هشتمین کنفرانس سالانه انجمن کامپیوتر ایران ۱۳۸۱

اتوماتای یادگیر بعنوان مدلی برای همکاری در یک تیم از عاملها

محمدرضا خجسته محمدرضا میبدی

آزمایشگاه سیستمهای نرم افزاری

دانشکدهٔ مهندسی کامپیوتر

دانشگاه صنعتی امیرکبیر

تهران، ایران

email : mrkhojas@ce.aku.ac.ir

چکیده

عاملها موجودات نرم افزاری هستند که بطور پیوسته و خود مختار و در محیط بخصوصی کار می کنند (که مخصوص کار در آن محیط طراحی شده اند). برای عاملها نیازهایی از قبیل واکنشی بودن نسبت به کنش های محیطی، خود مختاری در انتخاب مسیر و ادامه آن، قابلیت تطبیق و یادگیری و ... ضروری به نظر می رسد. امروزه مطالعه سیستمهای مبتنی برعاملها به یک موضوع مهم آکادمیک تبدیل شده است که کاربردهای تجاری و صنعتی فراوانی را نیز دربر دارد. در سیستمهای چندعامله، چندین عامل هوشمند با قابلیت برقراری ارتباط با یکدیگر، جهت رسیدن به مجموعه ای از اهداف، با هم همکاری می کنند. بدلیل پیچیدگیهای موجود در محیطهای چندعامله پویا و متغیر نیاز به روشهای یادگیری ماشین در چنین محیطهایی بخوبی احساس می شود. اتوماتاهای یادگیر، بعنوان مدلی برای یادگیری، در محیطی تصادفی عمل نموده و قادر هستند که بر اساس ورودیهای دریافت شده از محیط، احتمال انجام عملیات خود را بروز در آورده تا بتوانند از این طریق کارآیی خود را بهبود بخشند. اتوماتای یادگیر یک مدل انتزاعی است که تعداد محدودی عمل را می تواند انجام دهد. هر عمل انتخاب شده توسط محیطی احتمالی ارزیابی می گردد و پاسخی به اتوماتای یادگیر داده می شود. اتوماتای یادگیر از این پاسخ استفاده نموده و عمل خود برای مرحله بعد را انتخاب می کند. در این مقاله با استفاده از بستر تست شبیه سازی فوتبال روباتها به بررسی کارآیی اتوماتای یادگیر در همکاری بین عاملهای عضو یک تیم پرداخته شده است.

کلمات کلیدی: اتوماتای یادگیر، عامل، سیستمهای چندعامله، فوتبال شبیه سازی شده روباتها، تیم عاملها، همکاری

۱. مقدمه

عامل، موجود خودمختاری است که ویژگیهایی از قبیل اجتماعی بودن، واکنشی بودن و پیش فعال بودن را دارا می باشد. عاملها در محیطی زندگی می کنند که می تواند باز یا بسته باشد و نیز ممکن است که این محیط عاملهایی دیگر را در خود جای داده باشد. هرچند وضعیت هایی وجود دارد که یک عامل می تواند خودش به تنهایی عمل کند، رشد روزافزون اتصال کامپیوتر ها و شبکه سازی آنها باعث کم شدن این وضعیت ها شده است و دیگر، معمول آن است که یک عامل با دیگر عامل ها ارتباط متقابل داشته باشد. عاملها در جهت دست یافتن به اهداف بهتر خود یا جامعه ای که در آن زندگی می کنند با یکدیگر همکاری می کنند.وقتی که یک گروه از عاملها در یک سیستم چند عامله در یک هدف دراز مدت سهیم باشند، آنها در واقع تشکیل یک تیم را می دهند. در واقع، یک گروه از عاملها ، تشکیل یک تیم همکاری کننده[1] را می دهند وقتی که :

- همه عاملها یک هدف مشترک دارند.
- هر عامل باید سهم خود را در رسیدن به هدف مشترک گروه انجام دهد.
- هر عامل به یک تقاضا جهت گرفتن سهم خود جواب می دهد[۱].

اعضای یک تیم (یا هم تیمی ها) رفتار خود را براساس سازگارکردن فرآیندهای شناختی خود و نیز براساس تأثیر مستقیم بر ورودیهای قابل دریافت همدیگر (از طریق فعالیتهای ارتباطی)، با همدیگر هماهنگ می کنند. دیگرعاملهای موجود در محیط نیز که اهدافی متقابل با هدف درازمدت این تیم دارند، حریفان یا دشمنان این تیم محسوب می شوند.

بعنوان بستری برای تست و پیاده سازی سیستمهای چند عامله می توان به فوتبال روباتها یا روبوکاپ[2] اشاره کرد. فوتبال روباتها مثالی از یک محیط پیچیده است که در آن چند عامل باید جهت رسیدن به اهداف تیمی، با هم همکاری کنند. فرآیندهای رفتاری و تصمیم گیری می توانند از ساده ترین رفتارها، همانند حرکت مستقیم به طرف توپ تا پیچیده ترین استدلال ها که استراتژی های تیم خود و تیم مقابل را روشن می سازند، تشکیل شوند.

به عنوان مثالهایی در رفتارهای چند عامله در این بستر، می توان به این موارد اشاره کرد که یک روبات مشخص، چه موقع باید به سمت توپ حرکت کند یا کار دیگری را انجام دهد و یا این که وقتی که یک روبات مشخص توپ را در اختیار دارد، آیا باید حرکت کند، پاس دهد و یا این که عمل شوت را انجام دهد (و در اینصورت با چه سرعتی ؟) تا کل تیم را به سمت یک هدف مشخص دسته جمعی (که در این حالت گل زدن است) سوق دهد. بدلیل وجود پیچیدگیهای موجود در بسترهایی مانند بستر شبیه سازی فوتبال روباتها و رویارویی عاملهای موجود در چنین محیطهایی با حالات بسیار زیاد، متنوع و متغیر، ناگزیر به استفاده از روشهای یادگیری ماشین می باشیم.

اتوماتاهای یادگیر بعنوان مدلی برای یادگیری، در محیطی تصادفی عمل نموده و قادر هستند که بر اساس ورودیهای دریافت شده از محیط، احتمال انجام عملیات خود را بروز در آورند تا بتوانند از این طریق کارآیی خود را بهبود بخشند. در این مقاله با استفاده از بستر تست شبیه سازی فوتبال روباتها به بررسی کارآیی اتوماتای یادگیر در همکاری بین عاملهای عضو یک تیم پرداخته شده است و سعی شده است که با پیاده سازی تیمهای ۲ نفره، ۵ نفره و ۱۱ نفره از عاملهای دارای خاصیت یادگیری اتوماتا و مقایسه آن با تیم مشابه و بدون خاصیت یادگیری و یا تیمهای یادگیر دیگر، کارآیی اتوماتای یادگیر در یادگیری یک کار تیمی و جهت دست یافتن به یک هدف مشترک، مورد ارزیابی قرار بگیرد.

در ادامه مقاله، ابتدا به توصیفی از بستر تست شبیه سازی فوتبال و اتوماتای یادگیر بعنوان روشی برای یادگیری می پردازیم و سپس شبیه سازیها و نتایج خود را ارائه می دهیم.

۲. کارگزار فوتبال روباتها[۳] بعنوان بستر تستی برای سیستمهای چند عامله

فوتبال روباتها مثالی از محیط و وظایف پیچیده است که عاملهایی چند باید جهت رسیدن به اهداف تیمی، با هم همکاری کنند. فرآیندهای رفتاری و تصمیم گیری می توانند از ساده ترین رفتارها ، همانند حرکت مستقیم به طرف توپ تا پیچیده ترین استدلال ها که استراتژی های تیم خود و تیم مقابل را روشن می سازند، تشکیل شوند.

سیستمهای چند عامله در محیط های پیچیده و زمان واقعی[۴] نیاز به عاملهایی دارند که بتوانند بطور مؤثر هم بعنوان یک عامل خودمختار[۵] عمل کنند و هم بعنوان عضوی از یک تیم. تأکید ما در این مقاله بر روی سیستمهایی متشکل از چند عامل خودمختار است که می توانند در محیطهای زمان واقعی، نویزی[۶]، نیاز به همکاری[۷] و دارای دشمن با اهداف متقابل[۸] عمل کنند[۲].ذیلا چهار خاصیت محیطی فوق را توضیح می دهیم

زمان واقعی - محیطها و دامنه هایی را زمان واقعی گویند که در آنها موفقیت بستگی به عمل کردن مناسب و در پاسخ یک محیط متغیر دینامیک است.

نویزی - دامنه هایی را نویزی می گویند که در آنها عاملها نمی توانند بطور صحیح و دقیق دنیا را درک کنند و نیز نمی توانند بطور دقیق برروی آن تأثیر بگذارند.

همکاری گرا - دامنه هایی را گویند که در آنها یک گروه از عاملها در یک هدف مشترک سهیم هستند.

دارای دشمن - به دامنه هایی می گویند که در آنها عاملهایی با اهداف رقابتی وجود دارند.

به دلیل پیچیدگی ذاتی این نوع از سیستمهای چند عامله، یادگیری ماشین روشی جالب برای ترکیب با آن به شمار می رود. درحقیقت، یادگیری این قابلیت را دارد که عاملها را به تعداد زیادی فعالیت در رده های مختلف مجهز کند. این رده ها می توانند از رده رفتارهای فردی در یک تیم، تا رده رفتارهای همکاری و سطح بالای مشترک تیمی متغیر باشند. شاید اولین موردی که یک بستر تست را برای بررسی سیستمهای چند عامله ای که ترکیب هر چهار خاصیت فوق را در نظر دارند، مورد استفاده قرار می دهد [۲] باشد.

فوتبال روباتها، دامنه ای است که با خاصیت های در نظر گرفته شده فوق برای دامنهٔ کار، تناسب کامل دارد، در دسترس است و نیز به اندازهٔ کافی پیچیدگی دارد. همانگونه که هدف اساسی هر بستر تستی این است که این امکان را برای ما ایجاد کند که بتوانیم ایده های خود را در آن مورد ارزیابی قرار دهیم، فوتبال روباتها نیز این امکان را به ما می دهد که ایده های وجود در دنیای واقعی را در آن مورد پیاده سازی قرار دهیم و از این جهت یک بستر تست عالی برای پروژه هایی از این دست به شمار می رود [۲]. ما در این تحقیق از فوتبال شبیه سازی شده روبوتیک استفاده کرده ایم.

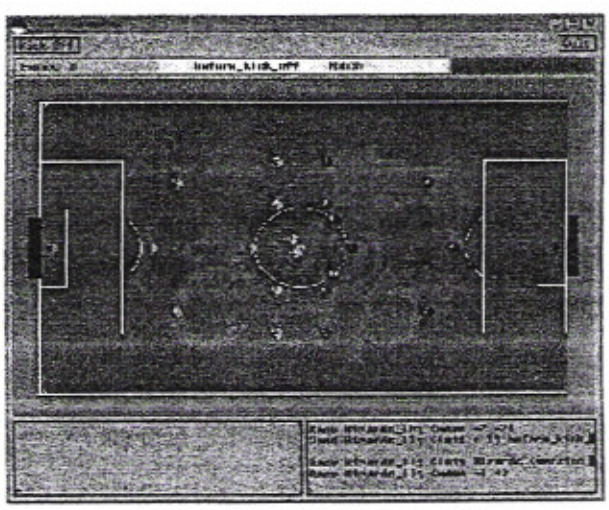
در دامنه هایی با چنین خواص پیچیده، عاملها قادر نخواهند بود که بطور مؤثر بیاموزند که چگونه از حسگرهای خود یک نگاشت مستقیم به محرک های خود ایجاد کنند، حتی اگر حالتهای گذشته دنیای خود را ذخیره کرده باشند [۲]. از آنجا که ما از کارگزار شبیه ساز فوتبال روباتها بعنوان بستر تست خود استفاده کرده ایم، ذیلا به شرح خلاصه ای از این بستر می پردازیم.

بسترفوتبال شبیه سازی شده روبوتیک، محیطی را فراهم می کند که کاملا"توزیع شده می باشد. در این محیط حالت مخفی[۹] وجود دارد، به این معنی که هر عامل تنها یک دید جزیی از دنیا را (در هر لحظه) دارا است. هم چنین عاملها دارای حسگرهای نویزی[۱۰] و نیز محرک های نویزی هستند، به این معنی که آنها نمی توانند دنیا را به شکل دقیق درک کنند و نیز نمی توانند بر روی آن نیز، دقیقا" به همان شکل می خواهند تأثیر بگذارند. بعلاوه، سیکل ها[۱۱] و زمان های دریافت[۱۲] و عمل[۱۳] به صورت آسنکرون هستند که این خود مانعی برای این مورد است که از کاربردهای سنتی هوش مصنوعی که در آنها از ورودی دریافتی برای تعیین عمل استفاده می شود ، استفاده کرد. در این محیط، فرصتهای ارتباطی محدود هستند، عاملها باید بصورت زمان واقعی تصمیم گیری نمایند و نیز اعمال

انجام شده توسط دیگر عاملها، چه هم تیمی و چه دشمن و نیز نتایج تغییر وضعیت حاصل از این اعمال برای هر عامل بخصوص، ناشناخته[14] هستند.

کارگزار فوتبال روباتها [۳] بعنوان پایه ای برای مسابقات بین المللی موفق [۴] و نیز چالشهای تحقیقاتی [۵] بکار رفته است. کارگزار فوتبال یک دامنه واقعیت گرا[15] و پیچیده است.برخلاف بسیاری از دامنه های هوش مصنوعی، محیط کارگزار فوتبال سعی نموده است که حداکثر ممکن پیچیدگیهای دنیای واقعی را درون خود مورد شبیه سازی قرار دهد. این کارگزار یک سیستم روباتیک فرضی را با ترکیب خصوصیاتی از سیستم های برنامه ریزی شده[16] و موجود[17] و نیز بازیکنان فوتبال انسانی پیاده سازی می نماید.

خصوصیت نویزی بودن حسگر و محرک در کارگزار از سیستم های روباتیک واقعی برداشت شده است، حال آن که خیلی از خصوصیات دیگر، از جمله قوت[18] محدود و نیز دید محدود، از پارامترهای انسانی برداشت شده اند. جزئیات موجود، مفاهیم و پارامترهای مهم در گزارش کارگزار فوتبال [۶] آمده است. محیط شبیه ساز از یک ابزار تصویری بصورتی که در شکل ۱ نشان داده شده است استفاده می کند.



شکل ۱- محیط شبیه سازی فوتبال روباتها

محیط شبیه ساز فوتبال از یک مدل عمل گسسته[19] استفاده می کند، بدین صورت که اعمال بازیکن را در طی یک سیکل شبیه سازی[20] بطول ۱۰۰ میلی ثانیه مورد جمع آوری قرار می دهد، ولی تنها در پایان سیکل، آنها را اجرا می کند و درحقیقت دنیا را به روز می سازد.

اگر یک کارفرما ، بیش از یک فرمان اجرایی را در یک سیکل شبیه سازی بفرستد، کارگزار یکی از آنها را بصورت تصادفی برای اجرا انتخاب می کند. بنابراین بجاست که هر کارفرما در حین هر سیکل شبیه سازی حداکثر یک فرمان را به کارگزار بفرستد. از سویی دیگر، اگر یک کارفرما در حین یک سیکل شبیه سازی هیچ فرمان اجرایی را به کارگزار نفرستد، فرصت عمل را در حین آن سیکل از دست می دهد که می تواند در یک محیط زمان واقعی و دارای دشمن ضرر زیادی باشد، زیرا در همین حین که عامل بیکار بوده است، حریفان ممکن است این فرصت را بدست آورند که برتر ظاهر شوند. در هر سیکل، شبیه ساز یک واحد به شمارندهٔ شبیه ساز اضافه می کند.

حس کردن و عمل نمودن بصورت آسنکرون، مخصوصا" وقتی که حس کردن می تواند در فواصل زمانی غیر قابل پیش بینی اتفاق بیفتد، برای عاملها موردی بسیار چالش برانگیز است. در حقیقت عاملها باید بین نیاز به عمل کردن بصورت منظم و سریع و نیز نیاز به جمع آوری اطلاعات از محیط، یک تعادل برقرار کنند. باید اضافه کرد که این مورد، یعنی حس کردن و عمل کردن آسنکرون، تنها یکی از موارد پیچیدگی دنیای واقعی است که شبیه ساز آن را پوشش می دهد.

کارگزار فوتبال یک شبیه ساز دو بعدی است. حداکثر ۲۲ بازیکن و یک توپ در هر لحظه می توانند در زمین وجود داشته باشند که همهٔ آنها بصورت دایره هایی مدل می شوند. همچنین چندین نشانه[21] قابل رؤیت در زمین و پیرامون آن وجود دارند، شامل پرچمها و خطهای کناری زمین که در اطراف زمین وجود دارند.

بازیکنان و توپ همگی اشیاء متحرک به شمار می روند. در سیکل شبیه ساز برای زمان t ، هر شیء در یک موقعیت مشخص (p_x^t , p_y^t) قرار دارد و سرعتی برابر با (v_x^t , v_y^t) نیز دارا است . همچنین هر بازیکن در لحظهٔ t به یک جهت مشخص θ^t رو کرده است. این موقعیت ها بصورت داخلی بصورت اعداد اعشاری نگهداری می شوند، ولی حسهایی که برای بازیکنان توسط کارگزار فرستاده می شود حداکثر دارای دقتی با یک رقم اعشارمی باشند. بنابراین فضای حالت دریافت شده با اغلب پارامترهای کارگزاری که ما استفاده می کنیم بیش از 10^{22} یا 10^{198} حالت است. بدینصورت که هرکدام از این ۲۲ بازیکن میتوانند در یکی از ۳۶۰۰ × ۱۰۵۰ × ۶۸۰ مکان باشند. با درنظرگرفتن توپ، سرعتها و حالتهای گذشته، فضای حالات واقعی حتی بسیار بزرگتر از این مقدار می باشد [۲].

کارگزار فوتبال، نویز با توزیع احتمال یکنواختی به حرکات همه اشیاء اضافه می کند. نیز کارگزار فوتبال مانع از آن می شود که یک بازیکن بتواند دائما" با حداکثر سرعت بدود و این کار را با انتساب یک قوت محدود به هر بازیکن انجام می دهد عاملها سه نوع اطلاعات درکی و حسی متفاوت از کارگزار دریافت می کنند: شنوایی[22] ، بینایی[23] و فیزیکی[24][۶]. ارتباطات در کارگزار فوتبال بدینصورت مدل شده است که از محیطی نویزی[25] و با پهنای باند کم استفاده شده است. در واقع همه ۲۲ عامل (هرتیم ۱۱ عامل) از یک کانال ارتباطی غیر مطمئن استفاده می کنند.

۳. اتوماتان یادگیر[26]

اتوماتاهای یادگیر در محیطی تصادفی عمل نموده و قادر هستند که بر اساس ورودیهای دریافت شده از محیط، احتمال انجام عملیات خود را بروز در آورده تا بتوانند از این طریق کارآیی خود را بهبود بخشند. اتوماتای یادگیر یک مدل

انتزاعی است که تعداد محدودی عمل را می تواند انجام دهد. هر عمل انتخاب شده توسط محیطی احتمالی ارزیابی می گردد و پاسخی به اتوماتای یادگیر داده می شود. اتوماتای یادگیر از این پاسخ استفاده نموده و عمل خود برای مرحله بعد انتخاب می کند [۷]، [۸]. اتوماتاهای یادگیر به دو گروه تقسیم می گردند:

آ- اتوماتای یادگیر با ساختار ثابت[۲۷]

ب- اتوماتای یادگیر با ساختار متغیر[۲۸]

محیط[۲۹]: محیط را می توان توسط سه تایی زیر تعریف نمود:

$$E \equiv \{\alpha, \beta, c\}$$

که

α ؛ $\{\alpha_1, \alpha_2, \ldots, \alpha_r\}$ مجموعه ورودیها

β ؛ $\{\beta_1, \beta_2, \ldots, \beta_m\}$ مجموعه خروجیها

c ؛ $\{c_1, c_2, \ldots, c_r\}$ مجموعه احتمالهای جریمه شدن

هرگاه β_i دو مقداری باشد، محیط از نوع P می باشد. در چنین محیطی $\beta_i = 1$ به عنوان جریمه و $\beta_i = 0$ به عنوان پاداش در نظر گرفته می شود. c_i احتمال اینکه عمل α_i نتیجه نامطلوب[۳۰] داشته باشد می باشد. در محیط پایدار[۳۱] مقادیر c_i بدون تغییر باقی می مانند، حال آنکه در محیط ناپایدار[۳۲] این مقادیر در طی زمان تغییر می کنند. شکل ۲ ارتباط بین اتوماتای یادگیر و محیط را نشان می دهد.

اتوماتای یادگیر با ساختار ثابت: اتوماتای یادگیر با ساختار ثابت توسط ۵ تایی زیر نشان داده مشود:

$$LA = \{\alpha, \beta, F, G, \phi\}$$

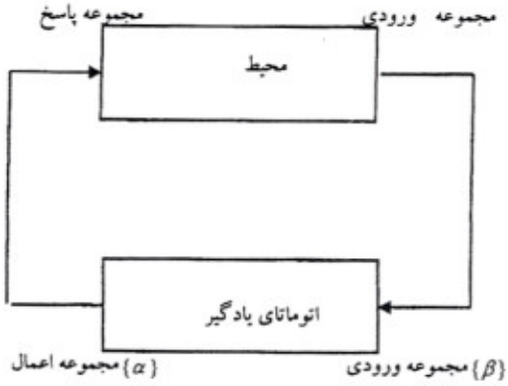
مجموعه عمل ها[۳۳]ی اتوماتا $\equiv \alpha = \{\alpha_1, \alpha_2, \cdots, \alpha_r\}$

مجموعه ورودیهای اتوماتا $\equiv \beta = \{\beta_1, \beta_2, \cdots, \beta_m\}$

تابعی که بر اساس پاسخ محیط، وضعیت جدید را می یابد $F \equiv \phi \times \beta \rightarrow \phi$

تابع خروجی که وضعیت کنونی را به خروجی بعدی می نگارد $G \equiv \phi \rightarrow \alpha$

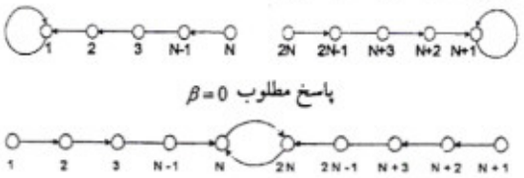
مجموعه وضعیت های داخلی اتوماتا $\equiv \phi(n) \equiv \{\phi_1, \phi_2, \cdots, \phi_k\}$



شکل ۲- ارتباط بین اتوماتان یادگیر و محیط

اتوماتا های با ساختار ثابت که ما در این مقاله از آنها استفاده کرده ایم در زیر آمده اند:

- **اتوماتای** $L_{2N,2}$: این اتوماتا تعداد پاداش ها و جریمه های دریافت شده برای هر عمل را نگهداری کرده و تنها زمانی که تعداد جریمه ها بیشتر از پاداش ها می گردد، عمل دیگر را انتخاب می کند. نمودار تغییر وضعیت این اتوماتای مطابق شکل ۳ می باشد.

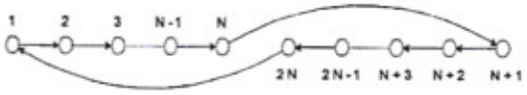


پاسخ مطلوب $\beta=0$

پاسخ نا مطلوب $\beta=1$

شکل ۳- نمودار تغییر وضعیت اتوماتای $L_{2N,2}$

- **اتوماتای** $G_{2N,2}$: در این اتوماتان بر خلاف $L_{2N,2}$، عمل α_2 حداقل N بار انجام می گردد (پس از گرفتن N جریمه) تا اینکه نهایتا عمل α_1 دوباره انتخاب شود. گراف تغییر وضعیت این اتوماتان برای پاسخ مطلوب مانند اتوماتان $L_{2N,2}$ بوده و برای پاسخ نامطلوب مطابق شکل ۴ می باشد.

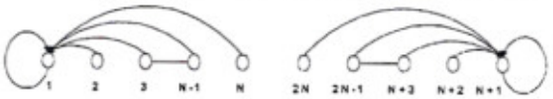


پاسخ نا مطلوب $\beta=1$

شکل ۴- نمودار تغییر وضعیت اتوماتان $G_{2N,2}$

بعنوان مثالهایی از دیگر اتوماتاهای معروف نیز می توان مثالهای زیر را بیان کرد:

- **اتوماتای Krinsky**: این اتوماتا زمانی که پاسخ محیط نامطلوب است، مانند $L_{2N,2}$ رفتار می کند. اما برای پاسخ مطلوب هر وضعیت $\phi(i = 1, 2, \ldots, N)$ به وضعیت ϕ_1 و هر وضعیت $\phi(i = N+1, N+2, \ldots, 2N)$ به وضعیت ϕ_{N+1} می رود. 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 ۵ می باشد.

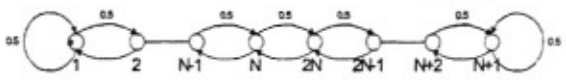


پاسخ مطلوب $\beta=0$

شکل۵- نمودار تغییر وضعیت اتوماتای Krinsky

- **اتوماتای Krylov**: در این اتوماتا زمانیکه پاسخ محیط مطلوب است، تغییر وضعیت مانند اتوماتان $L_{2N,2}$ می باشد. اما زمانیکه پاسخ محیط نامطلوب می باشد، هر وضعیت $\phi_i (i \neq 1, N, N+1, 2N)$ با

احتمال $\frac{1}{2}$ به وضعیت ϕ_{i+1} و با احتمال $\frac{1}{2}$ به وضعیت ϕ_{i-1} مطابق شکل ۶ منتقل می شود:



پاسخ نا مطلوب $\beta = 1$

شکل ۶- نمودار تغییر وضعیت اتوماتان Krylov

اتوماتای یاد گیر با ساختار متغیر[۳۴]: اتوماتای یادگیر با ساختار متغیر توسط ۴ تائی $\{\alpha, \beta, p, T\}$ نشان داده می شود که در آن $\alpha \equiv \{\alpha_1, \alpha_2, ..., \alpha_r\}$ مجموعه عمل های اتوماتا، $\beta = \{\beta_1, \beta_2, ..., \beta_m\}$ مجموعه ورودیهای اتوماتا، $p \equiv \{p_1, p_2, ..., p_r\}$ بردار احتمال انتخاب هر یک از عملها و $p(n+1) = T[\alpha(n), \beta(n), p(n)]$ الگوریتم یادگیری می باشد. 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 می یابند. و برای پاسخ نامطلوب احتمال $p_i(n)$ کاهش یافته و سایر احتمالها افزایش می یابند. در هر حال، تغییرات به گونه ای صورت می گیرد تا حاصل جمع $p_i(n)$ ها همواره ثابت و مساوی یک باقی بماند. الگوریتم زیر یک نمونه از الگوریتمهای یادگیری خطی در اتوماتای با ساختار ثابت است

الف- پاسخ مطلوب

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

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 می باشد. با توجه به مقادیر a و b سه حالت زیر را می توان در نظر گرفت. زمانیکه a و b با هم برابر باشند، الگوریتم را[۳۵] L_{RP} می نامیم. زمانیکه b از a خیلی کوچکتر باشد، الگوریتم را $L_{RεP}$[۳۶] می نامیم و زمانیکه b مساوی صفر باشد، الگوریتم را L_{RI}[۳۷] می نامیم. برای مطالعه بیشتر درباره اتوماتا های یادگیر می توان به [۷]، [۸]، [۹]، [۱۰]، و [۱۱] مراجعه نمود.

٤. همکاری بین اعضای یک تیم با استفاده از اتوماتای یادگیر

هدف ما در این قسمت، استفاده از اتوماتای یادگیر برای ایجاد همکاری در بین عاملهای موجود در یک تیم جهت رسیدن به یک هدف مشخص تیمی است. بستر انتخابی ما، محیط شبیه ساز فوتبال روباتها می باشد که با توجه به خصوصیات فوق الذکر، کلیه خصوصیات لازم جهت ایجاد یک محیط شبیه سازی پیچیده چند عامله را در اختیار ما قرار می دهد. در سری آزمایشهای انجام شده توسط ما، چند بازیکن فوتبال با خاصیت یادگیری از نوع اتوماتای یادگیر را در مقابل چند بازیکن بدون خاصیت یادگیری قرار داده ایم و تاثیر این نوع یادگیری را در آنها ارزیابی کرده ایم.

تا بحال از روشهای مختلفی برای یادگیری عاملهای فوتبالیست (چون انواع روشهای یادگیری تقویتی و یادگیری Q، الگوریتمهای ژنتیک، درختهای تصمیم گیری، یادگیریهای رفتاری و...[۲]) استفاده شده است.استفاده ما از اتوماتای یادگیر، اولین مورد استفاده از اتوماتای یادگیر در این زمینه است.

ما چندین سری شبیه سازی را انجام داده ایم. در اولین سری شبیه سازیهای خود، به پیاده سازی تیمهایی ۲ نفره از عاملها پرداختیم. این سری از شبیه سازیها با دو روش برای تعیین حالت هر عامل در محیط خود انجام شدند (یک روش صرفا یک عمومی سازی ساده و دیگر روش، تکنیک "بهترین گوشه در مربع حالت"[۱۲] بود)،

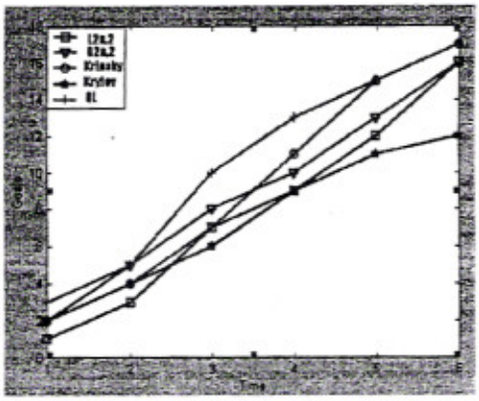
نتایج اولیه [۱۳] نشان دادند که تیم دارای اتوماتای یادگیری در قیاس با یک تیم بدون یادگیری بسرعت یاد می گیرد که در چه حالاتی، باید چه اعمالی را انجام دهد و بدین لحاظ براحتی بر حریف خود غلبه می کند. روش یادگیری همزمان نیز ما این حسن را داشت که به تیم ما اجازه می داد خود را با نحوه بازی تیم حریف تا حد زیادی وفق دهد.

در حقیقت در این قسمت، تعداد ۵۰ بازی را برای تیمهای اتوماتای با ساختار ثابت (و تیمی بر مبنای یادگیری Q که برای مقایسه پیاده سازی کرده بودیم) با تیم بدون یادگیری انجام دادیم که نتایج میانگین آنها در جدول زیر آمده است. در این جدول، عدد سمت راست نشان دهنده (جزء صحیح) تعداد گلهای زده توسط تیم یادگیر و عدد سمت چپ نشان دهنده (جزء صحیح) تعداد گلهای زده توسط تیم بدون یادگیری می باشد. همانگونه که دیده می شود، کلیه بازیها به سود تیمهای یادگیر به پایان رسیده اند. همچنین در مورد تیمهای اتوماتا در کلیه موارد، عمق حافظه برابر با ۳ در نظر گرفته شده است.

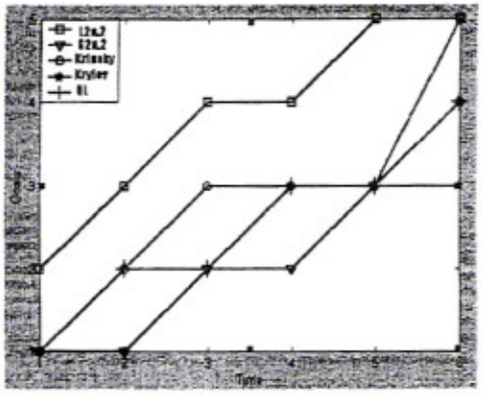
لازم بذکر است که در جدول زیر، ستون ۱ به معنی نتیجه تجمعی بازی از سیکل ۰ تا سیکل ۹۹۹، ستون ۲ به معنی نتیجه تجمعی بازی از سیکل ۱۰۰۰ تا سیکل ۱۹۹۹، ...، و ستون ۶ به معنی نتیجه تجمعی بازی از سیکل ۵۰۰۰ تا آخر بازی (سیکل ۵۹۹۹) می باشد. این مورد در محور افقی اشکال زیر (که بیانگر اطلاعات جدول هستند) نیز رعایت شده است.

	۱	۲	۳	۴	۵	۶
$L_{2N,2}$	۱-۲	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

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



شکل ۷- میانگین گلهای زده تیم های یادگیر ۲ نفره مختلف در مقابل تیم بدون یادگیری ۲ نفره در ۵۰ بازی در مقابل گذشت زمان بازی



شکل ۸- میانگین گلهای خورده تیم های یادگیر ۲ نفره مختلف در مقابل تیم بدون یادگیری ۲ نفره در ۵۰ بازی در مقابل گذشت زمان بازی

همانگونه که اشکال و نمودارهای فوق نشان می دهند، تفاوت زیادی بین اتوماتاهای یادگیر مختلف در سری شبیه سازیهای این فصل مشاهده نمی شود. هرچند که روشهای یادگیری Q و Krinsky اندکی بهتر از دیگر اتوماتاها می باشند.

باید خاطرنشان کرد که روش یادگیری Q قابل استفاده در دامنه ما، دارای محدودیتهایی است. به عنوان مثال، در این دامنه، هر عامل دارای دید محدود خود است و لهذا نمی تواند دیدی کامل از بقیه عاملها، محیط و همین طور تاثیر (بخصوص درازمدت) اعمال خود بر محیط و دیگر عاملها داشته باشد. از این رو برای پیاده سازی یادگیری Q، شبیه سازی ساده ای از روش یادگیری Q بر اساس روش بکار برده شده در [۲] (روش TPOT_RL) انجام شده است و در این شبیه سازی، تنها تاثیرات کوتاه مدت اعمال انجام گرفته توسط عامل، در تغییر مقادیر Q برای آن نقش دارند.

لازم بذکر است که ما در کلیه شبیه سازیهای خود، از خود عامل جهت قضاوت تاثیر عمل خود استفاده کرده ایم و از عامل ثالثی جهت مشاهده روند کار استفاده ننموده ایم که این خود می تواند دلیلی بر یادگیری چندعامله در شبیه سازیهای ما باشد. ضمن آنکه عامل همواره نمی تواند همه تغییرات در محیط خود را مشاهده کند و مقادیر زیادی حالات مخفی و پوشیده در محیط وجود دارند.

لهذا در اکثر موارد، اولین تاثیر قابل مشاهده عمل برای عامل انجام دهنده آن عمل، تخمین مناسبی از تاثیر درازمدت عمل انجام شده توسط عامل می باشد. بدین معنی که ممکن است تاثیر بلافاصله عمل انجام گرفته توسط عامل، منفی باشد ولی در چند سیکل بعد، حاصل انجام گرفتن عمل، در مجموع مثبت ارزیابی شود و عامل این تاثیر مثبت را مشاهده نماید.

بعنوان مثالی از این دست، می توان به انجام یک پاس در یک حالت بخصوص با توجه به قرارگیری بازیکن خودی و بازیکنان حریف اشاره کرد که ممکن است توسط بازیکنی از حریف که نزدیک بازیکن خودی قرار دارد دریافت شود، ولی بازیکن خودی در چند سیکل بعدی نهایتا توپ را صاحب شود. این مورد، یعنی قضاوت بر اساس اولین تاثیر قابل مشاهده توسط (دید محدود) خود بازیکن، در کلیه شبیه سازیهای انجام شده توسط ما رعایت شده است.

همانگونه که نتایج فوق نشان می دهند، تیم اتوماتای یادگیر موفق می شود در هر بازی، به مرور زمان، انجام عمل صحیح پاس و شوت را در ۴ حالت تعریف شده برای آن (در این سری از شبیه سازیها) فرا بگیرد و لهذا اکثر گلهای دریافتی آن در نیمه اول (۳۰۰۰ سیکل اول) می باشد و پس از آن، بازی را در دست می گیرد و اکثر گلهای خود را در نیمه دوم بازی (۳۰۰۰ سیکل دوم) به ثمر می رساند. شبیه سازیهای 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 در این قسمت آن بود که یادگیری چندعامله ما تا حد زیادی وابسته به حریف است. به این معنی که مقادیر حافظه ای بازیکنان یادگیر در هنگام بازی در برابر تیم ۳ نفره در این سری شبیه سازیها (بصورت میانگین) با مقادیر حافظه ای بازیکنان یادگیر در برابر تیم ۲ نفره تفاوت می کرد و این نشان دهنده آن است که بازیکنان ما یادگیری خود را بر اساس بازی حریف و بصورت همزمان انجام می دهند و لهذا از این نظر بر روشهای غیر همزمان ارجحیت دارند.

در واقع، مقدار از قبل تعیین شده ای برای همگرایی مقادیر حافظه بازیکنان ما وجود ندارد و شبیه سازیهای ما، در اغلب موارد، فرضهای ما را برای مقادیر حافظه رد می کنند. این بدان معنی است که بازیکنان ما در بعضی موارد که (از نظر ما)

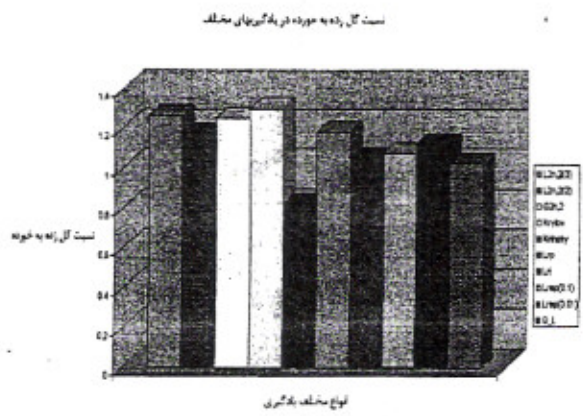
منطقا باید پاس بدهند، یاد گرفته اند که بهتر است به سمت دروازه حریف شوت کنند و این در واقع از محاسن یادگیری (بخصوص یادگیری همزمان) است.

هر چند، با توجه به وجود مقدار زیادی نویز (و حالات بسیار متفاوت پیش آمده در هر بازی) در محیط شبیه سازی ما نیز، مقادیر حافظه برای بازیکنان در بازیهای مختلف با هم تفاوت می کنند و مقادیر مشابه، بسیار کم پیش می آیند. لازم بذکر است که در سریهای بعدی شبیه سازی خود، برای عمومی سازی حالات محیطی، از تکنیک "بهترین گوشه در مربع حالت" [۱۲] استفاده نموده ایم. در این تکنیک، سعی بر آن است که ضمن آنکه عمومی سازی مناسبی از حالات بیشمار محیط به حالاتی محدود صورت بگیرد، تصمیم گیری هر عامل در محیط را به فضای محلی وی محدود کنیم. چنین فضایی را می توان فضای قابل اهمیت یک عامل نیز نامید.در واقع با افزایش حالات و اعمال برای عاملها در این قسمت، به پیاده سازی مجدد شبیه سازیهای ۲ بازیکن در مقابل ۲ بازیکن پرداختیم.

اگر تعداد گلهای زده به خورده در مقابل تیم بدون یادگیری (در مرحله آموزش توسط روشهای یادگیری مختلف) را از داده های این سری از شبیه سازیها جمع آوری کنیم، شاید بتوان نتایجی نسبی از تاثیر روشهای یادگیری متفاوت با استفاده از تکنیک بهترین گوشه در مربع حالت ارائه داد.

نام روش	نسبت گل زده به گل خورده در حین آموزش
اتوماتای $L_{2N,2}$ با عمق حافظه ۳ (۱۰ بازی)	۱/۲۶
اتوماتای $L_{2N,2}$ با عمق حافظه ۲ (۱۰ بازی)	۱/۱۷
اتوماتای $G_{2N,2}$ با عمق حافظه ۲ (۱۰ بازی)	۱/۲۵
اتوماتای Krylov با عمق حافظه ۲ (۱۰ بازی)	۱/۳۰
اتوماتای Krinsky با عمق حافظه ۲ (۱۰ بازی)	۰/۸۲
اتوماتای L_{RP} (۳۰ بازی)	۱/۱۸
اتوماتای L_{RI} (۱۵ بازی)	۱/۰۵
اتوماتای $L_{RεP}$ با نسبت پارامتر ۰/۱ (۱۵ بازی)	۱/۰۷
اتوماتای $L_{RεP}$ با نسبت پارامتر ۰/۰۱ (۱۵ بازی)	۱/۱۲
یادگیری Q (۱۰ بازی)	۱/۰۲

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)



شکل ۹- نسبت گل زده به خورده در یادگیریهای مختلف (در حین آموزش)

بعنوان یک نتیجه کلی از این قسمت می توان گفت که روش بهترین گوشه در مربع حالت، عمومی سازی مناسبی از حالات محیط و نگاشت حالات یکسان به یک حالت واحد انجام می دهد [۱۲].

در ادامه به بررسی شبیه سازیهای خود در همکاری عاملهای موجود در یک تیم ۵ نفره پرداختیم. در واقع، مهمترین تغییر نسبت به شبیه سازیهای قسمتهای قبل، دادن یک آرایش و شکل تیمی به تیم و استفاده از بازیکنان تیم در پستهای تخصصی بود که این موارد، با توجه به افزایش نفرات تیم و جهت برقراری نظم و همکاری هر چه بهتر در بین عاملهای موجود در یک تیم، ضروری بنظر می رسید.

هدف ما از انتخاب ۵ نفر بعنوان اعضای یک تیم، شبیه سازی مواردی چون مسابقات روبوکاپ با روباتهایی با اندازه های متوسط و بزرگ و نیز فوتبال داخل سالنی بود که در این دامنه ها از ۵ نفر عضو برای هر تیم استفاده می شود.

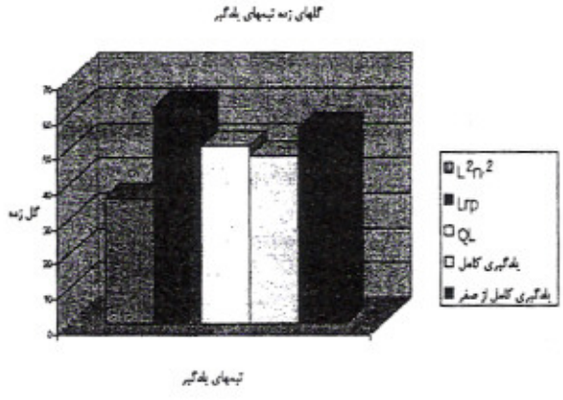
در سری شبیه سازیهای ۵ نفره، ما از تیمهای یادگیر $L_{2n,2}$ بعنوان نماینده تیمهای یادگیر با ساختار ثابت، L_{rp} بعنوان نماینده تیمهای یادگیر با ساختار متغیر، Q بعنوان نماینده ای از تیمهای دارای روش یادگیری غیر از اتوماتای یادگیر، و "تیم یادگیری کامل" [۱۲][۱۳] استفاده کردیم.

در شبیه سازیهای این قسمت، از ۱ دروازه بان، ۲ بازیکن کناری چپ و راست، یک دفاع عقب و یک بازیکن حمله استفاده شد و برای هر کدام از آنها، یک پست تخصصی خاص تعریف شد. این پستها که در واقع برای هر بازیکن، مجموعه ای از مربعهای حالت تعریف شده در تکنیک بهترین گوشه در مربع حالت می باشد، در مرزها دارای مربعهای مشترکی با هم هستند. این امر بدان جهت است که تعویض پستها به شکلی نرم و پیوسته انجام شود و نیز در قسمتهای حساس مرزی، دارای بیش از یک بازیکن باشیم.

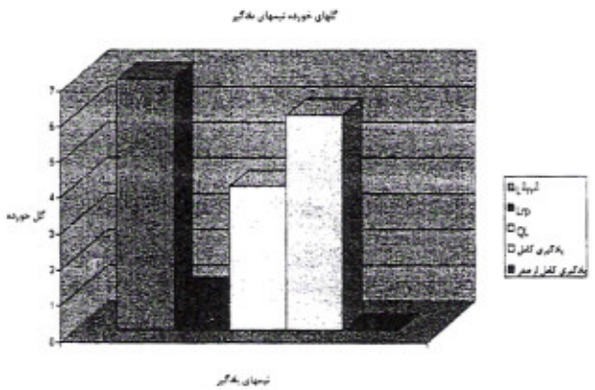
در بازیهای انجام شده در این قسمت نیز، تیمهای یادگیر، در هیچ دیداری بازنده نبودند و این مساله کارآیی مشاهده شده تیمهای اتوماتای یادگیر در شبیه سازیهای قبلی (۲ نفره) را برای ما قطعی تر کرد.

هر چند تعداد بازیهای انجام شده در این سری از شبیه سازیها محدود بود، شاید مقایسه تعداد گلهای زده و خورده در تیمهای یادگیر این فصل، ملاکی نسبی برای مقایسه آنها باشد. اشکال زیر مبین این نتایج هستند.

با توجه به نتایج موفقیت آمیز ابتدایی در شبیه سازیهای انجام گرفته، در سری شبیه سازیهای بعدی، به بررسی یادگیری در نحوه همکاری در بین عاملهای موجود در یک تیم کامل (با توجه به بستر تست ما، یک تیم ۱۱ نفره) و مقایسه و بررسی بیشتر اتوماتاهای یادگیر مختلف در همکاری بین عاملهای موجود در یک تیم پرداختیم.



شکل ۱۰- گلهای زده تیمهای یادگیر ۵ نفره در ۱۰ شبیه سازی (در حین آموزش)



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

این سری از شبیه سازیها مهمترین و کامل ترین سریهای شبیه سازیهای ما را در بردارد[۱۳]. این امر بدان جهت است که از تمام عاملهای مجاز در دامنه خود استفاده نموده ایم. منتها با توجه به اهمیت شبیه سازیهای این قسمت و این ایده که تنها تعداد گلهای زده و خورده نمی تواند ملاک تعیین صد در صد صحیحی برای شیوه همکاری در بین عاملهای موجود در یک تیم باشد (و با توجه به اینکه در بازیهای ۱۱ نفر در مقابل ۱۱ نفر، اصولا ممکن است تعداد گلهای رد و بدل شده خیلی زیاد نباشند)، سعی شد که جهت آنالیز نتایج شبیه سازی در این سری (و سریهای بعد)، از ملاکهای تعریف شده دیگری هم استفاده شود.

پس از جستجو در مورد ملاکهای مطرح در همکاری بین افراد یک تیم فوتبال، ملاکهای زیر به نظر ما مناسب تشخیص داده شده و مورد استفاده قرار گرفتند:

- درصد مالکیت توپ توسط تیم خودی در قیاس با مورد مشابه در تیم حریف در حین زمان بازی.
- درصد گردش توپ در ۱/۳ زمین خودی، ۱/۳ میانی زمین، و ۱/۳ زمین حریف در حین بازی.
- ماکزیموم زمان در اختیار داشتن توپ بصورت ممتد توسط تیم خودی در قیاس با مورد مشابه در تیم حریف در حین زمان بازی. این مورد بر حسب سیکل سنجیده می شود.
- ماکزیموم تعداد رد و بدلهای متوالی (و بدون برخورد به حریف) توپ توسط تیم خودی در قیاس با مورد مشابه در تیم حریف در حین زمان بازی.
- درصد خطای (اعمال) میانگین بازیکنان تیم خودی در حین بازی.

لازم بذکر است که ما حتی در پیاده سازی موارد فوق نیز، مساله چند عامله بودن کار را بطور کامل در نظر گرفته ایم. معمول این است که یک عامل که احاطه کاملی بر زمین بازی دارد (مانند عامل مربی در این دامنه)، موارد ثبت نتایج و آنالیز آنها را انجام دهد، ولی ما در این مورد نیز از خود عاملها جهت این امور استفاده کرده ایم.

این کارها بدین صورت انجام می گیرند که خود بازیکنان، در هر سیکل در طول بازی، در صورتی که توپ و/یا مالک فعلی آن را می بینند، این اطلاعات را در متغیرهای موقت خود ثبت می کنند. در پایان بازی، تمام عاملها این اطلاعات را درون فایلهایی که بهمین منظور در نظر گرفته شده اند می ریزند و با اجرای یک برنامه جنبی که برای این کار تهیه شده است، اطلاعات این فایلها با هم ترکیب می شوند و مورد آنالیز قرار می گیرند.

با توجه به این مساله که محل توپ و مالک فعلی آن همواره (و حداقل) توسط یک بازیکن از تیم قابل مشاهده است، تنها مساله ای که می ماند، ترکیب مناسب این فایلها و استفاده از اطلاعات آنهاست. بعنوان مثال برای استخراج فاکتور "ماکزیموم زمان در اختیار داشتن توپ بصورت ممتد توسط تیم خودی در حین زمان بازی"، برنامه فوق الذکر، فایل ترکیبی حاصل را از ابتدا به انتها مرور می کند و از هر نقطه زمانی که توپ در مالکیت یکی از بازیکنان خودی باشد، تا رسیدن به زمانی که توپ در اختیار یکی از بازیکنان حریف قرار بگیرد، زمان را ثبت می کند. این کار برای موارد مشابه بعدی نیز صورت می گیرد و در نهایت ماکزیموم چنین زمانهایی بعنوان فاکتور فوق ثبت می شود. بقیه فاکتورها نیز بطریقه های مشابه بدست می آیند.

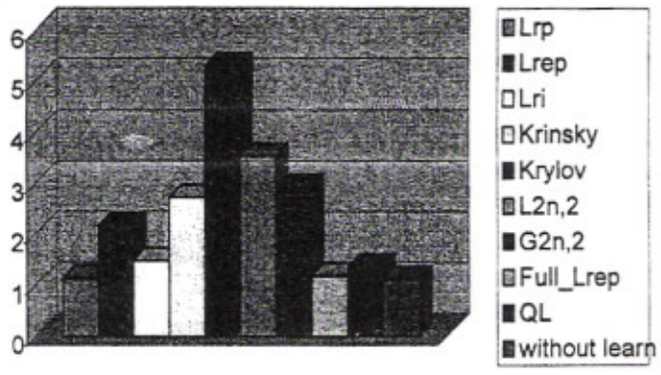
اما در هر صورت بدلیل مشکلاتی جزیی مانند احتمال نزدیکی بیش از حد ۲ یا بیشتر بازیکن به توپ در یک زمان خاص و لهذا ثبتهای مختلف توسط بازیکنان مختلف، احتمال مکث بیش از حد یک دروازه بان در ارسال توپ برای شروع

مجدد بازی و لهذا تاثیر در ثبت زمانهای ماکزیموم، و مواردی چون داشتن دید محلی (و نه لزوما پیوسته در زمان) و غیره، موارد ثبت شده برای این فاکتورها نمی توانند صد در صد صحیح باشند، هر چند تا جایی که امکان بوده است (و با توجه به مساله چند عامله بودن قضیه)، سعی شده است که این فاکتورها در اکثر مواقع بسیار نزدیک به واقعیت باشند.

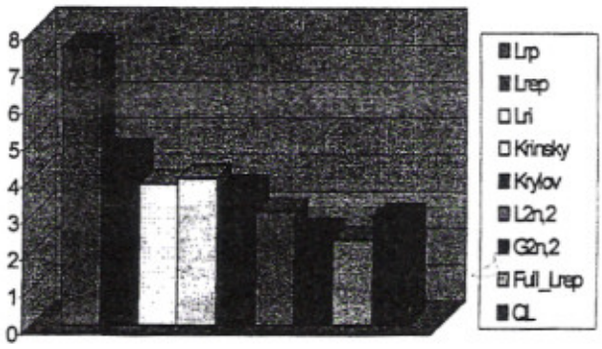
برای سازماندهی ۱۱ بازیکن درون زمین بازی نیز در سری شبیه سازیهای زیر(برای هر تیم) از یک شکل تیمی ۳-۳-۴ استفاده کرده ایم.بمانند شبیه سازیهای قبلیمان، در این سری نیز برای مقایسه روند یادگیری در تیمهای یادگیر و نیز بجهت داشتن یک پایه مناسب برای مقایسه اتوماتاهای یادگیر مختلف، یک تیم بدون یادگیری ایجاد کردیم. این تیم بدون یادگیری، جدا از مساله یادگیری بهترین عمل در هر حالت بر حسب تجربیات گذشته، در همه موارد دیگر (از جمله شکل تیمی فوق الذکر) مشابه تیمهای یادگیر است.

بعنوان نتیجه کلی از شبیه سازیهای این قسمت، تیمهای اتوماتای یادگیر موفق شدند با تعدادی محدود بازی آموزشی بر حریف بدون یادگیری خود غلبه کنند. آنها همچنین توانستند تیمی را که با دست کد شده بود و بهترین عمل ممکن در هر حالت را (بنا بر آنچه که به نظر می رسد در هر حالت بهترین باشد) انجام می داد، شکست دهند.

اشکال زیر، نتایج کلی شبیه سازیهای انجام گرفته و بازیهای انجام شده بین تیمهای یادگیر و تیم بدون یادگیری را در دوسری آموزشی (۱۵ بازی) و آزمایشی (۳ بازی پس از بازیهای آموزشی) خلاصه می کنند.



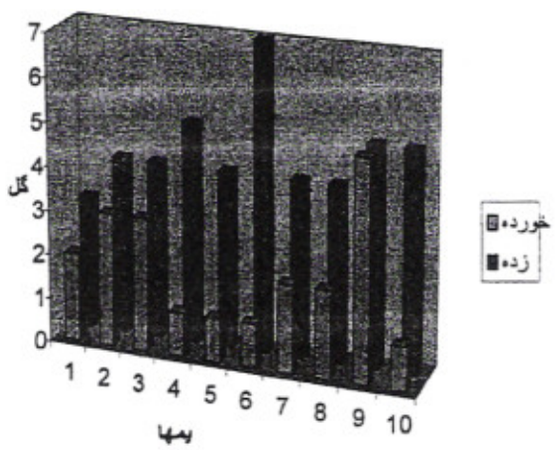
شکل ۱۲–مقایسه نسبت میانگین گلهای زده به خورده در هر بازی آموزشی و در مقابل تیم بدون یادگیری در تیمهای مختلف



شکل ۱۳– مقایسه نسبت میانگین گلهای زده به خورده در هر بازی پس از آموزش (آزمایشی) و در مقابل تیم بدون یادگیری در تیمهای مختلف

همانگونه که مشاهده می شود (حداقل در تعداد محدود بازیهای آموزشی ما) رابطه خاصی بین تعداد گلهای زده (یا خورده) در هر بازی آموزشی و تعداد گلهای زده (یا خورده) در هر بازی آزمایشی وجود ندارد. منتها با تعداد بازیهای محدود آموزشی ما، برتری نسبی در بازیهای آموزشی در مورد تیمهای اتوماتای یادگیر با ساختار ثابت دیده می شود.

در مورد بازیهای آزمایشی (پس از آموزش و تنها با استخراج مقادیر یادگرفته شده) نیز، به نظر می رسد که برتری نسبی (در غلبه بر تیم بدون یادگیری) با تیمهای اتوماتای با ساختار متغیر باشد. هر چند در [۱۳] نتایج بدست آمده برای فاکتورهای دیگر نیز (برای هر مورد و بصورت جزیی) آورده شده اند. شکل زیر نیز، گلهای زده و خورده تیمهای یادگیر را در مقابل تیم ثابت (کد شده با دست) را خلاصه می کند. این تیم بصورت دستی بهینه شده است.



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، یادگیری کامل، $L_{R\varepsilon P}$ ، L_{RP} ، L_{RI} ، $L_{R\varepsilon P}$ ، Krylov، Krinsky، $L_{2N,2}$ و $G_{2N,2}$.

۵. نتیجه گیری

هدف ما در این مقاله، بوجودآوردن تکنیک هایی برای تولید موفق سلسله فعالیتهایی برای عاملهای عضو یک تیم یادگیر بود بگونه ای که تیم حاصل بتواند در محیط هایی چند عامله، دارای دشمن، همکاری گرا، نویزی و زمان واقعی بخوبی عمل نماید. به دلیل پیچیدگی ذاتی چنین سیستمهایی، ناگزیر از استفاده روشهای یادگیری ماشین هستیم. روش یادگیری مورداستفاده ما در این پروژه، اتوماتای یادگیر می باشد.

اتوماتای یادگیر که قادر به عمل کردن در محیطی تصادفی است، بعنوان مدلی برای سیستمهای یادگیر در موارد مختلف و با موفقیتهای قابل ملاحظه ای بکار رفته است. تأکید ما در این مقاله، بر استفاده از اتوماتای یادگیر در همکاری بین عاملهای موجود در یک تیم بود و بر روی استفاده از این اتوماتاها برای موارد دیگری که می توانند از دیدگاه یادگیری در سیستم های چند عامله مطرح شوند، تأکیدی نداشته ایم. در واقع، با پیاده سازی تیمهای ۲ نفره، ۵ نفره و ۱۱ نفره از عاملهای دارای خاصیت یادگیری اتوماتا و مقایسه آن با تیم مشابه و بدون خاصیت یادگیری و یا تیمهای یادگیر دیگر[۱۴]، کارآیی اتوماتای یادگیر در یادگیری یک کار تیمی و همکاری در جهت دست یافتن به یک هدف مشترک، بسیار خوب ارزیابی می شود. از نظر فاکتورهای تعریف شده ما برای این سری از شبیه سازیها نیز، تیم های اتوماتای یادگیر بسیار بهتر از دیگر تیمها نشان دادند.

تحقیق ارائه شده در این مقاله، اولین تحقیق جدی در مورد استفاده از اتوماتاهای یادگیر در سیستم های چند عامله و بخصوص در همکاری در سیستم های چند عامله و نیز در محیط شبیه ساز فوتبال روباتها بشمار می رود. در اکثر مراجع چنین مواردی با استفاده از دیگر روشهای یادگیر متداول چون شبکه های عصبی، یادگیری Q ، الگوریتمهای ژنتیک، درخت تصمیم گیری و دیگر روشهای یادگیری تقویتی انجام شده اند، ولی هیچکدام از این مراجع، اتوماتاهای یادگیر را مورد استفاده قرار نداده اند. بدین جهت، تحقیق حاضر، کاربرد اتوماتای یادگیر را در دامنه های فوق الذکر، بصورتی بدیع فراهم نموده است. از جهت مقایسه با دیگر کارهای انجام شده در این رابطه نیز با توجه به بازیهای انجام شده بین تیم اتوماتای یادگیر با دیگر تیم های یادگیر [۱۴] و فاکتورهای ثبت شده ما در این بازیها، اتوماتای یادگیر نتایج بسیار خوبی را نشان می دهد.

مورد دیگری که باید به آن اشاره شود این است که روشهایی که ما برای تعیین حالت و انواع استفاده از اتوماتاهای یادگیر بجهت همکاری در بین عاملهای عضو تیم های یادگیر خود مورد استفاده قرار داده ایم، روشهایی عمومی و کلی محسوب می شوند و می توان با تغییراتی آنها را در دیگر دامنه ها و یا بسترهای تست نیز مورد پیاده سازی قرار داد.

ما تنها در موارد پیاده سازی و در مواردی که بستگی مستقیم به دامنه شبیه سازی انتخابی ما داشته است از خصوصیات وابسته به دامنه استفاده کرده ایم و در دیگر موارد، تنها بر نحوهٔ ایجاد همکاری در بین عاملهای موجود در یک تیم از عاملها با استفاده از اتوماتای یادگیر را مد نظر داشته ایم (لازم به تذکر است که ما در تمام موارد، چند عامله بودن پروژه را مد نظر داشته ایم و هیچ موردی که از پردازشهای متمرکز در این پروژه استفاده کرده باشد، وجود ندارد).

مراجع

1- Weiss G., Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence, The MIT Press, London, 1999.

2- Stone P., Layered Learning in Multi_Agent Systems, PhD thesis, School of Computer Science, Carnegie Mellon University, December 1998.

3- Noda I., Team GAMMA: Agent programming on gaea, In Kitano H., editor, RoboCup-97: Robot Soccer World Cup I, pages 500-507, Springer Verleg, Berlin, 1998.

4- RoboCup web page, 1997, At URL http://www.robocup.org.

5- Kitano H., editor, RoboCup-97: Robot Soccer World Cup I, Springer Verlag, Berlin, 1998.

6- Andre D., Corten E., Dorer K., Gugenberger P., Joldos M., Kummenje J., Navaratil P. A., Noda I., Riley P., Stone P., Takahashi R., and Yeap T., Soccer server manual, version 4.0, Technical Report RoboCup –1998-001, RoboCup, 1998.

7- Narendra K.S. and Thathachar M.A.L., Learning Automata: An Introduction, Prentice Hall, Inc., 1989.

8- Mars P., Chen J.R. and Nambir R., Learning Algorithms: Theory and Applications in Signal Processing, Control and Communications, CRC Press, Inc., pp. 5-24, 1996.

9- Lakshmivarahan S., Learning Algorithms: Theory and Applications, New York, Springer Verlag, 1981.

10- Meybodi M.R. and Lakshmivarahan S., ε — Optimality of a General Class of Absorbing Barrier Learning Algorithms, Information Sciences, Vol. 28, pp. 1-20, 1982.

11- Meybodi M.R. and Lakshmivarahan S., 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 – محمد رضا خجسته و محمد رضا میبدی/ تکنیک "بهترین گوشه در مربع حالت"برای عمومی سازی حالات محیطی در یک دامنه چند عامله همکاری گرا/ مرکز تحقیقات انفورماتیک/ آزمایشگاه محاسبات نرم/ دانشکده مهندسی کامپیوتر/ دانشگاه صنعتی امیرکبیر/ بهار ۱۳۸۱.

۱۳- محمد رضا خجسته/ همکاری در سیستمهای چند عامله با استفاده از اتوماتای یادگیر/ پایان نامه کارشناسی ارشد/ دانشکده مهندسی کامپیوتر/ دانشگاه صنعتی امیرکبیر/ بهار ۱۳۸۱.

۱٤- محمد رضا خجسته و محمد رضا میبدی/ ارزیابی اتوماتای یادگیر در همکاری بین عاملها در یک سیستم چند عامله پیچیده/ مرکز تحقیقات انفورماتیک/ آزمایشگاه محاسبات نرم/ دانشکده مهندسی کامپیوتر/ دانشگاه صنعتی امیرکبیر/ بهار ۱۳۸۱.

[1] Cooperative Team
[2] Robocup
[3] the robocup soccer server –
[4] real time –
[5] autonomous –
[6] noisy –
[7] collaborative –
[8] adversarial –
[9] hidden state –
[10] noisy –
[11] cycles –
[12] perception –
[13] action –
[14] unknown –
[15] realistic –
[16] Planned –
[17] existing –
[18] stamina –
[19] discrete –
[20] simulator cycle –
[21] marker –
[22] aural –
[23] visual –
[24] physical –
[25] crowded –
[26] Learning Automaton
[27] Fixed Structure
[28] Variable Structure
[29] Environment
[30] Unfavorable
[31] Stationary
[32] Non-Stationary
[33] Actions
[34] Variable Structure
[35] Linear Reward Penalty
[36] Linear Reward Epsilon Penalty
[37] Linear Reward Inaction