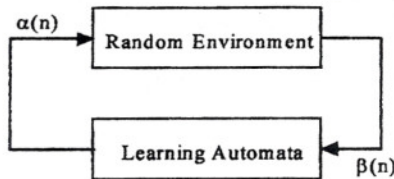


و عمل خود را برای مرحله بعد انتخاب می کند. شکل ۱ ارتباط بین اتوماتای یادگیر و محیط را نشان می دهد.



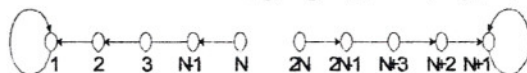
شکل ۱: ارتباط بین اتوماتای یادگیر و محیط

محیط^۲ را می توان توسط سه تایی $E \equiv \{\alpha, \beta, c\}$ نشان داد که در آن $\alpha \equiv \{\alpha_1, \alpha_2, \dots, \alpha_r\}$ مجموعه ورودیها، $\beta \equiv \{\beta_1, \beta_2, \dots, \beta_m\}$ مجموعه خروجیها و $c \equiv \{c_1, c_2, \dots, c_r\}$ مجموعه احتمالهای جریمه می باشد. هر گاه β مجموعه دو عضوی باشد، محیط از نوع P می باشد. در چنین محیطی $\beta_1 = 1$ به عنوان جریمه و $\beta_2 = 0$ به عنوان پاداش در نظر گرفته می شود. در محیط از نوع Q، $\beta(n)$ می تواند به طور گسسته یک مقدار از مقادیر محدود در فاصله $[0, 1]$ و در محیط از نوع S، $\beta(n)$ متغیر تصادفی در فاصله $[0, 1]$ است. c_i احتمال اینکه عمل α_i نتیجه نامطلوب^۳ داشته باشد می باشد. در محیط ایستا^۴ مقادیر c_i بدون تغییر می مانند، حال آنکه در محیط غیر ایستا^۵ این مقادیر در طی زمان تغییر می کنند. اتوماتای یادگیر به دو گروه با ساختار ثابت و با ساختار متغیر تقسیم می گردند.

اتوماتای یادگیر با ساختار ثابت^۶: اتوماتای یادگیر با

ساختار ثابت توسط ۵ تایی $\{\alpha, \beta, F, G, \phi\}$ نشان داده می شود که در آن $\alpha \equiv \{\alpha_1, \alpha_2, \dots, \alpha_r\}$ مجموعه اعمال اتوماتا، $\beta \equiv \{\beta_1, \beta_2, \dots, \beta_m\}$ مجموعه ورودیهای اتوماتا، $\phi \equiv \{\phi_1, \phi_2, \dots, \phi_s\}$ مجموعه وضعیتهای داخلی اتوماتا، $F: \phi \times \beta \rightarrow \phi$ تابع تولید وضعیت جدید اتوماتا، و $G: \phi \rightarrow \alpha$ تابع خروجی که وضعیت کنونی را به خروجی بعدی می نگارد می باشند. بعنوان مثال به بررسی دو نمونه از اتوماتاهای با ساختار ثابت می پردازیم.

اتوماتای $L_{2N,2}$: این اتوماتا تعداد پاداش ها و جریمه های در یافت شده برای هر عمل را نگهداری کرده و تنها زمانی که تعداد جریمه های پشت سر هم بیشتر از تعداد پاداش ها می گردد، عمل دیگر را انتخاب می کند. نمودار تغییر وضعیت این اتوماتا مطابق شکل ۲ می باشد.



پاسخ مطلوب $\beta=0$

² Environment
³ Unfavorable
⁴ Stationary
⁵ Non-Stationary
⁶ Fixed Structure

می کنند تا در طراحی و اجرای سیاستهای چند عامله امکانات بیشتری را فراهم نمایند. طراحی و پیاده سازی مهارتهای پیچیده برای عامل فوتبالیست وظیفه ای سخت است و علت آن علاوه بر عدم قطعیت محیط و دخالت عوامل دیگر، جنبه ها و حالتی است که تغییر آنها در عملکرد یک مهارت بخصوص تاثیر دارد. در طراحی یک مهارت پیچیده تا حد ممکن باید مهارت را مستقل از جنبه ها و حالتی خاص تعریف کرد در غیر اینصورت باید تمام حالتی ممکن که تعداد آنها بسیار زیاد است و همیشه قابل پیش بینی نمیشاند را مورد بررسی قرار داد. یک راه برای حل این مشکل استفاده از روشهای یادگیری می باشد. اکثر تیمها از روش یادگیری نظارتی [7] و یادگیری تقویتی [3] به عنوان روش یادگیری و آموزش عامل خود استفاده میکنند.

در این مقاله برای اولین کاربرد اتوماتای یادگیر در یادگیری مهارتهای پیچیده مانند در یافت توپ و دریبل در فوتبال رباتها مطرح میگردد. در قسمت اول این مقاله کاربرد اتوماتاهای یادگیری در یادگیری مهارت دریافت توپ ارائه می گردد. مسئله دریافت توپ در حالت عمومی عبارت از حرکت به سمت توپ در حال حرکت و در اختیار گرفتن توپ می باشد. اتوماتای یادگیر جهت مناسب برای حرکت عامل فوتبال را برای حرکت به سمت توپ در حال حرکت تعیین می کند. در روش پیشنهادی ابتدا عامل فوتبالیست را با یک آموزش جزئی غیر برخط آماده می کنیم. اتوماتای یادگیر سپس در حین بازی بصورت برخط سعی در بهبود کارایی عامل فوتبال از طریق سازگار نمودن عامل با تغییرات محیط مینماید. در قسمت دوم این مقاله کاربرد اتوماتاهای یادگیر در یادگیری مهارت پیچیده دریبل ارائه میگردد. برای یادگیری مهارت دریبل از دو اتوماتای یادگیر استفاده شده است. این دو اتوماتای یادگیر به طور همزمان عمل کرده و قدرت لازم و جهت مناسب را برای انجام عمل دریبل تعیین می کنند. در این روش عامل فوتبالیست در ابتدا با مقدار مناسبی آموزش غیر برخط آماده می شود. عامل فوتبال به طور برخط در حین بازی واقعی (با استفاده از اتوماتای یادگیری که برای این منظور طراحی شده است) همچنان به بهبود کارایی خود و سازگاری با تغییرات محیط ادامه میدهد.

ادامه مقاله بصورت زیر سازماندهی شده است. در بخش ۲ مقدماتی در باره اتوماتاهای یادگیر ارائه میگردد. در بخش ۳ کاربرد اتوماتای یادگیر در یادگیری در یافت توپ و در در بخش ۴ یادگیری دریبل شرح داده میشود. بخش ۵ مقاله نتیجه گیری می باشد.

۲- اتوماتای یادگیر^۱

اتوماتای یادگیر یک مدل انتزاعی است که تعداد معدودی عمل را می تواند انجام دهد. هر عمل انتخاب شده توسط محیطی احتمالی ارزیابی شده و پاسخی به اتوماتای یادگیر داده می شود. اتوماتای یادگیر از این پاسخ استفاده نموده

¹ Learning Automata