

# یک روش جدید مبتنی بر شاخصه های آماری توزیع، برای تنظیم خودکار نرخ یادگیری در آتاماتای یادگیر

محمد رضا ملاخلیلی میبیدی<sup>۱</sup>، محمد رضا میبیدی<sup>۲</sup>

<sup>۱</sup> دانشکده فنی و مهندسی دانشگاه آزاد اسلامی، واحد علوم و تحقیقات تهران، m.meybodi@srbiau.ac.ir

<sup>۲</sup> دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، mmeibodi@aut.ac.ir

چکیده - یکی از مسائل مطرح در طراحی سیستم های یادگیر نظیر آتاماتای یادگیر، تعیین نرخ یادگیری است. در بیشتر موارد از یک الگوریتم کاهش یابنده در طول زمان برای تنظیم نرخ یادگیری استفاده می شود تا به مرور زمان سیستم مورد آموزش به پایداری برسد. به این ترتیب نرخ یادگیری، پارامتری است که مشخص می کند یک سیستم یادگیری تا چه حد به تجربیات گذشته متکی، و به چه میزان از رویدادهای فعلی اثرپذیر باشد. این روش کارآمد است اما در محیطهای پویا عملکرد درستی ندارد.

در این مقاله روشی جدید برای تنظیم نرخ یادگیری در آتاماتای یادگیر پیشنهاد شده است. این روش جدید علاوه بر طول مدت یادگیری، از معیارهای آماری میانگین و واریانس توزیع فعلی به دست آمده برای بردار احتمالات اقدامهای آتاماتا به منظور تعیین افزایش یا کاهش نرخ یادگیری استفاده کرده و بر خلاف روش های موجود، در طول فرآیند یادگیری هم افزایش و هم کاهش مقدار نرخ یادگیری را انجام داده و نسبت به تغییرات محیط واکنش مناسب نشان میدهد.

بررسی عملکرد الگوریتم در محیط های پویای تصادفی نشان می دهد الگوریتم جدید ضمن انعطاف پذیری بیشتر در انطباق با محیطهای پویا، از نقاط بیشینه محلی فرار کرده و مقادیر یادگرفته شده به مقادیر حقیقی نزدیک تر هستند.

کلید واژه - آتاماتای یادگیر، نرخ یادگیری پویا، تنظیم نرخ یادگیری، میانگین، واریانس، شاخصه های آماری، محیط تصادفی پویا

از قدرت تطابق پذیری و یادگیری خوبی در برخورد با محیط های تصادفی گوناگون برخوردار نباشد. (3)

## ۱- مقدمه

پژوهش مستقلی که به بررسی تنظیم خودکار نرخ یادگیری در آتاماتای یادگیر پرداخته باشد، مشاهده نشده است. عموم پژوهش های این حوزه مربوط به شبکه های عصبی می باشند که برخی از آنها به لحاظ عمومی که در مبانی دارند برای سایر سیستم های یادگیر و از جمله آتاماتای یادگیر نیز قابل استفاده هستند.

بحث تنظیم خودکار نرخ یادگیری یکی از مسائلی است که در مباحث مرتبط با الگوریتم ها و سیستم های یادگیری و خصوصا در مباحث مربوط به شبکه های عصبی به کرات مورد بررسی قرار گرفته است (1) (2) (3). اکثر الگوریتم ها از یک نرخ یادگیری با مقدار بالا شروع کرده و در حین فرآیند آموزش به کمک یک تابع کاهش یابنده با زمان (مثلا  $\alpha(t) = \alpha(0)e^{-\frac{t}{T}}$ ) آن را کاهش میدهند (1). به این ترتیب پس از مدتی نرخ یادگیری مقدار معمولا کوچکی دارد.

## ۲- روش جدید پیشنهادی

برای ارائه روشی که بتواند تغییرات نرخ یادگیری را بر اساس تغییرات محیط و تغییرات بردار احتمال انتخاب اقدام های آتاماتا انجام دهد، نیاز به معیار مقایسه ای داریم تا بر اساس آن بتوان تشخیص داد تغییرات رخ داده در محیط از یک مقدار آستانه بیشتر است تا با افزایش نرخ یادگیری بتوان این تغییرات را منعکس کرد. ابزار ریاضی که این معیار را در اختیار می گذارد،

مشکل از اینجاست که اگر از این مرحله به بعد، نمونه هایی وارد شوند که مشخصه های آماری متفاوتی با مشخصه های آماری نمونه های قبلی داشته باشند، سیستم یادگیر قادر به یادگیری آنها نیست.

بنابراین نرخ یادگیری کاهش یابنده با زمان، با وجود مزیت هایی که در همگرایی سریع دارد، باعث می شود سیستم یادگیر

نامساوی مارکوف است.

**قضیه** (نامساوی مارکوف): اگر  $X$  یک متغیر تصادفی و  $a$  یک عدد حقیقی مثبت باشد، در این صورت

$$Pr(|X| > a) \leq \frac{E(X^2)}{a^2} \quad (1)$$

برای اثبات این قضیه می‌توانید به مرجع (1) مراجعه کنید.

با استفاده از نامساوی مارکوف می‌توان به نامساوی چبیشف رسید. اگر  $m$  نشان دهنده میانگین متغیر تصادفی  $X$  باشد با جایگذاری  $X-m$  در رابطه بالا به نامساوی، موسوم به نامساوی چبیشف خواهیم رسید.

$$Pr(|X - m| \geq a) \leq \frac{Var(X)}{a^2} \quad (2)$$

اگر  $Var(X)$  را با  $\delta$  نشان دهیم تفسیر دیگری از نامساوی

چبیشف به دست می‌آید

$$Pr(|X - m| \geq a\delta) \leq \frac{1}{a^2} \quad (3)$$

رابطه 3 در حقیقت بیان می‌کند که احتمال اینکه یک

متغیر تصادفی در خارج از بازه ای حول میانگین به شعاع  $a$  برابر واریانس باشد از  $\frac{1}{a^2}$  کمتر است. این رابطه برای مقادیر  $a > 1$  واجد اطلاعات مفید است.

یک آتاماتای یادگیر به عنوان مدلی از یک سیستم یادگیر است که در محیط های تصادفی ناشناخته عمل می‌کند. آتاماتا در هر دور یک اقدام از میان مجموعه محدود اقدام های خود انتخاب کرده و با بررسی عکس العمل محیط نسبت به این اقدام، احتمال انتخاب اقدام های بعدی را بهبود می بخشد. (5) (6)

آتاماتای یادگیر با ساختار متغیر، به کمک 6 تایی  $\langle \beta, \emptyset, a, P, G, T \rangle$  نمایش داده می شود.  $\beta$  مجموعه ورودی ها،  $\emptyset$  مجموعه حالات داخلی،  $a$  مجموعه خروجی ها است.  $G$  نگاشت خروجی و  $T$  الگوریتم یادگیری است. الگوریتم یادگیری یک رابطه بازگشتی است و برای اصلاح بردار احتمال حالت ها به کار می رود.

با این مقدمات فرض کنید یک آتاماتای یادگیر با  $n$  اقدام در یک محیط تصادفی فعالیت می کند. محیط تصادفی اقدام انجام شده توسط آتاماتا را ارزیابی می کند و آتاماتا بر اساس این ارزیابی، بردار احتمالات اقدام های خود را به روزرسانی می کند. فرض کنید  $P_i^t$  مقدار احتمال انتخاب اقدام  $i$  ام آتاماتا را در زمان  $t$  نشان دهد. ضمناً

$$\sum_{i=1}^n P_i^t = 1 \quad \forall t \in \{0, 1, \dots\}$$

بردار احتمال انتخاب اقدام های آتاماتا در زمان  $t$  را با  $P(t)$

$$\overline{P(t)} = [P_1^t, P_2^t, \dots, P_n^t]$$

علاوه بر این محیط تصادفی بر اساس بردار احتمالات  $\vec{C} = [C_1, C_2, \dots, C_n]$  اقدام های محیط را پاداش میدهد که در آن  $\sum_{i=1}^n C_i = 1$  است.

در این آتاماتای یادگیر هر اقدام آتاماتا دارای یک نرخ یادگیری است. این بردار را با  $\vec{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_n]$  نشان میدهیم. در این مقاله فرض می کنیم آتاماتای یادگیر برای به روزرسانی بردار احتمال انتخاب های خود از الگوریتم  $L_{RI}$  استفاده می کند.

**گزاره ۱:** رابطه به روزرسانی بردار احتمال انتخاب اقدام آتاماتای یادگیر برای زمانی که اقدام  $k$  ام توسط آتاماتا صورت گرفته است را می توان به صورت زیر نوشت:

$$P_k^{t+1} = P_k^t * (1 - \alpha_k) + \alpha_k R^{t+1} \quad (4)$$

در رابطه ۴،  $R$  پاسخ محیط به اقدام انجام شده توسط آتاماتا است و داریم

$$R^{t+1} = \begin{cases} 1 & \text{with probability } C_k \\ 0 & \text{with probability } 1 - C_k \end{cases} \quad (5)$$

**گزاره ۲:** فرض کنید  $X$  متغیری است که بر اساس رابطه ۶ به روز رسانی می شود

$$X^{t+1} = X^t * (1 - \alpha) + \alpha R^{t+1} \quad (6)$$

که در آن

$$R^{t+1} = \begin{cases} 1 & \text{with probability } C^* \\ 0 & \text{with probability } 1 - C^* \end{cases} \quad (7)$$

است. نشان دهید:

$$\lim_{t \rightarrow \infty} E[X^t] = \mu = C^* \quad \text{الف-}$$

ب- واریانس  $X^t$  محدود و دارای مقدار حدی

$$Var(X^t) = \delta^2 = \frac{\alpha}{2-\alpha} C^*(1 - C^*)$$

است.

ج- مطابق با نامساوی چبیشف

$$\forall q > 0 \quad P(|X^t - C^*| \geq q\delta) \leq \frac{1}{q^2}$$

اثبات:

الف-  $\langle R^{t+1} \rangle$  یک فرآیند تصادفی برنولی با پارامتر  $C^*$  است. علاوه بر این داریم:

$$E[X^{t+1}] = (1 - \alpha)E[X^t] + \alpha E[R^{t+1}] \quad (8) \\ = (1 - \alpha)E[X^t] + \alpha C^*$$

از رابطه بازگشتی (8) داریم:

$$E[X^{t+1}] = (1 - \alpha)^t E[X^0] + (1 - (1 - \alpha)^t) * C^* \quad (9)$$

با توجه به اینکه  $0 < \alpha < 1$  رابطه ۹ نشان می‌دهد که

$$\lim_{t \rightarrow \infty} E[X^{t+1}] = C^* \quad (10)$$

ب: به طریق مشابه میتوان نشان داد که

$$Var[X^{t+1}] = \frac{\alpha}{2-\alpha} C^*(1 - C^*)(1 - (1 - \alpha)^{2t+2}) \quad (11)$$

های قبلی استفاده می کنیم و به کمک آن در موارد لازم، نرخ یادگیری را افزایش میدهیم.

#### الگوریتم جدید پیشنهادی

۱- آتاماتای یادگیر برای یادگیری آماده سازی می شود

- بردار احتمالات انتخاب اقدام های آتاماتا را با توزیع یکنواخت مقاردهی می شود ( $p_i$  برای اقدام  $i$  ام)
  - برای هر یک از اقدام های آتاماتا نرخ یادگیری متناظر با یک مقدار بزرگ نزدیک به ۱ مقاردهی می شود ( $\alpha_i$  برای اقدام  $i$  ام)
  - پارامترهای کنترلی  $q > 1$  و  $Tsh < 1$  مقاردهی اولیه می شوند.
  - برای هر یک از اقدام های آتاماتا مقدار  $t$  با صفر مقاردهی می شود ( $t_i$  برای اقدام  $i$  ام)
  - برای هر یک از اقدام های آتاماتا مقدار اولیه میانگن و واریانس متناظر مقاردهی می شود ( $\mu_i$  و  $\delta_i$  برای اقدام  $i$  ام)
- ۲- آتاماتای یادگیر بر اساس بردار احتمال انتخاب اقدام های آتاماتا یک عمل را به صورت تصادفی انتخاب می کند.
- ۳- محیط اقدام انجام شده را ارزیابی می کند
- ۴- آتاماتا بر اساس الگوریتم  $L_{RI}$  بردار مقادیر احتمال انتخاب اقدام های خود را به روزسانی می کند.
- ۵- برای هر یک از اقدام های آتاماتا عملیات به روزسانی مقادیر  $t$  و میانگین و واریانس و نرخ یادگیری به صورت زیر انجام می شود
- مقادیر  $\mu_i$  و  $\delta_i$  را به روزسانی کن
  - اگر  $|p_i - \mu_i(p_i^{t+1})| > q \cdot \delta_i(p_i^{t+1})$  در این صورت  $t_i = 0$  و  $\alpha_i$  را افزایش بده. مقدار میانگین ریست می شود
  - وگرنه اگر  $Tsh < (1 - \alpha_i)^t$  در این صورت  $t_i = 0$  و  $\alpha_i$  را کاهش بده
  - وگرنه  $t_i = t_i + 1$
- ۶- برو به گام ۲

در این الگوریتم برای اجتناب از محاسبات، از یک کران بالا برای مقدار واریانس استفاده می کنیم. برای رسیدن به این کران بالا قسمت ب گزاره ۲ و اثبات آن را در نظر بگیرید. نشان داده شد که مقدار واریانس در زمان  $t+1$  برابر است با:

$$Var[X^{t+1}] = \frac{\alpha}{2-\alpha} C^*(1 - C^*)(1 - (1 - \alpha)^{2t+2}) \quad (13)$$

رابطه ۱۳ نشان میدهد که واریانس تابعی از میانگین ( $C^*$ ) و نرخ یادگیری ( $\alpha$ ) است.

شکل ۱ رابطه میان واریانس و میانگین را به ازای نرخ های مختلف یادگیری نشان میدهد. همانگونه که این شکل نیز نشان میدهد، هر چقدر میانگین به ۰ یا ۱ نزدیکتر باشد، واریانس کوچکتر است. برعکس در میانگین برابر با ۰.۵، واریانس بیشینه مقدار را دارد. علاوه بر این، می توان دید که هر چقدر نرخ یادگیری به ۱ نزدیکتر باشد (نمودار بالایی) واریانس مقدار بیشتری دارد و بالعکس، در مقادیر کوچکتر نرخ یادگیری (پایین ترین نمودار)، واریانس کمتری داریم.

با توجه به اینکه  $0 < \alpha < 1$  رابطه ۱۱ نشان میدهد که

$$\lim_{t \rightarrow \infty} Var[X^{t+1}] = \frac{\alpha}{2-\alpha} C^*(1 - C^*)$$

ج- با فرض  $m = E[X]$  برای متغیر تصادفی  $X$  نامساوی مارکوف

$$P(|X| \geq a) \leq \frac{E(X^2)}{a^2}$$

به صورت زیر تبدیل خواهد شد

$$P(|X - m| \geq a) \leq \frac{Var(X)}{a^2} \quad (12)$$

با جایگذاری  $m = E[X^t]$  و  $a = q \cdot \sqrt{Var(X^t)} = q \cdot \delta$

$C^*$  در مورد متغیر تصادفی  $X = X^t$  در رابطه ۱۲ نتیجه ج

حاصل خواهد شد ■

گزاره ۲ چند نکته را نشان میدهد:

اولا مقدار  $\alpha$  کنترل کننده نرخ همگرایی است. روابط بازگشتی مربوط به میانگین و واریانس نشان میدهند که اگر  $\alpha$  برابر با ۱ باشد سرعت همگرایی بیشینه است. اگرچه مقدار  $\alpha$  برابر با ۱ سرعت همگرایی بالایی دارد اما منجر به واریانس بزرگتری می شود (شکل ۱). مقادیر  $\alpha$  کوچک تر گرچه همگرایی کندتری را سبب می شوند، اما جواب نزدیکتری به مقدار احتمال واقعی دارند. نکته حائز اهمیت در آن است که واریانس به صفر نمی رسد.

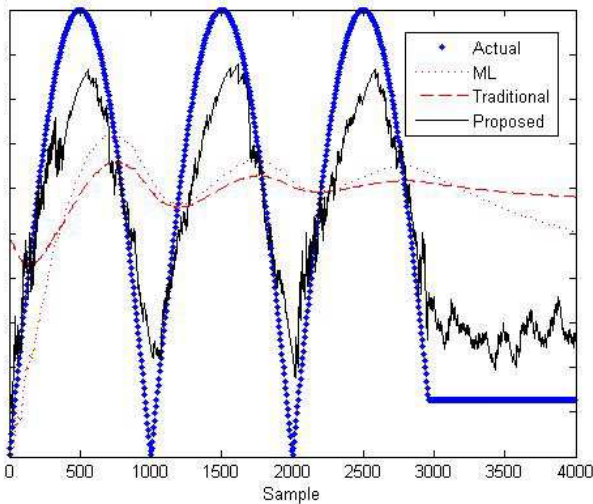
دوم اینکه گرچه وجود واریانس ناخوشایند به نظر می رسد، اما یک مقدار محدود و کوچک واریانس باعث می شود با مشاهده داده های جدید، الگوریتم از گیرافتادن در نقاط بیشینه محلی نجات یابد. علاوه بر اینکه باعث می شود قابلیت انطباق با محیط های متغیر را نیز پیدا کند. سرعت این انطباق پذیری به کمک  $\alpha$  یا همان نرخ یادگیری قابل کنترل است. ضمنا گیر نیفتادن در نقاط بیشینه محلی، تضمین نشده است و به مقدار  $\alpha$  و شکل تابع توزیع بستگی دارد.

سوم اینکه نامساوی چبیشف یک بازه اطمینان از مقدار احتمالی تخمین زده شده توسط آتاماتا نسبت به واریانس تخمین به دست می دهد. این ویژگی می تواند به انتخاب یک  $\alpha$  مناسب کمک کند.

تمام الگوریتم هایی که سعی می کنند به گونه ای پویا نرخ یادگیری را تنظیم کنند، از این نکته استفاده می کنند که در زمانی که همگرایی رخ میدهد، نرخ یادگیری را کاهش می دهند. به بیانی دیگر، چنانچه اختلاف میان احتمال تخمین جدید و میانگین تخمین های قبلی بزرگ باشد، نرخ یادگیری افزایش می یابد. این حالت زمانی رخ می دهد که یک بیشینه محلی وجود داشته باشد و یا محیط تصادفی، تغییر کند. در الگوریتم پیشنهادی جدید از ویژگی سوم مطرح در گزاره ۲ به عنوان عامل تشخیص دهنده اختلاف میان تخمین جدید و میانگین تخمین

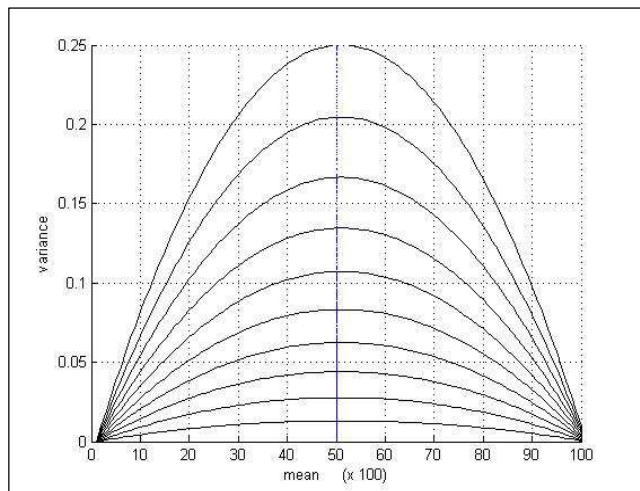
$$f(i) = \begin{cases} \left| \sin\left(\frac{4i\pi}{n}\right) \right| & i < 0.73n \\ \left| \sin\left(\frac{4t\pi}{n}\right) \right| & t = 0.73, i \geq 0.73n \end{cases} \quad (14)$$

در این رابطه،  $i$  معرف آمین نمونه و  $n$  نشاندهنده تعداد کل نمونه ها است. این محیط را محیط  $C$  نامیده ایم  
 برای بررسی عملکرد روش جدید و مقایسه آن، همزمان  
 آتاماتا را با روش  $L_{RP}$  و نرخ یادگیری کوچک  $\alpha = 0.002$  و  
 تکنیک سرد کردن تدریجی با ضریب کاهش  $0.001$  در هر  
 دور آموزش، آموزش داده ایم ( $\alpha^{t+1} = 0.999 * \alpha^t$ ). ضمناً از  
 روش شمارشی نیز برای تخمین میانگین احتمال متناسب به  
 اقدام ۱ آتاماتا استفاده کرده ایم (Maximum Likelihood :  
 ML). نتیجه مقایسه این سه در شکل ۲ آورده شده است



شکل ۲: نتایج آزمایش شماره ۱

همانگونه که مشاهده می کنید، روش جدید قادر به انطباق  
 با تغییر رفتارهای محیط است و نزدیک ترین رفتار به رفتار  
 محیط را نشان میدهد. در این آزمایش مشاهده شد که چنانچه  
 بخواهیم آتاماتا در حالت معمول خود (بدون الگوریتم پیشنهادی)  
 رفتاری منطبق تر با رفتار محیط داشته باشد، بایستی از کاهش  
 دادن نرخ یادگیری در طول زمان خودداری کرده و جهت  
 جلوگیری از نوسانی شدن مقادیر به دست آمده، مقدار پارامتر  
 یادگیری را روی یک مقدار کوچک تنظیم کرد. در این نمونه با  
 سعی و خطا، مقدار  $0.0125$  به عنوان تقریبی که میتواند رفتار  
 منطبق تری با محیط داشته باشد، استخراج شد. نتیجه را در  
 شکل ۳ ملاحظه می کنید. این مقادیر تا حدودی وابسته به حجم  
 نمونه ها و شکل تابع احتمالی رفتار محیط نیز می باشد، اما  
 تقریباً در بیشتر موارد مقادیر ثابت و کوچک نرخ یادگیری در  
 آتاماتا نیز می تواند رفتار دینامیک محیط را اقتباس کند.



شکل ۳: رابطه میان واریانس با میانگین و نرخ یادگیری

برای سهولت در محاسبه واریانس، از یک کران بالا برای آن  
 استفاده می کنیم. بدین صورت که مقدار واریانس به ازای  
 میانگین  $0.5$  را در محاسبات در نظر می گیریم. بدین ترتیب  
 محاسبه واریانس در هر دور تنها تابعی از نرخ یادگیری خواهد  
 بود. این کران بالا با مقدار  $q$  کوچکتر در الگوریتم جبران می  
 شود.

برای محاسبه میانگین نیز از یک میانگین گیری روی مقادیر  
 مربوط به احتمالات استفاده می کنیم. هر زمان که نرخ یادگیری  
 افزایش می یابد، میانگین های قبلی را در نظر نگرفته و میانگین  
 گیری را روی مقادیر جدید آغاز می کنیم. این کار باعث گریز از  
 بیشینه های محلی می شود. برای افزایش و کاهش مقدار نرخ  
 یادگیری آنها را در یک مقدار ثابت بزرگتر از ۱ ضرب (برای  
 افزایش) یا تقسیم (برای کاهش) می کنیم.

با این اصلاحات الگوریتم نهایی پیشنهادی، بدین شکل  
 خواهد بود

### ۳- بررسی عملکرد روش جدید پیشنهادی:

برای بررسی نحوه عملکرد روش پیشنهادی جدید، یک  
 آتاماتای تصادفی با ۲ اقدام در نظر گرفته ایم که توسط یک  
 محیط تصادفی مورد ارزیابی قرار می گیرد. محیط تصادفی، یک  
 محیط پویا در نظر گرفته شده است. بدین صورت که تابع ارزیابی  
 اقدام آتاماتا توسط محیط یک مقدار ثابت فرض نشده و در طول  
 زمان آموزش تغییر می کند. الگوریتم مورد استفاده توسط  
 آتاماتای یادگیر الگوریتم  $L_{RI}$  و  $L_{RI}$  پیشنهادی است.

آزمایش ۱: در اولین نمونه، از تابع رابطه ۱۴ به عنوان تابع  
 ارزیابی اولین اقدام آتاماتا استفاده کرده ایم:

#### ۴- نتیجه گیری:

در این مقاله، به کمک شاخص های آماری میانگین و واریانس حاصل از توزیع احتمال بردار انتخاب اقدام های آتاماتا، و با کمک نابرابری چبیشف روش جدیدی پیشنهاد شد که به کمک آن آتاماتا یادگیر ضمن تنظیم خودکار پارامتر یادگیری، قادر به یادگیری در محیطهای با پویایی بالا است

```

Proposed Algorithm  $L_{R-I}$ 
1: Parameters: Real  $TSH \ll 1, q > 1, \alpha$  Learning Rate Vector,
2: Initialization:  $p_j \leftarrow \frac{1}{K}, \mu_j \leftarrow \frac{1}{K}, \delta_j \leftarrow 0$ 
 $t_j \leftarrow 0$  for  $j \leftarrow 1$  to  $K$ 
3: loop
4: Draw randomly an action  $i$  according to probabilities  $p_0, \dots, p_K$ 
5: Receive feedback  $r$ 
6: for  $j \leftarrow 1$  to  $K$  do
7: if  $j \neq i$  then
8:  $p_j \leftarrow (1 - \alpha_i r) p_j$ 
9: else
10:  $p_i \leftarrow p_i + \alpha_i r (1 - p_i)$ 
11: end if
12: end for
13: for  $j \leftarrow 1$  to  $K$  do
14: update  $(\mu_j)$ 
15: update  $(\delta_j)$ 
16: if  $|p_j - \mu_j| > q \delta_j$ 
17: increment  $(\alpha_j); t_j \leftarrow 0$ ; reset  $\mu_j$ 
18: else if  $(1 - \alpha_j)^{t_j} < TSH$ 
19: decrement  $(\alpha_j); t_j \leftarrow 0$ 
20: else  $t_j \leftarrow t_j + 1$ 
21: end for
22: end loop
    
```

همچنین این الگوریتم، قادر به فرار از مینم های محلی است. بررسی ها نشان میدهد که الگوریتم جدید پیشنهادی قادر است رفتار محیطهای پویا را خیلی سریع فرا بگیرد و از این منظر، این الگوریتم برای کاربردهای واقعی مناسب تر به نظر می رسد. هم چنین به دلیل اینکه در حین آموزش، هم از افزایش و هم از کاهش نرخ یادگیری استفاده می شود، روش جدید پیشنهادی، تطبیق پذیری بیشتری نسبت به روش معمول دارد. بررسی ها نشان داد که به کمک پارامترهای مربوطه، میتوان رفتار یادگیرانه را انطباقی یا محافظه کارانه نمود. برای پیاده سازی الگوریتم نیز روش های محاسباتی ساده ای ارائه شد تا سر بار محاسباتی این روش جدید کاهش یافته و بهبودهایی نیز برای این الگوریتم پیشنهاد گردید.

برای بررسی و مقایسه الگوریتم جدید با الگوریتم معمول (و البته بهینه شده) نمونه های دیگری از محیط را در نظر گرفته و کار آموزش را برای آنها دنبال کرده ایم. همانند آزمایش شماره ۱ برای اقدام های آتاماتا دو بردار احتمال انتخاب اقدام ها در نظر گرفته ایم و این دو بردار را به دو گونه -روش معمول (Traditional) و روش جدید پیشنهادی (Proposed)- به روزرسانی کرده ایم. در روش معمول با نرخ یادگیری برابر با 0.2 آغاز کرده و در هر دور یادگیری 0.01 آن را کاهش داده ایم. (تا قبل از رسیدن به آستانه 0.001)

برای مقایسه این دو روش از مقایسه بردار احتمالی یادگرفته شده توسط هر یک از این الگوریتم ها و بردار احتمالی واقعی در هر دور استفاده کرده ایم. همچنین از نرم (norm) بردار احتمال انتخاب اقدام های یادگرفته شده و بردار احتمالی واقعی به عنوان معیاری جهت مقایسه میزان انطباق پذیری آتاماتا با محیط استفاده کرده ایم. نتایج را برای دو محیط نمونه دیگر، که محیط هایی با پویایی بالا هستند، در ادامه مشاهده می کنید. آزمایش شماره ۲:

در این آزمایش از یک محیط تصادفی استفاده کرده ایم که تابع ارزیابی آن در طول زمان به شکل پله ای تغییر می کند. این تابع  $f(i) = 0.25 * \left[ \frac{4i}{n} \right]$  است که  $i$  معرف آمین نمونه  $n$  و نشاندهنده تعداد کل نمونه ها می باشد. این محیط را محیط  $S$  نامیده ایم.

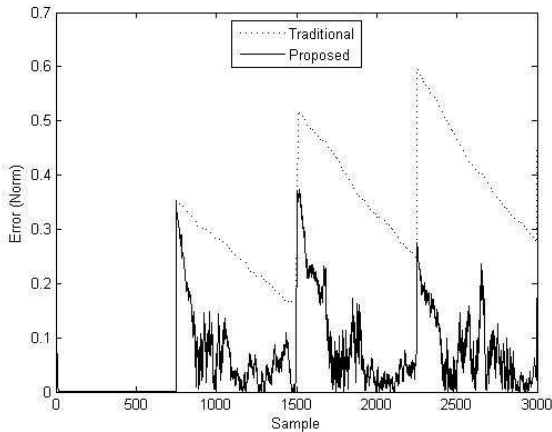
نتایج بررسی عملکرد مقایسه ای الگوریتم پیشنهادی و الگوریتم معمول بهینه شده، در تنظیم نرخ یادگیری را در شکل ۴ و شکل ۵ مشاهده می کنید. شکل ۴ نشان میدهد که الگوریتم پیشنهادی انطباق بیشتری با رفتار پویای محیط داشته و شکل ۵ نیز موید میزان خطای کمتر روش جدید در مقایسه با روش معمول است.

آزمایش ۳:

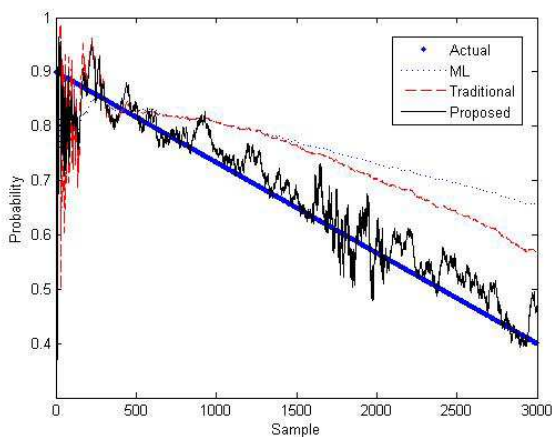
برای سومین آزمایش، محیطی را در نظر گرفته ایم که تابع ارزیابی آن به شکل پویا و بر اساس یک تابع خطی  $f(i) = 0.9 - c * i$  تغییر می کند. در این مثال ثابت کوچک  $c=1/6000$  در نظر گرفته شده است. این محیط را محیط  $L$  نامیده ایم نتایج بررسی را در شکل ۵ و شکل ۶ مشاهده می کنید. محیط  $L$  نسبت به محیط های قبلی از تغییرات ملایم تری برخوردار است. همچنانکه مشاهده می شود، روش جدید پیشنهادی، در مجموع، تقریبی نزدیک تر به مقدار واقعی ارائه می دهد.

## ۵- منابع و مراجع:

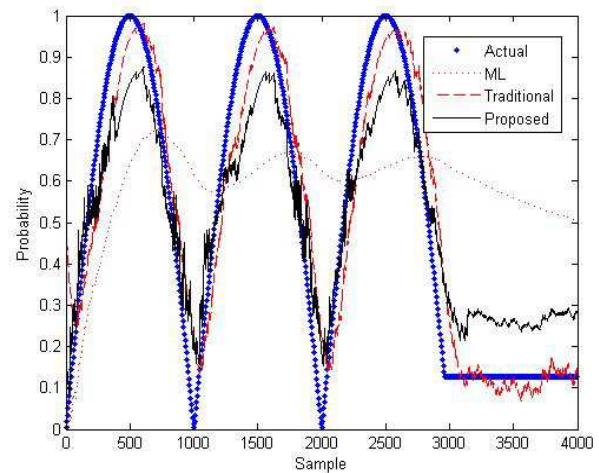
1. Automatic adjustment of learning rates of the self-organizing feature map. **Shah-Hosseini, Safabakhsh**. 4, s.l.: Scientia Iranica, 2001, Vol. 8.
2. Adaptation of Learning Rate in Back Propagation Algorithm using Fixed Structure Learning Automata. **Beigy, H., Meybodi, M. R. and Menhaj, M. B.** s.l.: 6th Iranian Conference on Electrical Engineering, 1998.
3. Optimal Adaptive K-Means Algorithm with Dynamic Adjustment of Learning Rate. **chedsada chinrungrueng, carlo H.sequin**. 1991.
4. Markov's inequality. *Wikipedia*. [Online] [Cited: 4 25, 2012.] [http://en.wikipedia.org/wiki/Markov%27s\\_inequality](http://en.wikipedia.org/wiki/Markov%27s_inequality).
5. **Tsetlin, Mikhail L'vovich**. *Automaton Theory and the Modelling of Biological Systems*. New York and London: Academic Press, 1973.
6. learning automata – a survey. **Narendra K., Thathachar M.A.L.** s.l.: IEEE Transactions on Systems, Man, and Cybernetics, July 1974, Vols. SMC-4.



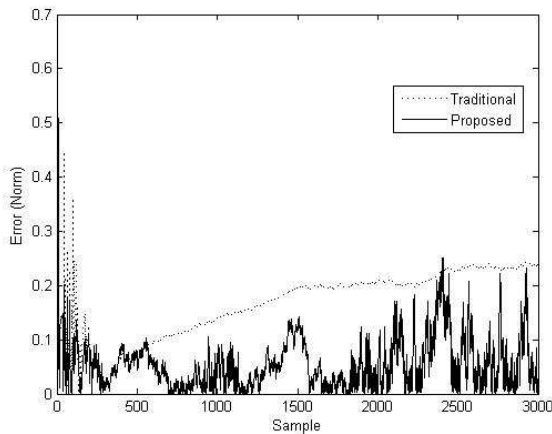
شکل ۵: مقایسه دقت روش پیشنهادی با روش معمول (آزمایش ۲)



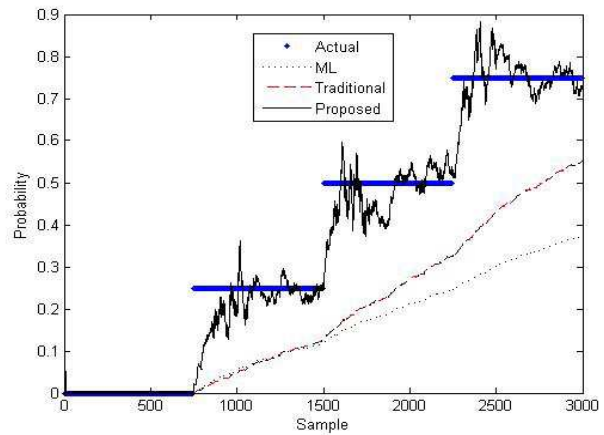
شکل ۶: مقایسه تغییرات احتمال (آزمایش ۳)



شکل ۳: آزمایش شماره ۱ و بررسی رفتار مقایسه ای اتاماتا



شکل ۷: مقایسه دقت روش پیشنهادی با روش معمول (آزمایش ۳)



شکل ۴: مقایسه تغییرات احتمال (آزمایش ۲)

Table 1: Thick line is the true probability (feedback distribution

from environment to action 1). Solid line is the proposed algorithm (adaptive learning rate adjustment) and dashed line is the traditional algorithm

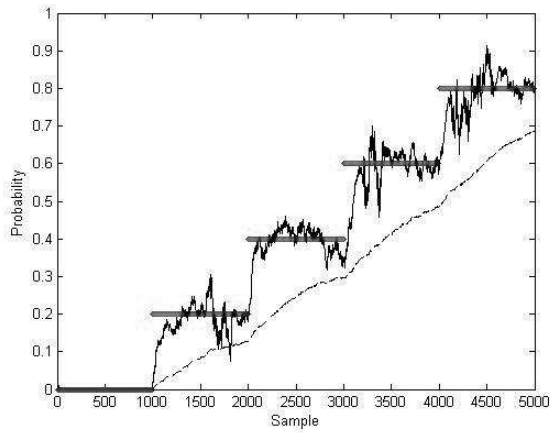


Figure 1

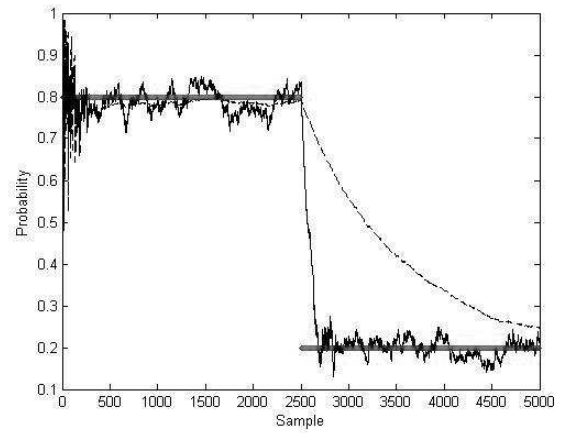


Figure 2

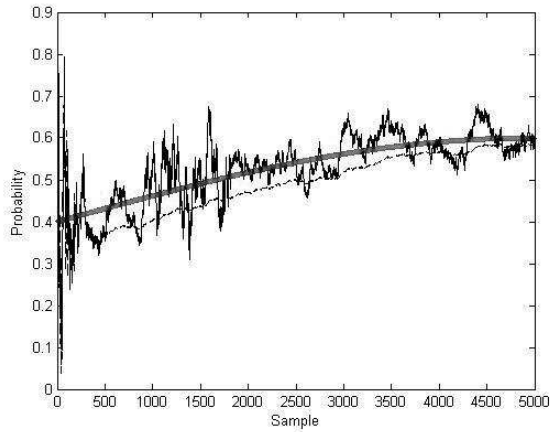


Figure 3

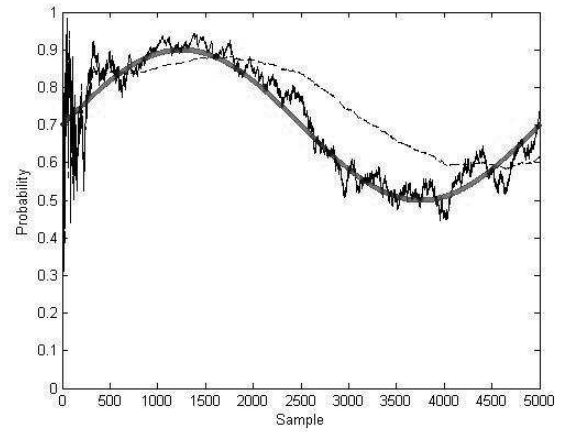


Figure 4

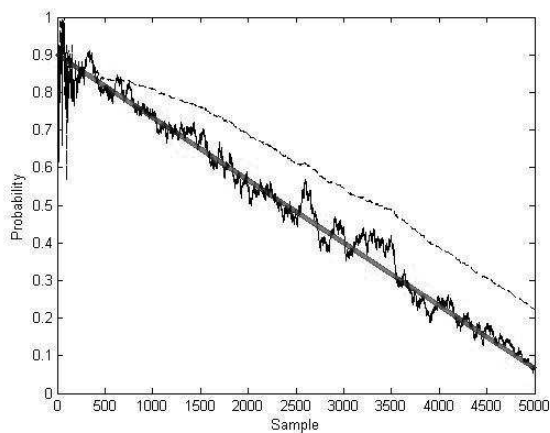


Figure 5

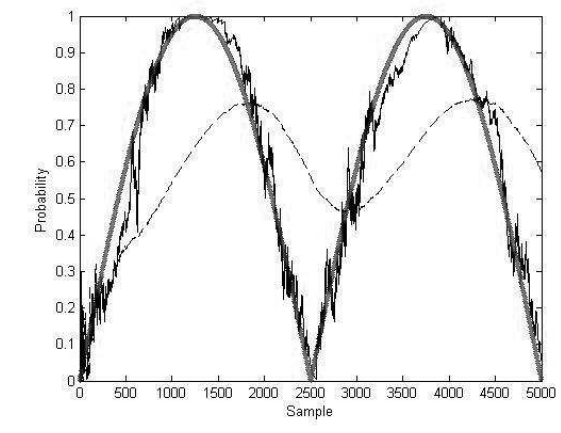


Figure 6