یک الگوریتم جدید برای پیشنهاد صفحات وب بر اساس قوانین انجمنی وزن دار

رعنا فرصتی

گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشکده فنی و مهندسی، دانشگاه آزاد اسلامي، کرج، ایران

forsati@kiau.ac.ir

محمد رضا میبدی

دانشکده مهندسي کامپيوتر و فناوري اطلاعات، دانشگاه صنعتي اميرکبير، تهران، ایران

mmeybodi@aut.ac.ir

چکیده:

مجموعه‌اي از عملیات که تجربه وب را براي یک کاربر خاص یا مجموعه‌اي از کاربران سازمان‌دهي مي‌کند و پیشنهادات پویا بر اساس الگوهاي رفتاري کاربران ارائه مي‌دهد، شخصی سازی وب نامیده می‌شود .اخیرا پیش‌بینی نیازهای آتی کاربران بر اساس تکنیک‌های داده‌کاوی، در زمینه شخصی‌سازی صفحات وب توجه زیادی به خود جلب کرده است. در این مقاله ما قوانین انجمنی معمول را با انتساب وزن به آیتم‌های موجود در نشت‌های کاربران برای نمایش میزان علایق آنها به آیتم‌ها توسعه داده و قوانین انجمنی جدیدی بر اساس آیتم‌های وزن‌دار تولید کرده و آن را قوانین انجمنی وزن دار نامیدیم. پس از معرفي معیار وزن‌دهي و ایجاد قوانین انجمنی وزن‌دار جدید، الگوریتمي که با استفاده از اطلاعات پیمایش کاربران و قوانین انجمنی وزن‌دار تولید شده، به پیشنهاد صفحات به کاربران می‌پردازد، ارائه شده است. معیار معرفي شده براي محاسبه وزن صفحات مشاهده شده توسط کاربران ازدو پارامتر "مدت زمان مشاهده صفحه" و "فرکانس مشاهده صفحه" استفاده مي‌کند که بر خلاف وزن باینری که در روش‌های قبلی استفاده شده است به خوبي میزان اهمیت و علاقه کاربران به آن صفحه را نشان مي‌دهد. الگوریتم ارائه شده مشکل کاهش دقت الگوریتمها با افزایش تعداد صفحات پیشنهادي را در حد قابل توجهی بهبود می‌دهد و به طور قابل ملاحظه اي کیفیت پیشنهادات را بهبود می‌بخشد. نتایج شبیه‌سازي الگوریتم در داده‌هاي واقعي کارایي بالای الگوریتم پیشنهادي را نشان می‌دهد.

واژه هاي کلیدي: شخصی سازی صفحات وب، قوانین انجمنی

۱- مقدمه

وب طي یک فرآیند آشفته و غیر متمرکز در حال رشد است و این روند منجر به تولید حجم وسیعي از مستندات متصل به یکدیگر گشته است که از هیچ گونه سازماندهي منطقي برخوردار نیستند. در حال حاضر موتور جستجويGoogleبیش از 3 بیلیون صفحه وب را شاخص گذاري کرده است که این تعداد با نرخ 7,3 میلیون صفحه در روز افزایش مي یابد. در واقع وب به مجموعه بزرگي از داده هاي ساخت یافته و نیمه ساخت یافته تبدیل شده است که کاربران آن از همپوشاني داده‌ها رنج مي‌برند. برای حل این مشکل، شخصي‌سازي[1] صفحات وب به یک پدیده محبوب به منظور سفارشی کردن محیط‌های وب تبدیل شده است. بر اساس این تکنولوژي سرویس‌هاي موردنیاز کاربران بر اساس علائق و ویژگي‌هاي کاربران، بدون آنکه کاربران به طور صریح آنها را بیان کنند، شخصي‌سازي مي‌شوند [۱]. شخصي‌سازي وب مجموعه‌اي از عملیات است که تجربه وب را براي یک کاربر خاص یا مجموعه اي از کاربران سازمان‌دهي مي‌کند و پیشنهادات پویا بر اساس الگوهاي رفتاري کاربران ارائه مي‌دهد. شخصي‌سازي وب مي‌تواند به عنوان فرآیندي براي سفارشي‌کردن محتوا و ساختار وب سایت بر حسب نیازهاي شخصي و ویژه‌ء هر کاربر باشد[۲].سایت از طریق پررنگ کردن لینکهاي موجود، ایجاد پویاي لینک‌هاي جدید که مورد نظر کاربر جاري باشد یا حتي ایجاد صفحات شاخص جدید شخصی مي‌شود.

کارهاي تحقیقاتي بسیاري در این حوزه انجام شده است که عمدتا بر مبناي تکنیك هاي وب کاوی[2] می باشند. یکي از انواع وب کاوي، کاوش کاربرد وب[3] است که با بررسي رفتارهاي کاربران در وب، به کشف دانش نهفته در نحوه تعامل کاربران با وب می پردازد و بر اساس اطلاعات موجود از رفتار کاربر (با استفاده از فایلهاي ثبت وقایع[4] در سرویس‌دهنده‌هاي وب یا برنامه‌هاي در سمت کاربر) در تعامل با وب به استخراج دانش موردنیاز پرداخته و الگوهای رفتاری استخراج شده را در شخصي‌سازي صفحات وب و پیشنهاد صفحات بکار می برد[۳-۸].

در کاوش استفاده از وب از تکنیک های عمده داده کاوی[5] از جمله کاوش قوانین انجمنی، استخراج الگوهای متوالی و خوشه بندی برای استخراج الگوهای هدایتی کاربران و ارئه پیشنهادات بر اساس آنها استفاده می کنند. قوانین انجمني به طور موفقیت آمیزي در سیستم‌هاي پیشنهاددهنده وب استفاده شده‌اند و نتایج تجربي خوبي براي قوانین انجمني به عنوان سیستم پیشنهاد دهنده صفحات وب گزارش شده است[۹].

---

[1] Personalization

[2] Web Mining

[3] Web Usage Mining

[4] Log files

[5] Data Mining

روش هایی که تاکنون برای پیشنهاد صفحات بر اساس قوانین انجمنی گزارش شده است برای کلیه آیتم های(صفحات) موجود در تراکنش ها بدون توجه و محاسبه وزن آنها در تراکنش‌ها ارزش و اهمیت یکساني در نظر گرفته‌اند و به وزن صفحات مشاهده شده در تراکنش ها در تولید قوانین انجمنی توجهی نکرده اند.

کاوش قوانین انجمنی وزن‌دار این امکان را می‌دهد که آیتم‌های موجود در تراکنش‌ها وزن‌های متفاوتی داشته باشند که این شیوه منجر به بهبود مدل قوانین انجمنی در زمینه شخصی‌سازی صفحات وب می‌شود. در[۱۰-۱۱] روشهايي براي انتساب وزن به آيتم‌هاي موجود در تراکنش‌ها برای نمایش میزان اهمیت آنها ارائه شده است که از مجموع وزن‌ها و یا میانگین وزن‌ها برای محاسبه وزن آیتم‌ها استفاده شده است و ضریب پشتیبانی وزن‌دار آیتم‌ها را حاصلضرب ضریب پشتیبانی آیتم ها در وزن آنها تعریف کرده اند. هر دو روش پیشنهادی وزن ثابتی برای هر آیتم در نظر می گیرند در حالیکه در زمینه سیستم‌های پیشنهاد دهنده صفحات وب، هر صفحه در نشست های متفاوت ممکن است وزن‌های متفاوتی داشته داشته باشد و میزان اهمیت هر صفحه در نشسته های متفاوت برای کاربران متفاوت است.

هیچ یک از کارهای انجام شده در زمینه شخصی‌سازی صفحات وب بر اساس قوانین انجمنی از مدت زمان مشاهده صفحه توسط کاربران برای تعیین میزان اهمیت صفحات برای کاربران استفاده نکرده‌اند. به عبارت دیگر آنها برای همه صفحاتی که توسط کاربران در لاگ‌های ثبت وقایع مورد دسترسی قرار گرفته‌اند ارزش و اهمیت یکسانی در نظر گرفته‌اند و اختلاف اهمیت صفحات برای کاربران را در نظر نگرفته و با همه آنها یکسان برخورد می‌کنند.

مقالاتي كه از قوانين انجمني براي كشف دانش رفتار پيمايشي كاربران و پيشنهاد صفحات استفاده مي كنند[۱۴] [۱۲][۹]، نشست جاری کاربر را با بخش مقدم مجموعه قوانين انجمني استخراج شده تطابق دقيق صفحه به صفحه می‌کنند تا مجموعه اي از صفحات را كه تا به حال كاربر مشاهده نكرده است به او پيشنهاد دهند. ولي هيچ كدام از آنها از وزن صفحات براي رتبه بندي صفحات كانديد استفاده نكرده‌اند. استفاده از این روش و تنها با اتکا به ضریب اطمینان به عنوان میزان تشابه باینری نشست جاری کاربر و قانون انجمنی، باعث می‌شود که کیفیت پیشنهاد پایین باشد، زیرا ممکن است دو قانون ضریب اطمینان برابری داشته باشند، اما کاربر قانون اول زمان بیشتری را در برخی صفحات خاص صرف کرده باشد که تطابق بیشتری با نشست کاربر جاری داشته باشد که در این موارد سیستم‌های موجود به صورت تصادفی عمل می‌کنند.

در این مقاله، ما قوانین انجمنی معمول را با انتساب وزن به آیتم های موجود در تراکنش‌ها برای تعیین میزان اهمیت هر آیتم در تراکنش توسعه داده و قوانین انجمنی وزن‌دار جدیدی معرفی کرده‌ایم. سپس الگوریتم جدیدی برای پیشنهاد صفحات وب بر اساس قوانین انجمنی وزن‌دار معرفی شده پیشنهاد داده‌ایم. در قوانین انجمنی وزن‌دار پیشنهاد شده از "*مدت زمان مشاهده صفحه*" توسط کاربر و "*تعداد دفعات*" مشاهده صفحه برای تعیین وزن صفحه به جای وزن باینری استفاده شده در الگوریتم‌های موجود استفاده کرده‌ایم. در الگوریتم پیشنهاد شده، ابتدا قوانین انجمنی وزن‌دار معرفی شده را از فایل‌های ثبت وقایع کاربران استخراج کرده و سپس میزان شباهت نشست جاری کاربر با هر یک از قوانین انجمنی وزن دار تولیده شده را به جای تطابق باینری محاسبه می‌کنیم. در نهایت موتور پیشنهادگر بر اساس دو پارامتر ضریب اطمینان وزن‌دار قوانین و نرخ شباهت نشست کاربر جاری با قوانین، صفحات موجود در قوانین با بالاترین امتیاز را یافته و به کاربر پیشنهاد می دهیم. همچنین الگوریتم جدیدی برای پالایش نشست کاربر جاری ارائه کردیم که صفحاتی را که کاربر در نشست به آنها علاقه‌ای نداشته و برای وی بی‌اهمیت بوده‌اند را فیلتر می‌کند. الگوریتم ارائه شده مشکل کاهش دقت الگوریتمها با افزایش تعداد صفحات پیشنهادي را در حد قابل توجهی بهبود می دهد.

الگوریتم ارائه شده بر روی داده‌های واقعی شبیه‌سازی شده و نتایج نشان مي‌دهد که دانش بدست آمده از سیستم مذکور به طور قابل ملاحظه اي کیفیت پیشنهادات را بهبود داده است. در ادامه در بخش ۲ معیار وزن دهی و قوانین انجمنی وزن دار جدید پیشنهاد شده و تعاریف مرتبط آنها ارائه مي‌گردد. در بخش ۳ الگوریتم معرفی شده برای پیشنهاد صفحات وب به کاربران آمده است. در بخش۴ پس از معرفي مدل استفاده شده براي شبيه‌سازي، نتايج شبيه‌سازي ارائه مي‌شود. در پایان نیز نتیجه‌گیری آورده شده است.

۲- استخراج قوانين انجمني وزن دار جديد از فايلهاي ثبت دسترسي كاربران

در اين بخش پس از معرفي معيار وزن دهي، قوانين انجمني وزن دار جديدی، كه از اطلاعات پيمايش كاربران براي توليد قوانين وزن دار استفاده می کند معرفی می شود. در اين مدل صفحات بر اساس معيار جديدي كه به خوبي ميزان علاقه كاربران و اهميت آن صفحه را نشان مي‌دهد، وزن‌دهي مي‌شوند. معيار معرفي شده براي محاسبه وزن صفحات مشاهده شده توسط كاربران از "مدت زمان مشاهده صفحه" و فركانس مشاهده صفحه" استفاده مي‌كند كه به خوبي ميزان اهميت و علاقه كاربران به آن صفحه را نشان مي‌دهد. مدل قوانين انجمني وزن دار توليد شده به همراه پارامترهای مرتبط آن در ادامه به تفصيل بررسي شده است.

۲-۱ تعيين وزن صفحات

در اين بخش، روشي براي وزن‌دهي صفحات در نشست‌هاي كاربران ارائه مي‌كنيم. مراحل پيش‌پردازش فايلهاي ثبت دسترسي كاربران در نهايت منجر به مجموعه اي از m صفحه $P=\{p_1,p_2,...,p_m\}$ ، يک

مجموعه از n تـراکنش کـاربر $T=\{t_1,t_2,...,t_n\}$ کـه هـر $t_i \in T$ یک زیرمجموعه از P مي‌باشد، مي‌شوند. هر تراکنش یک دنباله به طـول m از زوج هـــــــــــاي مرتـــــــــــب اســـــــــــت: $t=<(p_1^t,w(p_1^t)),(p_2^t,w(p_2^t)),...,(p_m^t,w(p_m^t))>$ کـــــه در آن $p_i^t = p_j$ بـراي یـک $j \in \{1,...,m\}$ و $w(p_i^t)$ وزن صـفحه‌ي p_i^t در تراکنش t است. وزن $w(p_i^t)$ در تـراکنش t_i لازم اسـت کـه بطور مناسبي ميزان علاقه کاربر به صفحه iام را نشان دهد.

روش هاي پيشنهاد صفحات که قبلا ارائـه شـده انـد تفـاوت اهميـت صفحات مختلف براي کاربران و تفاوت درجه علاقه کاربران به صـفحات را در نشست‌ها در نظر نمي‌گرفتنـد و بـا تمـام صـفحات مـشاهده شـده توسط کاربر کاملا يکسان رفتار مي‌کردند. در صـورتيکه احتمـال اينکـه تمام صفحات مشاهده شده توسط کاربر به يـک انـدازه بـراي او جـذاب باشد بسيار پايين است. در زمانی که کاربر بازديد کننده از سايت مسير خود ر ا در سايت گم کند نيز صـفحات بازديـد شـده توسـط کـاربر در فايلهاي دسترسي کاربران ثبت می‌شود، در حاليکه ايـن پديـده ممکـن است موجب گم راه کردن سيستم در توليد پيشنهادهاي آتي شـود. بـه عبارت ديگر، کاربر به دنبال يافتن اطلاعات موردنياز خود ممکـن اسـت صفحاتي را مشاهده کند که هيچ ارزشي براي او نداشته باشند کـه ايـن مسئله موجب مي شود صفحات نامربوط با نيـاز و علائـق کـاربر نيـز در فايل ثبت دسترسي کاربر ذخيره شود. بنـابراين در نظـر گـرفتن ارزش يکسان براي همه صفحاتي که کاربر به آنهـا مراجعـه کـرده در نمـايش ميزان علاقه کاربر به صفحات و پيش بيني رفتار کـاربران آتـي بـر ايـن اساس معايب بزرگي خواهد داشت. از طرفـي از آنجاييکـه در روش‌هـاي شخصي‌سازي بر اساس داده‌هاي کاربردي از کاربران انتظار نداريم علاقه ، عدم علاقه و يا ميزان علاقه خود به صفحات را صراحتا بيان کنند، ارائه معياري که درجـه علاقـه کـاربران بـه صـفحات را بـر اسـاس داده‌هـاي کاربردي و رفتار کاربران قبلي به خوبي برآورد کند يکـي از بخـش‌هـاي مهم در فرايند پيشنهاد صفحات مي‌باشد.

در اين مقاله برخلاف اکثر مقالات که از روش بـاينري بـه عنـوان وزن صفحه استفاده مي‌کنند معيـار جديـدي کـه بـر اسـاس فايلهـاي ثبـت دسترسي کاربران ميزان علاقه آنها به صفحات و اهميت صـفحات بـراي کاربران را به خوبي برآورد مي‌کند معرفي مي شود. در معيـار مـذکور از دو پارامتر "مدت زمان مشاهده صفحه" و "فرکانس صفحه" بـراي تعيـين اهميت هر صفحه در تراکنش هاي کاربران استفاده مي شود. معيار وزن دهي بر خلاف وزن باينري که در کارهاي قبلي استفاده شـده اسـت بـه طور دقيق تري علاقه کاربر به صفحات متفاوت را نمايش مـي دهـد. در اينجا ما سعي کرده ايم به صفحاتي که براي کاربر جالـب تـر بـوده انـد توجه بيشتري داشته باشيم تا اطلاعات موردنياز کاربر را بهتر تـشخيص داده و صفحات مفيدتري به او پيشنهاد دهيم.

مشاهدات زير استفاده از پارامتر" مدت زمان مشاهده صفحه" را در غالب يکي از پارامترهاي وزن دهي تائيد مي کنند:

الف: اين معيار اهميت نسبي هر صفحه براي کاربر را به خـوبي نمـايش مي دهد. زيرا کاربران معمولا مـدت زمـان بيـشتري را صـرف مـشاهده صفحاتي که نياز آنها را بيشتر براورده مـي کنـد و منطبـق بـا اطلاعـات موردنياز آنهاست و براي آنها جـذاب تـر اسـت مـي‌کننـد و زمـان قابـل ملاحظـه اي را بـه آن صـفحات اختـصاص مـي‌دهنـد. در حاليکـه اگـر صفحه‌اي براي کاربر جالب نباشد، زمان زيادي را براي مشاهده آن صرف نکـرده و صـفحه را رد کـرده و سـريعا بـه صـفحه ديگـري مراجعـه مي‌کند[۱۵]. البته بايد در نظر داشت که اگر طول صفحه وبي کم باشد، زمان مشاهده آن صفحه نيـز متناسـبا کمتـر مـي‌شـود. بنـابراين زمـان مشاهد هر صفحه متناسب با اندازه صفحه اسـت و در محاسـبه اهميـت صفحه اين تناسب بايد در نظر گرفته شود. به اين منظور ما مدت زمـان مشاهده هر صفحه را بر اساس طول آن نرمـال مـي‌کنـيم. رابطـه مـدت زمان مشاهده صفحه بر اساس طول صفحه در رابطه(۱) نشان داده شده است.

$$d_p(P)=\frac{\frac{Duration(P)}{Size(P)}}{\max_{Q\in T}(\frac{Duration(P)}{Size(P)}} \quad (۱)$$

ب: معمولا نسبت بدست آوردن اطلاعـات مفيـد از صـفحات وب بـراي کاربران مختلف تفاوت چنداني با يکديگر ندارند[۱۵] . بنابراين با فـرض آنکه نرخ بدست آوردن اطلاعات از هـر صـفحه بـراي کـاربران مختلـف مشابه است، مدت زمان سپري شده در هـر صـفحه متناسـب بـا ميـزان اهميت اطلاعات آن صفحه براي کاربران مي باشد.

ج: از آنجاييکه مدت زمان مشاهده هر صفحه براحتـي از روي فايلهـاي ثبت دسترسي کاربران به صفحات وب سايت قابـل محاسـبه اسـت، لـذا انتخاب مناسبي براي استنباط علاقه کاربران مي باشد.

در ادامـه بـه بررسـي دلايـل اسـتفاده از پـارامتر "فرکـانس صـفحه" در محاسبه وزن هر صفحه مي پردازيم. فرکانس هـر صـفحه بيـانگر تعـداد دفعات مراجعه به يك صفحه در يك تراکنش مي باشد. در يك نشـست امکان دارد کاربر به يك صفحه چند بار مراجعه کند که هر چقدر تعـداد اين ارجاع ها به يك صفحه در يك نشست بيـشتر باشـد آن صـفحه در نشست مذکور نسبت به ساير صـفحات نشـست مهمتـر اسـت و صـفحه مذکور براي کاربر جالب تر بوده است و نيازهاي کـاربر را بهتـر بـراورده کرده است. پارامتري که در تعيين تعداد دفعات مراجعه به يـك صـفحه تاثير بسزايي دارد درجه ورودي هر صفحه مي باشـد. تعـداد پيونـدهاي متمايزي که به يک صفحه اشـاره مـي‌کننـد، درجـه ورودي آن صـفحه ناميده مي شود. اين مسئله که صفحه اي که درجه ورودي بالاتري دارد ذاتا پتانسيل بيشتري براي مشاهده را دارد کاملا بديهي است. به عبارت ديگر احتمال مشاهده صفحه اي با درجه ورودي بالاتر بيشتر از احتمـال

مشاهده صفحه اي با درجه كمتـر مـي باشـد. همچنـين در مقايـسه دو صفحه‌اي كه تعداد دفعات يكـساني در يـک نشـست ملاقـات شـده‌انـد، صفحه‌اي كه به آن از تعداد صفحات كمتري پيونـد وجـود دارد (درجـه ورودي كوچكتر)، مهمتر است زيرا كه اين صفحه احتمال ملاقات بـالقوه پايين‌تري دارد. رابطه (۲) ارتباط فركانس مشاهده با درجه ورودي را بـه خوبي نمايش داده شده است.

$$f_p(P) = \frac{Visit(P)}{\sum_{Q \in T} Visit(Q)} * \frac{1}{In\deg ree(p)} \qquad (۲)$$

در اين بخش ما بر اساس دو معيار معرفي شده در بالا به هر صفحه در تراكنش مطابق رابطه(۳) وزني را نسبت مي‌دهيم. اهميت كل صـفحه از تركيب دو معيار فـوق بـه دسـت مـي‌آيـد كـه در رابطـه(۳)نـشان داده شده‌است. در اين رابطه α براي تنظيم اهميت پارامترها استفاده شـده است. ما نسبت اهميت "فركانس صفحه" به "مدت زمان مشاهده صفحه" را برابـر در نظـر گرفتـه و از ميـانگين هارمونيـك ايـن دو معيـار بـراي وزن‌دهي صفحات استفاده مي‌كنيم و α برابر دو قرار داده شده است.

$$W(p) = \frac{\alpha * f_p(P) * d_p(P)}{f_p(P) + d_p(P)} \qquad (۳)$$

در نهايت ما هـر تـراكنش را بـصورت بـردار m بعـدي از زوج مرتـب صـــــفحه و وزن متنـــــاظر آن صـــــفحه بـــــه شـــــكل $t = <(p_1, w_1), (p_2, w_2), ..., (p_m, w_m)>$ مدل مـي كنـيم. m تعداد صفحات مشاهده شده در ميان كل نشست هاي كاربران مي باشد.

۲-۲ قوانين انجمني

در مورد تـراكنش‌هـاي وب، قـوانين انجمنـي ارتباطـات بـين مـشاهده صفحه‌ها بر مبناي الگوهاي گردشي كاربران را نشان مي‌دهنـد. قـوانين انجمني در كاوش استفاده از وب، بر اي پيدا كردن ارتباط بـين صـفحا تي كه مكررا در كنار يكديگر در نشستهاي كاربر ظاهر ميشوند، به كـار ميروند.

يک قاعده‌ي انجمنـي r يـک عبـارت بـصورت $X \Rightarrow Y(S_r, C_r)$ است كه X و Y مجموعه اقـلام مـي‌باشـند و $S_r = S(X \cup Y)$ ضريب پشتيباني[6] $X \cup Y$ است كه احتمال آنكه X و Y با هم در يک تراكنش رخ دهند را نشان مي‌دهـد. ضـريب اطمينـان[7] C_r بـراي قـانون r ، از رابطـه‌ي $C_r = \frac{S(X \cup y)}{S(X)}$ داده مـي‌شـود و احتمـال شرطي اين‌كه Y در يک تراكنش رخ دهد به شـرط آنكـه X در آن رخ داده باشد را نشان مي‌دهد. ضريب پشتيباني اهميت قوانين استخراج شده و ضريب اطمينان قطعيت قوانين را نشان مي دهد.

كاوش قوانين انجمني شامل استخراج كليه قوانيني مي باشد كه ضريب اطمينان و ضريب پشتيباني آنها حد يک آستانه‌ي تعيـين شـده توسـط كاربر را ارضا مي‌كنند. بيشتر رويكردهاي كشف قواعد انجمني بر مبنـاي الگوريتم Apriori مي‌باشند [۱۶] كه يک استراتژي توليد و آزمايش را بكار مي‌برد. الگوريتم Apriori [۱۶] طـي دو فـاز قـوانين را اسـتخراج مي‌كند. در فاز اول آيتم‌هاي تكرار شونده توليد مي‌شوند. اين الگوريتم، گروه‌هاي اقلام (مشاهده صفحات ظاهر شده در ثبت پيش‌پردازش شده) را كه با يكديگر بطور مكرر در تراكنش‌هاي زيادي ظاهر شده‌اند (و يـک آستانه‌ي پشتيباني تعيين شده توسط كاربر را ارضا مي‌كنند) پيدا مـي-كند. چنين گروه هايي از اقلام به مجموعه اقلام مكرر معروفند.

فرض كنيد تراكنش T و مجموعه اقلام مكرر $I = \{I_1, I_2, ..., I_k\}$ داده شده است. پشتيباني يـک مجموعـه اقـلام $I_i \in I$ بـصورت زيـر تعريف مي‌شود:

$$S(I_i) = \frac{|\{t \in T : I_i \subseteq t\}|}{|T|} \qquad (۴)$$

يک ويژگي مهم پـشتيباني بـه نـام ويژگـي بـستاري رو بـه پـايين[8] در الگوريتم Apriori [۱۶] به اين شرح اسـت: اگـر يـک مجموعـه اقـلام، معيـار حـداقل پـشتيباني را بـرآورده نـسازد آن‌گـاه هـيچ يـک از ابرمجموعه‌هاي آن نيز اين معيار را بـرآورده نخواهنـد كـرد. يـا بعبـارت ديگر اگر آيتم ستي با اندازه K يك آيتم‌ست تكرار شونده باشد بنـابراين همه آيتم‌ست‌هاي زير اندازه (K-1) نيز بايد آيتم هاي تكرار شونده بوده باشند. اين ويژگي براي هرس كردن فـضاي حالـت در حـين هـر تكـرار الگوريتم Apriori اساسي است. در فاز دوم اين الگوريتم همـه قـوانين انجمني كه مقدار آستانه‌ي اطمينان را ارضا مي‌كنند، از روي مجموعـه اقلام مكرر توليد مي شوند.

در زمنيه وب هر صفحه وب را در غالب يك آيتم و مجموعه صفحاتي را كه توسط يك كاربر در مدت زمان كوتاهي مورد دسترسي قـرار گرفتـه اند در غالب يك تراكنش مدل مي‌كنيم. روش‌هاي پيشنهاد صفحات كـه بر پايه قوانين انجمني تا بحال ارائه شده‌اند تنها به حضور يا عدم حضور صفحات در تراكنش‌هاي كاربران دقت كرده‌اند و فـرض كـرده‌انـد تمـام صفحات ظاهر شده در تراكنش ها اهميت يكساني براي كاربر داشته اند و اهميت وزن آنها را در تراكنش ها ناديده گرفته اند و با همـه صـفحات ظاهر شده در تراكنش ها رفتار يكساني داشته اند [۱-۳] . اما از آنجا كه همه صفحاتي كه كاربر در يك نشست به آنها مراجعه كرده است براي او اهميت يكساني ندارند در ادامه بـه توصـيف قـوانين انجمنـي وزن دار و پارامترهاي مرتبط با آن مي پردازيم.

۳-۲ قوانين انجمني وزن دار جديد

[6] Support

[7] Confidence

[8] downward closure property

در این بخش، ما مدل قوانین انجمني رايج را با در نظر گرفتن وزن متناظر هر آيتم در تراكنش، كه بيانگر ميزان علاقه كاربر به آن آيتم در آن تراكنش مي باشد توسعه مي‌دهيم. در اين بخش با در نظر گرفتن پارامتر وزن در كنار هر آيتم در حين توليد قوانين انجمني، مجموعه اي از قوانين انجمني وزن‌دار توليد خواهيم كرد. كاوش قوانين انجمني وزن‌دار به ما اجازه مي‌دهد وزن هاي متفاوتي به آيتم‌هاي موجود در تراكنش‌ها نسبت دهيم كه روش مناسبي براي بهبود فرآيند شخصي‌سازي بر اساس قوانين انجمني مي‌باشد. در اين مدل، وزن‌هاي بزرگتر بيانگر آيتم‌هاي مهمتري مي‌باشند كه اين مساله استخراج آيتم‌هاي مهمتر با تعداد تكرار كمتر را فراهم مي‌كند. قوانين انجمني معمول وزن ثابتي براي همه آيتم ها در نظر مي‌گيرند در حاليكه در زمينه شخصي‌سازي بر اساس كاوش داده هاي كاربردي وب، احتمال آنكه ارزش يك صفحه در نشست هاي مختلف متغير باشد بسيار زياد است. در ادامه به تفصيل و توضيح قوانين انجمني وزن دار و تعريف پارامترهاي مرتبط با آن مي پردازيم. ما الگوريتم Apriori را با بسط آيتم هاي آن به آيتم هاي وزن دار و تعاريف مرتبط با آن توسعه داده و صحت آن را ثابت كرده ايم. در بخش نهايي از اين مدل براي پيشنهاد صفحات به كاربران استفاده خواهيم كرد.

۲-۳-۱ پارامترهاي وزن دار

ابتدا تراكنش كاربر را به صورت بردار m بعدي از زوج مرتب صفحه و وزن متناظر آن مدل مي كنيم. $t =< (p_1, w_1), (p_2, w_2), \ldots, (p_m, w_m) >$ كه $p_i \in P$ و w_i وزن متناظر با صفحه p_i مي باشد. سپس پارامترهاي مربوط به آن را تعريف مي كنيم.

تعريف ۱: قلم[9] وزن دار

وزن متناظر با هر قلم ($w(p_i)$) ميزان اهميت آن قلم را در آن تراكنش نشان مي دهد. در اين مدل وزن هر قلم همان وزن متناظر با هر صفحه مي باشد $w(p_i) = Weight(p_i)$ كه مطابق رابطه(۳)محاسبه شده است.

تعريف ۲: وزن اقلام[10] در هر تراكنش

وزن اقلام بر اساس وزن اعضايي(صفحاتي) $w(p_i)$ كه شامل آنها است تعيين مي گردد و با $w(X,t)$ نمايش داده مي شود. ساده ترين راه براي بدست آوردن وزن اقلام در نظر گرفتن مينيمم وزن عضو هايي كه شامل آنهاست مي باشد كه در رابطه (۵) نشان داده شده است.

$$\mathbf{w(X,t)} = \begin{cases} \min(w(p_1, p_2, \ldots, p_k) & X \subseteq t \\ 0 & X \not\subset t \end{cases} \qquad (۵)$$

K تعداد عضو هاي اقلام مي باشد. به طور مشابه مي توان ميانگين وزن اعضاي يك آيتم ست را نيز به عنوان وزن آن در نظر گرفت. آزمايشات ما نشان دادند كه مينيمم وزن نتايج بهتري را در مقايسه با ميانگين وزنها ارائه مي‌دهد.

تعريف ۳: وزن تراكنش

با اختصاص وزن به اقلام، ما مي‌توانيم به هر تراكنش نيز وزني نسبت دهيم. نسبت دادن وزن به هر تراكنش به ما اجازه مي دهد تفاوت تراكنش هاي مختلف را بهتر تشخيص بدهيم. بدين ترتيب تراكنشي كه وزن و ارزش بالاتري دارد در كاوش نتايج حاصله نقش بيشتري ايفا كرده است. ساده ترين راه براي محاسبه وزن هر تراكنش، بدست آوردن ميانگين وزن اقلامي مي باشد كه تراكنش شامل آنها بوده است. بدين ترتيب وزن هر تراكنش مطابق رابطه(۶)محاسبه مي شود.

$$\mathbf{w(t_k)} = \frac{\sum_{i=1}^{|t_k|} w(p_i)}{|t_k|} \qquad (۶)$$

تعريف ۴ : ضريب پشتيباني وزن دار در ميان همه تراكنش ها

ما ضريب پشتيباني قوانين انجمني را مطابق روبرو تغيير داده ايم. ضريب پشتيباني وزن دار آيتم ستX در ميان همه تراكنش ها مطابق رابطه زير تعريف مي شود.

$$wsp(X) = \frac{\sum_{t_i \in T} w(t_i) * w(X, t_i)}{\overline{w} * \sum_{k=1}^{|T|} w(t_k)} \qquad (۷)$$

كه در آن $\overline{w}$ ميانگين وزن همه آيتم ها در كل تراكنش ها مي باشد و T نيز مجموعه اي از همه تراكنش ها مي باشد.

تعريف ۵ : اقلام مكرر وزن دار

مسئله كشف الگوهاي تكرار شونده در كاوش قوانين انجمني معمول، يافتن مجموعه كاملي از اقلام مي باشد كه حد آستانه مينيمم ضريب پشتيباني را در پايگاه داده داشته باشند. در مدل معرفي شده در اين بخش، گروه هايي از اقلام را مجموعه اقلام مكرر مي ناميم كه آستانه‌ي ضريب پشتيباني وزن دار تعيين شده توسط كاربر را ارضا مي‌كنند. در مدل ارائه شده نيز كاوش اقلام تكرار شونده بر اساس الگوريتم Apriori انجام مي شود. براي هرس الگوهاي غير تكرار شونده يا الگوهايي با مرتبه تكرار پايين از ويژگي بستاري رو به پايين در الگوريتم Apriori استفاده مي كنيم. ويژگي بستاري رو به پايين در الگوريتم Apriori به اين شرح است كه هر زير مجموعه اي از اقلام مكرر خود به تنهايي نيز جز اقلام مكرر خواهد بود. به عبارت ديگر اگر يک مجموعه اقلام غير تكرار شونده باشد، معيار حداقل پشتيباني را برآورده نسازد، آن‌گاه هيچ يک از ابرمجموعه‌هاي آن نيز اين معيار را برآورده نخواهند كرد.

از اين ويژگي براي هرس كردن فضاي حالت در حين هر تكراراستفاده مي شود. در مدلي كه در اين بخش ارائه داديم و با توجه به تعاريف ضريب پشتيباني وزن دار و اقلام مكرر وزن دار ، مفهوم جديد ويژگي "بستاري رو به پايين وزن دار "را نيز معرفي مي كنيم. اين ويژگي

---

[9] Item

[10] Itemset

همانند ویژگي بستاري رو به پایین در قوانین انجمني معمول بیـانگر آن است که هر زیر مجموعه اي از اقلام مکرر وزن دار نیز اقـلام مکـرر وزن دار خواهد بود. قضیه بعدي اثبات این ویژگي را براي اقلام مکرر وزن‌دار نشان مي‌دهد.

قضیه ۱: شماي وزن دهي پیشنهاد شده ویژگي بستاري رو بـه پـایین را برآورده مي کند و براي اقلام مکررکاندید همه زیر مجموعه هاي آن نیز کاندید خواهند بود.

اثبات: I_1 و I_2 مجموعه اي از اقلام در یـک مجموعـه از تـراکنش مـي باشند به طوریکه $I_1 \subset I_2$ مي باشد به عبارت دیگر I_2 بـالا مجموعـه I_1 است. براي اثبات ویژگي "بستاري رو به پایین" در الگوریتم ارائه شده فرض مي کنیم I_1 عضو مجموعه اقلام مکرر نیست در حالیکه I_2 عضو مجموعه اقلام مکرر مي باشـد. T_1 را مجموعـه اي از تـراکنش هـا کـه شامل I_1 هستند تعریف کـرده و مـشابه آن T_2 رانیـز مجموعـه اي از تراکنش هایي که شامل I_2 هستند قرار مي دهیم. از انجا کـه I_2 بـالا مجموعـه I_1 اسـت بنـابراین $T_2 \subset T_1$ و $\sum_{t\in T_1} w(t) \geq \sum_{t\in T_2} w(t)$.

مطابق تعریفي که بـراي ضـریب پـشتیباني وزن دار اقـلام ارئـه دادیـم، ضـــــریب پـــــشتیباني وزن دار I_1 و I_2 بـــــه ترتیـــــب $wsp(I_1)=\frac{\sum_{t\in T_1} w(t)* w(I_1,t)}{\overline{w}*\sum_{t\in T_1} w(t)}$ و $wsp(I_2)=\frac{\sum_{t\in T_2} w(t)* w(I_1,t)}{\overline{w}*\sum_{t\in T_2} w(t)}$ است. بـا مقایسه $wsp(I_1)$ و $wsp(I_2)$ و در نظر گرفتن این حقیقت کـه $\sum_{t\in T_1} w(t) \geq \sum_{t\in T_2} w(t)$ بـه $wsp(I_1) \geq wsp(I_2)$ مـي رسـیم. از آنجا که I_1 عضو مجموعه اقلام مکرر نیست ضـریب پـشتیباني وزن دار آن از حد آستانه مینیمم ضریب پـشتیباني کمتـر اسـت و بـا توجـه بـه $wsp(I_1) \geq wsp(I_2)$ ، $wsp(I_2)$ نیز از حد آستانه مینـیمم ضریب پشتیباني کمتر خواهد بود بنابراین I_2 نیز عضو مجموعـه اقـلام مکرر نخواهد بود. بنابراین براي اقلام مکررکاندید همه زیر مجموعه هاي آن نیز عضو مجموعه اقلام مکرر خواهند بود. این مساله صـحت ویژگـي "بستاري رو به پایین وزن دار" را در الگوریتم ارائه شده اثبات مي کند.

■

تعریف ۶ : ضریب اطمینان وزن دار براي قوانین انجمني وزن دار

در مدل ارائه شده ضریب اطمینان وزن دار را براي قوانین انجمني وزن دار طبق رابطه(۸)تعریف مي کنیم.

$$wconf(X \Rightarrow Y) = \frac{wsp(X \cup Y)}{wsp(X)} \qquad (۸)$$

تعریف ۷ : قوانین انجمني وزن دار

در این مدل در کنار دو پارامتر ضریب پشتیباني و ضریب اطمینـان وزن دار، وزن متناظر هر صفحه نیز نمایش داده مـي شـود. بنـابراین قـوانین انجمني وزن دار را به شکل زیر تعریف مي کنیم.

$r =< (p_1,p_2,...,p_k),(q_{k+1},q_2,...,q_{k+l}),(w_1,w_2,...,w_{k+l}),\delta,\alpha > \in R$

به طوریکه $(p_1,p_2,...,p_k)$ ،$(q_{k+1},q_2,...,q_{k+l})$ سرآمد و بدنه قانون وزن‌دار را به ترتیب نمایش مي‌دهند و δ بیانگر ضریب پشتیباني وزن‌دار و α ضریب اطمینان وزن‌دار و $(w_1,w_2,...,w_{k+l})$ بیانگر وزن متناظر هر یك از صفحات مي‌باشد.

۳- پیشنهاد صفحات بر اساس قوانین انجمني وزن دار

هدف از شخصي سازي بر اساس داده هاي کاربردي وب، تولیـد و ارائـه مجموعه پیشنهادات بـه کـاربر جـاري شـامل (لینـك هـا، محـصولات، صفحات و ...) مي باشد که بیشترین شباهت را به نشـست جـاري کـاربر داشته باشد. براي پیشنهاد صـفحات بـه کـاربران بـا اسـتفاده از قـوانین انجمنی، دنباله صفحات مشاهده شده در نشست جاری کاربر بـا قـوانین استخراج شده منطبق می شود و صفحاتی که تا کنـون کـاربر مـشاهده نکرده است به وی پیشنهاد می شود. محاسبه مجموعه پیشنهادي، rs، براي نشست کاربر جاري که بیشترین تطابق را با علایـق کـاربر داشـته باشد تنها جـز بـرخط سیـستم بـوده و بایـد از کـارایي و دقـت بـالایي برخوردار باشد. فرض کنیم که کاربري در حال گردش در سایت اسـت و مسیر $p_1 \rightarrow p_2 \rightarrow p_3 \rightarrow ... \rightarrow p_k$ را پیمـوده اسـت. تعـداد آخـرین صفحاتي را که توسط کـاربر در نشـست جـاري مـشاهده شـده و بـراي پیشنهاد صفحات جدید مورد استفاده قرارمـي‌گیـرد را *پنجـره پیـشنهاد* مي‌نامیم و آن را بـا w و انـدازه آن را بـا $|w|$ نـشان مـي‌دهـیم کـه حداکثر برابر با تمام صفحات مشاهده شـده و حـداقل برابـر بـا آخـرین صفحه مشاهده شده مـي‌باشـد. در پنجـره مـشاهده، صـفحات ابتـدایي نشست جاري کاربر یعني صفحات اولیه مشاهده شده توسط کاربر، نیـاز جاري کاربر را به خوبي بیان نمي‌کنند و صفحات انتهایي نشست بیـشتر گویاي نیاز آتي کاربر هستند. به عبارت دیگر صفحاتی که اخیـرا توسـط کاربر مـشاهده شـده بـرای اسـتفاده در الگـوریتم پیـشنهاد مناسـب‌تـر می‌باشند.

همچنانکه در تحقیقات دیگر نیز نشان داده شده است [۱۸-۱۷]، تنهـا در نظر گرفتن آخرین صفحه، موجب از دست رفتن اطلاعات سیستم در مورد نیاز اطلاعاتي کاربر خواهد شد. نتیجه این مسئله، اتخاذ تصمیمات نادقیق(بر اساس اطلاعات ناکافي) در ارائه پیشنهاد خواهد بـود. نکتـه مهم دیگر این که، در سیستمهاي پیشنهادگر وب، معمولا مطلوب است که با دیدن بخش کوچکي از صفحات بازدید شده توسط کاربر قادر بـه تولید پیشنهاد باشیم. بنابراین متدهاي پیشنهاد صفحات کـه از قـوانین انجمني براي ارائه پیشنهادات به کاربران اسـتفاده مـي کننـد از مفهـوم N-Gram براي کنترل تعداد صفحات نشست جاري کـاربر کـه بایـد بـا بخش مقدم قوانین انجمني سنجیده شود استفاده مي‌کنند و تنهـا N-1 صفحه آخر از نشست جاری کاربر در پیشنهاد صفحات جدید مورد نظـر قرار می گیرد. در این متدها با استفاده از یك پنجره به طول ثابت w بر روی دنباله صفحات بازدید شده توسط کاربر، صفحات اخیري که کـاربر در نشست جاري مشاهده کرده است با بخـش سـرآمد قـوانین انجمنـي

استخراج شده منطبق شده و در صورت تطابق، بخش بدنه قوانين كه شامل صفحاتی می باشد که کاربر تا بحال مشاهده نکرده است به وی پیشنهاد مي شود. با انتخاب اين روش، در واقع پيش‌بيني صفحه مورد توجه كاربر را بر اساس دانستن آخرين w صفحه بازديد شده انجام مي‌دهيم. بنابراين نحوه نگهداري دنباله صفحات بازديد شده توسط كاربر در ارائه پيشنهادات مفيد و ارزنده موثر خواهد بود. در ادامه مكانيزم جديدي كه صفحات مهم در نشست كاربر را به خوبي نگهداري مي كند ارائه مي دهيم.

۳-۱ بهبود نگهداري نشست جاري كاربر

نگهداري طول مناسبي از دنباله صفحات مشاهده شده در نشست جاري كاربر يكي از چالش‌های مهم در ارائه پيشنهادات مفيد به كاربر مي‌باشد. كاربر در حين جستجوي اطلاعات مورد نياز خود ممكن است مسيرهاي كاملا متفاوتي را طي كند و به صفحات غير مرتبط مراجعه كند كه اين مسئله منجر به ذخيره اطلاعات مستقل در نشست كاربر مي شود. روش هاي قبلي كه تا به‌حال ارائه شده اند از يك پنجره لغزان با طول ثابت بر روي نشست جاري كاربر استفاده كرده و بر اساس آن پيشنهادات خود را ارائه مي دادند. مكانيزم استفاده از پنجره لغزان در اين روش ها به شرح زير است: پنجره لغزاني به طول ثابت w در نظر گرفته مي شود. w طول پنجره ناميده می شود که از قبل توسط سيستم تعريف شده است[۱۹]. اين پنجره تنها آخرين w صفحه بازديد شده در سمت راست نشست را كه كاربر مشاهده كرده است در نظر مي گيرد و تنها اجازه مي دهد اين w صفحه آخر در ارائه پيشنهادات به كاربر نقش داشته باشند. شهود اين متد بر اين اساس است كه بيشتر كاربران به منظور يافتن اطلاعات مورد نياز خود در يك سايت حركات رو به جلو و عقب زيادي دارند صفحاتي كه در ابتداي نشست مشاهده شده اند نمي توانند نياز جاري كاربر را به درستي بيان كنند بنابراين صفحات آخري كه كاربر اخيرا به آنها مراجعه كرده است ارزش بيشتري دارند و بيشتر گوياي رفتار جاري كاربر هستند. اما استفاده از اين متد به تنهايي مناسب نيست. زيرا تفاوت اهميت صفحات متفاوت براي كاربر در اين متد در نظر گرفته نمي شود و همه w صفحه اخيری كه توسط كاربر بازديد شده است ارزش يكساني دارند حتي اگر اين صفحات براي كاربر مفيد نباشند. روش مناسب تر آن است كه صفحاتي كه براي كاربر جالب نبوده و اطلاعات موردنياز او را براورده نكرده اند از نشست جاري كاربر حذف شده و فرايند شخصي سازي و پيشنهاد صفحات بر اساس صفحاتي كه براي كاربر اهميت دارند و نيازهای او را تامين می كنند انجام شود.

پارامتر مهمي كه مي توان به عنوان يك معيار ديگر براي تعيين اهميت هر صفحه در دنباله صفحات مشاهده شده توسط كاربر در نظر گرفت "وزن هر صفحه" است. هر چند به نظر مي رسد صفحاتي كه كاربر اخيرا بازديد كرده است براي تعيين پيشنهادات مناسب تر هستند اما در در برخي موارد، كاربران در وب سرگردان مي‌شوند و بدون داشتن هدف مشخص بر روي صفحات مختلف كليک مي كنند. وقتي كاربر بازديد كننده از سايت مسير خود ر ا در سايت گم ميكند، صفحات بازديد شده توسط كاربر در فايلهاي دسترسي كاربران ثبت شده و ممكن است موجب گم راه كردن سيستم در توليد پيشنهادهاي آتي شود. به عنوان مثال يک كاربر ممكن است به يک صفحه رجوع كند تا بفهمد كه آيا به آن علاقه دارد يا خير، سپس بلافاصله به بخش ديگري از سايت مراجعه كند. از اين رو مي‌توان اين نوع دسترسي كاربر به صفحه را يک دسترسي غيرارزشمند به حساب آورد. به حذف صفحات يا اقلامي كه توسط كاربر درخواست شده اند و غيرارزشمند بوده اند، فيلتركردن ارزشي گفته مي‌شود. بنابراين مكان صفحه در نشست جاري كاربر به تنهايي پارامتر مناسبي در تعيين صفحات پيشنهادي نمي باشد. به اين منظور ما از دو پارامتر مكان صفحه و وزن صفحه در نشست جاري كاربر همزمان براي انتخاب صفحه اي كه بايد پيشنهاد شود استفاده مي كنيم.

در مقابل روش هاي استفاده از پنجره لغزان كه از اطلاعات اخير نشست براي تطابق قوانين استفاده مي كند معيار جديدي براي تعيين علايق كاربر جاري و حذف صفحاتي كه براي او جذاب نبوده و ناآگاهانه بر روي آنها كليك كرده است معرفي مي كنيم.

ما دو معيار "تازگي" و "وزن" را معرفي كرده و همزمان از تركيب هر دوي آنها طبق رابطه زير براي تعيين صفحاتي كه در فرايند پيشنهاد دخالت دارند استفاده كرديم. به اين منظور ابتدا نشست جاري كاربر(شامل دنباله صفحات بازديد شده)را مطابق عملياتي كه بر روي تراكنش ها انجام شد وزن دار مي كنيم(تا مدت زمان مشاهده صفحه و دفعات بازديد هر صفحه ارزش واقعي هر صفحه را در غالب وزن آن تاييد كند). براي تعيين اهميت هر صفحه بر اساس تازگي، اين پارامتر را مطابق رابطه(۹) براي هر صفحه محاسبه مي كنيم.

$$Fresh(p_i) = \frac{i}{|w|}\ ,\ i = 1, 2, ..., w \qquad (۹)$$

كه $|w|$ اندازه پنجره لغزان و i مكان صفحه موردنظر در پنجره است. عدد يك به اولين صفحه بازديد شده نسبت داده مي شود و آخرين صفحه نيز تازه ترين صفحه خواهد بود. پارامتر بعدي وزن هر صفحه است. براي آنكه پارامتر "تازگي" اثر خود را بهتر منعكس كند لازم است پارامتر وزن مطابق رابطه(۱۰) نرمال سازي شود.

$$W_{normalized}(p_i) = \frac{w(p_i)}{\sum_{j=1}^{n} w(p_j)} \qquad (۱۰)$$

از آنجا كه در معيار معرفي شده "تازگي" و "وزن" هر صفحه اهميت يكساني دارند از ميانگين هارمونيك انها استفاده شده است. اين رابطه

تضمين مي كند يك صفحه تنها در صورتي كه هر دو معيار "وزن" و "تازگي" آن همزمان بالا باشند صفحه با ارزشي است.

$$Interest(p_i) = \frac{2 * Fresh(p_i) * W_{normalized}(p_i)}{Fresh(p_i) + W_{normalized}(p_i)} \quad (11)$$

فرض كنيد نشست جاري كاربر پس از محاسبه وزن صفحات به شكل $S = <(A,30),(B,20),(C,5),(D,5),(E,4),(F,10)>$ نمايش داده شود. شكل (۱) مقايسه اي از متد ارائه شده در اين بخش و متد پنجره لغزان معمول را نشان مي دهد.

از آنجا كه اندازه پنجره لغزان را روي ۳ تنظيم كرده ايم متد پنجره لغزان معمول ۳ صفحه آخر را از نشست جاري كاربر انتخاب می‌كند($X = \{D, E, F\}$) در حاليكه متد ارائه شده مجموعه صفحات $X = \{A, B, F\}$ را برگزيده است. اگرچه صفحه A اولين (قديمي ترين) صفحه اي است كه كاربر مشاهده كرده است اما چون وزن بالاتري در مقايسه با صفحات D, E دارد بنابراين براي كاربر جالب تر بوده است و بر خلاف متد پنجره لغزان كه آنرا حذف مي كند در پنجره ما قرار مي گيرد.

Current Session	A	30	B	20	C	5	D	5	E	4	F	10
Fresh	A	1/6	B	2/6	C	3/6	D	4/6	E	5/6	F	6/6
Normalized Weight	A	30/74	B	20/74	C	5/74	D	5/74	E	4/74	F	10/74
Interest	A	0.22	B	0.29	C	0.10	D	0.09	E	0.094	F	0.23

Traditional Slide Window Scheme for Current Session: D E F

Proposed Silde Window Scheme for Current Session: A B F

شكل (۱): متد ارائه شده در مقايسه با متد پنجره لغزان

۲-۳ مكانيزم پيشنهاد صفحات

در كارهای قبلی پس از بررسی تطابق دنباله صفحات بازديد شده توسط كاربر و قوانين استخراج شده صفحات تنها بر اساس معيار ضريب اطمينان مرتب می شدند و صفحاتی كه بالاترين امتياز را كسب می كردند به كاربر پيشنهاد می شدند. در مواردي كه دو صفحه با امتياز يكسان موجود بود يكي از آنها بصورت تصادفي انتخاب مي شد.

فرايند پيشنهاد صفحات بر اساس قوانين انجمني وزن دار در الگوريتم پيشنهاد شده در اين بخش شامل دو فاز مي باشد. در فاز اول قوانين انجمني وزن دار از فايلهاي ثبت دسترسي كاربران به صفحات وب سايت استخراج می‌شوند. اين قوانين رفتار هدايتي كاربران را بر روي سايت نمايش می‌دهند. در فاز دوم موتور پيشنهاد دهنده صفحات، پس از بهبود نشست جاري كاربر m تا از شبيه ترين قوانين انجمني وزن دار را كه بيشترين تطابق با نشست كاربر را دارند مطابق معيار شباهتي كه معرفي مي شود پيدا كرده و اميتاز پيشنهاد آنها را نه فقط بر اساس ضريب اطمينان بلكه بر اساس تركيبي از آن و درجه شباهت محاسبه كرده و در نهايت صفحه اي با بالاترين امتياز به كاربر پيشنهاد مي شود.

در حين اجراي فاز دوم، براي ارائه پيشنهادات دقيق تر همان طور كه اشاره شد ما به جاي بررسي شباهت باينري نشست كاربر با بخش مقدم مجموعه قوانين انجمني استخراج شده كه در كارهاي قبلي انجام گرفته است از يك معيار شباهت كه ميزان شباهت w صفحه ديده شده توسط كاربر را با مجموعه قوانين توليد شده محاسبه كرده سپس مناسب ترين قوانين را كه به رفتار جاري كاربر شباهت بيشتري دارند انتخاب كرده و از آنها براي ارائه پيشنهادات استفاده مي كنيم.

۱-۲-۳ بررسی شباهت دنباله صفحات بازديد شده توسط كاربر و قوانين انجمنی

روشهايي كه تا به حال گزارش شده‌اند و از قوانين انجمني براي كشف دانش رفتار پيمايشي كاربران استفاده مي‌كنند[۱۲-۱۴]، نشست جاری كاربر را با بخش مقدم مجموعه قوانين انجمني استخراج شده تطابق دقيق صفحه به صفحه می‌كنند تا مجموعه اي از صفحات را كه تا به حال كاربر مشاهده نكرده است به او پيشنهاد دهند. ولي هيچ كدام از آنها از وزن صفحات براي رتبه بندي صفحات كانديد استفاده نكرده‌اند. استفاده از اين روش و تنها با اتكا به ضريب اطمينان به عنوان ميزان تشابه نشست جاری كاربر و قانون انجمنی، باعث می‌شود كه كيفيت پيشنهاد پايين باشد زيرا ممكن است دو قانون ضريب اطمينان برابری داشته باشند، اما كاربر قانون اول زمان بيشتری را در برخی صفحات خاص صرف كرده باشد كه تطابق بيشتری با نشست كاربر جاری داشته باشد كه در اين موارد سيستم‌های موجود به صورت تصادفی عمل می‌كنند. در الگوريتم پيشنهادی، براي پيشنهاد صفحاتي با بيشترين شباهت به نشست جاري كاربر، شباهت w (اندازه پنجره پيشنهاد) صفحه ديده شده توسط كاربر را با مجموعه قوانين انجمني وزن دار توليد شده بررسي كرده سپس مجموعه اي از صفحات مشابه كه تا به حال توسط كاربر ملاقات نشده اند را انتخاب مي كنيم.

براي تعيين ميزان شباهت نشست جاري كاربر و قوانين توليد شده، پس از نگهداري دنباله مناسبي از مجموعه صفحات بازديد شده توسط كاربر مطابق مكانيزم معرفی شده، نشست جاري و قوانين انجمني وزن دار $r = <(p_1, p_2, ..., p_k), (q_{k+1}, q_2, ..., q_{k+l}), (w_1, w_2, ..., w_{k+l}), \delta, \alpha> \in R$ استخراج شده را بصورت بردار m تايي از زوج مرتب صفحه و وزن متناظر آنها نمايش مي دهيم. به طور مثال براي قانون انجمني وزن دار r، بخش بدنه اين قانون r_L را بصورت بردار m بعدي مطابق رابطه(۱۲)نمايش مي دهيم.

$$r_L = \{w_1, w_2, ..., w_m\} \quad (12)$$

$$w_i = \begin{cases} weight(p_i, r_{Li}), & if \quad p_i \in r_L \\ 0, & otherwise \end{cases}$$

مشابه بالا دنباله صفحات بازديد شده توسط كاربر در نشست جاري به صورت بردارm بعدي $S = \{s_1, s_2, ... s_m\}$ مدل مي شود كه s_i

بيانگر وزن متناظر با صفحه iام در نشست جاري كاربر، در صورت بازديد صفحه توسط كاربر مي باشد و در غير اين صورت $s_i = 0$ خواهد بود.

ما در ادامه به محاسبه درجه شباهت نشست كاربر هدف با هر يك از قوانين انجمني وزن دار مي پردازيم و درجه تطابق آنها طبق رابطه(۱۳) محاسبه مي شود.

$$Match\ Score(S, r_L) = 1 - \frac{1}{4}\sqrt{\frac{Dissimilarity(S, r_{Li})}{\sum_{i: r_{Li} > 0} 1}} \qquad (۱۳)$$

$$Dissimilarity(S, r_L) = \sum_{i: r_{Li} > 0} \left(\frac{2 * (w(s_i) - w(r_{Li}))}{w(s_i) + w(r_{Li})}\right)$$

كه S و r_L به ترتيب بيانگر دنباله صفحات مشاهده شده توسط كاربر و بخش بدنه قوانين انجمني وزن دار مي باشد.

توجيه استفاده از اين رابطه بر اين اساس است كه ابتدا عدم شباهت آنها($Dissimilarity(S, r_L)$) را بدست مي آوريم. معيار معرفي شده براي عدم شباهت از معيارهاي رايجي مي باشدكه در تحقيقات بسياري استفاده شده است[۱۹] و در حل مسائل مختلف موفقيت آميز بوده است. مطابق رابطه(۱۳)بديهي است در صورتيكه تطابق كاملي بين نشست جاري كاربر و قانون انجمني وزن دار وجود داشته باشد نرخ تطابق آنها يك خواهد بود.

۳-۲-۲ امتياز بندي صفحات با استفاده از قوانين انجمني وزن دار

هدف از سيستم هاي شخصي سازي محاسبه يك مجموعه پيشنهادي براي نشست جاري كاربر مي باشد كه بيشترين تطابق را با علايق كاربر داشته باشد. اين جز تنها جز بر خط سيستم بوده كه بر اساس رفتار جاري كاربران پيشنهادات پويايي ايجاد مي كند و بايد از كارايي و دقت بالايي برخوردار باشد. در بخش بر خط، وب سرور نشست فعال كاربران را از طريق درخواستهاي HTTP دنبال كرده و موتورهاي پيشنهاد شباهت اين نشستها را با الگوهای استخراج شده بررسي كرده و براي مجموعه اي از URL ها كه تا اكنون توسط كاربر مشاهده نشده است پيشنهاد مي كنند. مجموعه صفحات پيشنهاد شده بصورت لينک به درخواست صفحه اي كه هنوز به مرورگر كلاينت ارسال نشده است الحاق می شوند.

به اين منظور، به ازاي همه صفحاتي كه كاربر تابحال مشاهده نكرده است امتياز پيشنهاد محاسبه مي شود. در تعيين امتياز پيشنهاد دو فاكتور نرخ تطابق نشست جاري كاربر با قوانين انجمني مذكور كه در رابطه(۱۳) بررسي شد و ضريب اطمينان وزن دار قوانين انجمني وزن دار دخالت دارند. امتياز هر پيشنهاد از حاصلضرب اين دو پارامتر تعيين مي شود. به ازاي قانون انجمني وزن دار موجود و نشست جاري كاربر امتياز پيشنهاد براي صفحه P مطابق رابطه (۱۴)محاسبه مي شود.

$$Rec(S, X \Rightarrow p) = Match\ Score(S, X) * wconf\ (X \Rightarrow p) \qquad (۱۴)$$

پس از محاسبه امتياز پيشنهاد به ازاي تمام صفحاتي كه در بخش سرآمد قوانين منطبق با نشست جاري كاربر قرار داشتند در نهايت m تا از بالاترين امتيازات را رتبه بندي كرده و سرانجام صفحه اي كه بالاترين امتياز را به خود اختصاص داده است به كاربر پيشنهاد مي شود.

بهبود اين روش در آن است كه بر خلاف روش هاي قبلي كه تنها شباهت باينري دنباله صفحات بازديد شده توسط كابر و قوانين را در نظر گرفته و تنها با استناد به ضريب اطمينان قوانين پيشنهادات خود را ارئه مي دادند در اين مدل از پارامترهاي مهم نرخ تطابق نشست جاري كاربر با قوانين انجمني و ضريب اطمينان وزن دار قوانين انجمني براي ارائه پيشنهادات استفاده شده است.

۴- ارزيابي پيشنهادات به دست آمده با استفاده از قوانين انجمني وزن دار

۴-۱ مدل شبيه سازي

دو روش عمده براي ارزيابي الگوريتم‌هايي كه از اطلاعات پيمايش كاربران استفاده مي‌كنند وجود دارد. روش اول، استفاده از صفحات وب واقعي و داده‌هاي واقعي كاربران وب موجود در فايل‌هاي ثبت رخداد سايت‌ها مي‌باشد. براي استفاده از اين روش مجموعه داده‌هاي استانداردي كه از چند سايت معتبر استخراج شده‌اند در دسترس مي‌باشد. روش دوم مدل ارائه شده در [۲۰] مي‌باشد. در اين روش Lui و همكارانش نظم موجود در رفتارهاي كاربران در محيط وب را با استفاده از يک مدل مبتني بر عامل، مشخص و اعتبار مدل خود را با استفاده از اطلاعات استفاده از وب چندين سايت وب بزرگ مانند مايكروسافت، تاييد كرده‌اند. در اين مقاله ما از داده‌هاي استاندارد سايت CTI DePaul استفاده مي‌كنيم. اين مجموعه داده اطلاعات نشست كاربران را به مدت ۲ هفته در سايت CTI DePaul در سال ۲۰۰۲ شامل مي‌شود[۲۱]. اين اطلاعات پيش‌پردازش شده و نشست‌هاي با اندازه ۱ و غير استاندارد از آن حذف شده‌اند و در نهايت اطلاعات ۱۳۷۴۵ كاربر كه از ۶۸۳ صفحه ديدن كرده‌اند در فايل‌هاي جداگانه قرار داده شده است.

۴-۲ معيار و متدلوژي ارزيابي

پارامترهاي تاثيرگذار در كارايي الگوريتم عبارتند از: $|w|$ اندازه پنجره پيشنهاد، rt حد آستانه پيشنهاد. براي بررسي دقت الگوريتم ارائه شده رويه زير اتخاذ شده است. ابتدا مجموعه داده‌ها را به دو دسته به نسبت ۲ به ۱ تقسيم مي‌كنيم كه مجموعه اول براي يادگيري و مجموعه دوم براي تست و ارزيابي مورد استفاده قرار مي‌گيرد. سپس با استفاده از

مجموعه يادگيري الگوريتم را اجرا مي‌كنيم. بر اساس مقدار $|w|$ ، از هر نشست در مجموعه تست كه اندازه آن حداقل $w+rt$ مي‌باشد، $|w|$ صفحه متوالي را انتخاب كرده و به الگوريتم مي‌دهيم. معيار ارزيابي روابط معرفي شده در [۲۲] است. فرض كنيم مجموعه $rp=\{x_{w+1},x_{w+2},...,x_{w+|rs|}\}$ صفحات مشاهده شده توسط كاربر در ادامه نشست واقعي باشد. درجه شباهت مجموعه پيشنهادي و مجموعه صفحات واقعي و درصد پوشش آنها از روابط زير به دست مي‌آيند:

$$\Pr ecision(rs,rp)=\frac{|rs\cap rp|}{|rs|} \quad (۱۵)$$

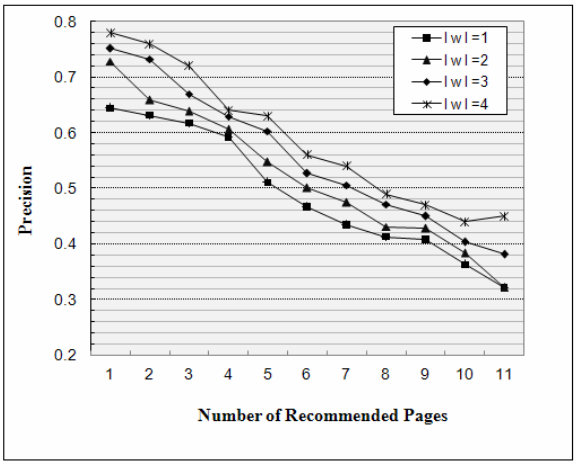
$$Coverage(rs,rp)=\frac{|rs\cap rp|}{|rp|} \quad (۱۶)$$

معيار *Precision* همپوشاني دو مجموعه يعني نسبت پيشنهادات مناسب را به تعداد كل پيشنهادات نشان مي دهد. هدف اين است كه چند تا از صفحات پيشنهادي در مجموعه تست براي آن كاربر وجود دارد. در اين رابطه ترتيبي براي صفحات پيشنهادي در نظر گرفته نشده است. در صورتي كه تعداد صفحات پيشنهادي را به يک صفحه محدود كنيم، كافي است تا صفحه پيشنهادي الگوريتم را با صفحه $w+1$ ام در نشست كاربر مقايسه مي‌كنيم. در اين صورت دقت الگوريتم درصد نشست‌هايي مي‌باشد كه در آنها صفحه پيشنهادي الگوريتم با صفحه واقعي يكسان مي‌باشد. معيار *Coverage* توانايي سيستم را براي پيشنهاد همه صفحاتي كه بايدتوسط كاربر ملاقات شوند يا درصدي از كل صفحاتي كه سيستم قادر است پيشنهاد بدهد را نشان مي دهد . معيار پوشش نسبت پيشنهادات مناسب را به كل تعداد صفحات مشاهده شده توسط كاربر كه بايد پيشنهاد شود نشان مي دهد.

۳-۴ حساسيت سيستم به تغيير در طول پنجره فعال بـر دنباله صفحات بازديد

در تعريف حالات، با استفاده از مفهومN-Gram ، پنجره اي به طول ثابت w بر روي دنباله صفحات بازديد شده توسط كاربر، قرار داديم. با انتخاب اين روش، در واقع پيش بيني صفحه مورد توجه كاربر را بر اساس دانستن آخرين صفحات بازديد شده انجام مي دهيم. انتخاب مقدار w از چند جهت مي تواند بر روي عملكرد سيستم تاثير گذار باشد. پنجره اي با طول زياد، معادل داشتن حافظه اي طولاني تر براي سيستم است و اطلاعات بيشتري در اختيار سيستم قرار مي دهد.

جهت بررسي تاثير انتخاب w ، سيستم را با مقادير مختلف w بررسي كرديم. شكل(۲) نتايج دقت الگوريتم را نسبت به تعداد صفحات پيشنهادی برای اندازه پنجره نشان می‌دهد. همانطور كه از شكل (۲) پيداست با افزايش تعداد صفحات در پنجره پيشنهاد، w ، از ۱ به ۴ ، دقت پيشنهاد بالا می‌رود. همانطور كه در اين شكل مشخص است ، بهترين نتايج در عملكرد سيستم، هنگام استفاده از ۴ = w بدست آمده است. از اين نتايج مي توان استنتاج كرد كه طول پنجره ۱، اطلاعات كافي در مورد پيمايش كاربر و نياز اطلاعاتي او در حافظه سيستم نگهداري نمي كند. در واقع در اين حالت، اطلاعات كافي براي ارائه پيشنهادهاي دقيق در اختيار سيستم قرار ندارد. مشاهده مي شود كه با افزايش w ، كيفيت پيشنهادها بهبود يافته و با پنجره اي به طول ۴ بهترين نتايج به دست مي آيد. نتايج نشان می دهد كه با افزايش اندازه پنجره پيشنهاد دقت الگوريتم در حد قابل توجهی بالا می رود. زيرا با افزايش اندازه پنجره، اطلاعات بيشتری از تاريخچه رفتاری كاربر برای توليد پيشنهادات موجود می باشد.. البته در درصد پوشش الگوريتم با تغيير اندازه پنجره تغيير محسوسی مشاهده نمی شود.



شكل(۲):تاثير اندازه پنجره لغزان بر دقت الگوريتم مبتنی بر قوانين انجمنی وزن دار

۴-۴ حساسيت سيستم به تغيير در آستانه‌ي پشتيباني

جهت بررسي تاثير مقادير متفاوت آستانه پشتيبانی، سيستم را با مقادير مختلف ضريب پشتيبانی بررسی كرديم. نتايج اين آزمايشها در جدول(۱) قابل مشاهده است. نتايج نشان می دهد كه هر چه مينيمم حد آستانه ضريب پشتيبانی وزن‌دار بالاتر باشد پوشش الگوريتم در حد قابل توجهی كاهش می‌يابد و در مقابل دقت الگوريتم تنها مقدار اندكی افزايش می يابد. مطلوب آن است كه حد آستانه ضريب پشتيبانی بالا باشد تا سايز مدل را كوچک نگه داريم و از توسعه پذيری قوانين انجمنی مطمئن باشيم. يک مشكل در رابطه با موافقی كه مينيمم آستانه ضريب پشتيبانی بالا می باشد آن است كه برخی از صفحاتی كه پتانسيل (استعداد نهانی) بالايي دارند ولی در مجموعه تراكنش ها كمتر ظاهر شده اند را در مجموعه صفحات پيشنهادی از دست می دهيم و مجموعه پيشنهادی صفحاتی كه به ندرت ظاهر مي‌شوند اما مهم هستند را شامل نمي‌شوند. اين مساله خصوصا در هنگام مواجهه با داده هاي كاربرد وب اهميت پيدا مي‌كند. اغلب، ارجاعات به محتواي در عمق

بیشتر یا صفحات محصولات، بسیار کم‌تکرارتـر از صـفحات گردشـي در عمق کمتر هستند.

جدول(۱): بررسي تاثير مينيمم آستانه پشتيبانی بر کارايي الگوريتم مبتنی بر قوانين وزن دار

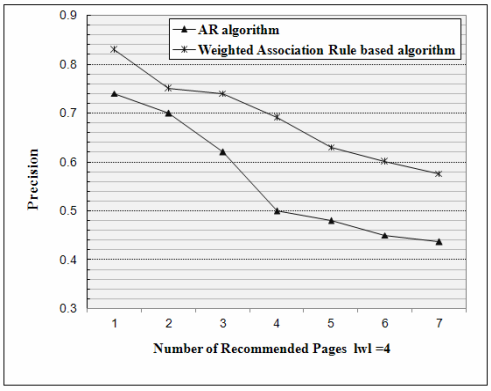
$\delta_{\min}$	پوشش پیشنهاد ارائه شده	دقت پیشنهاد ارائه شده
۰,۰۱	۰,٦۹	۰,۷۳
۰,۰۲	۰,٦٤	۰,۷۳۹
۰,۰۳	۰,٦۰	۰,۷٤۲
۰,۰۵	۰,۵٤	۰,۷٦
۰,۰٦	۰,۵۲	۰,۷۸
۰,۰۷	۰,٤۸	۰,۷۸۵
۰,۰۸	۰,٤۷	۰,۷۸۹
۰,۰۹	۰,٤٦	۰,۸۰
۰,۱	۰,٤٦	۰,۸۲۳

جدول(۲): پارامترهاي ثابت در پيشنهاد صفحات

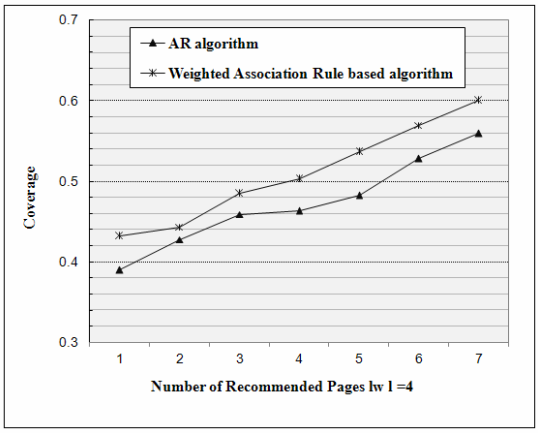
حد آستانه مینیمم ضریب پشتیبانی	۰,۰۶
w اندازه پنجره لغزان	۴

۵-۴ مقايسه روش ارائه شده بـا ديگـر روشـهاي توليـد پيشنهاد

در اين بخش، روش توليد پيشنهاد بر اساس قوانين انجمني وزن دار را با يکی از روش های پيشنهاد متداول بر اساس قوانين انجمنی معمول مقايسه مي کنيم[۱۳]. در تمامی نمودارهاي اين بخش، از جمله نمودارهای شکل (۳) و (۴)، از نمايش تغييرات يکي از دو معيار دقت و يا درصد پوشش (محور عمود ي) بر اساس تعداد صفحات پيشنهادی (محور افقي) براي نمايش عملکرد سيستم استفاده شده است. پارامترهاي استفاده شده در روش پيشنهادي در جدول(۲) نشان داده شده است. در شکل های (۳) و (۴) مي توان خلاصه عملکرد اين ۲ روش را بر مبناي دقت پيشنهادها و پوشش پيشنهادها با تغيير تعداد صفحات پيشنهاد شده مشاهده کرد.



شکل(۳): مقايسه دقت الگوريتم با اندازه پنجره پيشنهاد ۴ و تعداد صفحات پيشنهادي متفاوت



شکل(۴): مقايسه درصد پوشش با اندازه پنجره پيشنهاد ۴ و تعداد صفحات پيشنهادي متفاوت

با توجه به اين دو نمودار ، مشخص مي شود که روش مبتني بر قوانين انجمنی وزن دار، نتايج بهتري از روش مبتني بر قوانين انجمني معمول، کسب مي کند. با افزايش پوشش، بطور طبيعي و همانطور که انتظار می رود، دقت پيشنهاد ها در هر ۲ روش کاهش پيدا مي کند. در نتيجه، روش پيشنهاد مبتني بر قوانين وزن دار، قادر به ارائه عملکردي بهتر(در قالب دقت، پوشش) در مقايسه با روش مبناي استفاده شده است. با بررسي پيشنهاد هاي انجام شده توسط اين دو روش، مشخص شد که در مقادير تعداد صفحات پيشنهادی پايين، پيشنهادها عموما مربوط به صفحه هايي مي شوند که بلافاصله بعد از صفحه فعلي توسط کاربر بازديد مي شوند. در واقع، هر يک از روشها مطمئن ترين پيشنهادهاي خود (پيشنهادها با بالاترين امتياز)، را انجام مي دهند. در نتيجه، اين دسته از پيشنهادها از دقت نسبتًا بالايي نيز برخوردار هستند در مقابل، با افزايش تعداد صفحات پيشنهاد ها، هر يک از روشها، پيشنهادهاي بعدي خود را با امتياز پايين تر، ارئه مي کنند. در اين حالت، روش پيشنهاد ما با دقت بيشتر شروع به پيشنهاد صفحاتي مي کند که در قدمهاي بعدي(دورتر)، در جلسات کاربر بازديد مي شوند. در مجموع،

عملکرد این روش با افزایش تعداد صفحات، دچار افت شدید تري مي شود. اما روش مبتني بر قوانین وزن دار، موفق به ارائه مدل کامل تري از رفتار کاربران شده است. این روش، حتي در تعداد صفحات بالا، موفق به ارائه عملکرد بهتري از روش دیگر پیشنهاد، در هردو معیار دقت و پوشش شده است.

۵- نتیجه گیری

در این مقاله، پس از معرفي یک معیار وزن‌دهي که میزان علاقه کاربران به صفحات را به خوبي مدل می‌کند، الگوریتم جدیدي معرفي شده است که از اطلاعات پیمایش کاربران، براي استخراج قوانین انجمنی وزن دار استفاده مي‌کند. ما مدل قوانین انجمني معمول را با در نظر گرفتن وزن متناظر هر آیتم در تراکنش، که بیانگر میزان علاقه کاربر به آن آیتم در آن تراکنش مي باشد توسعه دادیم. کاوش قوانین انجمني وزن دار به ما اجازه مي دهد وزن‌هاي متفاوتي به آیتم‌هاي موجود در تراکنش ها نسبت دهیم که روش مناسبي براي بهبود فرایند شخصي سازي بر اساس قوانین انجمني مي باشد. همچنین مدل جدیدی برای انتخاب پنجره نشست جاری کاربر ارائه نمودیم. نتاج بدست آمده کارایي قوانین انجمنی وزن دار را در مقایسه با قوانین انجمنی رایج تایید می‌کنند.

۶- مراجع

[1]B. Mobasher, R. Cooley, J. Srivastava, "Automatic Personalization Based on Web Usage Mining", Communications of the ACM, vol. 43, no.8, 2000, pp. 142-151.

[2]B. Mobasher, H. Dai, T. Luo, M. Nakagawa, "Effective Personalization based on Association Rule Discovery from Web Usage Data", Proceedings of the 3rd ACM Workshop on Web Information and Data Management, 2001.

[3]J. Srivastava, R. Cooley, M. Deshpande, P. N. Tan, "Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data", SIGKDD Explorations, vol. 1, no. 2, 2000, pp.12-23.

[4]B. Mobasher, H. Dai, T. Luo, M. Nakagawa, "Discovery and Evaluation of Aggregate Usage Profiles for Web Personalization", Data Mining and Knowledge Discovery, vol. 6, no. 1, 2002, pp. 61-82.

[5]G. Pierrakos, C. Paliouras, C. D. Spyropoulos, "Web Usage Mining as a Tool for Personalization", A Survey, User Modeling and User-Adapted Interaction, vol. 13, no. 4, 2003, pp. 311-372.

[6]P. Kazienko, M. Adamski, "AdROSA - Adaptive Personalization of Web Advertising", Information Sciences 177(11), 2007, pp. 2269-2295.

[7]P. Kazienko, M. Kiewra, "Personalized Recommendation of Web Pages", In: Nguyen, T. (ed.) Intelligent Technologies for Inconsistent Knowledge Processing, Advanced Knowledge International, International, Adelaide, South Australia, 2004, pp. 163-183.

[8]P. Kazienko, "Filtering of Web Recommendation Lists Using Positive Positive and Negative Usage Patterns", Springer-Verlag Berlin Heidelberg, 2007.

[9]A. Demiriz, "Enhancing Product Recommender Systems on Sparse Binary Data", published in the Journal of Data Mining and Knowledge Discovery, 2003.

[10]C. H. Cai, A. W. C. Fu, C. H. Cheng, W. W. Kwong, "Mining Association Rules with Weighted Items", In Database Engineering and Applications Symposium, Proceedings IDEAS'98, 1998, pp. 68 – 77.

[11]F. Tao, F. Murtagh, M. Farid, "Weighted Association Rule Mining using Weighted Support and Significance Framework", In Proceedings of the 9th SIGKDD Conference, 2003.

[12]B. Mobasher, "Web Usage Mining and Personalization", In Practical Handbook of Internet Computing, Munindar P. Singh (ed.), CRC Press, 2005.

[13]B. Mobasher, H. Dai, T. Luo, M. Nakagawa, "Using Sequential and Non-Sequential Patterns for Predictive Web Usage Mining Tasks", Proceedings of the IEEE International Conference on Data Mining, Japan, 2002.

[14]B. Mobasher, H. Dai, T. Luo, M. Nakagawa, "Improving the Effectiveness of Collaborative Filtering on Anonymous Web Usage Data", In Proceedings of the IJCAI workshop on Intelligent Techniques for Web Personalization (ITWP01), 2001.

[15]Y. Liang, L. Chunping, "Incorporating Page View Weight into an Association-Rule-Based Web Recommendation System", Springer-Verlag Berlin Heidelberg, AI 2006, LNAI 4304, 2006, pp. 577 – 586.

[16]R. Agrawal, R. Srikant, "Fast Algorithms for Mining Association Rules in Large Databases", In Proceedings of the 20th International Conference on Very Large Data Bases VLDB'94, Santiago, Chile, 1994, pp. 487-499.

[17]G. Linden, B. Smith, G. York, Item-to-Item Collaborative Filtering", IEEE Internet Computing, 2003.

[18]B. N. Miller, I. Albert, S. Lam, J. Konstan, J. Riedl, "MovieLens Unplugged: Experiences with an Occasionally Connected Recommender System", Proc. Int'l Conf, Intelligent User Interfaces, 2003.

[19]A. Tomovic, P. Janicic, V. Keselj, "N-gram-based Classification and Hierarchical Clustering of Genome Sequences", Computer Methods and Programs in Biomedicine, 2005.

[20]J. Mori, Y. Matsuo, M. Ishizuka, B. Faltings, "Keyword Extraction from the Web for FOAF Metadata", Proceeding of 1st International Workshop on Friend of a Friend, Social Networking and the Semantic Web, 2004.

[21]J. Manuel Barrueco Cruz, T. Krichel, "Automated Extraction of Citation Data in a Distributed Digital Library", Proceedings of the 2nd International Workshop on New Developments in Digital Libraries, 2002.

[22]T. Haveliwala, "Topic-Sensitive PageRank", in Proceedings of the 11thInternational Conference on World Wide Web, New York: ACM Press, 2002, pp. 517–526.