

روشی مبتنی بر آتاماتای یادگیر توزیع شده برای مدل‌سازی کاربران در محیط‌های فراپیوندی

محمد رضا ملاخلیلی میبیدی^۱ محمد رضا میبیدی^۲

^۱ دانشکده فنی و مهندسی، دانشگاه آزاد اسلامی، واحد علوم و تحقیقات، تهران، ایران
^۲ دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، تهران، ایران
mmeybodi@srbiau.ac.ir, mmeybodi@aut.ac.ir

چکیده - مدل‌سازی کاربر در وب یکی از موضوعات مهمی است که در سالهای اخیر توجه زیادی را به خود جلب کرده است. در این مقاله راه حلی مبتنی بر آتاماتاهای یادگیر برای محیط‌های فراپیوندی مانند وب پیشنهاد می‌گردد. در راه حل پیشنهادی از آتاماتای یادگیر توزیع شده به عنوان ابزاری جهت استخراج دانش نهفته در تعامل کاربران، برای یادگیری حرکت کاربران وب استفاده شده است. اطلاعات استخراج شده از این طریق می‌تواند برای مدل کردن کاربران مورد استفاده قرار گیرد. نتایج شبیه‌سازی‌ها نشان داده است که آتاماتای یادگیر توزیع شده قادر است توزیع احتمالی نحوه حرکت کاربران که برآیند انگیزه‌ها، علائق، اهداف، دانش و پیش‌دانسته‌های کاربران است را استخراج کند. کلید واژه - مدل‌سازی کاربر، وب، آتاماتاهای یادگیر، آتاماتای یادگیر توزیع شده.

۱. مقدمه

یک مدل از کاربر تطبیق می‌دهد، صورت می‌گیرد. بر این اساس عنصر کلیدی در یک کاربرد مبتنی بر فرارسانه‌های انطباق‌پذیر، مدل کاربر است. هر چه قدر مدل کاربر جامع‌تر و حاوی اطلاعات بیشتری باشد، محتوا و نمایش اطلاعات، بهتر شخصی‌سازی می‌گردد.

در این مقاله از آتاماتای یادگیر توزیع شده، به عنوان ابزاری جهت استخراج دانش نهفته در حرکت کاربران وب استفاده شده است. این دانش در قالب یک توزیع احتمالی ارائه می‌گردد که معرف برآیند خط مشی رفتار کاوشگرانه کاربران در وب است و به عنوان مدلی از کاربر در وب در نظر گرفته می‌شود. علاوه بر ارائه مبانی ریاضی این ایده، بر اساس مدل معرفی شده در [1] که قادر به ایجاد ساختاری مشابه با وب است، توزیع احتمالاتی رفتار کاربران ایجاد و با انجام شبیه‌سازی، این الگوی رفتاری توسط آتاماتای یادگیر توزیع شده استخراج شده است. نتایج شبیه‌سازی‌ها نیز نشان می‌دهد که آتاماتای یادگیر توزیع شده قادر است توزیع احتمالات نحوه حرکت کاربران که برآیند انگیزه‌ها، علائق، اهداف، دانش و پیش‌دانسته‌های کاربران است را استخراج کند. مدل پیشنهادی دارای کاربردهای متعددی است که برخی از اینها در [۳][۴][۱۴] آمده است.

بررسی رفتارهای کاربران در وب، به عنوان روشی جهت کشف دانش نهفته در نحوه تعامل آنها با وب، یکی از ابزارهای مهم در حوزه کاوش در وب شناخته می‌شود که در متون مربوطه از آن با عنوان داده کاوی استفاده از وب نام می‌برند. کارهای تحقیقاتی بسیاری در این حوزه انجام شده است که عمدتاً بر مبنای اطلاعات موجود از رفتار کاربر در تعامل با وب و با فرضیات مختلف در مورد ذهنیات و پیش‌دانسته‌های کاربر، به استخراج این دانش و استفاده از آن در کاربردهای مختلف در وب نظیر خودسازمانده کردن وب [8][7]، بهبود کارایی وب با حدس زدن حرکت بعدی کاربر [14]، بهبود در نتایج حاصل از الگوریتم‌های رتبه‌دهی [12][11][10][9]، کمک به کاربران در تجارت الکترونیک و موارد دیگری از این نوع می‌پردازند.

یکی از کاربردهایی که بر مبنای داده کاوی استفاده از وب در کانون توجه قرار دارد، تکنولوژی فرارسانه‌های انطباق‌پذیر است [2]. این تکنولوژی اجازه می‌دهد نرم‌افزارهای کاربردی در محیط‌های فرارسانه‌ای بر اساس علائق و ویژگی‌های کاربران شخصی‌سازی شوند. این فرآیند بر اساس یک موتور شخصی‌سازی و یک سیستم تصمیم‌یار که محتویات را بر اساس

صورت می‌گیرد. بر این اساس عنصر کلیدی در یک کاربرد مبتنی بر فرارسانه‌های انطباق‌پذیر، مدل کاربر است. هر چه قدر مدل کاربر جامع تر و حاوی اطلاعات بیشتری باشد، محتوا و نمایش اطلاعات بهتر شخصی‌سازی می‌شود. یک مدل کاربر را مجموعه‌ای از ساختارهای اطلاعاتی تعریف می‌کنیم که برای نمایش یک یا چند مورد از موارد زیر طراحی شده باشد:

الف- نمایش فرضیات درباره دانش، اهداف، ترجیحات، وظایف یا توانایی‌های یک نوع یا بیشتر، از انواع مختلف کاربران.
ب- نمایش ویژگی‌های مشترک شخصیتی کاربران در قالب گروه‌بندی آنها

ج- دسته‌بندی کاربران در یک یا بیشتر از زیرگروه‌ها
د- ذخیره‌سازی رفتار کاربر

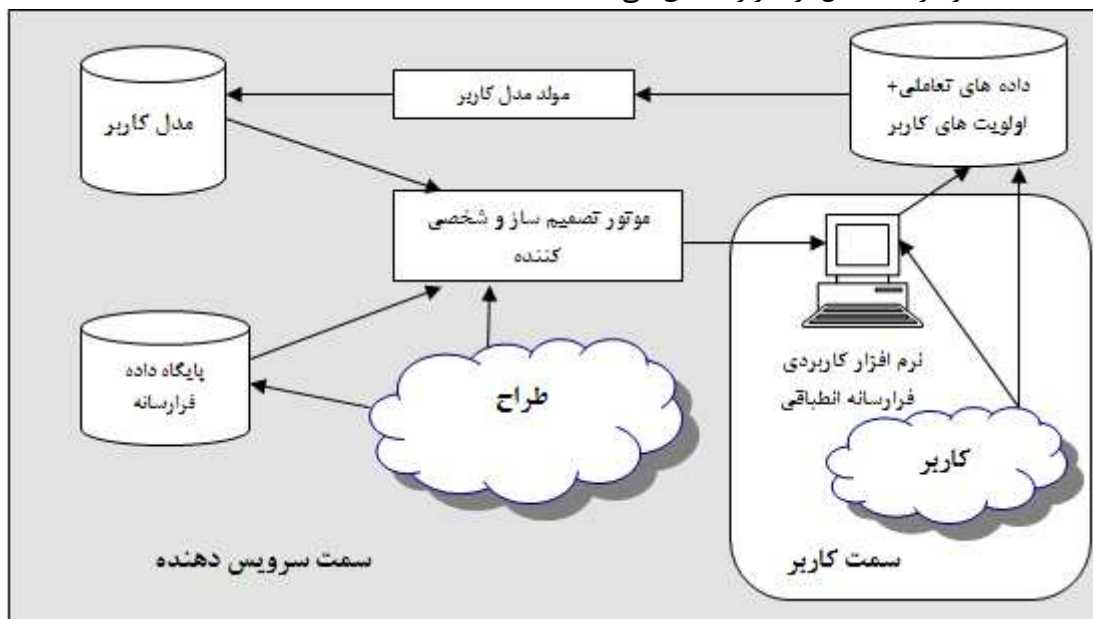
ه- ساخت فرضیات درباره کاربر بر اساس سابقه تعامل
و- تعمیم سوابق تعامل تعداد زیادی از کاربران به یک گروه.

معماری عمومی یک سیستم AH را در شکل ۱ می‌بینید [2]. این معماری به دو بخش تقسیم شده است. طرف سرویس دهنده و طرف سرویس گیرنده. طرف سرویس‌دهنده مدل‌های کاربر را بر اساس اطلاعات تراکنش کاربران با سیستم (که در بانک‌های اطلاعاتی سیستم نظیر Log فایل‌های موجود در سرویس‌دهنده‌های وب موجود است) و اطلاعات/تمایلات شخصی مرتبط با هر کاربر که به سیستم داده شده است را استخراج می‌کند. این مدل‌های کاربر در ترکیب با پایگاه داده فرارسانه، توسط یک "موتور تصمیم‌ساز و شخصی‌کننده" مورد استفاده قرار می‌گیرد تا نیازهای کاربران را مشخص کند.

ادامه این مقاله بدین صورت سازماندهی شده است. در بخش دوم، به بررسی مفهوم مدلسازی کاربر و شیوه‌های مرسوم مدلسازی کاربر مبتنی بر روشهای محاسبات نرم می‌پردازیم. در بخش سوم وب را به عنوان بزرگترین فرارسانه موجود، به کمک گراف مدل‌سازی کرده و ویژگی‌های مدل گراف وب را بررسی می‌کنیم. بر اساس این اطلاعات، مدل معرفی شده در [1] را که برای مدلسازی وب و نشست‌های کاربران در وب ایجاد و یافته‌های آن بر اساس اطلاعات حاصل از برخی از وبسایت‌های بزرگ مورد بررسی قرار گرفته است معرفی خواهیم کرد. در بخش سوم، ضمن معرفی آتاماتای یادگیر توزیع شده به بررسی برخی از کاربردهای آن خواهیم پرداخت. در بخش چهارم، الگوریتم استخراج رفتار کاربران در وب بر مبنای آتاماتای یادگیر توزیع شده ارائه و نشان خواهیم داد که این الگوریتم قادر است توزیع احتمالاتی مربوط به حرکت کاربران در وب را استخراج کند. در بخش پنجم، بر اساس اطلاعات مدل معرفی شده در بخش ۲، با استفاده از شبیه‌سازی الگوریتم ارائه شده در بخش چهارم در مدل ایجاد شده، نتایج عملی این شبیه‌سازی را ارائه خواهیم داد. بخش نهایی مقاله نتیجه‌گیری می‌باشد.

۲. مدلسازی کاربر در فرارسانه‌های انطباق‌پذیر

تکنولوژی فرارسانه‌های انطباق‌پذیر (AH) اجازه شخصی‌سازی نرم‌افزارهای کاربردی مبتنی بر فرارسانه‌ها را بر اساس علائق و ویژگی‌های کاربران، به آنها می‌دهد [2]. فرآیند شخصی‌سازی یک کاربرد ابررسانه‌ای از طریق یک موتور شخصی‌سازی و تصمیم‌یار که اطلاعات را بر یک مدل از کاربر تطبیق می‌دهد



شکل ۱: معماری عمومی یک کاربرد مبتنی بر فرارسانه انطباق‌پذیر

یک کاربر در دسترسی به یک سیستم فرارسانه الگوهایی را به نمایش می‌گذارد. تکنیک‌های یادگیری ماشین برای تشخیص نظم‌های موجود در اثرات باقی‌مانده از تعامل کاربر به کار گرفته می‌شود. این الگوها با یکدیگر یکپارچه شده و بخشی از مدل کاربر را تشکیل می‌دهد. محدودیت تکنیک‌های یادگیری ماشین فعلی برای مدل‌سازی رفتار انسانی، ما را به استفاده از شیوه‌های محاسبات نرم برای مدل‌سازی کاربر سوق می‌دهند. تکنولوژی محاسبات نرم، یک جواب تقریبی برای مسائل بد تعریف ارائه می‌کنند. این راه‌حل‌ها می‌توانند مدل‌های کاربر را در محیط‌هایی نظیر فرارسانه که در آن کاربران تمایلی به بازخورد اطلاعات در هنگام تعامل‌شان ندارند و طراحان نیز قادر به تعریف کامل تمام تعاملات ممکن نیستند، ایجاد کنند. در [2][5] خلاصه‌ای از تکنیک‌های مختلف محاسبات نرم (شامل شبکه‌های عصبی، الگوریتم‌های ژنتیک، منطق فازی و سیستم‌های نروفازی) موجود برای مدل‌سازی رفتار کاربران ارائه شده است.

مدل‌های کاربر ایجاد شده را می‌توان از دو منظر الف- میزان بزرگی یا دانه درشتی مدل و ب- نحوه استفاده از مدل ایجاد شده و کاربرد آن رده‌بندی کرد. از نظر بزرگی مدل، مدل‌ها را می‌توان به دو گروه تقسیم کرد: مدل‌هایی که برای یک کاربر منفرد ایجاد می‌گردد (که آنها را *Content-based Modeling* یا مدل‌سازی مبتنی بر محتوا نیز می‌گویند) و مدل‌هایی که برای گروه‌هایی از کاربران ایجاد می‌شوند (که آنها را *Collaborative Modeling* یا مدل‌سازی گروهی نیز می‌گویند). از جنبه کاربرد چهار نوع کاربرد برای مدل‌ها تعریف می‌شود: پیش‌بینی، پیشنهاد، دسته‌بندی و پالایش. هر مدل قادر است برخی یا تمامی این کاربردها را پوشش دهد.

۳. مدل گراف وب

گراف وب یک گراف پویا است که شامل میلیون‌ها گره و یال می‌باشد. هر گره در گراف وب نشان دهنده یکی از صفحات وب و یال‌های این گراف، فرایوندهای موجود میان صفحات وب را نمایش می‌دهد. در سال‌های اخیر تحقیقات بسیاری به بررسی ویژگی‌های ساختاری گراف وب، ارائه مدل‌های تصادفی برای گراف وب و روش‌هایی برای استخراج اطلاعات از این گراف پرداخته است. بخشی از این تحقیقات در باره استخراج ویژگی‌های این گراف با استفاده از *Crawler* ها و داده‌های حاصل

از *Crawling* در وب می‌باشد. بعضی از ویژگی‌های گراف وب به اختصار در ادامه این شرح داده شده است.

توزیع قانون توانی: مهمترین ویژگی مشاهده شده در وب، وجود توزیع‌های قانون توانی است. گراف غیرجهت‌دار G و عدد صحیح غیرمنفی K داده شده است. عدد $P_G(k)$ را طبق رابطه ۱ زیر تعریف می‌کنیم:

$$P_G(k) = \frac{|\{x \in V(G) : deg_G(x) = k\}|}{|V(G)|} \quad (1)$$

عدد $P_G(k)$ درصد گره‌های با درجه K را در گراف G مشخص می‌کند. گوییم توزیع درجه G ، از یک قانون توانی (*Power-Low*) تبعیت می‌کند اگر برای هر درجه K ، داشته باشیم:

$$P_G(k) = ck^{-\beta} \quad (2)$$

در رابطه ۲ مقادیر c و β اعداد ثابت مثبت هستند.

گراف‌های مدل دنیای واقعی نظیر گراف وب با توزیع درجه قانون توانی معمولاً گراف مستقل از اندازه نامیده می‌شوند. نتایج مطالعات نشان داده است که گراف وب دارای توزیع درجه ورودی قانون توانی با مقدار $\beta = 2.1$ و درجه خروجی با مقدار $\beta = 2.45$ و یا $\beta = 2.7$ است.

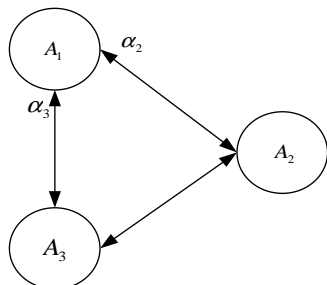
ویژگی *Small World*: در گراف G ، فاصله میان گره‌های u و v را با $d(u, v)$ نشان می‌دهیم و آن را برابر با تعداد یال‌های یک کوتاهترین مسیر میان u و v تعریف می‌کنیم. (در صورت عدم وجود این مسیر، این مقدار برابر با ∞ است). مقدار $L(G)$ را به شکل رابطه ۳ تعریف می‌کنیم:

$$L(G) = \sum_{\{u,v\} \in S} \frac{d(u,v)}{|S|} \quad (3)$$

که S مجموعه‌ای از جفت گره‌های مجزای u و v از گراف G با این ویژگی است که $d(u, v)$ متناهی است. نسبت $L(G)$ ، فاصله متوسط G است. ویژگی *Small World* ایجاب می‌کند که $L(G)$ می‌بایست بسیار کوچک‌تر از مرتبه گراف باشد. به‌عنوان مثال $L(G) \in O(\log|V(G)|)$. بررسی‌های گزارش شده در مورد وب نشان می‌دهد که $L_d(W) = 19$ و یا $L_d(W) = 16$ و $L_d(W) = 6.8$.

قطر گراف G که با $diam(G)$ نشان داده می‌شود بیشترین مقدار $d(u, v)$ است که به ازای تمام گره‌های مجزای u و v در گراف G است. نتایج بررسی‌ها نشان داده است که قطر گراف وب از ۹۰۰ بیشتر است.

A_j با p^j نشان داده شود، آنگاه p_m^j احتمال انتخاب عمل α_m از آتاماتای یادگیر A_j را نشان می‌دهد که احتمال انتخاب لبه خروجی (j,m) از میان لبه‌های خروجی گره j می‌باشد [6][15][16].



شکل ۲: آتاماتای یادگیر توزیع شده

۵. استفاده از آتاماتای یادگیر توزیع شده جهت یادگیری حرکت کاربران

قبل از استفاده از آتاماتای یادگیر توزیع شده جهت استخراج تابع توزیع حرکت کاربران، یک مدل ریاضی برای حرکت کاربران در وب ارائه میشود.

فرض کنید یک سیستم دینامیک زمان گسسته با n حالت $(1, 2, \dots, n)$ داریم (بخشی از وب با n سند وب). در هر حالت i (در هر سند i) مجموعه‌ای از تصمیم‌ها یا کنترل‌های $U(i)$ وجود دارد. (این کنترل، ناشناخته فرض می‌شود و ما در این مقاله آن را رفتار کاربر می‌نامیم که برآیندی است از انگیزه‌ها، ذهنیت‌ها، علایق، شناخت‌های پیشین، معماری اطلاعات در مستند وب و هر عامل دیگری که توسط کاربر در انجام حرکت‌ها و انتخاب فرایوندهای وب موثر است). در یک حالت i سیستم با کنترل $U(i)$ (کاربر) و با احتمال $p_{ij}(u)$ به حالت j تغییر حالت می‌دهد. (کنترل و احتمال مورد بحث هیچ کدام شناخته شده نیستند).

مجموعه توابع $M = \{\mu | \mu(i) \in U(i) \quad i = 1, 2, 3, \dots, n\}$ را در نظر بگیرید. دنباله $\{\mu_0, \mu_1, \mu_2, \dots\}$ با شرط $\mu_t \in M$ یک خط‌مشی نامیده می‌شود. فرض کنید $P(\mu)$ ماتریس احتمال گذر $\mu \in M$ باشد.

$$[P(\mu)]_{ij} = p_{ij}(\mu(t)) \quad i, j = 1, 2, 3, \dots, n$$

برای هر خط‌مشی $\pi = \{\mu_0, \mu_1, \mu_2, \dots\}$ داریم:

$$\Pr(\text{State is } j \text{ at time } t | \text{Initial State is } i \text{ and } \pi \text{ is used}) = [P(\mu_0), P(\mu_1), \dots, P(\mu_{t-1})]_{ij}$$

هدف از استفاده از آتاماتای یادگیر توزیع شده، یادگیری بردار احتمال مربوط به انتخاب لینک‌ها بر اساس یک خط‌مشی خاص

مدلسازی وب: با توجه به ویژگی‌های مذکور، یکی از چالش‌های مهم در تحقیقات در حوزه وب نحوه به دست آوردن اطلاعات از تعامل کاربران با وب می‌باشد. یکی از مهمترین روش‌هایی که برای این منظور وجود دارد، استفاده از مدل کاربر در وب است. در [1] مدلی ارائه شده است که بر اساس ویژگی‌های مدل گراف وب، می‌تواند نشست‌های کاربران در وب را شبیه‌سازی کند. به صورت خلاصه در این مدل، توزیع پیوندهای خروجی هر گره در گراف وب، تابع توزیع قانون توانی است. اتصالات بر اساس میزان شباهت میان اسناد ایجاد می‌گردند و شباهت میان اسناد تابعی از فاصله اقلیدسی میان اسناد است. هر صفحه وب در این مدل، دارای برداری است که هر مولفه آن میزان ارتباط این صفحه با موضوع متناظر با این مولفه را نشان می‌دهد. (تعداد موضوعات ثابت و قابل تعریف است). میزان ارتباط هر صفحه با یک موضوع به صورت عددی بین صفر و یک بیان می‌شود به طوری که مجموع آنها برای همه موضوعات برابر با یک است. همچنین هر صفحه دارای پیوندهایی با صفحات دیگر است. تعداد این پیوندها تابع توزیع قانون توانی است.

۴. آتاماتای یادگیر توزیع شده

آتاماتای یادگیر توزیع شده شبکه‌ای از تعدادی آتاماتای یادگیر است که برای حل یک مساله مشخص با یکدیگر همکاری می‌کنند. [15][16] یک آتاماتای یادگیر توزیع شده را می‌توان با یک گراف جهت‌دار مدل کرد. به صورتی که مجموعه گره‌های آن را مجموعه‌ای از آتاماتاهای یادگیر و یال‌های خروجی هر گره مجموعه اعمال متناظر با آتاماتای یادگیر متناظر با آن گره است. هنگامی که آتاماتا یکی از اعمال خود را انتخاب می‌کند، آتاماتایی که در دیگر انتهای یال متناظر با آن عمل قرار دارد، فعال می‌شود. بعنوان مثال در شکل ۲ هر آتاماتا ۲ اقدام دارد. اگر آتاماتای A_1 اقدام α_3 خود را انتخاب کند، آنگاه آتاماتای A_3 فعال خواهد شد. در گام بعد، آتاماتای A_3 یکی از اعمال خود را انتخاب می‌کند که منجر به فعال شدن یکی از اوتوماتاهای یادگیر متصل به A_3 می‌شود. در هر لحظه فقط یک آتاماتای یادگیر در آتاماتای یادگیر توزیع شده فعال می‌باشد. بصورت رسمی، یک آتاماتای یادگیر توزیع شده با n آتاماتای یادگیر توسط یک گراف (A, E) تعریف می‌شود که $A = \{A_1, A_2, \dots, A_n\}$ مجموعه آتاماتا و $E \subset A \times A$ مجموعه لبه‌های گراف است بطوری که لبه (i, j) متناظر با اقدام a_j از آتاماتای A_i است. اگر بردار احتمال اعمال آتاماتای یادگیر

با استفاده از نمونه‌گیری است. (اگر فرض کنیم، خطی مشی کاربر تغییر نمی‌کند).

$$\forall n > M_0 \quad \|LA_i^n - \vec{P}(X)\|_1 < \delta$$

اثبات: برای اثبات این قضیه مقدار متوسط مقادیر اضافه شده (یا کاسته شده) در فرآیند یادگیری از بردار اقدام‌های هر یک از آتاماتاهای یادگیر را در آتاماتای یادگیر توزیع شده بررسی خواهیم کرد. فرض کنید LA_i متناظر با LA_i ی i از DLA ی مساله باشد و α_i^j نمایشگر z امین اقدام قابل انجام توسط آتاماتای LA_i که منجر به فعال شدن LA_j می‌شود، باشد. تعداد اقدام‌های قابل انجام توسط آتاماتای LA_i در DLA را r_i در نظر می‌گیریم. فرض می‌کنیم p_{ij}^G مقدار احتمال مربوط به توزیع احتمال یال V_{ij}^G در گراف متناظر با آتاماتای یادگیر توزیع شده مورد بحث باشد. امید ریاضی مقدار پاداش یا خطای گرفته شده توسط LA_i را محاسبه می‌کنیم. برای محاسبه این مقدار دو حالت متصور است:

۱- اقدام A_i را انتخاب کرده و این اقدام مطابق با الگوریتم ۱ پاداش گرفته است.

۲- LA_i اقدام A_j $k \neq j$ را انتخاب کرده و این اقدام مطابق با الگوریتم ۱ پاداش گرفته است (و در نتیجه اقدام A_i خطا گرفته است)

بنابر این اگر میزان احتمال این دو حالت را به همراه مقدار پاداش یا خطای گرفته شده در هر یک از این دو حالت در نظر بگیریم مقدار متوسط پاداش یا خطای داده شده در هر دور آموزش به هر یک از اقدام‌های آتاماتاهای یادگیر در آتاماتای یادگیر توزیع شده برابر خواهد بود با:

در یک حالت می‌توانیم فرض کنیم که هیچ شناختی از محرک‌های کاربر نداریم (حالتی که فرض آن معقولتر به نظر می‌رسد) بنابراین آتاماتای یادگیر توزیع شده در نهایت توزیع احتمالی انتخاب لینک‌ها را بر اساس برآیند علائق کاربران مختلف با خط مشی‌های متفاوت یاد می‌گیرد. الگوریتم پیشنهادی برای این مورد در ادامه آمده است (الگوریتم ۱).

در یک حالت می‌توانیم فرض کنیم که هیچ شناختی از محرک‌های کاربر نداریم (حالتی که فرض آن معقولتر به نظر می‌رسد) بنابراین آتاماتای یادگیر توزیع شده در نهایت توزیع احتمالی انتخاب لینک‌ها را بر اساس برآیند علائق کاربران مختلف با خط مشی‌های متفاوت یاد می‌گیرد. الگوریتم پیشنهادی برای این مورد در ادامه آمده است (الگوریتم ۱).

دلیل استفاده از آتاماتای یادگیر توزیع شده، توانایی آن در استخراج رفتار تعاملی کاربر با وب است. به عبارت دیگر، آتاماتای یادگیر باید بتواند توزیع احتمال انتخاب لینک‌های موجود در هر صفحه توسط کاربر را فرا بگیرد. این ایده بر پایه این فرض استوار است که چنانچه از آتاماتاهای یادگیر مشابه با الگوریتم‌های استفاده شده در [۳] [۴] استفاده کنیم، در این صورت، بردار احتمالات مربوط به انتخاب اقدام‌ها توسط آتاماتای یادگیر توزیع شده در بی نهایت، با توزیع احتمالی انتخاب لینک‌ها توسط کاربر برابر خواهد شد. این موضوع را در قالب قضیه ۱ بیان می‌کنیم.

قضیه ۱: فرض کنید که $\vec{P}(x)$ بردار توزیع احتمال انتخاب لینک‌های صفحه P_i توسط کاربرها باشد. در این صورت اگر LA_i^k نشان دهنده بردار احتمال مربوط به اقدام‌های آتاماتای LA_i پس از بازدید k امین کاربر و به روز شده مطابق الگوریتم ۱ باشد، اعداد صحیح $M_0 > 0$ و $\delta > 0$ به گونه‌ای

الگوریتم ۱:

- ۱- یک DLA یکریخت با ساختار گراف سایت ایجاد کن.
- ۲- برای هر آتاماتای یادگیر در آتاماتای یادگیر توزیع شده اقدامات زیر را انجام بده
 - a. مقدار احتمال مربوط به اقدام‌های هر یک از آتاماتاهای یادگیر را مطابق با توزیع یکنواخت (قدم زدن تصادفی) مقداردهی کن $(p_j^0 = \frac{1}{outdegree(P_j)})$
 - b. مقداردهی اولیه مربوط به پارامتر یادگیری را انجام بده
 - c. $k=0$
- ۳- مراحل زیر را تا زمانیکه که کاربری وجود دارد تکرار کن
 - a. با ورود هر کاربر به یک سند i و انتخاب لینک مربوط به سند j در سند i مراحل زیر را انجام بده
 - i. به اقدام A_i^j مطابق روابط زیر پاداش بده
 - ii. پارامتر یادگیری مربوط به آتاماتای یادگیر LA_i را تنظیم کن $(\alpha_i^{new} < \alpha_i^{old})$

$$p_j^{k+1} = p_j^k + \alpha_i(1 - p_j^k)$$

$$\forall t \neq j \quad p_t^{k+1} = (1 - \alpha_i)p_t^k$$

$$\alpha_i^{new} = f(\alpha_i^{old}, k)$$

$$k=k+1 \quad b$$

آن مشابه با محیط واقعی گردد. در این مقاله پروفایل علاقه کاربران بصورت توزیع قانون-توانی و توزیع محتوای اسناد بصورت توزیع نرمال در نظر گرفته شده است. سایر پارامترهای استفاده شده در این مدل برای شبیه‌سازیهای انجام شده در این قسمت در جدول ۱ نشان داده شده است.

جدول ۱: پارامترهای استفاده شده در مدل شبیه‌سازی

حد آستانه ایجاد اتصال	۰/۷
تعداد کاربران	۱۰۰۰۰
تعداد اسناد	۲۰
تعداد موضوعها	۵
T_c مقدار ثابت سند اولیه (صفحه اولیه سایت) در موضوعات مختلف	۰/۲
α_{it} پارامتر توزیع قانون-توانی توزیع احتمال علائق کاربران	۱
ϕ ضریب پاداش دریافتی از مشاهده یک سند	۱/۲
λ ضریب جذب اطلاعات از یک سند توسط یک کاربر	۰/۵
μ_m میانگین توزیع نرمال ΔM_t^v	۵/۹۷
σ_m واریانس توزیع نرمال ΔM_t^v	۰/۲۵
α_p پارامتر توزیع قانون-توانی توزیع احتمال وزنه‌های مطالب برای هر سند	۳
σ_t واریانس توزیع نرمال برای مقدار افزایش یک گره برای یک موضوع خاص	۰/۲۵
θ ضریب کاهش علاقه کاربر	۱
حداقل اشتیاق کاربر برای ادامه جستجو	۰/۲

پس از ایجاد گراف مدل وب، بر اساس نشست‌های کاربران، یک توزیع علاقه برای کاربران ایجاد شده است. (بردار $\vec{P}(X)$ در قضیه ۱). پس از این دو مرحله، کاربران شبیه‌سازی شده بر اساس توزیع علاقه به دست آمده برای کاربران و بر اساس الگوریتم DLA_WUM حرکت در وب را آغاز می‌کنند.

برای مقایسه بردارهای $\vec{P}(X)$ و بردار احتمال اقدام‌های آتاماتا‌های یادگیر در آتاماتا‌ی یادگیر توزیع شده، DLA ، از دو معیار استفاده شده است. معیار اول، مجموع تفاضل مقادیر مربوط به هریک از لینک‌ها در گراف وب و گراف DLA همریخت با آن. این معیار به عنوان شاخصی جهت تعیین میزان همگرایی بردار DLA به بردار $\vec{P}(X)$ مورد استفاده قرار گرفته است. معیار دوم، همبستگی بین بردارهای $\vec{P}(X)$ و بردار احتمال اقدام‌های آتاماتا‌های یادگیر میباشد که میزان همبستگی خطی این دو بردار را تعیین میکند. همبستگی میان دو بردار X و Y طبق رابطه زیر محاسبه میشود.

$$\rho_{X,Y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)}\sqrt{E(Y^2) - E^2(Y)}} \quad (9)$$

شبیه‌سازی برای تعداد ۲۰۰۰ کاربر صورت گرفته است.

$$P_{reward} \times V_{reward} + P_{penalty} \times V_{penalty} = \quad (4)$$

$$p_{ij}^G \times \alpha_i^t (1 - p_j^t) - \alpha_i^t \times p_j^t (1 - p_{ij}^G) = \alpha_i^t (p_{ij}^G - p_j^t)$$

رابطه ۴ را به صورت برداری می‌توان به شکل زیر نمایش داد:

$$LA_i^{t+1} = LA_i^t + \alpha_i^t (\vec{P}(x) - LA_i^t) \quad (5)$$

لم:۱: دوبردار $X_{k \times 1}$ و $W_{k \times 1}$ مفروضند. بردار $W_{k \times 1}$ طول ثابت w دارد ($|W| = w$). نیز $|X|_1 = |W|_1 = 1$ (هر دو بردار بردارهای احتمالاتی هستند). اگر طی یک فرآیند یادگیری در هر گام مطابق رابطه ۶ عمل کنیم:

$$X^{t+1} = X^t + \alpha^t (W - X^t) \quad 0 < \alpha^t < 1 \quad (6)$$

$$\lim_{t \rightarrow \infty} X^t = W$$

اثبات: با تفسیر هندسی روابط بالا، در حقیقت باید نشان دهیم، زاویه بین بردارهای W و X در حال کاهش و رسیدن به صفر است. اما از نقطه نظر جبری داریم:

$$W - X^{t+1} = W - (X^t + \alpha^t (W - X^t)) = (1 - \alpha^t)(W - X^t) \quad (7)$$

لذا داریم:

$$\|W - X^n\|_1 \leq (1 - \alpha^0)^n \|W - X^0\| \quad (8)$$

رابطه ۸ نشان میدهد که X در نهایت به سمت W میل می‌کند. اما از نقطه نظر هندسی نیز می‌توان نشان داد که زاویه بین بردارهای \vec{W} و \vec{X} در فرآیند یادگیری در حال کاهش است.

۶. نتایج شبیه‌سازی‌ها

برای انجام شبیه‌سازی‌ها بر اساس $ALG2$ عمل شده است:

ALG2: Procedure DLA_WUM

Input: Graph $W = (V, E)$, Iteration;

Output: User_Action_Probability;

Begin

Construct a DLA from W ;

Iteration=0;

Repeat

PathLen=K;

Node=Home_page;

While PathLen > 0 do

nextNode=node.getNextNode(node);

Reward($DLA[node][nextNode]$);

PathLen=PathLen-1;

node=nextNode;

end while

evaluate Convergence, Correlation;

Iteration=Iteration+1;

End DLA_WUM ;

مدل معرفی شده در [1] محیطی شامل صفحات وب و کاربران آن را فراهم می‌کند. مزیت استفاده از این مدل در این است که تشخیص کاربران و بازدیدهای انجام شده از صفحات وب با استفاده از این مدل بسیار دقیق‌تر می‌باشد و به عملیات پالایش داده‌ها نیز احتیاجی نخواهد بود. البته باید پارامترهای معرفی شده در این مدل به دقت تنظیم گردند تا نتیجه حاصل از

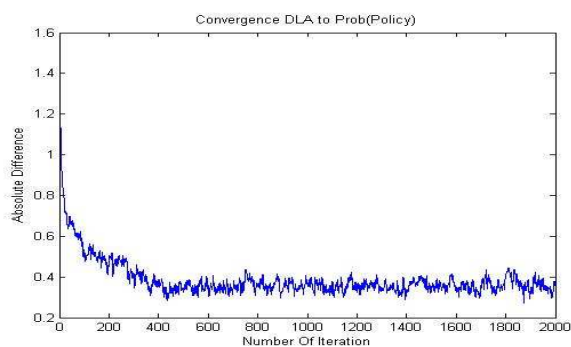
۸. مراجع

- [1] Liu, J., Zhang, S. and Yang, J., "Characterizing Web Usage Regularities with Information Foraging Agents," IEEE Transactions on Knowledge and Data Engineering, vol. 16, no. 4, April 2004, pp. 566-584
- [2] Frias-Martinez, E., Magoulas, G., Chen, S. and Macredie, R., "Recent Soft Computing Approaches to User Modeling in Adaptive Hypermedia", 2003, retrieved from <http://citeseer.ist.psu.edu/710382.html>
- [۳] ساعتی، سعید، "حوشه بندی اسناد در کتابخانه های دیجیتال با استفاده از کلونی مورچه ها و اتوماتهای توزیع شده"، پایان نامه کارشناسی ارشد، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیر کبیر، ۱۳۸۴.
- [۴] برادران هاشمی، علی، میبیدی، محمدرضا، "تعیین ساختار اسناد وب با استفاده از آتاماتای یادگیر توزیع شده"، مجموعه مقالات دوازدهمین کنفرانس انجمن کامپیوتر ایران، دانشگاه شهید بهشتی، صفحات ۵۵۳-۵۶۰، تهران، ایران، ۱۳۸۵.
- [5] Pal, S. K., Tawler, V. and Mitra, P., "Web Mining in Softcomputing Framework: Relevance, state of the Art and Future Directions", IEEE Transactions on Neural Networks, Vol. 13 5, September 2002
- [6] Beigy, H. and Meybodi, M. R., "Utilizing Distributed Learning Automata to Solve Stochastic Shortest Path Problem", International Journal of Uncertainty, Fuzziness and Knowledge-based Systems, Vol. 14, No. 5, pp. 591-615, 2006.
- [7] Zhu, J., Hong, J., and Hughes, J.G., "Mining Conceptual Link Hierarchies from Web Log Files for Adaptive Web Site Navigation," ACM Transactions on internet Technology (TOIT), in press, 2003.
- [8] Zhu, J., Hong, J., and Hughes, J.G., "Using Markov Chains for Link Prediction in Adaptive Web Sites," Proc. of Soft-Ware: First International Conference on Computing in an Imperfect World, PP.60-73, Lecture Notes in Computer Science, Springer, Belfast, April 2002.
- [9] Brin, S., and Page, L., "The Anatomy of a Large-scale Hyper Textual Web Search Engine," Computer Networks and ISDN System, 30(0 1-7):107-117, 1998.
- [10] Richardson, M. and Domingos, P., "The Intelligent Surfer: Probabilistic Combination of Link and Content Information in PageRank", Advances in Neural Information Processing Systems, 14:1441-1448, 2002.
- [11] Maseglier, F., Poncelet, P. and Cicchetti, R., "An Efficient Algorithm for Web Usage Mining", Networking and Information System Journal (NIS), Vol. 2, No 5-6, pp. 571-603, 1999.
- [12] Nanopoulos, A. and Manolopoulos, Y., "Mining Patterns from Graph traversals", Data and Knowledge Engineering, Vol. 37, No. 3, pp. 243-266, 2001.
- [13] Nygren, E., and Allard, A., "Between the Clicks Skilled Users Scanning of Pages. Designing for the Web", Empirical Studies. Human Factors and the Web/HTML Conference. Sandia National Laboratories, Albuquerque, NM. 1996.
- [۱۴] ملاخلیلی، محمدرضا، میبیدی، محمدرضا، "استفاده از آتاماتای یادگیر توزیع شده در پیش‌بینی حرکت کاربران در وب"، مجموعه مقالات سیزدهمین کنفرانس انجمن کامپیوتر ایران، اسفند ۸۶، کیش.
- [15] Beigy, H. and Meybodi M. R., "Solving Stochastic Shortest Path Problem Using Distributed Learning Automata", Proceedings of CSICC-2001, Isfahan, Iran, pp. 70-86, 2001.
- [16] Beigy, H. and Meybodi, M. R., "A New Distributed Learning Automata for Solving Stochastic Shortest Path Problem", Proceedings of the Sixth International Joint Conference on Information Science, Durham, USA, pp. 339-343, 2002.

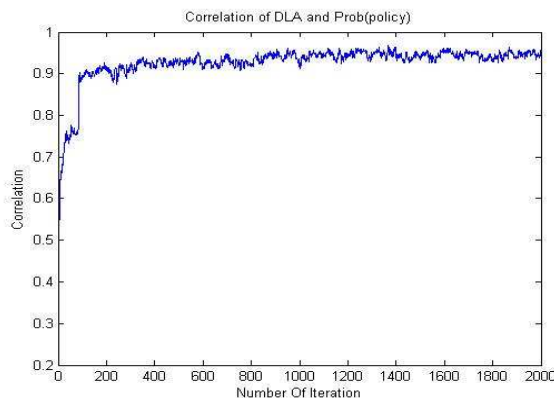
مقدار پارامتر یادگیری برای آتاماتاهای یادگیر ۰,۱ انتخاب شده است. شکل های ۴ و ۵ نتایج این دو معیار را برای حرکت ۲۰۰۰ کاربر شبیه سازی شده در وب نشان می‌دهد. همانگونه که این شکل‌ها نشان می‌دهند در طی فرآیند یادگیری آتاماتاهای یادگیر در آتاماتای یادگیر توزیع شده مطابق با الگوریتم ۱، بردار احتمال اقدام‌های آتاماتاهای یادگیر به بردار توزیع احتمال انتخاب پیوندها بر روی هر گره نزدیک می‌شود (تفاضل مقادیر کوچک تر و هم‌بستگی این دو بردار به عدد ۱ نزدیک تر می‌شود)

۷. نتیجه گیری

در این مقاله رهیافتی مبتنی بر آتاماتای یادگیر توزیع شده جهت مدلسازی کاربران در محیط‌های فرایوندی ارائه گردید. نتایج تئوری و شبیه‌سازی موید آن بود که آتاماتای یادگیر توزیع شده قادر به استخراج رفتار کاربران در وب و مدلسازی کاربر بر اساس نحوه تعاملشان با ساختار وب می‌باشد. مدل پیشنهادی نوعی دانش نهفته در ساختار وب و رفتار کاربر را استخراج می‌کند و آن را در قالب یک توزیع احتمالی بر روی صفحات وب، در دسترس می‌گذارد.



شکل ۳: معیار تفاضل مقدار مربوط به بردارهای $\vec{P}(X)$ و بردار احتمال اقدام‌های آتاماتاهای یادگیر در هر گره



شکل ۴: معیار هم‌بستگی خطی مقدار مربوط به بردارهای $\vec{P}(X)$ و بردار احتمال اقدام‌های آتاماتاهای یادگیر در هر گره