

بهبود چارچوب تعیین مدل دانشجو در سیستم شبه آموزشی با استفاده از

N-Armed bandit و اتوماتای یادگیر

سیده لیلا جواد^۱، بهروز معصومی^۲، محمدرضا میبیدی^۳

^۱ کارشناسی ارشد، گرایش نرم افزار، دانشگاه آزاد اسلامی قزوین.

Leila_javadi@yahoo.com

^۲ استادیار دانشکده مهندسی کامپیوتر و فناوری اطلاعات دانشگاه آزاد اسلامی قزوین.

Masoumi@Qiau.ac.ir

^۳ استاد دانشکده مهندسی کامپیوتر و فناوری اطلاعات دانشگاه صنعتی امیرکبیر.

mmeybodi@aut.ac.ir

چکیده

در جامعه کنونی به کارگیری کامپیوتر جهت آموزش امری ضروری است و در این میان سیستم های آموزشی هوشمند توجه محققان را به خود جلب کرده است. مدلسازی دانشجو نقطه تمرکز اصلی و یکی از اجزاء سیستم آموزشی هوشمند و نمایشی از رفتار و وضعیت دانشجو است. در هنگام استفاده دانشجو از سیستم، دنباله ای از اعمال انجام شده او ثبت می شود و سیستم تلاش می کند تا بر اساس آن، مدل یادگیری دانشجو را تعیین کند. این مقاله یک دیدگاه جدید برای بهبود چارچوب تعیین مدل دانشجو در سیستم شبه آموزشی با استفاده از ترکیب اتوماتای یادگیر و N-Armed Bandit بیان می کند. در مدل پیشنهادی، مدل دانشجو توسط عامل تعیین کننده سطح به نام $LDA-Bandit$ یا LA که مدل یادگیری دانشجو را مشخص می کند، تعیین می شود. $LDA-Bandit$ در حقیقت الگوریتم یادگیری N-Armed Bandit را به عنوان مکانیزم یادگیری برای تعیین اینکه دانشجو یادگیرنده کند، نرمال یا سریع می باشد، بکار می گیرد و نشان داده شد با بکارگیری این الگوریتم سرعت و دقت تعیین مدل یادگیری دانشجو افزایش یافته است.

کلمات کلیدی اتوماتای یادگیر، مدلسازی دانشجو، یادگیری N-Armed Bandit، سیستم های شبه آموزشی

۱- مقدمه

هوشمندسازی و نیز حوزه های کاربردی مختلف، معماری ها و اجزاء گوناگونی پیشنهاد شد. در حالت کلی این سیستم ها شامل سه مولفه اصلی هستند و گاهی مولفه واسط کاربری نیز به آن اضافه می شود [۴،۳]. سه مولفه تشکیل دهنده سیستم آموزشی هوشمند شامل مدل دامنه، مدل دانشجو و مدل آموزشی^۲ می باشد [۵].

به کارگیری یادگیری ماشین در بهبود سیستم های آموزشی در تعدادی از تحقیقات گذشته مورد مطالعه قرار گرفت. Lelouche در سال ۲۰۰۰ مجموعه ای از عامل های تعاملی را برای مدلسازی اصلی دانش، آموزشی در سیستم آموزشی هوشمند بکار گرفت [۶]. Legaspi و Sison در سال ۲۰۰۰ آموزش را در سیستم آموزشی هوشمند با استفاده از یادگیری تقویتی به عنوان رویه یادگیری مرکزی مدل کردند [۷]. Baffes و Mooney در سال ۱۹۹۶، ASSERT را پیاده

در طول بیست سال اخیر حضور گسترده سیستم های کامپیوتری در امور آموزشی هرروز بیشتر شده است. سازگاری و تطابق در آموزش، به منظور افزایش کارایی و تاثیر بر فرایند یادگیری ضروری است. اما تعداد کمی از سیستم های آموزشی وجود دارند که پویا بوده و قادرند مطالب آموزشی را به صورت شخصی متناسب با نیازها و سبک یادگیری هر دانش آموز فراهم نمایند [۱]. برنامه هایی را که با استفاده از تکنیک های هوش مصنوعی به افراد در یادگیری کمک می کنند، "سیستم های آموزشی هوشمند" می نامند.

سامانه های آموزشی هوشمند، نسل جدیدی از سامانه های آموزشی هستند که شامل تکنیک های هوش مصنوعی و متدهای ارائه دانش بوده که با هدف بهبود آموزش و یادگیری انسان ها بوجود آمدند [۲،۱]. برای سیستم های آموزشی هوشمند با توجه به تعابیر مختلف از مفهوم

در مساله bandit، هر عمل دارای یک پاداش موردانتظار و یا میانگینی از آن می‌باشد که به هنگام انتخاب عمل موردنظر، برگردانده می‌شود که به این پاداش، ارزش عمل گفته می‌شود [۱۷].

یکی از روش‌های یادگیری کارآمد برای مساله n-armed bandit روش‌های پیگیری می‌باشد. این روش هر دو مقدار تخمین ارزش-عمل و اولویت‌های عمل را به صورت دنبال کردن پیوسته اولویت‌های عمل حریصانه نگهداری می‌کند. در ساده‌ترین روش پیگیری، اولویت هر عمل $P_t(a)$ احتمالاتی است که بر اساس آن عمل a در بازی t ام انتخاب می‌شود.

در مرحله قبل از انتخاب عمل، این احتمالات به منظور ایجاد عمل حریصانه و افزایش احتمال انتخاب آن، بروزرسانی می‌شوند. فرض کنید $a_t^* = \operatorname{argmax}_a Q_{t-1}(a)$ عمل حریصانه‌ی نمونه تصادفی از اعمال حریصانه، پیش از انتخاب عمل a_t باشد، در اینصورت احتمال انتخاب $a_t = a_t^*$ به صورت کسری از β ، افزایش می‌یابد.

$$\pi_t(a_t^*) = \pi_{t-1}(a_t^*) + \beta(1 - \pi_{t-1}(a_t^*)) \quad (1)$$

در حالیکه احتمالات انتخاب اعمال دیگر به سمت صفر، کاهش می‌یابند.

$$\pi_t(a) = \pi_{t-1}(a) + \beta(0 - \pi_{t-1}(a)) \text{ for all } a \neq a_t^* \quad (2)$$

پس از انجام عمل a_t و مشاهده پاداش جدید، ارزش عمل $Q_t(a_t)$ به میانگین نمونه پاداش‌های مشاهده شده بروزرسانی می‌شود [۱۷].

۲-۲ آتوماتای یادگیر

آتوماتای یادگیر یکی از مدل‌های یادگیری تقویتی است که در آن آتوماتا یک عمل^۴ بهینه را از میان یک مجموعه از عمل‌های مجاز آتوماتا انتخاب می‌کند. این عمل روی یک محیط تصادفی اعمال می‌شود و محیط به این عمل آتوماتا بوسیله یک پاسخ تصادفی از مجموعه پاسخ‌های مجاز جواب می‌دهد. پاسخ محیط بصورت آماری به عمل آتوماتا وابسته است. اصطلاح محیط شامل تمام شرایط خارجی و تاثیرات آنها روی عملکرد آتوماتا است. ساختار یک آتوماتای یادگیر با محیط در شکل (۱) نشان داده شد.

سازی کردند که یادگیری تقویتی را در مدلسازی دانشجو برای بدست آوردن خطاهای دانشجو جدید فقط با استفاده از دانش دامنه صحیح استفاده کردند [۸]. hashem در سال ۲۰۰۷ استفاده از آتوماتای یادگیر را در سیستم آموزشی هوشمند بکار گرفت [۹].

مدلسازی دانشجو نقطه تمرکز اصلی و یکی از اجزاء سیستم آموزشی است. این مدل، نمایشی از رفتار و وضعیت دانشجو بوده و مبنایی برای نمایش وضعیت دانشجو است. هدف این مقاله بهبود چارچوب تعیین مدل دانشجو در سیستم یادگیری است. و نشان داده شد که چگونه با استفاده از الگوریتم یادگیری N-Armed Bandit در چارچوب سیستم شبه آموزشی، این مدلسازی می‌تواند موفقیت آمیز باشد. برای بهبود تعیین مدل دانشجو از الگوریتم یادگیری N-Armed Bandit به جای الگوریتم $DLRI^3$ در مدل مرجع [۹] استفاده شد که با توجه به آزمایش‌های مختلف نشان داده شد که مدل بهبودیافته با سرعت بیشتری مدل دانشجو را تشخیص می‌دهد [۱۰، ۹]. برای ارزیابی مدل مطرح شده از محیط‌های مختلفی شامل محک‌های ثابت تست استفاده شد و نتایج بدست آمده از شبیه‌سازی نشان داد که مدل پیشنهادی در تعیین مدل یادگیری بر اساس مشاهده اعمال دانشجو و محیط تدریس موفق است. در ادامه مقاله در بخش ۲ به مفهوم یادگیری N-Armed Bandit و آتوماتای یادگیری و انواع آن پرداخته شد. در بخش ۳ سیستم شبه آموزشی توضیح داده شد و در بخش ۴ مدل پیشنهادی بهبود یافته ارائه شد. در بخش ۵ به ارزیابی مدل پیشنهادی پرداخته و در بخش ۶ نتیجه گیری بیان شد.

۲- یادگیری N-armed Bandit و

آتوماتای یادگیر

در این قسمت یادگیری Q و آتوماتای یادگیر به عنوان دو مدل یادگیری توصیف شدند.

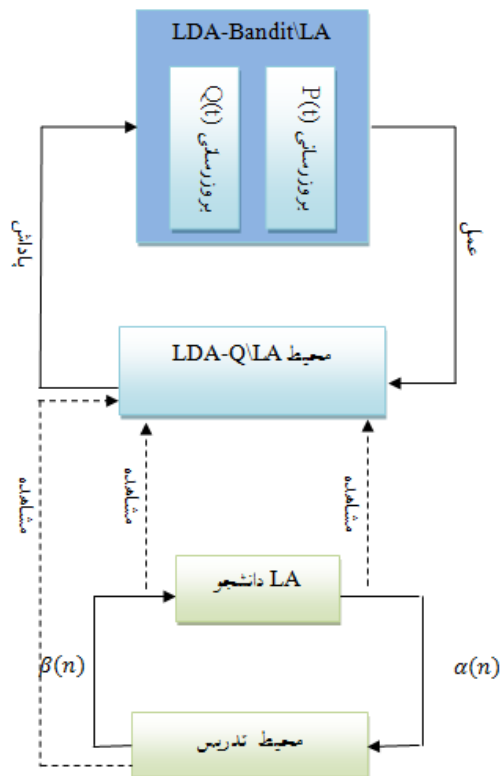
۲-۱ N-armed Bandit

مساله یادگیری را در نظر بگیرید که در آن به طور مکرر با انتخاب یک عمل یا گزینه از میان n عمل دیگر مواجه بوده و پس از هر بار انتخاب، بر اساس عمل انتخاب شده، پاداش عددی با استفاده از توزیع احتمالی ثابت دریافت می‌کنید. هدف این مساله، بیشینه ساختن پاداش موردانتظار کل در طول زمان است. به فرایند انتخاب هر عمل، یک بازی گفته می‌شود. مفهوم فوق، مساله اصلی bandit با n بازو نامیده می‌شود که در حالت کلی دارای n بازو(هرم) عملگر است.

۴- مدل پیشنهادی LDA-Bandit

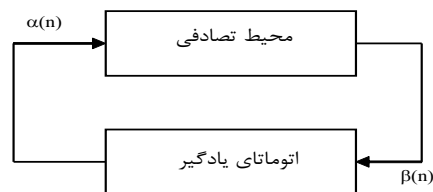
Bandit\LA

این مدل ترکیبی از آتوماتای یادگیر و N-armed Bandit است بطوریکه چند آتوماتای یادگیر به عنوان شبیه ساز دانشجو و یک عامل تعیین کننده سطح دانشجو به نام LDA-Bandit\LA بطور همزمان به هم متصل شدند، چگونگی ارتباطات این مدل در شکل ۲ نشان داده شد. این مدل از دو سطح تشکیل شده، در سطح بالا عامل تعیین کننده سطح که برای تعیین نوع دانشجو از N-armed Bandit می‌کند و بر اساس بررسی کارایی، آتوماتای یادگیر سطح پایین به عمل انتخابی پاداش یا جریمه می‌دهد.



شکل (۲) مدل پیشنهادی LDA-Bandit\LA

همگرایی عامل سطح بالا به رفتار آتوماتای سطح پایین بستگی دارد بطوریکه عامل تعیین کننده سطح، می‌تواند زمانیکه هنوز آتوماتای سطح پایین همگرا نشده، همگرا شود، در حالیکه برعکس آن نادرست است. محیط DLA-Bandit LA ترکیبی از مشاهدات است، که باید بطور ضمنی بعد از انجام پردازش پیام‌های مشاهده شده، دریافت شود، DLA-



شکل (۱) اتصال آتوماتای یادگیر با محیط [۱۲]

آتوماتای یادگیر به دو گروه با ساختار ثابت و با ساختار متغیر تقسیم می‌شود [۱۲، ۱۴، ۱۳]. آتوماتای یادگیر با ساختار ثابت با احتمالات انتقال ثابت مشخص^۵ می‌شود. آتوماتای یادگیر با ساختار متغیر^۶ توسط ۴ تایی $\{\alpha, \beta, p, T\}$ نشان داده می‌شود، که در آن $\alpha \equiv \{\alpha_1, \alpha_2, \dots, \alpha_r\}$ مجموعه اعمال آتوماتا، مجموعه ورودی‌های آتوماتا $\beta \equiv \{\beta_1, \beta_2, \dots, \beta_m\}$ است و $p = \{p_1, p_2, \dots, p_n\}$ بردار احتمال انتخاب هر یک از اعمال و T الگوریتم بروزرسانی است. در این نوع از آتوماتاها، اگر عمل α_i در مرحله n ام انتخاب شود و این عمل، پاسخ مطلوب از محیط دریافت کند، احتمال $p_i(n)$ افزایش یافته و سایر احتمالات کاهش می‌یابد. برای پاسخ نامطلوب احتمال $p_i(n)$ کاهش یافته و سایر احتمالات افزایش می‌یابد. در هر حال، تغییرات به گونه‌ای صورت می‌گیرد تا حاصل جمع $p_i(n)$ ها همواره ثابت و مساوی یک باقی بماند.

۳- سیستم شبه آموزشی^۷

سیستم‌های شبه آموزشی، نمایشی از وضعیت دانشجو می‌باشد. این سیستم، توانایی کارکردن و تست بدون نیاز به اشخاص واقعی را دارا می‌باشد. یکی از مشکلات هر سیستم real-life با دانشجویان real-life این است که سیستم باید برای رسیدن به یک استراتژی تدریس بهینه از قبل توسط دانشجویان بسیاری، تست شود. استفاده از سیستم شبیه‌ساز دانشجو، بطور امیدوارکننده‌ای، این مشکل را برطرف می‌کند و سیستم شبه آموزشی را برای رسیدن به استراتژی بهینه بدون نیاز به اشخاص برای تعامل با یک سیستم آموزشی بهینه، توانا می‌سازد [۱۶].

در اینجا از شبیه‌ساز دانشجویی که الگوریتم یادگیری N-Armed Bandit را بعنوان مکانیزم یادگیری برای تقلید چگونگی رفتار دانشجو نشان می‌دهد، استفاده شد. با به کارگیری این الگوریتم مدل شبیه ساز می‌تواند خصوصیتی از دانشجوی واقعی را تشخیص و به همان اندازه تعاملات دانشجو با سیستم شبه-آموزشی را افزایش دهد [۱۶].

اگر سرعت یادگیری مشاهده شده جاری از شبیه ساز دانشجو کمتر از θ_1 باشد، ادعا می کنیم که شبیه ساز دانشجو یادگیرنده کند است. حال اگر $LDA-Bandit \setminus LA$ ، α_1 را انتخاب کند، این پاداش را انتخاب می کند در غیر این صورت جریمه انتخاب می شود.

جدول ۵-۱ آستانه مورد استفاده توسط $LDA-Bandit \setminus LA$ برای مدل یادگیری [۱۵]

انواع دانشجو	حد آستانه
یادگیرنده سریع	$\theta(t) < \theta_1$
یادگیرنده نرمال	$\theta_1 \leq \theta(t) < \theta_2$
یادگیرنده کند	$\theta_2 \leq \theta(t)$

اگر سرعت یادگیری مشاهده شده جاری از شبیه ساز دانشجو مساوی یا بزرگتر از θ_1 باشد و کمتر از θ_2 ، ادعا می کنیم که شبیه ساز دانشجو یادگیرنده نرمال است. حال اگر $LDA-Bandit \setminus LA$ ، α_2 را انتخاب کند، این پاداش را انتخاب می کند در غیر این صورت جریمه انتخاب می شود.

اگر سرعت یادگیری مشاهده شده جاری از شبیه ساز دانشجو مساوی یا بزرگتر از θ_2 باشد، ادعا می کنیم که شبیه ساز دانشجو خصوصیات یادگیرنده سریع را دارد. حال اگر $LDA-Bandit \setminus LA$ ، α_3 را انتخاب کند، این پاداش را انتخاب می کند در غیر این صورت جریمه انتخاب می شود.

۱. Q بردار ارزش عمل است که پس از انتخاب عمل و

دریافت پاداش از محیط $LDA-Bandit \setminus LA$ توسط ۵-

۳ بروزرسانی می شود.

$$Q_{t+1}(a) \leftarrow Q_t(a) + \lambda(r - Q_t(a)) \quad (3)$$

در آغاز $LDA-Bandit \setminus LA$ ، مقدار مساوی $[0,0,0]^T$

برای هر یک از سه مدل یادگیری در نظر میگیرد.

۲. p بردار احتمال عمل است که شامل احتمالاتی است

که $LDA-Bandit \setminus LA$ به هر یک از اعمالش می دهد.

در لحظه t ، $LDA-Bandit \setminus LA$ بطور تصادفی

عمل $\alpha(t) = \alpha_i$ را انتخاب می کند. احتمالی که

آتوماتا عمل α_i را در زمان t انتخاب کند، احتمال

عمل است.

$Bandit/LA$ به محیط سطح پایین که شامل مجموعه ای از احتمالات جریمه می باشد و برای آتوماتای یادگیر سطح پایین ناشناخته است، دسترسی دارد، و همچنین به اعمال انتخاب شده توسط آتوماتای یادگیر سطح پایین و نیز پاسخ های ایجاد شده توسط محیط سطح پایین دسترسی دارد و نیز این محیط شامل احتمالات پاداش برای هریک از اعمالش می باشد. در مجموع، همانطور که در شکل مشخص است، این نتایج مجموعاً محیط $DLA-Bandit/LA$ را تشکیل می دهد.

۴-۱ پیاده سازی مدل دانشجو

در این مدل نیز مجموعه ای متناهی از مدل های یادگیری ممکن را برای هر دانشجو داریم و دانشجویان را در سه سطح کند، نرمال و سریع در نظر گرفتیم، هریک به صورت یک عمل برای $LDA-Bandit \setminus LA$ سطح بالا مورد استفاده قرار می گیرد. در مدل ارائه شده $LDA-Bandit \setminus LA$ برای یادگرفتن بهترین مدل ممکن برای شبیه ساز دانشجو از پنج تایی $\{\underline{\alpha}, \underline{\beta}, P, T, Q\}$ استفاده می کند.

$\alpha_1, \underline{\alpha} = \{\alpha_1, \alpha_2, \alpha_3\}$ نشان می دهد عمل با

یادگیرنده کند و α_2 نشان می دهد عمل با یادگیرنده نرمال و در

نهایت، α_3 نشان می دهد عمل با یادگیرنده سریع مطابق است.

$\underline{\beta} = \{0,1\}$ ، $\underline{\beta} = 0$ به پاداش برای عمل انتخاب شده و

$\underline{\beta} = 1$ به در نظر گرفتن جریمه برای عمل جاری اشاره دارد.

$P = [p_1, p_2, p_3]^T$ ، $p_i(n)$ احتمالات جاری شبیه ساز

دانشجو که با α_i نشان داده شد و T قانون بروزرسانی

احتمالات است که بصورت $T : (P, \alpha, \beta) \rightarrow P$ نگاشت

می شود و نهایتاً، $Q = [q_1, q_2, q_3]^T$ ، $Q_i(n)$ ارزش انجام

عمل جاری شبیه ساز دانشجویی که با α_i نشان داده شده، می-

باشد. هریک از اینها در ادامه بصورت کامل توضیح داده شدند:

β ورودی دریافتی $LDA-Bandit \setminus LA$ است. $LDA-$

$Bandit \setminus LA$ باید دنباله ای از تصمیمات اتخاذ شده توسط شبیه

ساز دانشجو را مشاهده کند و براساس این دنباله، باید نتیجه-

گیری کند که آیا مدل جاری اش درست است یا نه؟ فرض -

کنیم، $\theta(t)$ سرعت یادگیری آتوماتای یادگیر پایین در زمان t

باشد. توسط بررسی انتخاب های شبیه ساز دانشجو که در مدت

$window$ یاد می گیرد، استنتاج می شود که آیا مدل جاری باید

پاداش بگیرد یا جریمه شود؟ توسط دو آستانه

$\{\theta_j | 1 \leq j \leq 2\}$ به هدفمان می رسیم:

یادگیرنده سریع: برای شبیه سازی دانشجو از این نوع، از اتوماتای یادگیر Pursuit استفاده شد که در این پایان نامه با الگوی PLRI و با $\lambda = 0.05$ پیاده سازی شد.

برای اندازه گیری θ ، نرخ یادگیری دانشجو، محیط LDA-Bandit از فرمول (۷) استفاده می کند:

$$\frac{1}{n} \sum n_i c_i \quad (۷)$$

که n_i تعداد دفعاتی است که عمل α_i در طول انجام یک تعداد ثابت از پرس و جوها انتخاب می شود و c_i احتمال جریمه وابسته با عمل α_i اتوماتای یادگیر سطح پایین است. نرخ یادگیری دانشجو تابعی از زمان است یعنی با گذشت زمان این مقدار برای هر سه نوع دانشجو کاهش می یابد زیرا دانش دانشجو با گذشت زمان افزایش می یابد.

۵- ارزیابی مدل

در این قسمت نتایج بدست آمده از شبیه سازی مدل LDA-Bandit نشان داده شد. برای بدست آوردن این نتایج شبیه سازی های مختلفی را اجرا کردیم تا مشخص کنیم مدل کننده دانشجو تا چه اندازه می تواند مدل یادگیری دانشجو را درست تشخیص دهد و نتایج بدست آمده با نتایج مدل قبلی [۱۰] مقایسه شد. شبیه سازی ها در انواع مختلفی از محیط ها تست شد که این محیط ها مسئله تدریس است که شامل مواد درسی نامعلوم از مدل دامنه وابسته با سامانه شبه آموزشی است. در همه آزمایش های اجرا شده، اگر احتمال انتخاب یک عمل بزرگتر یا مساوی به آستانه $(0 < T \leq 1)$ شود، الگوریتم همگرا در نظر گرفته می شود. اگر اتوماتا به بهترین عمل (عملی با بیشترین احتمال پاداش) همگرا شود آن را همگرایی صحیح بیان کردیم.

شبیه سازی ها برای محیط های محک موجود تست شد و حد آستانه، ۰.۹۹، تنظیم شد و ۷۵ بار آزمایشات اجرا شد و نتیجه، میانگین ۷۵ اجرا است. در شبیه سازی ها فرض شد که هنگامی که دانشجو استفاده از سامانه را آغاز می کند، نتایج در طول یک فاز یادگیری کوتاه نادرست است و بنابراین ۲۰۰ تکرار اول (تمام تکرارها، تصمیم گیری ها و پاسخ های اتوماتای سطح پایین) توسط LDA-Bandit صرف نظر شد. نتایج این شبیه سازی ها در ادامه توضیح داده شد.

۵-۱ آزمایش در محیط با ۴ عمل

$$p_i(t) = pr[\alpha(t) = \alpha_i],$$

$$where, \sum_{i=1}^r p_i(t) = 1, \forall t, i = 1, 2, 3 \quad (۴)$$

در آغاز LDA-Bandit\LA، احتمال مساوی $[0.333, 0.333, 0.333]^T$ برای هر یک از سه مدل یادگیری در نظر می گیرد.

۳. الگوریتم بروزرسانی ۵-۶ کی است که برای بروزرسانی بردار احتمال عمل LDA-Bandit\LA بکار گرفته می شود و بصورت (۵) نمایش داده می شود.

$$p(t+1) = T[p(t), \alpha(t), \beta(t)] \quad (۵)$$

این الگوریتم بروزرسانی می تواند هر یک از الگوهای یادگیری LA باشد. در روش پیشنهادی از الگوی DPRI استفاده شد که در این الگو هنگامی که عملی پاداش دریافت می کند، همه اعمالی که منطبق بر بیشترین تخمین نیستند، کاهش می یابند و عمل با بیشترین تخمین افزایش می یابد. مجموع بردار احتمال همیشه برابر با یک است. هنگامی که عمل انتخابی جریمه می شود، بردار احتمال عمل بروزرسانی نمی شود [۱۲]. الگوی DPRI در ۶ بیان شد.

$$\text{If } \beta(t) = 0 \text{ and } p_m(t) \neq 1 \text{ Then}$$

$$p_j(t+1) = \max_{j \neq m} \{p_j(t) - \Delta, 0\}$$

$$p_m(t+1) = 1 - \sum_{j \neq m} \{p_j(t+1)\}$$

Else

$$p_j(t+1) = p_j(t) \text{ for all } 1 \leq j \leq r$$

LDA-Bandit\LA به عنوان یک اتوماتای یادگیر سطح بالا پیاده سازی شد و به محیط یادگیری، اعمال و جریمه های اتوماتای یادگیر سطح پایین وابسته به شبیه ساز دانشجو و نیز محیط خودش دسترسی دارد بطوریکه بعدا مواد تدریسی را یاد می گیرد. LDA-Bandit\LA با الگوریتم پیگیری N-armed bandit پیاده سازی شد. شبیه ساز دانشجو برای تقلید سه نوع از دانشجویان بصورت زیر شبیه سازی شد. یادگیرنده کند: این نوع دانشجویان، با LR از الگوریتم های VSSA است، شبیه سازی شد. در اصل از اتوماتای LRI با نرخ یادگیری ۰.۰۲۵ استفاده شد.

یادگیرنده نرمال: برای این گروه از دانشجویان، برای شبیه سازی رفتار دانشجو از VSSA استفاده کردیم که در این پایان نامه با الگوی LRI و با $\lambda = 0.05$ پیاده سازی شد.

این مقدار در الگوریتم پیشنهاد شده در مرجع [۱۰]، ۱۰۰٪ است ولی با تعداد تکرار ۶۰۷. همچنین در زمان استفاده از آتوماتای LRI، درصد همگرایی صحیح در مدل پیشنهادی ۹۷٪ و در الگوریتم پیشنهاد شده در مرجع [۱۰]، ۸۷٪ و همینطور در زمان تعیین یادگیرنده کند، این مقادیر به ترتیب ۹۹٪ و ۹۱٪ است.

محیط با ۴ عمل یک سوال ۴ گزینه ای را نشان می دهد. دانشجو باید محتویات این سوال را یاد بگیرد. بر اساس کارایی دانشجو -LDA Bandit/LA مدل یادگیری دانشجو را تعیین می کند. دو تنظیم مختلف برای محیط ها به همراه احتمال های پاداش در جداول مشخص شد. نتایج بدست آمده در محیط EA که بهترین عمل α_1 است در جدول ۲ نشان داده شد. در زمانیکه آتوماتای سطح پایین تر از الگوریتم pursuit استفاده می کند، مدل Bandit/LA -LDA قادر به تشخیص مدل یادگیری درست دانشجو با دقت ۱۰۰٪ است، در حالیکه

جدول ۲- مقایسه همگرایی مدل LDA-Bandit/LA و مدل پیشنهاد شده در مرجع [۱۰] در محیط E(A,ε)

محیط $E(A, \varepsilon)$	θ_1	θ_2	یادگیرنده سریع		یادگیرنده نرمال		یادگیرنده کند	
			تعداد تکرار	همگرایی صحیح	تعداد تکرار	همگرایی صحیح	تعداد تکرار	همگرایی صحیح
مدل پیشنهادشده در مرجع [۱۰]	۰٫۴۶-۰٫۳۵	۰٫۵۷-۰٫۴۰	۶۰۷	٪۱۰۰	۱۰۳۷	٪۸۷	۱۲۷۳	٪۹۱
مدل پیشنهادی LDA-Bandit/LA			۴۸۱	٪۱۰۰	۵۵۴	٪۹۷	۵۹۹	٪۹۹
احتمال پاداش	$E_A = \{0.7, 0.5, 0.3, 0.2\}$							

جدول ۳- مقایسه همگرایی مدل LDA-Bandit/LA و مدل پیشنهاد شده در مرجع [۱۰] در محیط E(B,ε)

محیط E(B,ε)	θ_1	θ_2	یادگیرنده سریع		یادگیرنده نرمال		یادگیرنده کند	
			تعداد تکرار	همگرایی صحیح	تعداد تکرار	همگرایی صحیح	تعداد تکرار	همگرایی صحیح
مدل پیشنهادشده مرجع [۱۰]	۰,۴۶-۰,۳۵	۰,۵۷-۰,۴۰	۷۹۰	٪۹۷	۱۳۰۶	٪۸۹	۱۴۸۹	٪۸۵
مدل پیشنهادی-LDA Bandit/LA			۴۹۵	٪۱۰۰	۵۷۹	٪۱۰۰	۵۳۵	٪۱۰۰
احتمال پاداش	$E_B=\{0.1 \quad 0.45 \quad 0.84 \quad 0.76 \}$							

در مدل LDA-Bandit/LA قادر به تشخیص مدل یادگیری درست دانشجو در ۱۰۰٪ زمان است، در حالیکه این مقدار در مدل

نتایج بدست آمده در محیط EB که بهترین عمل α_3 است و پیچیدگی بیشتری نسبت به محیط EA دارد در جدول ۳ نشان داده شد. همانطور که در جدول مشخص شد، مدل پیشنهادی در تعیین دانشجو کند، نرمال و سریع در ۱۰۰٪ مواقع موفق بود و با تعداد تکرار بسیار کمتر از مدل مرجع.

۲-۵ آزمایش در محیط با ۱۰ عمل

پیشنهاد شده در مرجع [۱۰]، ۹۵٪ است. همچنین در زمان استفاده از آتوماتای LRI، درصد همگرایی صحیح در مدل ما ۱۰۰٪ و در الگوریتم پیشنهاد شده در مرجع [۱۰]، ۹۳٪ بود و در این حالت مدل ما با تعداد تکرار کمتری به این نتیجه رسید و همینطور در مورد یادگیرنده کند، این مقادیر به ترتیب ۱۰۰٪ و ۹۶٪ می باشند.

محیط با ۱۰ عمل، یک سوال با ۱۰ گزینه را نمایش می دهد. احتمالات پاداش محیط در جدول ۴ و ۵ مشخص شد. نتایج بدست آمده در محیط EA که بهترین عمل α_1 است در جدول ۴-۵ نشان داده شد. در زمانیکه آتوماتای سطح پایین از PLRI استفاده می کند،

نتایج بدست آمده در محیط EB که بهترین عمل α_3 است و پیچیدگی بیشتری نسبت به محیط EA دارد در جدول ۵ نشان داده

شد. همانطور که در جدول مشخص شد، مدل پیشنهادی در تعیین

دانشجو سریع و نرمال در ۱۰۰٪ مواقع موفق بوده و در تعیین

جدول ۴ مقایسه همگرایی LDA-Bandit\LA و Meta LA [۱۰] در محیط $E(A, 10)$

محیط E(A,۱۰)	θ_1	θ_2	یادگیرنده سریع		یادگیرنده نرمال		یادگیرنده کند	
			تعداد تکرار	همگرایی صحیح	تعداد تکرار	همگرایی صحیح	تعداد تکرار	همگرایی صحیح
مدل پیشنهادشده درمرجع [۱۰]	۰٫۴۶ – ۰٫۳۵	۰٫۵۷ – ۰٫۴۰	۶۶۸	٪۹۵	۷۴۸	٪۹۳	۱۳۶۹	٪۹۶
مدل پیشنهادی- <i>LDA</i> <i>Bandit/LA</i>			۴۴۴	٪۱۰۰	۴۹۶	٪۱۰۰	۷۰۲	٪۱۰۰
احتمال پاداش	$E_A=\{0.7 \quad 0.5 \quad 0.3 \quad 0.2 \quad 0.4 \quad 0.5 \quad 0.4 \quad 0.3 \quad 0.5 \quad 0.2 \}$							

جدول ۵ مقایسه همگرایی مدل LDA-Bandit\LA و مدل پیشنهادشده در مرجع [۱۰] در محیط $E(B, 10)$

محیط $E(B, ۱۰)$	θ_1	θ_2	یادگیرنده سریع		یادگیرنده نرمال		یادگیرنده کند	
			تعداد تکرار	همگرایی صحیح	تعداد تکرار	همگرایی صحیح	تعداد تکرار	همگرایی صحیح
مدل پیشنهادشده در مرجع [۱۰]	۰٫۴۶ - ۰٫۳۵	۰٫۵۷ - ۰٫۴۰	۶۱۶	%۱۰۰	۹۷۰	%۸۵	۱۱۴۰	%۹۶
مدل پیشنهادی- <i>LDA</i> <i>Bandit/LA</i>			۴۵۷	%۱۰۰	۵۰۵	%۱۰۰	۶۵۷	%۹۹
احتمال پاداش	$E_B = \{0, ۱, ۰, ۴۵, ۰, ۸۴, ۰, ۷۶, ۰, ۲, ۰, ۴, ۰, ۶, ۰, ۷, ۰, ۵, ۰, ۳\}$							

دانشجو کمک کرد. نتایج شبیه سازی‌ها نشان داد که این روش مکانیزم معتبر و امکان پذیر برای اجرای فرایند یادگیری دانشجو است و همچنین نشان داد که مدل دانشجو در تعیین مدل یادگیری دانشجو موفق است و استفاده از الگوریتم یادگیری N-Armed Bandit در همه حالت‌ها، با دقت بالای ۹۵٪ مدل یادگیری را مشخص کرد و تعداد تکرارها بطور چشمگیری کاهش یافت و دقت افزایش یافت.

مراجع

- [۱] V.J.Shute; J. Psotka; "Intelligent Tutoring Systems: Past, Present, and Future", Handbook of Research on Educational Communications and Technology, Scholastic Publications, ۱۹۹۵.
- [۲] E. Wenger; "Artificial Intelligence and Tutoring Systems: Computational and Cognitive Approaches to the Communication of Knowledge", Los Altos, CA: Morgan Kaufmann Publishers, Inc, ۱۹۸۷.
- [۳] E. Fischetti ; A. Gisolfi; "From computer-aided instruction to intelligent tutoring systems", Educ. Technol, vol, ۳۰, no. ۸, pp. ۷-۱۷, ۱۹۹۰.

۵- جمع بندی

این مقاله استراتژی جدیدی برای سیستم‌شبه‌آموزشی بیان می‌کند که به مدلسازی دانشجو اشاره دارد. مدل کننده دانشجو خودش از الگوریتم یادگیری N-Armed Bandit به عنوان یک مکانیزم درونی برای تعیین مدل یادگیری دانشجو استفاده می‌کند بطوریکه می‌تواند در سیستم‌شبه‌آموزشی جهت تعیین تمرین یادگیری برای هر دانشجو مورد استفاده قرار گیرد. برای رسیدن به این، مدل کننده دانشجو از الگوریتم یادگیری N-Armed Bandit سطح بالا به نام LDA-Bandit\LA استفاده می‌کند که اعمال اتوماتهای شبیه ساز دانشجو و محیط تدریس را مشاهده می‌کند و تلاش می‌کند تا مدل یادگیری دانشجو را تعیین کند. LDA-Bandit\LA مشخص می‌کند که دانشجو در سوال، یادگیرنده کند، سریع و یا نرمال است.

این مقاله با استفاده از بکارگیری الگوریتم یادگیری N-Armed Bandit سرعت و دقت همگرایی و در نتیجه تعیین مدل یادگیری دانشجو را بهبود داد و به سیستم شبه آموزشی برای تعیین بهتر مدل

۲۰۰۷ International Conference of Machine Learning and Cybernetics, Hong Kong, pp. ۱۱۲–۱۱۸, August ۲۰۰۷.

[۱۸] A. G. Barto and R. S. Sutton, *Reinforcement Learning: an introduction*, MIT Press, Cambridge, MA, ۱۹۹۸.

^۱ Level determinant Agent

^۲ Pedagogical

^۳ Discrete Linear Reward Inaction

^۴ Action

^۵ fixed state transition probabilities

^۶ Variable Structure Learning Automata

^۷ Tutorial-like system

[۴] R. Winkels ; J. Breuker; *What's in an ITS? a functional decomposition*, in New Directions for Intelligent Tutoring systems, E. Costa, Ed. Berlin, Germany: Springer-Verlag, ۱۹۹۰.

[۵] J. Self; “*The defining characteristics of intelligent tutoring systems research: ITSs care, precisely*”, Int. J. Artif. Intell. Educ, vol. ۱۰, pp. ۳۵۰–۳۶۴, ۱۹۹۹.

[۶] R. Lelouche; “*A collection of pedagogical agents for intelligent educational systems*”, In intelligent tutoring systems: ۶th international Conference, ITS ۲۰۰۰: Lecture Notes in computer Science, pages ۱۴۳–۱۵۲. Springer-verlag, Berlin, ۲۰۰۰.

[۷] R. S. Legaspi ; R. C. Sison; “*Modeling the tutor using reinforcement learning*”, in Proc. PCSC, pp. ۱۹۴–۱۹۶, ۲۰۰۰.

[۸] P. Baffes ; R. Mooney; “*Refinement-based student modeling and automated bug library construction*”, J. Artif. Intell. Educ, vol. ۷, no. ۱, pp. ۷۵–۱۱۶, ۱۹۹۶.

[۹] M.K., Hashem; B.J. Oommen; “*Modeling a Student's Behavior in a Tutorial-Like System Using Learning Automata* ”, IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics., ۱۰۸۳-۴۴۱۹/\$۲۶,۰۰, ۲۰۰۹ IEEE.

[۱۰] M. K. Hashem; “*Learning automata based intelligent tutorial-like systems*”, Ph.D. dissertation, School Comput. Sci., Carleton Univ., Ottawa, ON, Canada, ۲۰۰۷.

[۱۱] R.S. Sutton; A.C. Barto; “*Reinforcement Learning: An introduction*”, MIT Press, Cambridge, MA, Bradford Book, ۱۹۹۸A.

[۱۲] K. S. Narendra ; M. A. L. Thathachar; *Learning Automata: An Introduction*. Englewood Cliffs, NJ: Prentice-Hall, ۱۹۸۹.

[۱۳] M. A. L. Thathachar ; P. S. Sastry; “*Varieties of learning automata: An overview*”, IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. ۲۲, no. ۶, pp. ۷۱۱–۷۲۲, Dec. ۲۰۰۲.

[۱۴] S. Lakshmivarahan; “*Learning Algorithms Theory and Applications*” New York: Springer-Verlag, ۱۹۸۱.

[۱۵] M.K., Hashem; B.J. Oommen; “*On using learning automata to model a student's behavior in a tutorial-like system*”, In: Okuno, H.G, Ali, M. (eds.) IEA/AIE ۲۰۰۷. LNCS (LNAI), vol. ۴۵۷۰, pp. ۸۱۳–۸۲۲, Springer, Heidelberg (۲۰۰۷).

[۱۶] M.K., Hashem; B.J. Oommen; “*Learning Automata Based Intelligent Tutorial-like System*”, J.D. Velásquez et al. (Eds.): KES ۲۰۰۹, Part I, LNAI ۵۷۱۱, pp. ۳۶۱–۳۷۴, ۲۰۰۹. Springer-Verlag Berlin Heidelberg ۲۰۰۹.

[۱۷] M.K., Hashem; B.J. Oommen; “*Using learning automata to model a domain in a tutorial-like system*”, In: Proceedings of ICMLC ۲۰۰۷, the