوماتاهای یادگیر می‌پردازیم. در بخش ۳ "تکنیک بهترین گوشه در مربع حالت" برای عمومی‌سازی فضای حالات محیطی معرفی 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- اتوماتاهای یادگیر

اتوماتاهای 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 بخشند. یک اتوماتای یادگیر 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 انتخاب می‌کند [۷] و [۸].

اتوماتاهای یادگیر به دو گروه تقسیم می‌گردند:

الف- اتوماتای یادگیر با ساختار ثابت[1]

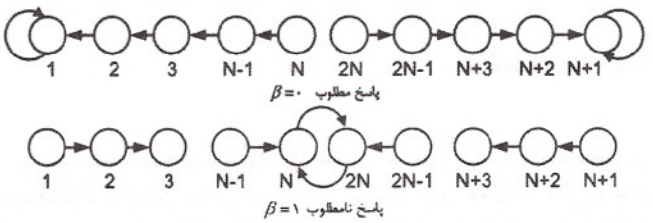
ب- اتوماتای یادگیر با ساختار متغیر[2].

۲-۱ محیط

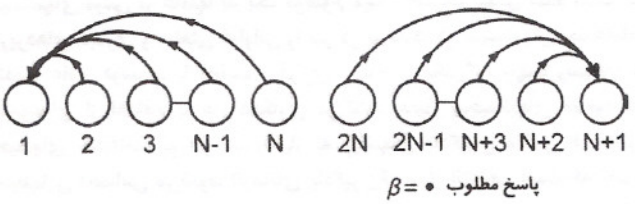
محیط را می‌توان توسط سه‌تایی $E \equiv \{\alpha, \beta, c\}$ تعریف نمود که $\alpha = \{\alpha_1, \alpha_2, \cdots, \alpha_r\}$ مجموعه ورودی‌ها، $\beta = \{\beta_1, \beta_2, \cdots, \beta_r\}$ مجموعه خروجی‌ها و $c = \{c_1, c_2, \cdots, c_r\}$ مجموعه احتمالهای جریمه شدن می‌باشند.

اهرگاه β_i دو مقداری باشد، محیط از نوع P می‌باشد. در چنین محیطی $\beta_i = 1$ به عنوان جریمه و $\beta_i = 0$ به عنوان پاداش در نظر گرفته می‌شود. c_i احتمال این است که عمل α_i نتیجه نامطلوب داشته باشد می‌باشد. در محیط پایدار[3] مقادیر c_i بدون تغییر باقی می‌مانند، حال آن که در محیط ناپایدار[4] این مقادیر در طی زمان تغییر می‌کنند. شکل ۱

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



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



شکل ۴: نمودار تغییر وضعیت اتوماتای Krinsky.

ارتباط بین اتوماتای یادگیر و محیط را نشان می‌دهد.

۲-۲ اتوماتای یادگیر با ساختار ثابت

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

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

- اتوماتای $L_{2N,2}$: این اتوماتا 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 اتوماتای مطابق شکل ۲ می‌باشد.
- اتوماتای $G_{2N,2}$: در این اتوماتا بر خلاف اتوماتای $L_{2N,2}$، عمل α_2 حداقل N بار انجام گیرد (پس از گرفتن N جریمه) تا اینکه عمل α_1 دوباره انتخاب شود. گراف تغییر وضعیت این اتوماتا برای پاسخ مطلوب مانند اتوماتای $L_{2N,2}$ بوده و برای پاسخ نامطلوب مطابق شکل ۳ می‌باشد.
- اتوماتای Krinsky: این اتوماتا زمانی که پاسخ محیط نامطلوب است، مانند $L_{2N,2}$ رفتار می‌کند. اما برای پاسخ مطلوب هر وضعیت $\phi_i (i = 1,2,..., N)$ به وضعیت ϕ_1 و هر وضعیت $\phi (i = N+1, N+2, ..., 2N)$ به وضعیت ϕ_{N+1} می‌رود. بنابراین همیشه N پاسخ نامطلوب متوالی لازم است تا اتوماتا عمل خود را عوض کند. نمودار تغییر وضعیت این اتوماتا برای پاسخ نامطلوب مانند اتوماتا $L_{2N,2}$ بوده و برای پاسخ مطلوب مطابق شکل ۴ می‌باشد.
- اتوماتای Krylov: در این اتوماتا زمانی که پاسخ محیط مطلوب است، تغییر وضعیت مانند اتوماتا $L_{2N,2}$ می‌باشد. اما زمانی که پاسخ محیط نامطلوب باشد، هر وضعیت $\phi_i (i \neq 1, N, N+1, 2N)$ با احتمال ۰/۵ به وضعیت ϕ_{i+1} و با احتمال ۰/۵ به وضعیت ϕ_{i-1} مطابق شکل ۵ منتقل می‌شود.

SW	SW		NW	NW
SW	SW		NW	NW
		●		
SE	SE		NE	NE
SE	SE		NE	NE

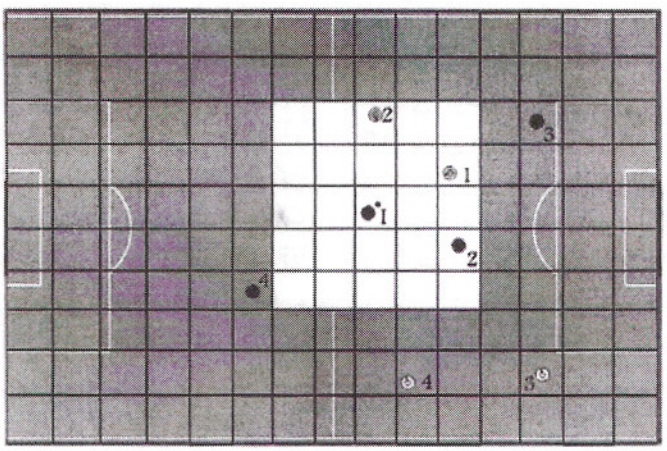
W
S N
E

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.

	W	W	W	
S		W		N
S	S	●	N	N
S		E		N
	E	E	E	

W
S N
E

شکل ۸: مربعهای شمال، شرق، غرب، و جنوب مربع دربردارنده عامل با توجه به موقعیت عامل مورد نظر و چهار جهت اصلی.



شکل ۹: وضعیت بازیکنان هم تیمی و حریف با توجه به فضای محلی عامل صاحب توپ در یک بازی ۴ در مقابل ۴.

مربع‌های جنوب غرب خود (مربع‌های $X-21$ و یا $X-22$ و یا $X-11$ و یا $X-12$) 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 توپ) در آن جهت، متناسب است. در واقع، عامل‌های نزدیک‌تر (و درون فاصله محلی تا ۲۴ مربع حالت پیرامون عامل صاحب توپ) تاثیر بیشتر و عاملهای دورتر (و درون فاصله محلی تا ۲۴ مربع حالت پیرامون عامل صاحب توپ) تاثیر کمتری را بر روی ضرایب اطمینان عامل صاحب توپ خواهند داشت و عامل‌های خارج از فاصله محلی (خارج از ۲۴ مربع حالت پیرامون عامل)، تاثیری در تغییر ضرایب اطمینان عامل صاحب توپ 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 توپ، عامل صاحب توپ دارای ۸ حالت و ۸ عمل (که متناظر با ارسال توپ به سمت مرکز ۸ مربع بلافاصله پیرامون عامل می‌باشند) خواهد بود. لازم بذکر است که تقسیم زمین به مربعهای ۷×۷ متر مربعی براساس تجربه کسب شده در کار با محیط شبیه‌ساز و میزان جابجایی توپ در اثر هر ضربه حاصل شده است و عملاً وابسته به دامنه انتخابی است. بدین ترتیب، فضای حالات پیرامون عامل به ۸ حالت کاهش می‌یابد و مشکل نگاشت حالات متفاوت به یک حالت بخصوص تا حد زیادی از بین می‌رود.

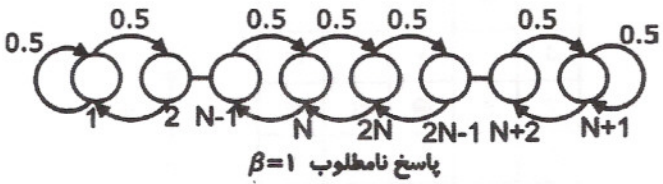
ممکن است در یک لحظه چندین عامل هم تیمی و یا حریف در درون ۲۴ مربع حالت پیرامون مربع حالت عامل وجود داشته باشند. حال اگر عامل یا عامل‌هایی درون همان مربعی قرار داشته باشند که عامل صاحب توپ در آن قرار دارد، عامل صاحب توپ بر حسب آنکه عامل‌های فوق در کدام یک از ۸ جهت خود قرار دارند کمیت‌های عددی خود را تغییر می‌دهد. یعنی، عامل صاحب توپ 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 فضا 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 توپ، بازیکن ۲ هم تیمی و بازیکنان 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 توپ 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 توپ، 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 خودی را صاحب توپ ببیند، موقعیت خود و دیگر بازیکنان موجود در حوزه دید خود را به بازیکن صاحب توپ اعلام می‌کند تا بازیکن صاحب توپ، حتی‌المقدور، بهترین تخمین را از حالتی که در آن قرار دارد داشته باشد. این امر با توجه به این نکته که دید هر بازیکن



شکل ۵: نمودار تغییر وضعیت اتوماتای Krylov.

$X-۲۲$	$X-۱۲$	$X-۲$	$X+۸$	$X+۱۸$
$X-۲۱$	$X-۱۱$	$X-۱$	$X+۹$	$X+۱۹$
$X-۲۰$	$X-۱۰$	X	$X+۱۰$	$X+۲۰$
$X-۱۹$	$X-۹$	$X+۱$	$X+۱۱$	$X+۲۱$
$X-۱۸$	$X-۸$	$X+۲$	$X+۱۲$	$X+۲۲$

شکل ۶: ۲۴ مربع اطراف مربع دربردارنده عامل مورد نظر در مربع X.

۲-۳ اتوماتای یادگیر با ساختار متغیر

اتوماتای یادگیر با ساختار متغیر توسط ۴تائی $\{\alpha,\beta,p,T\}$ نشان داده می‌شود که در آن $\alpha=\{\alpha_۱,\alpha_۲,...,\alpha_r\}$ مجموعه عمل‌های اتوماتا، $\beta=\{\beta_۱,\beta_۲,...,\beta_m\}$ مجموعه ورودی‌های اتوماتا، $p=\{p_۱,p_۲,...,p_r\}$ بردار احتمال انتخاب هر یک از عملها و $p(n+۱)=T[\alpha(n),\beta(n),p(n)]$ الگوریتم یادگیری می‌باشد. در این نوع از اتوماتاها، اگر عمل α_i در مرحله nام انتخاب شود و پاسخ مطلوب از محیط دریافت نماید، احتمال $p_i(n)$ افزایش یافته و سایر احتمالها کاهش می‌یابند و برای پاسخ نامطلوب 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 اتوماتای با ساختار متغیر است

الف- پاسخ مطلوب

$$p_i(n+۱)=p_i(n)+a[۱-p_i(n)]$$
$$\forall j \quad j\neq i \qquad p_j(n+۱)=(۱-a)p_j(n)$$

ب- پاسخ نامطلوب

$$p_i(n+۱)=(۱-b)p_i(n)$$
$$p_j(n+۱)=\frac{b}{r-۱}+(۱-b)p_j(n) \quad \forall j \quad j\neq i$$

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} [1] می‌نامیم. وقتی که b از a خیلی کوچکتر باشد، الگوریتم را $L_{R\varepsilon P}$ [2] می‌نامیم و هنگامی که b مساوی صفر باشد، الگوریتم را L_{RI} [3] می‌نامیم. حافظه و زمان مورد نیاز برای پیاده‌سازی اتوماتاهای یادگیر با ساختار ثابت $O(۱)$ و برای اتوماتاهای یادگیر با ساختار متغیر $O(m)$ می‌باشد که m تعداد اعمال اتوماتا است. برای مطالعه بیشتر درباره اتوماتاهای 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 فوتبال 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 باشند، هدف پیدا کردن تابع $f:S\rightarrow V$ 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 مثالهای 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 باشند 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 فوتبال که محیط دامنه چندعامله محسوب می‌شود به ۱۵۰ مربع یکسان با ضلع ۷ متر تقسیم می‌شود. از این طریق محیط پیوسته پیرامون عامل به محیطی گسسته 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 بعنوان شعاع دید آن عامل در نظر می‌گیریم. به شکل ۶ توجه کنید.

در این شکل، مربعهای $X+۱۰$، $X+۱۱$، $X+۱$، $X-۹$، $X-۱۰$، $X-۱۱$، $X-۱$، و $X+۹$ را مربع‌های بلافاصله اطراف مربع X می‌نامیم. شماره‌گذاری مربع‌ها از چپ به راست و بصورت ستونی و با شروع از شماره صفر برای اولین مربع (منتهاالیه چپ و بالا) تا شماره ۱۴۹ برای آخرین مربع (منتهاالیه راست و پایین) انجام گرفته است. هر ستون دارای ۱۰ مربع و هر سطر دارای ۱۵ مربع می‌باشند. 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 $X+۹$، $X+۸$، $X+۱۸$، و $X+۱۹$) ببیند، 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 توپ (واقع در مربع حالت X) یکی از عامل‌های خودی را در درون یکی از

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

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

	۱	۲	۳	۴	۵	۶
$L_{2N,2}$	۲-۱	۳-۳	۴-۷	۴-۹	۵-۱۲	۵-۱۶
$G_{2N,2}$	۱-۲	۱-۵	۲-۸	۲-۱۰	۳-۱۳	۵-۱۶
Krinsky	۱-۲	۲-۴	۳-۷	۳-۱۱	۳-۱۵	۴-۱۷
Krylov	۱-۲	۱-۴	۲-۶	۳-۹	۳-۱۱	۳-۱۲
Q	۱-۳	۲-۵	۲-۱۰	۳-۱۳	۳-۱۵	۴-۱۷

محدود است و هر بازیکن کنار و پشت خود را نمی‌بیند، برای بازیکن صاحب توپ 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 توپ، مدل دنیای خود را به بازیکن صاحب توپ هم تیمی خود اعلام می‌کند و وی را در تعیین حالت و به تبع آن، تعیین عمل مناسب در آن حالت، یاری می‌کند. بازیکن بدون توپ نیز، به مرکز مربع بلافاصله‌ی (از مربع حالت در بردارنده خود) می‌رود که به منظور دریافت توپ (در صورت لزوم) از بازیکن هم تیمی صاحب توپ، مناسب تشخیص می‌دهد. بدین نحو، در شبیه‌سازیهای انجام گرفته، بازیکن بدون توپ نیز می‌تواند دارای حالات و اعمال خاص (حرکت به سمت یکی از ۸ جهت اطراف خود به منظور دریافت توپ) باشد. یعنی بازیکن می‌تواند "حرکت بدون توپ" انجام دهد. باید توجه داشت که عمل مناسب انتخابی عامل صاحب توپ در هر حالت، الزاماً به معنی پاس و یا شوت نمی‌باشد و در بعضی موارد، ممکن است دریبل و یا نگهداشتن موقت توپ به منظور یافتن روزنه‌ای 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 مفاهیم اوت کردن توپ، پاس به عقب، و حتی بازکردن بازی و ارسال توپ به فضای خالی (با احتمال دریافت مناسب توپ برای بازیکن هم تیمی و یا خود در چند سیکل بعد) نیز استفاده می‌کند [۱۳].

لازم به ذکر است که یادگیری در کلیه شبیه‌سازیهای انجام گرفته کاملاً چندعامله و توزیع شده است و برخلاف روشهای موجود، بازیکن صاحب توپ (بدون توپ)، قصد خود را از ارسال (دریافت) توپ به (از) بازیکن بدون توپ (صاحب توپ) اعلام نمی‌دارد و بدین ترتیب، هر عامل در انتخاب عمل بهینه خود کاملاً خودمختار است و درحین انجام بازی بدون ایجاد ارتباط با دیگر بازیکنان سعی می‌کنند که با آنها همکاری کنند. شبیه‌سازیهای 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

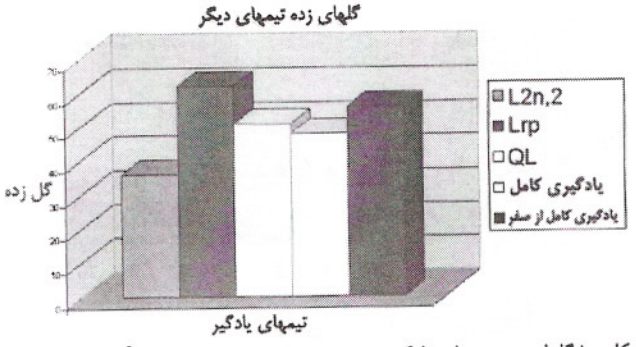
تا بحال از روشهای مختلفی ازجمله یادگیری Q، الگوریتمهای ژنتیک، درختهای تصمیم‌گیری و یادگیریهای رفتاری [۲] برای یادگیری عامل‌های فوتبالیست استفاده شده است. در این بخش از مقاله، توانایی اتوماتای یادگیر بمنظور ایجاد همکاری بین عامل‌های موجود در یک تیم روبوکاپ برای رسیدن به یک هدف مشخص تیمی که همانا برد باشد را مورد بررسی قرار می‌دهیم. در آزمایشهای انجام شده در این مقاله، چند بازیکن فوتبال که هرکدام مجهز به یک اتوماتای یادگیر می‌باشند را در مقابل چند بازیکن بدون توانایی یادگیری و یا دارای روشهای دیگر یادگیری قرار داده‌ایم تا از این طریق مدل اتوماتای یادگیر را در ایجاد همکاری مورد ارزیابی قرار دهیم.

۴-۱ شبیه‌سازیها برای تیم ۲ نفره

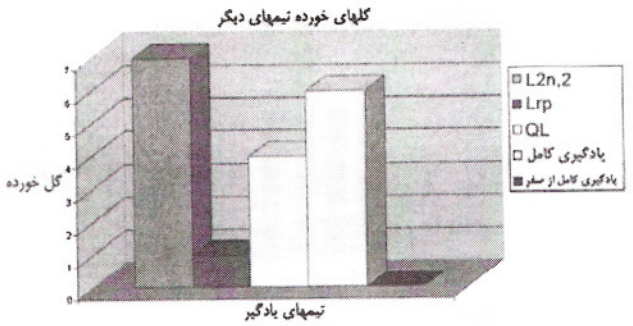
چندین سری شبیه‌سازی انجام گرفته است. در اولین سری شبیه‌سازیها، به پیاده‌سازی تیمهایی ۲ نفره از عاملها می‌پردازیم. این سری از شبیه‌سازیها با دو روش برای تعیین حالت هر عامل در محیط خود انجام میگیرد. یکی از این دو روش یک عمومی‌سازی ساده و دیگر روش، تکنیک "بهترین گوشه در مربع حالت" که قبلاً در این مقاله به آن اشاره شده است می‌باشد. در روش عمومی‌سازی ساده، کلیه حالات محیط به ۴ حالت برای بازیکن دارای توپ و ۴ حالت برای بازیکن بدون توپ خلاصه 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، بایستی 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 ۹۹۹، ستون ۲ 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.
همان گونه که این جدول نشان می‌دهد، تفاوت زیادی بین اتوماتاهای یادگیر مختلف در سری شبیه‌سازیهای انجام شده مشاهده نمی‌شود، هرچند که روشهای یادگیری Q و اتوماتای Krinsky اندکی بهتر از دیگر اتوماتاها عمل کرده است.

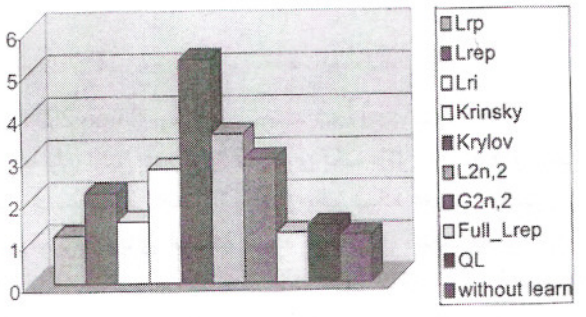
همان گونه که جدول ۱ نشان می‌دهد، تیم اتوماتای یادگیر موفق شده است تدریجاً در حین بازی، انجام عمل صحیح پاس و شوت را در ۴ حالت تعریف شده برای آن (در این سری از شبیه‌سازیها) فرا بگیرد. به همین دلیل اکثر گلهای دریافتی تیم اتوماتای یادگیر در نیمه اول (۳۰۰۰ سیکل اول) و اکثر گلهای زده شده توسط تیم اتوماتای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



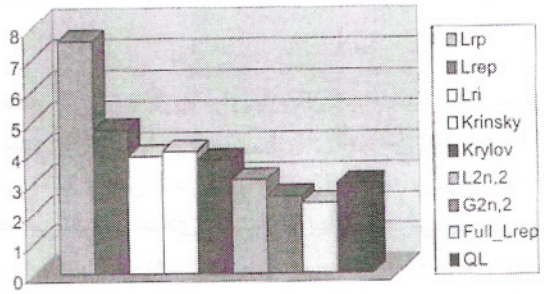
شکل ۱۰:گلهای زده تیمهای یادگیر ۵ نفره در ۱۰ شبیه‌سازی (در حین آموزش).



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



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



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

صورتی پاداش می‌گیرد که تیم صاحب توپ بعد از خاتمه این عمل هنوز تیم اوست و توپ به سود تیم او جلوتر رفته است (بطور مثال تیم او گل زده است) و در غیر این صورت بازیکن جریمه می‌شود. همچنین اگر انجام دهنده آخرین عمل، خود من بوده‌ام (یا اگر هنوز نتیجه آخرین عمل خود را ندیده‌ام) و تیم صاحب توپ، تیم حریف است و اگر فاصله موقعیت قبلی توپ تا موقعیت فعلی توپ خیلی زیاد نیست (بطور مثال به تیم حریف گل نزده‌ام) بازیکن جریمه می‌شود و در غیر این صورت به او پاداش داده می‌شود. نتیجه عمل که پاداش یا جریمه می‌باشد توسط اتوماتای یادگیر استفاده می‌شود تا حالت داخلی خود را به روز نماید. چگونگی به روز در 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.

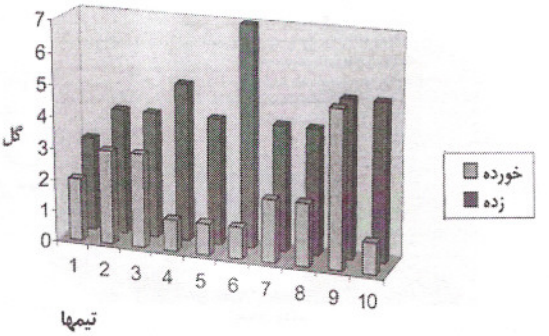
در ادامه به بررسی شبیه‌سازیها برای تیمهای ۵ نفره می‌پردازیم. مهمترین تغییری که در این شبیه‌سازیها نسبت به شبیه‌سازیهای قسمتهای قبل انجام گرفته است دادن شکل تیمی به تیم و استفاده از بازیکنان تیم در پستهای تخصصی بوده است. دادن شکل تیمی با توجه به افزایش نفرات تیم جهت برقراری نظم و همکاری هر چه بهتر در بین عاملها بنظر ضروری می‌رسید. علت انتخاب ۵ نفر برای یک تیم، شبیه‌سازی مواردی چون مسابقات روبوکاپ با روباتهایی با اندازه‌های متوسط و بزرگ و نیز فوتبال داخل سالنی بوده است. در شبیه‌سازیهای ۵ نفره، تیمهای یادگیر $L_{2N,2}$ بعنوان نماینده تیمهای یادگیر با ساختار ثابت، 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 شبیه‌سازیهای 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 بازیهای ۲ نفره در هیچ دیداری بازنده نبودند. نتایج آزمایشها در اشکال ۱۰و ۱۱ آمده است.

۴-۳ شبیه‌سازیها برای تیم ۱۱ نفره

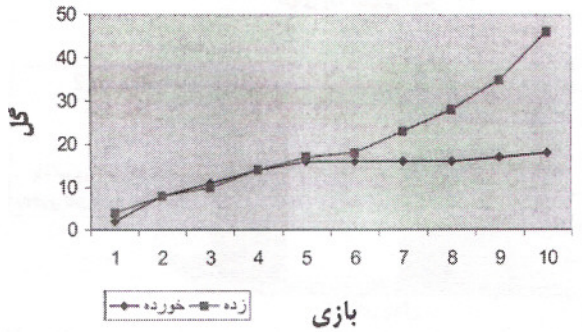
شبیه‌سازیهای ارایه شده در این قسمت به بررسی یادگیری در نحوه همکاری در بین عاملهای موجود در یک تیم ۱۱ نفره و مقایسه ان با دیگر تیمها می‌پردازد. با توجه به این که در بازیهای ۱۱ نفر در مقابل ۱۱ نفر، ممکن است تعداد گلهای رد و بدل شده خیلی زیاد نباشند به منظور بررسی نتایج شبیه‌سازی‌ها در این سری (و سریهای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

- درصد مالکیت توپ توسط تیم خودی در مقایسه با مورد مشابه در تیم حریف در حین بازی.
- درصد گردش توپ در ۱/۳ زمین خودی، ۱/۳ میانی زمین، و ۱/۳ زمین حریف در حین بازی.
- حداکثر زمان در اختیار داشتن توپ بصورت ممتد توسط تیم که بر حسب سیکل سنجیده می‌شود.
- حداکثر تعداد رد و بدلهای متوالی توپ 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 ۳-۳-۴ انتخاب گردیده است. همانند شبیه‌سازیهای 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 شبیه‌سازیهای انجام شده تیمهای اتوماتای 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 ۱۴: 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، کامل $L_{R\varepsilon P}$، L_{RP}، L_{RI}، $L_{R\varepsilon P}$، $L_{\Upsilon N,\Upsilon}$، Krylov، Krinsky، و $G_{\Upsilon N,\Upsilon}$.



شکل ۱۵: مقایسه گلهای زده و خورده در طی ۱۰ بازی آموزشی متوالی با پارامترهای تصادفی اضافه شده زمانیکه تیم اتوماتای $L_{R\varepsilon P}$ با تیم بدون یادگیری بازی می‌کند.

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.

گر چه تیم اتوماتای یادگیر بدلیل 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 روبوکاب 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.

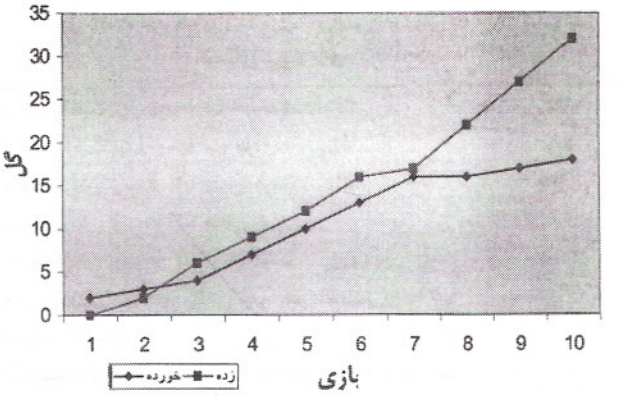
یادآوری می‌شود که کد پایه که برای پیاده‌سازی تیمهای اتواتاهای یادگیر مورد استفاده قرار گرفته است، کد CMUnited98 [۲] می‌باشد. لازم به ذکر است که این تیم در مقایسه با تیمهای جدیدتر از مهارتهایی فردی و گروهی بالایی بر خوردار نیست. تیمهای Sharif Arvand 2000، Yberoos 2000، FuzzyFoo 2001 و Saloo 2001 که از نظر فردی اختلاف زیادی با تیم ما نداشتند (هر چند بهتر از تیم ما بودند) می‌توانند بعنوان حریف بازیها انتخاب شوند. از بین چهار تیم فوق، تیم Saloo 2001 بدلیل 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.

تیم Saloo 2001 توسط نودا نوشته شده است. نودا در شرح تیم خود تقلید را بعنوان 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 بعنوان یک تمیزدهنده بین انواع بازیها آموزش داده شده مورد استفاده قرار گیرد. وی نتیجه‌گیری کرده است که این شبکه به همراه معماری وی می‌توانند انواع گوناگونی از بازیها را مجدداً تولید کنند [۱۶].

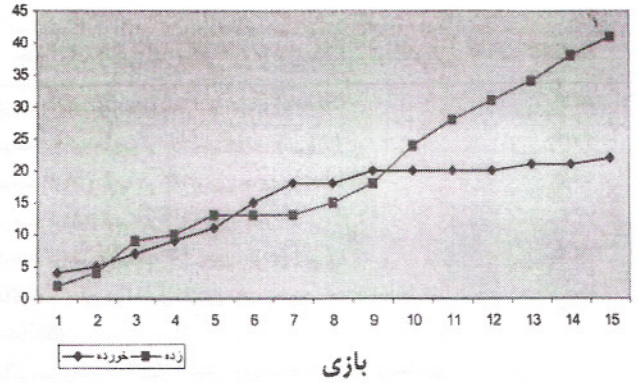
در آزمایشهای این قسمت از تیم اتوماتای $L_{R\varepsilon P}$ و تیم یادگیر کامل $L_{R\varepsilon P}$ استفاده شده است. این انتخاب بدین دلیل بوده است که این تیمها در شبیه‌سازیهای قبلی [۱۲] تا [۱۵] نتایج خوبی را تولید کرده است. در اولین سری از شبیه‌سازیها به ارزیابی تیم اتوماتای یادگیر در شرایطی که نویز در محیط وجود داشته باشد می‌پردازیم. پارامترهای متعددی در کارگزار شبیه‌سازی فوتبال 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 ۰/۱ تا ۰/۲، پارامتر ball_rand را از ۰/۰۵ تا ۰/۱، و بالاخره پارامتر kick_rand را از ۰ تا ۰/۱ تغییر دادیم. پارامتر اول باعث ایجاد اختلال (نویز) توسط کارگزار در حرکت بازیکن می‌شود و پارامترهای دوم و سوم، همین نقش را درمورد حرکت توپ در زمین و زدن ضربه به توپ بر عهده دارند. ۱۰ بازی متوالی بین تیم $L_{R\varepsilon P}$ (با یادگیری از صفر) و تیم بدون یادگیری برگزار نمودیم. هدف از این شبیه‌سازی این بود که ببینیم آیا روشهای یادگیری پیشنهادی در این مقاله می‌توانند در این گونه شرایط نیز موثر واقع شوند. شکل ۱۵ نتایج را (بصورت تجمعی) نشان می‌دهد. در این شکل، کارآیی روش یادگیری پیشنهادی (اتوماتای یادگیر $L_{R\varepsilon P}$) به صورت روشن مشاهده می‌شود. همانگونه که شکل نشان می‌دهد، با افزایش تعداد بازیها، تیم اتوماتای یادگیر موفق می‌شود که رفته 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.

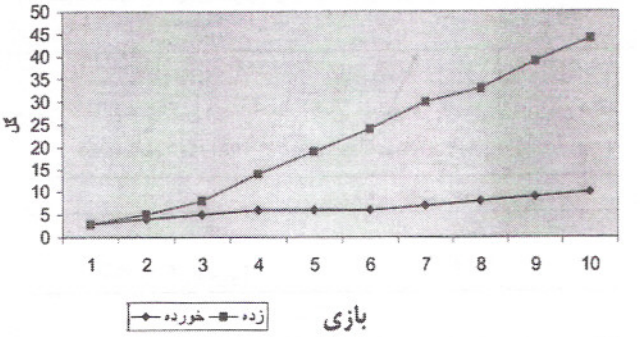
گر چه بدلیل تاکید بر خصوصیت خودمختاری عامل‌های تیمهای پیشنهادی از حداقل ارتباطات بین بازیکنان در حین بازی سود می‌برد ولی بدلیل اینکه کانال ارتباطی بین بازیکنان شلوغ و فاقد اطمینان است، برای همین حجم اندک ارسال (و نه تبادل) اطلاعات اتکا به ارتباطات بین عامل‌ها در چنین دامنه‌هایی قابل اطمینان نیست. تیمهای اتوماتاهای



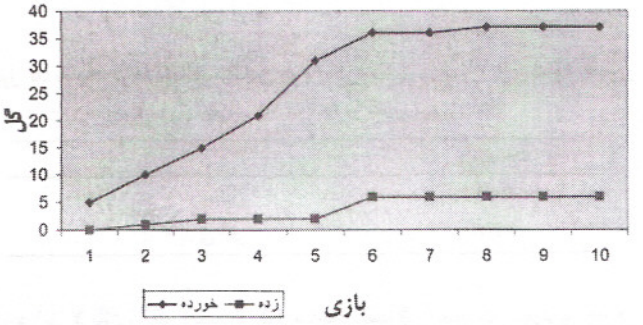
شکل ۱۶: مقایسه گلهای زده و خورده در طی ۱۰ بازی آموزشی متوالی بدون امکان ارتباط مابین بازیکنان بین تیم اتوماتای $L_{R\varepsilon P}$ زمانیکه با تیم بدون یادگیری بازی می‌کند.



شکل ۱۷: مقایسه گلهای زده و خورده در طی ۱۰ بازی آموزشی متوالی با برداشتن ۳ بازیکن جناح چپ از تیم یادگیر زمانیکه تیم اتوماتای $L_{R\varepsilon P}$ با تیم بدون یادگیری بازی می‌کند.



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



شکل ۱۹: مقایسه گلهای زده و خورده در طی ۱۰ بازی آموزشی متوالی بین تیم اتوماتای $L_{R\varepsilon P}$ 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).

یادگیر حتی در چنین دامنه‌هایی می‌توانند موفق باشند. بدین منظور، ۱۰ بازی متوالی آموزشی بین تیم اتوماتای یادگیر $L_{R\varepsilon P}$ (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 $L_{R\varepsilon P}$) 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 آزمایشهای دیگری ارزیابی عملکرد تیم در صورت مواجه 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 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 آزمایشهای دیگری، به بررسی اثر جریان باد در زمین بازی بر روی یادگیری عامل‌ها پرداختیم. نتایج حاصله درشکلهای ۱۷ و۱۸ 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 تدریجا بازیها را به سود خود به پایان برساند.

در شبیه‌سازیهای سری بعد، تیم اتوماتای $L_{R\varepsilon P}$ 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 برنده بازی بوده است و با تفاضل گل نسبتاً بالایی (میانگین گل زده ۵/۷ گل در مقابل گل خورده ۰/۳ در هر بازی) تیم اتوماتای یادگیر را شکست داده است. آمار 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 رفته بازی 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 ساعته انجام شده بین تیم اتوماتای یادگیر و تیم Saloo در جدول ۴ آمده است. در این ۷ بازی، تیم اتوماتای یادگیر میانگین گل زده ۳/۶ گل در مقابل گل خورده ۰/۱ 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.

یکی 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

جدول ۳: میانگین آمار برای ۷ بازی اول بین تیم اتوماتای $L_{R\varepsilon P}$ با تیم SALOO.

درصد مالکیت توپ توسط تیم حریف (Saloo)	۵۲/۶
درصد مالکیت توپ توسط تیم خودی ($L_{R\varepsilon P}$)	۴۷/۴
درصد گردش توپ در ۱/۳ زمین حریف (Saloo)	۱۰/۵
درصد گردش توپ در ۱/۳ میانی زمین	۴۷
درصد گردش توپ در ۱/۳ زمین خودی ($L_{R\varepsilon P}$)	۴۲/۵
ماکزیمم زمان در اختیار داشتن توپ بصورت ممتد توسط تیم حریف (Saloo)	۴۳۵
ماکزیمم زمان در اختیار داشتن توپ بصورت ممتد توسط تیم خودی ($L_{R\varepsilon P}$)	۱۱۲
ماکزیمم تعداد رد و بدلهای متوالی توپ توسط تیم حریف (Saloo)	۱۴
ماکزیمم تعداد رد و بدلهای متوالی توپ توسط تیم خودی ($L_{R\varepsilon P}$)	۸

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

	درصد خطا
در ۷ بازی اول	۴۰/۱
در ۷ بازی آخر	۲۴/۶

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

در تیم CMUnited98 دو لایه برای رفتار چند عامله (ارزیابی پاس که بصورت غیرهمزمان 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 اتوماتاهای 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 [۲]، بایستی درصد اعمال درست و نادرست بازیکنان تیم خودی را با درصدهای پاس درست و غلط توسط عاملهای تیم CMUnited9 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 مرتکب می‌شوند.

نتایج بازیهای تیم اتوماتای یادگیر با تیمهای دیگر ذکر شده در ابتدای این بخش در [۱۳] آمده است. نتایج شبیه‌سازیها برای تیمهایی که از اتوماتانهای یادگیر دیگری مانند الگوریتمهای تخمین‌زن [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

جدول ۴: میانگین آمار برای ۷ بازی آخر (پس از آموزش ۲۵ ساعته) بین تیم اتوماتای $L_{R\varepsilon P}$ با تیم SALOO.

درصد مالکیت توپ توسط تیم حریف (Saloo)	۴۵
درصد مالکیت توپ توسط تیم خودی ($L_{R\varepsilon P}$)	۵۵
درصد گردش توپ در ۱/۳ زمین حریف (Saloo)	۲۴/۵
درصد گردش توپ در ۱/۳ میانی زمین	۴۲
درصد گردش توپ در ۱/۳ زمین خودی ($L_{R\varepsilon P}$)	۳۳/۵
ماکزیمم زمان در اختیار داشتن توپ بصورت ممتد توسط تیم حریف (Saloo)	۱۱۲/۷
ماکزیمم زمان در اختیار داشتن توپ بصورت ممتد توسط تیم خودی ($L_{R\varepsilon P}$)	۱۳۴/۲
ماکزیمم تعداد رد و بدلهای متوالی توپ توسط تیم حریف (Saloo)	۸/۸
ماکزیمم تعداد رد و بدلهای متوالی توپ توسط تیم خودی ($L_{R\varepsilon P}$)	۱۲

جدول ۶: مقایسه میانگین درصد اعمال (پاس) صحیح هر بازیکن در دو روش درخت تصمیم (تیم CMUnited98) و اتوماتای یادگیر (تیم پیشنهادی).

	درصد درستی اعمال (ارزیابی پاس)
درخت تصمیم گیری	۶۵
اتوماتای یادگیر ($L_{R\varepsilon P}$)	۷۶/۴

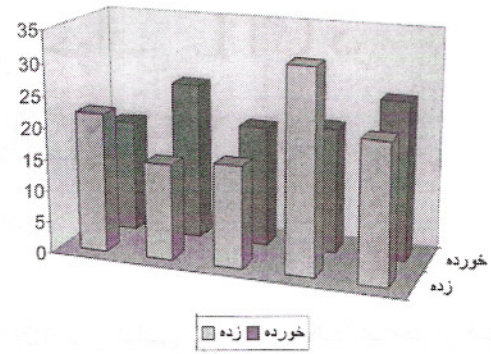
زیادی در بالا بردن سرعت همگرایی مؤثر هستند.

۵-۱ 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 ($L_{R\varepsilon P}$) با آرایشهای ۲-۴-۴، ۱-۶-۳، ۲-۵-۳، 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 امده، بیشترین تعداد گل زده مربوط به شکل تیمی ۲-۵-۳ و کمترین تعداد گل خورده مربوط به شکل تیمی ۲-۴-۴ 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 شبه‌سازیهای این مقاله استفاده شده است).

۶- نتیجه گیری

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



شکل ۲۰: مقایسه گل‌های زده و خورده توسط تیمهای یادگیر اتوماتای $L_{R\varepsilon P}$ با شکلهای تیمی متفاوت در برابر همدیگر (از چپ به راست ۲-۴-۴، ۳-۳-۴، ۱-۶-۳، ۲-۵-۳، ۳-۴-۳).

شبیه‌ساز فوتبال روباتها بشمار می‌رود. از طریق پیاده‌سازی تیمهای ۲ نفره، ۵ نفره و ۱۱ نفره از عامل‌هایی که هر کدام از آنها مجهز به یک اتوماتای یادگیر می‌باشد توانایی اتوماتاهای یادگیر در ایجاد همکاری در سیستمهای چندعامله مورد بررسی و ارزیابی قرار گرفت. یک روش جدید عمومی‌سازی حالات محیط به نام "بهترین گوشه در مربع حالت" نیز معرفی و پیاده‌سازی 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 Verleg, 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, " ε -optimality of a general class of absorbing barrier learning algorithms", *Information Sciences*, vol. 28, pp. 1-20, 1982.

[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.

[۱۲] محمدرضا خجسته و محمدرضا میبدی/ تکنیک "بهترین گوشه در مربع حالت برای عمومی‌سازی حالات محیطی در یک دامنه چند عامله همکاری گرا"، مجموعه مقالات هشتمین کنفرانس سالانه انجمن کامپیوتر ایران، صفحات ۴۵۵-۴۴۶، دانشگاه فردوسی مشهد، مشهد، ، اسفند ۱۳۸۱.

[۱۳] محمدرضا خجسته، "همکاری در سیستمهای چند عامله با استفاده از اتوماتای یادگیر"، پایان‌نامه کارشناسی ارشد، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، بهار ۱۳۸۱.

[۱۴] محمدرضا خجسته و محمدرضا میبدی، "ارزیابی 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.

[۱۵] محمدرضا خجسته و محمدرضا میبدی، "اتوماتای یادگیر بعنوان مدلی برای همکاری در یک تیم از عاملها"، مجموعه مقالات هشتمین کنفرانس سالانه انجمن 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.

محمدرضا خجسته شرح حال ایشان در زمان انتشار نشریه در دسترس نبود.

محمدرضا میبدی تحصیلات خود را در مقاطع کارشناسی و کارشناسی ارشد اقتصاد برترتیب در سالهای ۱۳۵۲ و ۱۳۵۶ از دانشگاه شهید بهشتی و در مقاطع کارشناسی ارشد و دکتری علوم کامپیوتر برترتیب در سالهای ۱۳۵۹ و ۱۳۶۲ از دانشگاه اوکلاهمای آمریکا به پایان رسانده است و هم اکنون استاد دانشکده مهندسی کامپیوتر دانشگاه صنعتی امیرکبیر می‌باشد. نامبرده قبل از پیوستنش به دانشگاه صنعتی امیرکبیر در سالهای ۱۳۶۲ الی ۱۳۶۴ استادیار دانشگاه میشیگان غربی و در سالهای ۱۳۶۴ الی ۱۳۷۰ دانشیار دانشگاه اوهایو در ایالات متحده آمریکا بوده است. زمینه‌های تحقیقاتی مورد علاقه ایشان عبارتند از: الگوریتمهای موازی، پردازش موازی، محاسبات نرم و کاربردهای آن، شبکه های کامپیوتری و مهندسی نرم افزار.