

# A New Automatic Learning Rate Adjustment Method based on Distribution Metrics for Using in non-Stationary Environments by Learning Automata

## یک روش جدید مبتنی بر معیارهای آماری توزیع برای تنظیم خودکار نرخ یادگیری آتاماتای یادگیر در محیط های پویا

Mohammad Reza Mollakhalili Meybodi, Mohammad Reza Meybodi

**Abstract** — *The value of learning rate and its change mechanisms is one of the issues in designing learning systems such as learning automata. In most cases a time-based reduction function is used to adjust the learning rate aim at reaching stability in training system. So the learning rate is a parameter that determines to what extent a learning system is based on past experiences, and the impact of current events on it. This method is efficient but does not properly function in dynamic and non-stationary environments.*

*In this paper, a new method for adaptive learning rate adjustment in learning automata is proposed. In this method, in addition to the length of time to learn, some statistical characteristics of actions probability vector of Learning Automata are used to determine the increase or decrease of learning rate. Furthermore, unlike existing methods, during the process of learning, both increase and decrease of the learning rate is done and Learning Automata responds effectively to changes in the dynamic random environment.*

*Empirical studies show that the proposed method has more flexibility in compatibility to the non-stationary dynamic environments and get out of local maximum points and the learned values are closer to the true values<sup>1</sup>.*

**Keywords** — learning automata, dynamic learning rate, learning rate adjustment, mean statistical characteristics, variance, dynamic random environment.

تنظیم می‌کند. ضمن تشریح مبانی ریاضی این الگوریتم جدید، عملکرد این الگوریتم را در محیط های تصادفی نمونه بررسی کرده و با مقایسه نتایج به دست آمده نشان داده‌ایم، الگوریتم جدید به دلیل اینکه در طول زمان یادگیری، هم زمان و بر اساس معیارهای تعیین شده افزایش و کاهش نرخ یادگیری را انجام می‌دهد، از انعطاف پذیری بیشتری نسبت به روشهای قبلی برای انطباق با محیطهای تصادفی پویا برخوردار است و مقادیر یادگرفته شده به مقادیر حقیقی نزدیکتر هستند.

کلمات کلیدی

آتاماتای یادگیر، نرخ یادگیری پویا، تنظیم نرخ یادگیری، نابرابری چبیشف.

### ۲. مقدمه

بحث تنظیم خودکار نرخ یادگیری یکی از مسائلی است که در مباحث مرتبط

### ۱. چکیده

یکی از مسائل مطرح در ساخت سیستم‌های یادگیر نظیر آتاماتای یادگیر، تعیین نرخ یادگیری است. در اکثر موارد از یک الگوریتم کاهش یابنده در طول زمان برای تنظیم نرخ یادگیری استفاده می‌شود. در این مقاله یک روش جدید برای تغییر نرخ یادگیری و انطباق سیستم یادگیرنده با وضعیت محیط، برای استفاده در آتاماتای یادگیر پیشنهاد شده است. این روش جدید از برخی معیارهای آماری مربوط به توزیع فعلی به دست آمده برای بردار احتمالات متناظر با اقدام های آتاماتا به منظور تعیین افزایش یا کاهش نرخ یادگیری استفاده می‌کند. مزیت این روش در آن است که برخلاف روش های پیشنهادی قبلی، در طول فرآیند یادگیری هم افزایش و هم کاهش مقدار نرخ یادگیری را، بسته به نتایج مقایسه معیارهای آماری، انجام می‌دهد و به صورت خودکار نرخ یادگیری را

<sup>1</sup> M.R. Mollakhalili Meybodi is with Department of Computer Engineering, Islamic Azad University, Science and Research branch, Tehran, Iran (email: m.meybodi@srbiau.ac.ir).

M.R. Meybodi is with Department of Computer Engineering and IT, Amirkabir University of Technology, Tehran, Iran (email: mmeybodi@aut.ac.ir).

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. (6) (5) آتاماتاهاى یادگیر، دو دسته هستند آتاماتاهاى ergodic که با توزیعی مستقل از حالت اولیه اشان همگرا می شوند و absorbing که بعد از یک تعداد گام اولیه روی یک حالت خاص قفل می شوند. در محیط هایی که احتمال پاداش گرفتن با زمان تغییر می کند، آتاماتاهاى گروه اول ارجحیت دارند؛ چرا که قادرند با تغییرات محیط هماهنگ شوند. آتاماتاى یادگیر با ساختار متغیر، به کمک ۶ تایی  $\langle \beta, \theta, a, P, G, T \rangle$  نمایش داده می شود.  $\beta$  مجموعه ورودی ها،  $\theta$  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.

عامل موثر در کارآیی آتاماتاى ساختار متغیر، الگوریتم های یادگیری هستند که برای به روزرسانی احتمال اقدام ها استفاده می شود. الگوریتم های مختلفی تاکنون در این مورد گزارش شده است. یکی از این الگوریتم ها، الگوریتم LR-I است. فرض کنید بر اساس توزیع  $P(t)$  در زمان  $t+1$ ، اقدام  $\alpha_k$  انتخاب شده باشد. در این صورت الگوریتم یادگیری LR-I مطابق رابطه ۲ اقدام به روزرسانی بردار احتمال انتخاب اقدام های خود می کند

$$P_j^{t+1} = \begin{cases} P_j^{t+1} + \alpha(1 - P_j^{t+1}) & \text{if } k = j \\ P_j^{t+1}(1 - \alpha) & \text{if } k \neq j \end{cases} \quad (2)$$

در این رابطه  $\alpha$  را نرخ یادگیری یا گام نیز می گویند. رابطه به روزرسانی بردار احتمال انتخاب اقدام آتاماتاى یادگیر برای زمانی که اقدام  $k$  ام توسط آتاماتا صورت گرفته است را می توان به صورت رابطه ۳ نیز نوشت:

$$P_k^{t+1} = P_k^t * (1 - \alpha) + \alpha R^{t+1} \quad (3)$$

که در آن  $R$  پاسخ محیط به اقدام انجام شده توسط آتاماتا است و داریم

$$R^{t+1} = \begin{cases} 1 & \text{with probability } C_k \\ 0 & \text{with probability } 1 - C_k \end{cases} \quad (4)$$

و به سادگی قابل اثبات است که

$$\lim_{t \rightarrow \infty} E[P_k^{t+1}] = \mu = C_k \quad (5)$$

رابطه یادگیری آتاماتا را به شکل برداری (رابطه ۶) میتوان نشان داد

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

رابطه ۶ در حقیقت نشان میدهد که بردار احتمال انتخاب اقدام های آتاماتاى یادگیر در طول فرآیند یادگیری، به بردار میانگین پاسخ های محیط همگرا می شود.

در اکثر سیستم های یادگیر و از جمله در آتاماتاى یادگیر، نرخ یادگیری در مدت آموزش، به شکل پویا -کاهش یابنده با زمان- تغییر می کند. دلیل این امر هم واضح است. در ابتدای یادگیری، مقادیر بزرگتر نرخ یادگیری، باعث آموزش سریع تر می گردند. به تدریج و با افزایش تعداد نمونه ها، سیستم یادگیرنده، سعی می کند بیشتر متکی بر تجربیات آموزشی گذشته باقی بماند تا اینکه سعی کند از نمونه های جدید برای یادگیری استفاده کند. این منطبق با عمل می شود، در ابتدای فرآیند یادگیری، سیستم یادگیرنده جسورانه تر و با گذر زمان محافظه کارانه تر عمل کند.

با الگوریتم ها و سیستم های یادگیری و خصوصا در مباحث مربوط به شبکه های عصبی به کرات مورد بررسی قرار گرفته است (1) (3) (2). 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 معمولاً کوچکی دارد. مشکل از اینجا است 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 از قدرت تطابق پذیری و یادگیری خوبی در برخورد با محیط های تصادفی گوناگون برخوردار نباشد. (3) پژوهش مستقلی که به بررسی تنظیم خودکار نرخ یادگیری در آتاماتاى 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 آتاماتاى یادگیر نیز قابل استفاده هستند.

### ۳. روش جدید پیشنهادی

#### ۱. مبانی ریاضی:

برای تنظیم پویای نرخ یادگیری، به منظور انطباق در محیطهای پویا، نیاز به معیاری داریم تا بر اساس آن معیار میزان توانمندی نرخ یادگیری فعلی سیستم یادگیر را در رصد کردن تغییرات محیط بسنجیم. یکی از ابزارهای آماری مناسب در این مورد نابرابری موسوم به نابرابری چیشف است. در نظریه احتمالات نابرابری چیشف تضمین می کند که در هر نمونه تصادفی یا در هر توزیع احتمال، "تقریباً تمامی" مقادیر در نزدیکی میانگین خواهند بود. به طور دقیق تر این قضیه بیان می کند که حداکثر مقادیری که در هر توزیع می توانند بیش از  $k$  برابر انحراف معیار با میانگین فاصله داشته باشند  $\frac{1}{k^2}$  است. (4) نابرابری چندبعدی چیشف تعمیمی از نابرابری چیشف است که به کمک آن میتوان مرزی را برای اینکه یک بردار تصادفی از بردار مقادیر میانگین اش بیش از یک مقدار معین فاصله داشته باشد را تعیین کرد. این نابرابری بدین شکل فرمول بندی می شود

نابرابری چیشف (تعمیم یافته):

فرض کنید  $X$  یک بردار تصادفی با میانگین  $\mu = E[X]$  و ماتریس کوواریانس  $V = E[(X - \mu)(X - \mu)^T]$  باشد. اگر  $V$  یک ماتریس تعریف شده مثبت باشد در این صورت برای هر عدد حقیقی  $t > 0$  داریم

$$\Pr \left( (X - \mu)V^{-1}(X - \mu)^T > qN \right) \leq \frac{1}{q^2} \quad (1)$$

که  $N = \text{trace}(V^{-1}V)$  است.

#### ب. بررسی رفتار یادگیری در آتاماتاى یادگیر و تاثیر نرخ یادگیری:

یک آتاماتاى یادگیر به عنوان مدلی از یک سیستم یادگیر است که در محیط های تصادفی ناشناخته عمل می کند. آتاماتا در هر دور یک اقدام از میان

آزمایش ۱: محیط پویایی که برای این بررسی انتخاب شده است. تا حدود ۳۰٪ زمان آموزش از تابع  $|\sin(\frac{9i\pi}{n})|$  برای ارزیابی اقدام آتاماتا استفاده می کند (در  $i$  امین نمونه که  $n$  تعداد کل نمونه ها است). اما از این مرحله به بعد، تابع به شکل خطی و بدون تغییر، اقدام آتاماتا را مورد ارزیابی قرار میدهد. نتایج مقایسه ای را در شکل ۳ مشاهده می کنید. در **Error! Reference source not found.** نیز نحوه تغییرات نرخ یادگیری نشان داده شده است. تلاش آتاماتا برای انطباق با محیط در بخش اول نمونه ها که محیط از پویایی بالایی برخوردار است، از طریق افزایش نرخ یادگیری قابل مشاهده است. آزمایش شماره ۲: برای این نمونه، محیط دیگری در نظر گرفته ایم که تابع ارزیابی محیط (احتمال پاداش به اقدام) از یک رفتار سینوسی برخوردار است. نتیجه مقایسه ای عملکرد الگوریتم یادگیری LR-I و الگوریتم جدید را در **Error! Reference source not found.** مشاهده می کنید.

**Proposed Algorithm LR-I**  
 1: Parameters: Real  $q > 1$ ,  $\alpha < 1$  Learning Rate  
 2: Initial:  $p_j \leftarrow \frac{1}{K}$ ,  $\mu_j \leftarrow p_j$ ,  $dt \leftarrow 0$ ,  $t \leftarrow 0$  for  $j \leftarrow 1$  to  $K$   
 3: loop  
 4: Draw randomly an action  $i$  according to probabilities  $P_0, \dots, P_K$   
 5: Receive either reward or penalty  
 6: if reward then  
 7: for  $j \leftarrow 1$  to  $K$  do  
 8: if  $j \neq i$  then  
 9:  $p_j \leftarrow (1 - \alpha)p_j$  //penalt  
 10: else  
 11:  $p_i \leftarrow p_i + \alpha(1 - p_i)$  //reward 12: end if  
 13: end for  
 14:  $t \leftarrow t + 1$ ;  
 15:  $dt \leftarrow dt + 1$ ;  
 16:  $\bar{\mu} = \text{mean}(\bar{P}(i - dt - 1), \dots, \bar{P}(i))$   
 17:  $\bar{V} = \text{Cov}(\bar{\mu}, \bar{P})$   
 18:  $N = \text{trace}(\bar{V}^{-1})$   
 19: if  $(\bar{P} - \bar{V}) \bar{V}^{-1} (\bar{P} - \bar{V}) > qN$   
 20: increment  $(\alpha)$ ;  $t \leftarrow 0$ ;  $dt \leftarrow 0$ ;  
 21: else if  $(1 - \alpha)^t < \text{TSH}$   
 22: decrement  $(\alpha)$ ;  $t \leftarrow 0$ ;  
 23: end if  
 24: end if

شکل ۱. شبه کد الگوریتم پیشنهادی

## ۵. نتیجه گیری

در این مقاله یک الگوریتم به کمک یک تابع مبتنی بر فاصله میان میانگین توزیع بدست آمده برای احتمال انتخاب اقدام های یک آتاماتا و بردار جدید به روز شده از اقدام های آتاماتا در یک دور آموزش و با استفاده از نابرابری چبیشف تعمیم یافته به شکل برداری، معیاری برای تعیین میزان تغییرات در محیط تصادفی استخراج شد. این الگوریتم جدید قادر به تنظیم اتوماتیک نرخ یادگیری است و می تواند بر حسب میزان تغییرات در محیط و به منظور انطباق با آن، اقدام به کاهش یا افزایش نرخ یادگیری نماید. استفاده از این الگوریتم جدید در محیط های تصادفی پویا (non-stationary) و بررسی های نمونه در چندین محیط تصادفی پویا، نشان داد که الگوریتم جدید از عملکرد بهتری در این گونه محیط ها نسبت به روش های یادگیری معمول برخوردار است.

اکثر سیستم ها یادگیری، از یک نرخ یادگیر پویای کاهش یابنده در حین فرآیند یادگیری استفاده می کنند. نرخ یادگیری با این مفهوم پارامتری است که میزان فراموشکاری سیستم را تعریف می کند. مقادیر کوچکتر این پارامتر، یعنی اتکای بیشتر سیستم به تجربیات گذشته و مقادیر بزرگتر به معنای نادیده گرفتن تجربیات گذشته است.

در بیشتر سیستم ها، این منطق پاسخگوی نیازها می باشد. مساله اینجاست که کاهش نرخ یادگیری باعث می شود به مرور زمان انطباق پذیری سیستم یادگیر کاهش یابد و نسبت به تغییرات در محیط عکس العمل مناسب نشان ندهد. بنابراین برای تنظیم اتوماتیک نرخ یادگیری در طول مدت آموزش بایستی معیاری داشته باشیم تا بر اساس آن نسبت به افزایش (در صورت بروز تغییرات جدی در محیط) یا کاهش (در صورت یکنواخت بودن پاسخ محیط و نزدیک شدن به همگرایی) نرخ یادگیری اقدام کنیم.

## ج. الگوریتم پیشنهادی:

با توجه به رابطه یادگیری مورد استفاده توسط آتاماتای یادگیر و تفسیر برداری آن، میتوان از نابرابری چبیشف برای تنظیم اتوماتیک نرخ یادگیری استفاده کرد. منطق کار بدین صورت است که:

هر زمان که بردار جدید یادگرفته شده توسط آتاماتای یادگیر، بیش از یک میزان مشخص از بردار میانگین مقادیر قبلی فراگرفته شده توسط آتاماتا، فاصله داشته باشد، به معنای وجود تغییرات گسترده در محیط است. برای اینکه بتوان اثر این تغییرات را منعکس کرد آتاماتا بایستی نرخ یادگیری را افزایش دهد. ملاک مورد استفاده در این قسمت نابرابری چبیشف است.

اگر تعداد دفعاتی که آتاماتا توسط یک نرخ یادگیری ثابت، عملیات یادگیری را انجام داده است، از مقدار معینی بیشتر باشد، نشان میدهد که تقریباً همگرایی صورت گرفته است و بردار احتمال انتخاب اقدام آتاماتا به بردار واقعی نزدیک شده است. برای تسریع در یادگیری، میتوان نرخ یادگیری را کاهش داد. ملاک برای این کاهش (با توجه به رابطه میان بردار میانگین اصلی و میانگین بردارهای یادگرفته شده توسط آتاماتا) تابعی از  $(1 - \alpha)^t$  خواهد بود که در آن  $\alpha$  نرخ یادگیری و  $t$  تعداد دفعاتی است که نرخ یادگیری بدون تغییر برای آموزش مورد استفاده قرار گرفته است. به این ترتیب میتوان الگوریتم پیشنهادی جدید را به شکل زیر فرموله کرد.

## ۴. بررسی نتایج شبیه سازی

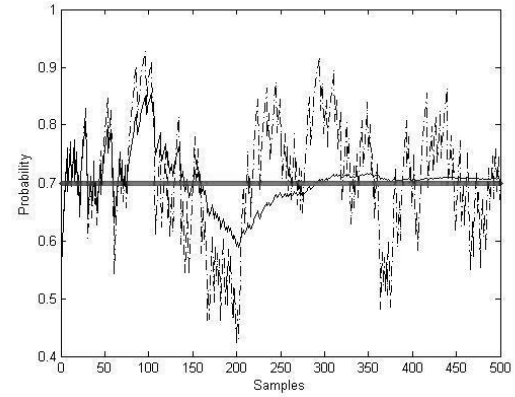
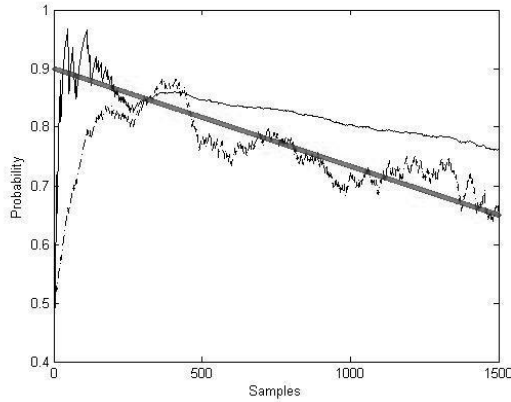
به منظور بررسی عملکرد الگوریتم پیشنهادی، تعدادی محیط پویا در نظر گرفته و آتاماتای یادگیری با دو اقدام را همزمان به دو شیوه آموزش داده ایم. روش اول، همان روش معمول است به این صورت که با توجه به پویایی محیط، آتاماتای یادگیر از یک نرخ یادگیری کوچک بین 0.1 و 0.2 - آغاز کرده و با کاهش آن به اندازه 0.01 در هر دور اجرای الگوریتم (تا رسیدن به آستانه 0.001) فرآیند یادگیری را انجام داده است.

الگوریتم جدید نیز از یک نرخ یادگیری دلخواه (و غالباً بزرگ نزدیک به 1) آغاز می کند. در هر دور اجرا بر اساس میزان فاصله بردار جدید احتمال انتخاب بدست آمده برای اقدام های آتاماتا از بردار میانگین، اقدام به کاهش یا افزایش نرخ یادگیری نموده است. نتیجه را در مورد نمونه هایی از محیط ها در ادامه مشاهده می کنید.

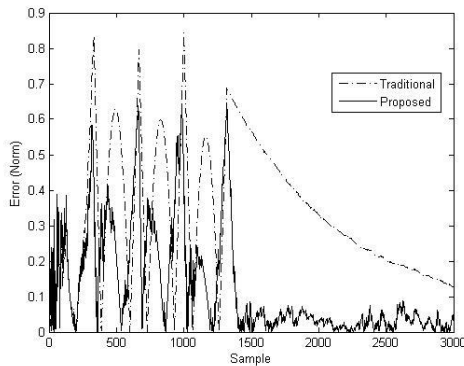
- 4 Markov's Inequality. Wikipedia. [Online] [Cited: 4 25, 2012.] Available from: [http://en.wikipedia.org/wiki/Markov%27s\\_inequality](http://en.wikipedia.org/wiki/Markov%27s_inequality).
- 5 Narendra K, Thathachar MAL. Learning Automata: A survey. IEEE Transactions on Systems. 1947; 5.
- 6 Setlin T, L'Vovich M. Automaton Theory and the Modelling of Biological Systems. New York and London : Academic Press, 1973.

## REFERENCES

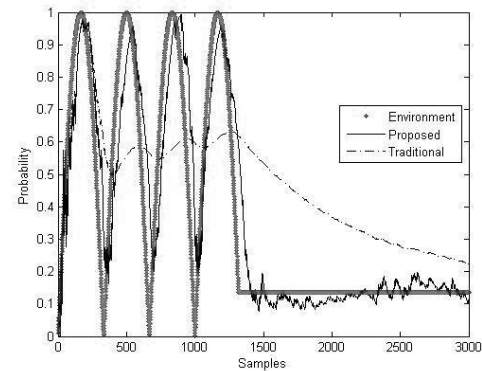
- 1 Shah-Hosseini, Safabakhsh. Automatic adjustment of learning rates of the self-organizing feature map. Scientia Iranica. 2001; 8(1).
- 2 Beigy H, Meybodi MR, Menhaj MB. Adaptation of Learning Rate in Back Propagation Algorithm using Fixed Structure Learning Automata. 6th Iranian Conference on Electrical Engineering, 1998.
- 3 Chinrungrueng C, Sequin CH. Optimal Adaptive K-Means Algorithm with Dynamic Adjustment of Learning Rate. 1991.



شکل ۲. (سمت راست) اثر کاهش نرخ یادگیری در حین آموزش (نمودار ممتد) در مقایسه با نرخ یادگیری ثابت (نمودار نقطه چین). خط ممتد درشت مقداری است که در نهایت می بایستی فراگرفته شود. در سمت چپ، نمونه ای از یک محیط پویا را مشاهده می کنید. بر خلاف نمودار سمت راست، مشاهده می شود که نرخ یادگیری با مقدار ثابت و کوچک (نمودار خط چین) نسبت به نرخ کاهش یابنده (نمودار ممتد) عملکرد بهتری دارد

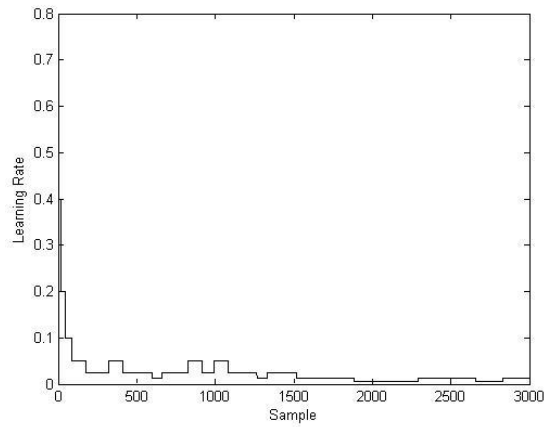


b

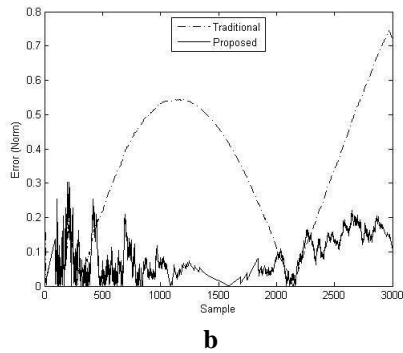


a

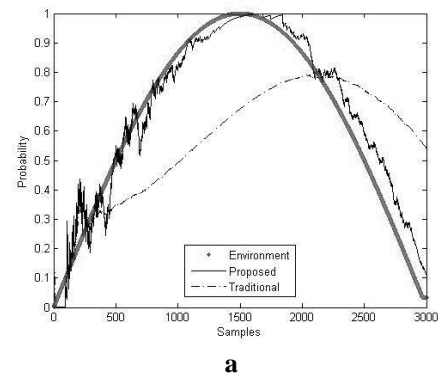
شکل ۳. در شکل **a** خط تیره رنگ درشت نحوه محیط را نشان میدهد. خط تیره رنگ نازک نحوه تغییرات بردار احتمال مربوط به اقدام آتاماتای آموزش داده شده به شیوه جدید و خطوط نقطه چین، همان بردار، آموزش داده شده به شیوه معمول را نشان میدهد. شکل **b** به بررسی فاصله بردارهای آموزش داده شده به هریک از دو شیوه و بردار واقعی پرداخته است.



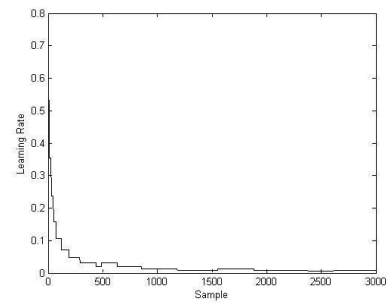
شکل ۴. تغییرات افزایشی و کاهش‌ی نرخ یادگیری در روش پیشنهادی (آزمایش شماره ۱)



**b**



**a**



**c**

شکل ۵. در نمودار **a** نحوه انطباق روش جدید با بردار احتمالی محیط را در مقایسه با روش معمول مشاهده می کنید. شکل **b** میزان دقت روش ها را به شکل مقایسه ای مشاهده می کنید. شکل **c** نحوه تغییر نرخ یادگیری (افزایش-کاهش) را در طول یادگیری نشان داده است.