

# سیستم پیشنهاددهنده وب ترکیبی مبتنی بر گراف دو لایه و پارتبیشن بندی گراف

هانیه محمدی دوستدار<sup>۱</sup>؛ رعنا فرصتی<sup>۲</sup>؛ محمدرضا میبدی<sup>۳</sup>

## چکیده

در سال های اخیر، با رشد سریع وب، کاربران از مشکل گرانباری اطلاعات رنج می برند. سیستم های پیشنهاددهنده وب جهت غلبه بر این مشکل ایجاد شده اند و به کاربران کمک می کنند تا تصمیمات مناسبی را در این فضای اطلاعاتی گسترش دهند. این سیستم ها نیازهای اطلاعاتی آنها را پیشگویی نموده و پیشنهادهای را برای تسهیل در هدایتشان فراهم می نمایند. اخیراً، کارهای تحقیقاتی بسیاری در این حوزه بر مبنای تکنیک های وب کاوی صورت گرفته است. اکثر سیستم های پیشنهاددهنده موجود از رویکرد مبتنی بر محتوا، یا پالایش گروهی جهت ارائه پیشنهادات استفاده می کنند. تخمین بهترین رویکرد برای پیشنهاد کار دشواری است؛ هر یک از این روش ها به تنهایی نقاط قوت و البته محدودیت ها و ضعف هایی دارند؛ بنابراین استفاده از رویکردهای ترکیبی می تواند بر این نقاط ضعف غلبه نماید و دقت سیستم پیشنهاددهنده را افزایش دهد.

از این رو، در این مقاله، جهت افزایش کارایی، یک سیستم پیشنهاددهنده ترکیبی ارائه می دهیم که نتایج این دو رویکرد را در قالب گرافی دو لایه بر مبنای پارتبیشن بندی گراف ترکیب می کند. از طریق روش هایی نوبن، شباهت میان صفحات وب و شباهت میان کاربران را بدست آورده ایم و پارتبیشن بندی گراف را برای دسته بندی کاربران و صفحات وب به کار بردهیم. نتایج شبیه سازی ها نشان می دهد که رویکرد ترکیبی در مقایسه با دو رویکرد مبتنی بر محتوا و پالایش گروهی از دقت بالاتری برخوردار است و پیشنهادات مناسب تری را به کاربران ارائه می دهد.

## کلمات کلیدی

سیستم های پیشنهاددهنده، کاوش وب، رویکرد مبتنی بر محتوا، رویکرد پالایش گروهی، گراف دو لایه، پارتبیشن بندی گراف.

## The Hybrid Web Recommender System based on Bi-section Graph Model and Graph Partitioning

Hanieh Mohammadi Doustdar; Rana Forsati; Mohammad Reza Meybodi

### ABSTRACT

In recent years, with the rapid growth of the World Wide Web, users usually suffer from the information overload problem. Web recommender systems have been introduced to deal with this problem and help users make decisions in this complex information space. These systems predict the information needs of users and provide them with recommendations to facilitate their navigation. Recently, Many researches accomplished in this area are based on techniques of web mining. Most existing recommender system use either content-based or collaborative filtering approach for recommendation process. It could be difficult to estimate a single best model for recommendation; Each of these approaches has its own strengths and limitations.

Therefore, in this paper, in order to have a better performance, we investigate a hybrid recommender system, which combines the results of these approaches in a bi-section graph model based on graph partitioning. We have gained web page similarity and user similarity by the new methods and have used

۱. دانشگاه آزاد اسلامی، واحد قزوین، دانشکده برق رایانه و فن آوری اطلاعات، گروه مهندسی کامپیوتر، قزوین، ایران. h.m.doustdar@gmail.com

۲. دانشگاه آزاد اسلامی، واحد کرج، دانشکده فنی و مهندسی، گروه مهندسی کامپیوتر، کرج، ایران. forsati@kiau.ac.ir

۳. دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، تهران. mmeybodi@aut.ac.ir

graph partitioning for clustering web pages and users. Evaluation results show that the hybrid approach achieved more accurate predictions and relevant recommendations than using only one of approaches.

## KEYWORDS

Recommender systems, Web mining, Content-based approach, Collaborative approach, Bi-section graph, Graph partitioning.

## ۱. مقدمه

با رشد سریع وب، بدست آوردن اطلاعات مفید از اینبوه اطلاعات در وب با مشکل مواجه شده است و موجب مسئله جدیدی بنام گرانیاری اطلاعات شده است که تلاش‌های بسیاری جهت غلبه بر این مشکل صورت گرفته است. کاربران به علت حجم زیاد اطلاعات، قادر به اتخاذ تصمیم، انتخاب یا به روز نگهداری اطلاعات خود راجع به یک موضوع خاص نخواهند بود. سیستم های پیشنهادگر<sup>۱</sup> به کاربران کمک می کنند تا تصمیمات مناسبی را در این فضای اطلاعاتی گستردۀ اتخاذ نمایند. در واقع سیستم پیشنهادگر به هر سیستمی گفته می شود که کاربر را در فضای بسیار بزرگ از اختیابهای ممکن به سوی اشیا مفید و جالب برای او هدایت کند. محققان به جهت غلبه بر مشکل گرانیاری اطلاعات سیستم های پیشنهاددهنده را به کاربران ارائه می کنند و موجب افزایش رضایتمندی کاربران از وب سایت می شوند.

اخیراً برای ساخت سیستم های پیشنهاددهنده توجه بسیاری به استفاده از تکنیک های وب کاوی شده است. وب کاوی فرآیند استخراج اطلاعات و الگوهای جالب و مفید از وب است که به سه شاخه کاوش محتوای وب، کاوش ساختار وب و کاوش استفاده از وب تقسیم می شود<sup>[۲]</sup>. کاوش محتوای وب شامل فرآیند استخراج اطلاعات مفید از محتوای مستندات وب می باشد؛ کاوش ساختار وب فرآیند استخراج اطلاعات ساختاری از وب است و کاوش استفاده از وب کاربرد تکنیک های داده کاوی برای کشف الگوهای استفاده از وب به منظور درک و برآوردن بهتر نیازهای کاربران می باشد. در روش پیشنهادی ما از هر سه شاخه وب کاوی برای ارائه پیشنهادات کاراتر استفاده نموده ایم. تکنیکها و روش های وب کاوی از کاربرد وسیعی در حوزه های مختلف همچون شخصی سازی صفحات وب، تعیین ارتباط بین اسناد وب<sup>[۳]</sup>، بهبود در نتایج حاصل از الگوریتم های رتبه دهی و خود سازمانده کردن وب<sup>[۴]</sup> برخوردارند.

در [۵]، تکنیک های پیشنهاد بر مبنای منبع دانش به شش دسته تقسیم بندی شده اند: پالایش گروهی، مبتنی بر محتوا، مبتنی بر دانش، مبتنی بر اطلاعات دموگرافیک، مبتنی بر سوددهی، ترکیبی.

سیستم های مبتنی بر پالایش گروهی، موفق ترین سیستم های پیشنهاددهنده هستند که به صورت وسیعی در ساخت این سیستمهای کاربرده می شوند. در این روش پیشنهادات بر اساس شباهت میان پروفایل علاقه کاربر هدف - کاربری که قرار است به او پیشنهاد کنیم - با دیگر کاربران ارائه می شود. در واقع در این روش تمرکز بر روی یافتن شباهت بین کاربران است. رویکردهای موجود در [۶] از این روش جهت ارائه پیشنهادات استفاده نموده اند.

سیستم های مبتنی بر پالایش گروهی با مشکلات زیر روبرو هستند<sup>[۱۴]</sup>:

- مشکل شروع سرد: این روش هنگامیکه یک محصول جدیدی معرفی شود یا هنگامیکه کاربر جدیدی شروع به استفاده از سیستم کند به خوبی کار نمی کند بدلیل اینکه سیستم اطلاعات رتبه بندی زیادی روزی محصول و کاربر در اختیار ندارد.
- مشکل پراکندگی: هنگامیکه اقلام زیاد و کاربران نسبتاً کم هستند، پالایش گروهی نمی تواند به سادگی پیشنهادات مناسبی را فراهم سازد؛ بدلیل اینکه رتبه بندی های کمی در اختیار است.
- برخی کاربرانی که علایقشان متفاوت با علایق گروه است از پالایش گروهی سود نمیبرند.<sup>۴</sup>

در سیستم های مبتنی بر محتوا، اقلام پیشنهادی، نه بر اساس تشابه رفتاری کاربرفعال با دیگر کاربران، بلکه به دلیل شباهت با اقلامی که کاربر هدف نسبت به آنها ابزار علاقه کرده است، به کاربر توصیه می شوند. در این روش، تمرکز بر روی یافتن شباهت میان اقلام است. [۹,۸] از رویکردهای مبتنی بر محتوا جهت ارائه پیشنهادات استفاده نموده اند.

سیستم های مبتنی بر محتوا نیز چندین محدودیت اساسی دارند<sup>[۱۴]</sup>:

- در این سیستم ها فقط اطلاعات محتوای متنی می تواند گرفته شود نه محتوای سمعی یا بصری.
- فقط اقلام مشابه می توانند پیشنهاد شوند ( عدم توانایی در ارائه نتایجی متفاوت با اقلام پیشین<sup>۵</sup>).
- فقط علاقه کاربر هدف در این رویکرد استفاده می شود هر چند که علاقه کاربر ممکن است بوسیله سلایق دیگر کاربران مورد تاثیر واقع شود.

سیستم های مبتنی بر دانش، از بحث های اکتشافی و تلاش های انسان استفاده می کنند تا فاکتورهایی را که بر سلایق کاربران در انتخاب محصول تاثیر می گذارند را کشف نمایند. این سیستمها براساس ادراکی که از نیازهای مشتری و ویژگی های کالاها پیدا نموده اند، توصیه هایی را ارائه می دهند. به عبارتی در این گونه سیستم ها، مواد اولیه مورد استفاده برای تولید لیست پیشنهادات، دانش سیستم در مورد مشتری و کالا است. در سیستم های مبتنی بر اطلاعات دموگرافیک، از اطلاعاتی مثل سن، جنسیت یا شغل افراد جهت شناسایی نوع افرادی که به شیء خاصی توجه دارند، استفاده می شود و کاربران را بر اساس اطلاعات دموگرافیک دسته بندی کرده و با استفاده از کلاسهای دموگرافیک، پیشنهادها را تولید می کنند. این سیستم ها نیز اساساً از وابستگی و تشابه بین کاربران برای تولید پیشنهاد استفاده می کنند.

در سیستم های مبتنی بر سوددهی، پیشنهاد بر اساس ارزیابی سیستم از تطابق انتخابهای ممکن با نیاز کاربر انجام می شود. در واقع پیشنهادها بر اساس تابعی بیانگر میزان سوددهی هر آیتم برای کاربر انجام می شوند. مسئله اصلی در این سیستم ها، چگونگی طراحی و محاسبه تابع سوددهی برای هر کاربر خاص است.

کارایی یک سیستم پیشنهاددهنده به ساختار وب سایت و همچنین تکنیک خاصی که آن سیستم استفاده می کند، بستگی دارد. به علاوه کاربران استراتژی های حرکتی متفاوتی دارند، بنابراین تخمین بهترین مدل برای پیشنهاد کاری دشوار است. هر یک از رویکردهای پیشنهاد به تهایی، مزایا و البته محدودیت ها و کمبودهایی دارند. بنابراین، ترکیب روش های مختلف می تواند بر این نقاط ضعف و محدودیت ها غلبه نماید و کارایی سیستم پیشنهاددهنده را تا حدود زیادی بهبود بخشد. سیستم های پیشنهاددهنده ترکیبی، دو یا چند روش را برای بهبود نتایج پیشنهاد با یکدیگر ترکیب می نمایند [۱۰، ۱۱].

در [۱۲] سیستمی پیشنهاددهنده مبتنی بر گراف دو لایه برای کتابخانه های دیجیتال ارائه شده است. در این مقاله شباهت میان مشتریان با استفاده از اطلاعات دموگرافیک آنان بدست می آید و شباهت میان کتاب ها از طریق اطلاعات محتوا و خصوصیات کتاب محاسبه می شوند. سپس، ارتباطات میان دو لایه با استفاده از اطلاعات تراکنشی مشتریان بدست آمده و فرآیند پیشنهاد بر اساس روابط قوی میان مشتریان و کتابها ارائه می گردد.

در [۱۳]، یک سیستم پیشنهاددهنده مبتنی بر گراف ارائه شده است که در آن الگوهای هدایتی بازدید کنندگان قبلی وب سایت مورد استفاده قرار گرفته اند تا برای کاربران جدید تولید پیشنهاد کنند.

در این مقاله پس از معرفی معیاری نوین برای یافتن شباهت صفحات وب و شباهت کاربران، معماری جدیدی را جهت ترکیب دو رویکرد مبتنی بر محتوا و پالایش گروهی در یک مدل گراف دو لایه ارائه می نماییم. این مدل با ترکیب داده های استفاده کاربران و داده های ساختاری و محتوایی صفحات وب به ارائه پیشنهادات می پردازد. الگوریتم ارائه شده بر روی داده های واقعی شبیه سازی شده و نتایج نشان می دهد که دانش بدست آمده از سیستم ترکیبی به طور قابل ملاحظه ای کیفیت پیشنهادات را بهبود بخشیده است. در ادامه در بخش ۲ الگوریتم پیشنهادی ارائه می گردد. در بخش ۳ پس از معرفی مدل استفاده شده برای شبیه سازی، نتایج حاصل از شبیه سازی ارائه می شود. در پایان نیز نتیجه گیری آورده شده است.

## ۲. الگوریتم پیشنهادی

در این مقاله ما رویکردی مبتنی بر گراف دو لایه را جهت ترکیب دو روش مبتنی بر محتوا و پالایش گروهی معرفی می نماییم. این گراف شامل لایه کاربران و لایه صفحات وب است و ارتباطات کاربر-کاربر، صفحه وب-صفحه وب، و کاربر-صفحه وب را در بر می گیرد. هر گره در لایه صفحات وب، نشاندهنده یک صفحه وب و هر گره در لایه کاربران نشاندهنده یک کاربر است و یال های میان این گره ها نیز میزان ارتباط میان آنها را نشان می دهد. ارتباطات میان دو لایه نیز از طریق میزان علاقه هر کاربر به هر صفحه وب، بدست می آید [۱۲]. رویکردهای نوینی جهت یافتن این شباهت ها ارائه خواهد شد. شکل (۱)، این گراف دو لایه را نشان می دهد.

رویکرد ما برای ایجاد گراف دو لایه شامل مراحل محاسباتی زیر است:

۱. محاسبه شباهت صفحات وب و صفحات وب و ایجاد لایه صفحات وب.
۲. محاسبه شباهت میان کاربران و ایجاد لایه کاربران.
۳. ایجاد ارتباط میان لایه ای بر مبنای کاوش استفاده از وب.
۴. پارچه شدن بندی هر یک از لایه ها به صورت جداگانه.

۵. ارائه پیشنهادات بر اساس پارسیشن بندی صورت گرفته.



## ۱.۱. ایجاد لایه صفحات وب

برای اندازه‌گیری شباهت صفحات وب، ما از شباهت صفحات وب بر مبنای کلمات کلیدی و وزن هر صفحه وب داخل هر نشست استفاده می‌کنیم. جهت محاسبه شباهت صفحات وب بر مبنای کلمات کلیدی، هر صفحه وب را به صورت برداری از این کلمات نمایش می‌دهیم. بدین صورت که کلیه کلمات کلیدی داخل هر صفحه وب استخراج شده و وزن هر یک از آنها داخل هر صفحه بست می‌آید. بنابراین بردار  $p_i = (w_{i1}, w_{i2}, \dots, w_{in})$  نمایانگر یک صفحه وب است که  $w_{ij}$  نشانده‌نده تعداد دفعات تکرار کلمه کلیدی  $i$  در صفحه وب  $j$  است و  $n$  تعداد کل کلمات کلیدی کلیه صفحات وب است. در این کاربردها، غالباً از مدل ساده TF-IDF<sup>۹</sup> در مدل فضای بردار استفاده می‌شود. وزن کلمه کلیدی  $i$  در صفحه وب  $j$  از فرمول ارائه شده در [۱۵] بدست می‌آید.

بنابراین، شباهت مبتنی بر کلمه کلیدی صفحات وب به صورت زیر خواهد بود:

$$Similarity_{(keyword-based)}(P_i, P_j) = \frac{P_i \cdot P_j}{\|P_i\| \|P_j\|} \quad (1)$$

که مقادیر  $\|P_i\|$  و  $\|P_j\|$  برابر است با:

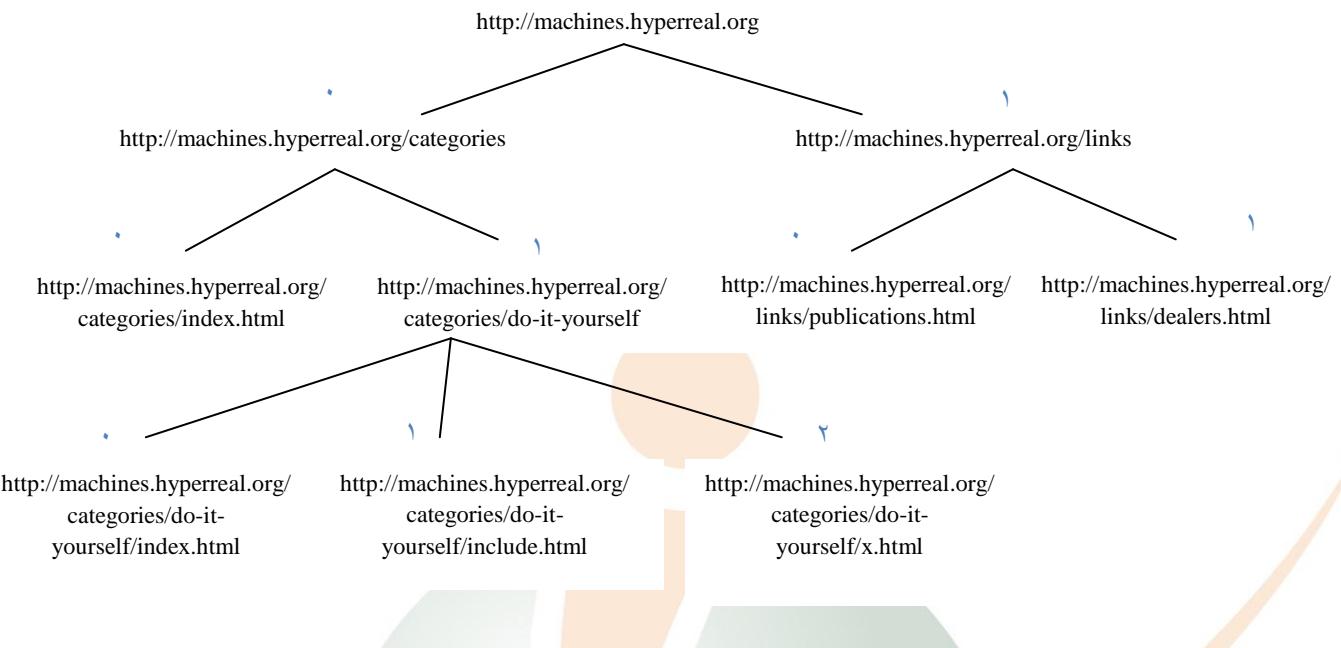
$$\|P_i\| = \sqrt{\sum_{m=1}^n (w_{im})^2}, \|P_j\| = \sqrt{\sum_{m=1}^n (w_{jm})^2} \quad (2)$$

حال به محاسبه شباهت موضوعی صفحات وب بر اساس جایگاه آنها در ساختار سلسله مراتبی وب سایت می پردازیم. با تجزیه و تحلیل صفحات وب، در می یابیم که صفحات به صورت تصادفی در وب سایت جای نگرفته اند، بلکه آنها در یک ساختار سلسله مراتبی که سلسله مراتب صفحات نامیده می شود، سازماندهی شده اند. سلسله مراتب صفحات، ساختاری از صفحات وب است که هر گره برگ در این ساختار، صفحه وی را نشان می دهد که متناظر با یک فایل در سرور است و هر گره غیربرگ در این ساختار نمایانگر یک دایرکتوری<sup>۱۰</sup> در سرور است؛ در این ساختار ما گره های برگ را "صفحات ساده" و گره های غیربرگ را "صفحات عمومی" می نامیم. برای مثال ما در این مقاله برای انجام آزمایشات از داده های واقعی وب سایت<sup>۱۱</sup> Music Machine استفاده کرده ایم. ساختار سلسله مراتبی برای برخی از صفحات این وب سایت در شکل (۲) آمده است. در این شکل، ۶ صفحه وب ساده و ۴ صفحه وب عمومی داریم و صفحه ریشه این ساختار همان صفحه ابتدایی<sup>۱۲</sup> وب سایت است. سلسله مراتب صفحات می تواند به صورت خودکار بر اساس URL صفحات ایجاد شود. برای مثال URL#۱، URL#۲ و URL#۳ را در جدول (۱) در نظر بگیرید. URL#۱ و URL#۲ از نظر موضوعی بسیار به یکدیگر شباهت دارند. ولی این دو URL با URL#۳ تفاوت بیشتری دارند. برای اندازه گیری شباهت موضوعی هر دو صفحه بر اساس جایگاهشان، ابتدا این ساختار سلسله مراتبی را کاراکتر گذاری نموده، سپس هر صفحه وب را به صورتی رشته ای از کاراکترها از ریشه تا خود صفحه نشان می دهیم. برای مثال در شکل (۲)، رشته مورد نظر برای URL#۱، "۰۰۱۰"؛ برای URL#۲، "۰۰۱۱" و برای URL#۳، "۰۰۱۱۰۰" خواهد بود.

URL شماره	آدرس URL	رشته استخراج شده از ساختار سلسله مراتبی
URL#۱	http://machines.hyperreal.org/categories/do-it-yourself/index.html	۰۰۱۰
URL#۲	http://machines.hyperreal.org/categories/do-it-yourself/include.html	۰۰۱۱
URL#۳	http://machines.hyperreal.org/categories/index.html	۰۰۱۱۰۰

جدول (۱): رشته های استخراج شده از ساختار سلسله مراتبی وب سایت

# کنفرانس داده کاوی ایران



شکل (۲): مثالی از سلسله مراتب صفحات برای بخشی از وب سایت Music Machine

گذشته از شباهت مبتنی بر کلمات کلیدی صفحات وب که در بخش قبلی مطرح شد، می توان با استفاده از سلسله مراتب صفحات وب که در بالا مطرح شد نیز، شباهت موضوعی صفحات را بدست آورد. در این قسمت بعد از تشکیل سلسله مراتب صفحات بر اساس URL آنها، محاسبه شباهت صفحات وب، بر مبنای مقایسه رشته های متناظر با این URL ها صورت می گیرد. برای مقایسه هر دو رشته باید دو گام محاسباتی صورت پذیرد [۱۶]:

۱. از ابتدای هر دو رشته کاراکترهای متناظر را با هم مقایسه نموده و این فرآیند تا جایی که نخستین جفت کارکترها با یکدیگر تفاوت داشته باشند، ادامه می دهیم.

۲. در میان این دو رشته، طول رشته بزرگتر را در نظر می گیریم؛ سپس به هر سطح از کاراکترها، وزنی اختصاص می دهیم. به آخرین سطح رشته بلندتر، وزن ۱، به سطح بعدی آن وزن ۲، و به همین ترتیب تا سطح ۱ عملیات وزن دهی را انجام می دهیم. سپس شباهت میان دو رشته به صورت جمع وزن های کاراکترهایی که با یکدیگر شبیه بوده اند تقسیم بر جمع وزن کلیه کاراکترها بدست می آید. اگر هر دو صفحه کاملاً متفاوت باشند یعنی هیچ کاراکتر مشابهی نداشته باشند شباهت آنها برابر صفر و اگر هر دو صفحه کاملاً مشابه باشند، یعنی همه کاراکترهای آنها با یکدیگر یکسان باشند، شباهت آنها برابر ۱ خواهد بود. به عنوان مثال برای URL#۱ و URL#۳ میزان شباهت به صورت زیر محاسبه می گردد.

وزن هر کاراکتر	۴	۳	۲	۱
رشته URL#۱ استخراج شده از ساختار سلسله مراتبی شکل (۲)	۰	۰	۱	۰
رشته URL#۳ استخراج شده از ساختار سلسله مراتبی شکل (۲)	۰	۰	۰	
شباهت URL#۱ و URL#۳				$۳+۴/۱+۲+۳+۴ = ۰,۷$

جدول (۲): محاسبه شباهت میان URL#۱ و URL#۳

حال با توجه با شباهت مبتنی بر کلمات کلیدی و مبتنی بر جایگاه صفحات وب، میانگین وزنی این دو شباهت به عنوان شباهت کلی صفحات وب در نظر گرفته، همچنین با توجه به پیوندهای موجود میان صفحات وب، کوتاهترین پیوند میان هر دو صفحه را برای اندازه گیری شباهت میان آن دو، در فرمول محاسبه شباهت وارد می کنیم. بنابراین فرمول کلی محاسبه شباهت صفحات وب به صورت زیر خواهد بود.

$$\text{Similarity}_{(Total)}(P_i, P_j) = \frac{1}{\text{LinkMin}(P_i, P_j)} * (\eta * \text{Sim}_{(\text{Keyword-based})}(p_i, p_j) + \lambda * \text{Sim}_{(\text{Position-based})}(P_i, P_j)). \quad (3)$$

## ۲.۲. ایجاد لایه کاربران

تشخیص نشست های کاربران در پروژه های وب کاوی کاربرد وب اهمیت ویژه ای دارد، چون نشست ها رفتار گردشی کاربران را نمایش می دهند و از این رو برای کشف الگو بسیار اهمیت دارند. نشست کاربر، مجموعه ای از صفحات شده ملاقات شده آن کاربر در طول یک بازدید خاص او از وب سایت است. مسئله محاسبه شباهت میان نشست های وب در این مقاله، با شباهت میان صفحات وب، شباهت توالی ها، زمان صرف شده روی هر صفحه وب، سایز صفحه وب و تعداد دفعات ملاقات آن صفحه وب در آن نشست روبرو خواهد بود. چنانچه زمان ملاقات بین دو صفحه متوالی بیش از ۳۰ دقیقه باشد، صفحه دوم را به عنوان شروع یک نشست جدید در نظر می گیریم. علاوه بر این چنانچه میزان شباهت میان دو صفحه متوالی از یک حد آستانه ای بیشتر باشد نشاندهنده اینست که احتمالاً کاربر دچار سردرگمی شده و از یک صفحه به صفحه دیگری که شباهتی به صفحه قبل نداشته مراجعه نموده است. بنابراین در این حالت نیز صفحه دوم را به عنوان شروع یک نشست جدید در نظر می گیریم. در نظر گرفتن این تعریف برای نشست ها موجب افزایش کارایی سیستم پیشنهاددهنده می شود.

برای بهبود کیفیت پیشنهادات ارائه شده، ما از اهمیت صفحات وب در نشست ها استفاده می نماییم. در واقع تمام صفحاتی را که کاربران مورد بازدید قرار می دهند برای آنها از اهمیت و ارزش یکسانی برخوردار نیستند. برخی از صفحاتی که کاربران آنها را بازدید می نمایند ممکن است اصلاً برای آنها جذاب نباشند. بنابراین ما سعی می کنیم تا درجه اهمیت هر صفحه را برای هر کاربر محاسبه نماییم؛ بدین ترتیب که هر نشست را به صورت برداری  $m$  بعدی از صفحات وب در نظر می گیریم ( $S = \langle (p_1, w_1), (p_2, w_2), \dots, (p_m, w_m) \rangle$ ، که  $w_i$  نشاندهنده وزن  $i$  امین ( $1 \leq i \leq m$ ) صفحه ای است که در نشست  $S$  بازدید شده است. وزن صفحات وب از طریق روشی که در [۱۷، ۱۸] مطرح شده است، محاسبه می شود. بنابراین شباهت نشست ها بر مبنای وزن صفحات وب در نشست ها به صورت زیر محاسبه می گردد:

$$\text{Similarity}(S_a, S_b) = \frac{S_a \cdot S_b}{\|S_a\| \|S_b\|}. \quad (4)$$

که مقادیر  $\|S_a\|$  و  $\|S_b\|$  برابر است با:

$$\|S_a\| = \sqrt{\sum_{l=1}^m (w_{al})^2}, \|S_b\| = \sqrt{\sum_{l=1}^m (w_{bl})^2} \quad (5)$$

در مرحله بعد، نشست ها را به صورت رشته ای از صفحات وب درنظر می گیریم و از فاصله لون اشتاین وزن دهی شده<sup>۱۳</sup>، برای یافتن بهترین حالت انطباق دو نشست استفاده می کنیم و از شباهت صفحات وب که در بخش قبل بدست آوردهیم به عنوان یکی از معیارهای شباهت نشست ها در فاصله لون اشتاین استفاده می نماییم [۲۰]. سپس میانگین وزنی شباهت نشست ها را، به عنوان شباهت کلی آن ها در نظر می گیریم.

## ۳. ایجاد ارتباطات میان لایه ای

پس از ایجاد لایه صفحات وب و لایه کاربران و ارتباطات میان آنها، به یافتن ارتباطات میان لایه ای می پردازیم. با توجه به نحوه حرکت کاربران در وب سایت می توان از اطلاعات استفاده از وب برای محاسبه شباهت میان این دو لایه بهره برد. با توجه به تعریف نشست در بخش قبل، می توان به

سادگی از طریق محاسبه وزن هر صفحه داخل هر نشست که در فرمول (۴) بیان شد، ارتباطات میان لایه ای را بدست آورد. هر یال در این گراف وزنی بین ۰ و ۱ خواهد داشت.

## ۴.۲. پارتبیشن بندی گراف

پارتبیشن بندی گراف، یک گراف را به چندین زیرگراف تقسیم می‌کند. هدف اصلی یک پارتبیشن بندی می‌نیم کردن هزینه‌های تعریف شده روی لبه‌های متصل پارتبیشن‌ها با کمترین برش لبه<sup>۱۴</sup>، می‌باشد. روش‌های زیادی برای تعیین هزینه خارجی پارتبیشن‌بندی گراف وجود دارد اما دو نکته که اغلب بطور وسیعی بکار برده می‌شوند، ماکریتم وزن لبه‌های بین رئوس که روش پارتبیشن‌های مختلف قرار می‌گیرد و وزن کلی همه لبه‌های متعلق به پارتبیشن‌های مشخص است. پارتبیشن بندی گراف جزء مسائل NP می‌باشد بهمین دلیل روش‌های مختلفی برای حل و بهینه کردن این مساله مطرح شده است [۱].

یکی از بهترین الگوریتم‌های پارتبیشن بندی گراف، الگوریتم مطرح شده در [۲۱] می‌باشد بطوریکه پارتبیشن بندی با یک پارتبیشن دلخواه شروع می‌گردد و سپس برای کاهش هزینه خارجی توسط یک سری از تغییرات داخلی در زیرمجموعه‌های پارتبیشن‌ها اقدام می‌کند این کار تا وقتی که دیگر امکان بهبودی وجود ندارد، ادامه می‌یابد. برای اجتناب از بهینه سازی محلی، الگوریتم بطور متوالی برای بدست آوردن تعدادی از پارتبیشن‌های بهینه محلی از میان آنها درخواست می‌شود تا بهترین پارتبیشن انتخاب شود.

در پارتبیشن بندی سنتی، پارتبیشن بندی بر روی کل گراف انجام می‌شود. این الگوریتم‌ها با افزایش رئوس بسیار کند بوده و پارتبیشن‌هایی با کیفیت پایین تولید می‌کردند. پارتبیشن بندی چند سطحی روشی کاملاً متفاوت نسبت به روش‌های سنتی دارند [۲۲، ۲۳، ۲۴]. روش کار این الگوریتم‌ها به این صورت می‌باشد که گراف را بوسیله فروریختن<sup>۱۵</sup> رئوس و لبه‌ها کاهش می‌دهند. پس از انجام این عمل پارتبیشن بندی بر روی گراف کوچک شده انجام می‌شود. پارتبیشن بدست آمده، کوچک‌تر از گراف اصلی می‌باشد. به همین منظور جهت بازیابی گراف اصلی عملیات بازیابی گراف انجام می‌گیرد که یک پارتبیشن برای گراف اصلی می‌سازد.

در واقع این نوع از الگوریتم‌های پارتبیشن بندی در ۳ فاز اجرا می‌شوند.

- در فاز اول سایز گراف و ب فرو ریختن رئوس و لبه‌ها کوچک می‌شود.<sup>۱۶</sup>
- در فاز دوم<sup>۱۷</sup> گراف تبدیل شده در فاز ۱ بر اساس روش‌های سنتی پارتبیشن بندی می‌گردد.
- در فاز سوم گراف به حالت اولیه تبدیل می‌شود که به آن فاز پالایش<sup>۱۸</sup> گفته می‌شود.

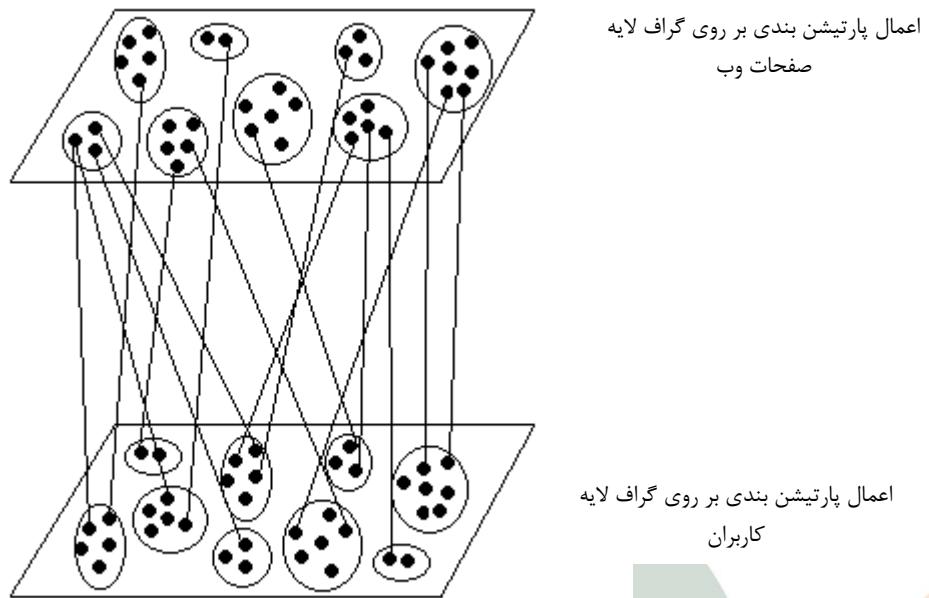
همچنین پارتبیشن بندی گراف توسط نرم افزار پارتبیشن بندی با استفاده از روش‌های چند سطحی انجام می‌شود. Metis از رویکردهای نوینی استفاده می‌کند تا به صورت پی در پی، سایز گراف را کاهش دهد و نیز، پالایش پارتبیشن را در طول فاز پالایش جلو برد. در طول فاز ۱، الگوریتم هایی را به کار می‌برد که یافتن پارتبیشن‌هایی با کیفیت بالاتر را در گراف حاصل از آن مرحله ساده تر سازد. در طول فاز پالایش Metis روی مز پارتبیشن نیست و عمدهاً روی بخشی از گراف که بسته است، می‌باشد. این الگوریتم‌ها به Metis این امکان را می‌دهند تا در سریعترین زمان، زیر گراف‌هایی با کیفیت بالا برای گراف‌های بزرگ ایجاد کند.

مزایای Metis در مقایسه با دیگر نرم افزارهای مشابه به صورت زیر است [۲۷]:

- فراهم کردن پارتبیشن‌هایی با کیفیت بالا: آزمایشات روی تعداد زیادی گراف‌هایی که در دامنه‌های مختلف شامل روش المان محدود<sup>۱۹</sup>، برنامه نویسی خطی<sup>۲۰</sup> VLSI<sup>۲۱</sup> بوجود آمده اند، نشان می‌دهد که پارتبیشن‌هایی که توسط Metis تولید می‌شوند، به طور حتم، بهتر از پارتبیشن تولید شده توسط روش‌هایی است که به صورت مکرر استفاده می‌شوند.
- سرعت اجرای بالا: آزمایشات روی محدوده وسیعی از گراف‌ها، نشان داده است که Metis در رتبه ۱ و ۲ سرعت عمل نسبت به دیگر الگوریتم‌هایی است که در پارتبیشن بندی گراف زیاد مورد استفاده قرار می‌گیرند. زمان اجرای Metis اغلب برابر یا حتی کمتر از زمان اجرای برخی الگوریتم‌های پارتبیشن بندی هندسی است که اغلب پارتبیشن‌های بدتری را تولید می‌نمایند.

در این مقاله، عملیات پارتبیشن بندی را همانند شکل (۳) بر روی هر دو لایه گراف اعمال می‌کنیم. در گام نخست، عملیات پارتبیشن بندی را روی لایه نخست گراف اعمال می‌کنیم و صفحات وب را پارتبیشن بندی می‌نماییم؛ بدین ترتیب صفحات وب شبیه به یکدیگر در یک پارتبیشن قرار می‌لایه

گیرند. در گام دوم، لایه دوم گراف را پارتبیشن بندی نموده، بنابراین کاربران مشابه نیز در یک دسته قرار می‌گیرند. سپس با توجه به این دسته بندی‌ها، فرآیند پیشنهاد صورت می‌گیرد.



شکل (۳): انجام پارتبیشن بندی گراف بر روی گراف صفحات وب و گراف کاربران

## ۵.۲. فرآیند پیشنهاد

با رویکرد گراف دو لایه، برای تولید پیشنهادات می‌توان از هر سه رویکرد پیشنهاد (پالایش گروهی، مبتنی بر محتوا، و روش ترکیبی)، از طریق انتخاب انواع مختلف پیوندها استفاده نمود. چنانچه فقط از اطلاعات صفحات وب استفاده نماییم، این نشانگر اینست که فقط لایه صفحات وب فعال شده است یعنی فقط از رویکرد مبتنی بر محتوا استفاده کرده ایم. اگر پیوندهای لایه کاربران و لایه میانی فعال شده باشند، یعنی رویکرد پالایش گروهی را به کار بدهد ایم؛ و اگر همه پیوندها فعال شوند یعنی از رویکرد ترکیبی استفاده نموده ایم.

برای انجام فرآیند پیشنهاد، در رویکرد مبتنی بر محتوا، با توجه به صفحاتی که قبلًا مورد بازدید کاربر قرار گرفته اند و پارتبیشن بندی صورت گرفته، صفحاتی که شباهت بیشتری به این صفحات دارند، به کاربر پیشنهاد می‌گردد.

در رویکرد پالایش گروهی نیز، با توجه به پارتبیشن بندی انجام شده، در ابتدا کاربرانی که مشابه با کاربر هدف هستند استخراج شده، سپس با توجه به پیوندهای میان لایه ای، صفحاتی که مورد علاقه این کاربران هستند، به کاربر هدف پیشنهاد می‌شود.

پیشنهادات ارائه شده در رویکرد ترکیبی از طریق ترکیب نتایج پیشنهادات دو روش مبتنی بر محتوا و پالایش گروهی بدست خواهد آمد.

## ۳. ارزیابی الگوریتم پیشنهادی

### ۱.۳. مدل شبیه سازی

دو روش عمده برای ارزیابی الگوریتم‌هایی که از اطلاعات پیمایش کاربران استفاده می‌کنند وجود دارد. روش اول، استفاده از صفحات وب واقعی و داده‌های واقعی کاربران وب موجود در فایل‌های ثبت رخداد سایتها می‌باشد. برای استفاده از این روش مجموعه داده‌های استانداردی که از چند سایت معتبر استخراج شده‌اند در دسترس می‌باشند. روش دوم مدل ارائه شده در [۱۹] می‌باشد. در این روش Luii و همکارانش نظم موجود در رفتارهای کاربران در محیط وب را با استفاده از یک مدل مبتنی بر عامل، مشخص و اعتبار مدل خود را با استفاده از اطلاعات استفاده از وب چندین سایت وب

بزرگ مانند مایکروسافت، تایید کرده‌اند. در این مقاله ما از داده‌های استاندارد سایت Music Machine استفاده می‌کنیم. این مجموعه داده، در سال ۱۹۹۷ جمع آوری شده است. امکان دسترسی کامل به تمامی اسناد و لاغ فایل‌های آن وجود دارد. هر وب سایت Music Machine حاوی اطلاعاتی در مورد انواع مختلف تجهیزات موسیقی الکترونیکی است که توسط کارخانه‌ها گروه‌بندی شده‌اند. هر فایل لاغ شامل برچسب کاربر<sup>۲۲</sup>، روش درخواست، URL مورد دسترسی قرار گرفته، پروتکل انتقال داده<sup>۲۳</sup>، زمان دسترسی، و مرورگری که برای دسترسی به سایت استفاده شده است، می‌باشد. فایل‌های لاغ سورپالایش شده‌اند و تمامی مدخل‌هایی که برای تحلیل نامعتبرند، حذف شده‌اند. اطلاعات مورد استفاده در این مقاله بر مبنای اطلاعات ۱ هفته فایل‌های لاغ این وب سایت از تاریخ ۱۶-۰۵-۱۹۹۷ ماه فوریه سال ۱۹۹۷ می‌باشد. داده‌های پالایش شده شامل ۱۶۴۲۷ نشست و ۸۹۲ صفحه وب می‌باشند. مجموعه نتایج تراکنش‌ها به دو مجموعه آموزش (۰/۸۰)، و مجموعه آزمایش (۰/۲۰) تقسیم شده‌اند. نتایج شبیه سازی‌ها نشان داد که پیشنهاده‌دهنده ترکیبی نتایج دقیق تر و پیشنهادات مناسب تری را نسبت به دو روش مبتنی بر محتوا و پالایش گروهی ارائه داده است.

## ۲.۳. معیار و متولوژی ارزیابی

با در دست داشتن عملکرد سیستم، قدم بعدی تعیین معیارهای برای ارزیابی سیستم است. در این قسمت از معیارهای مطرح شده در کارهای مشابه مثل [۲۶، ۲۵] برای ارزیابی سیستم استفاده می‌کنیم. با توجه به شباهت صورت مسئله ما با مقالات یاد شده، این معیارها می‌توانند جهت ارزیابی چنین سیستمی مناسب باشند. در مقالات یاد شده، دو معیار پوشش<sup>۲۴</sup> و دقت<sup>۲۵</sup> معرفی می‌شوند. این دو معیار، بسیار شبیه به معیارهای متداول در بازیابی اطلاعات یعنی فراخوانی<sup>۲۶</sup> و دقت<sup>۲۷</sup> بازیابی اسناد هستند.

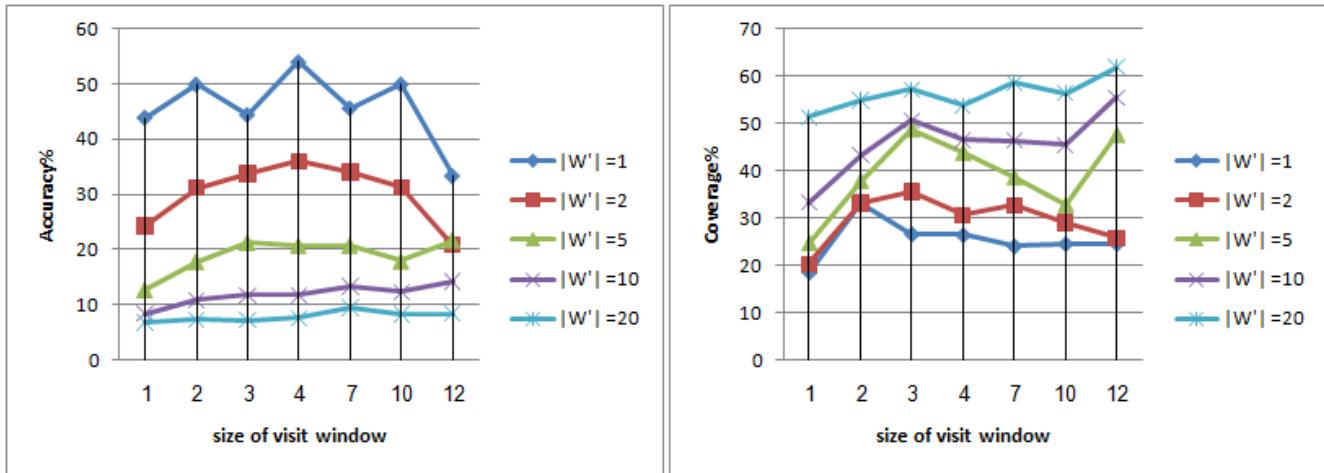
دقت پیشنهادها برابر با "نسبت پیشنهادهای درست به کل پیشنهاد" است. منظور از پیشنهاد درست، پیشنهادی است که با توجه به بخش دیده شده (پیشوند) یک جلسه کاربر تولید شده و در ادامه جلسه کاربر (پسوند) رخ دهد. پوشش پیشنهادها، قدرت سیستم در پیش بینی تمام صفحاتی که ممکن است مورد نظر کاربران باشد را اندازه گیری می‌کند.

مسلمًا در عملکرد هر سیستم پیشنهادگر، رابطه‌ای بین معیارهای دقت و پوشش برقرار بوده و نتایج عملکرد روش پیشنهاد، در رابطه با این دو معیار، از هم مستقل نیست. مثلاً، اگر ما در هر قدم تعداد صفحات بسیار زیاد، یا در حالت فرضی، کل صفحات سایت را پیشنهاد کنیم مسلمًا پوشش بسیار بالا (یا ۱۰۰٪) خواهیم داشت؛ در طرف مقابل در این حالت مسلمًا دقت بسیار پایینی نصیب روش پیشنهاد خواهد شد. از این رو، بررسی هر یک از این دو معیار، بدون توجه به دیگری، چنان‌با معنی نمی‌باشد در نتیجه از هر دو معیار به صورت همزمان برای ارزیابی سیستم استفاده می‌شود.

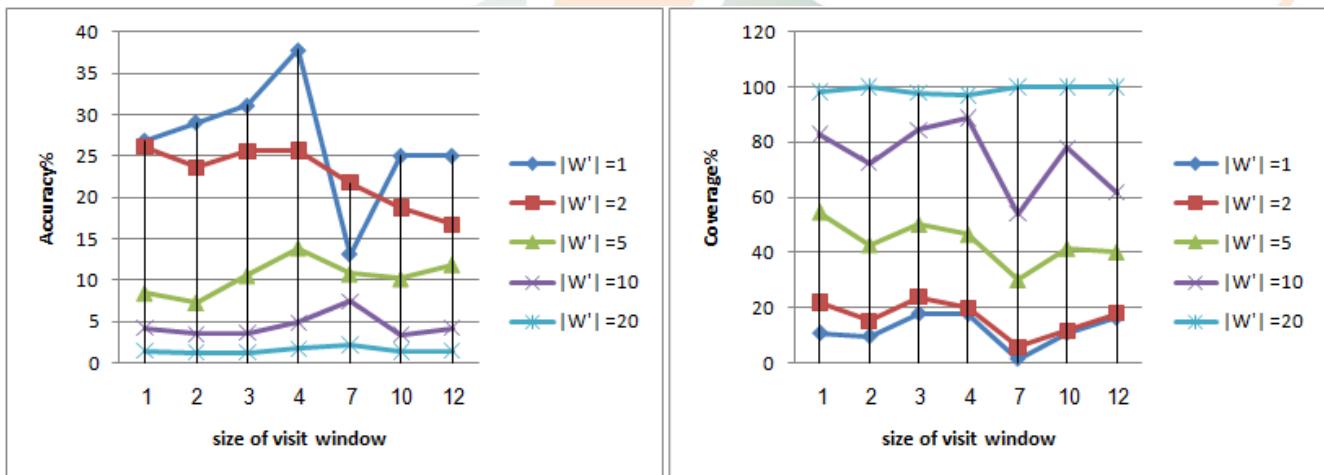
## ۳.۳. نتایج شبیه سازی

ما رویکرد پیشنهادیمان را تحت شرایط مختلف آزمایش می‌نماییم. در اولین مجموعه از آزمایش‌ها، تاثیر انتخابهای مختلف در رابطه با پارامترهای دخیل در سیستم را بررسی می‌کنیم. پارامترهای تاثیرگذار در کارایی الگوریتم عبارتند از: اندازه پنجره بازدید<sup>۲۸</sup> ( $|W|$ ) و اندازه پنجره پیشنهاد<sup>۲۹</sup> ( $|W'|$ ) یا تعداد صفحات پیشنهاد شده. ما تاثیر یک از این پارامترها را در کارایی سیستم پیشنهادیمان نشان می‌دهیم.

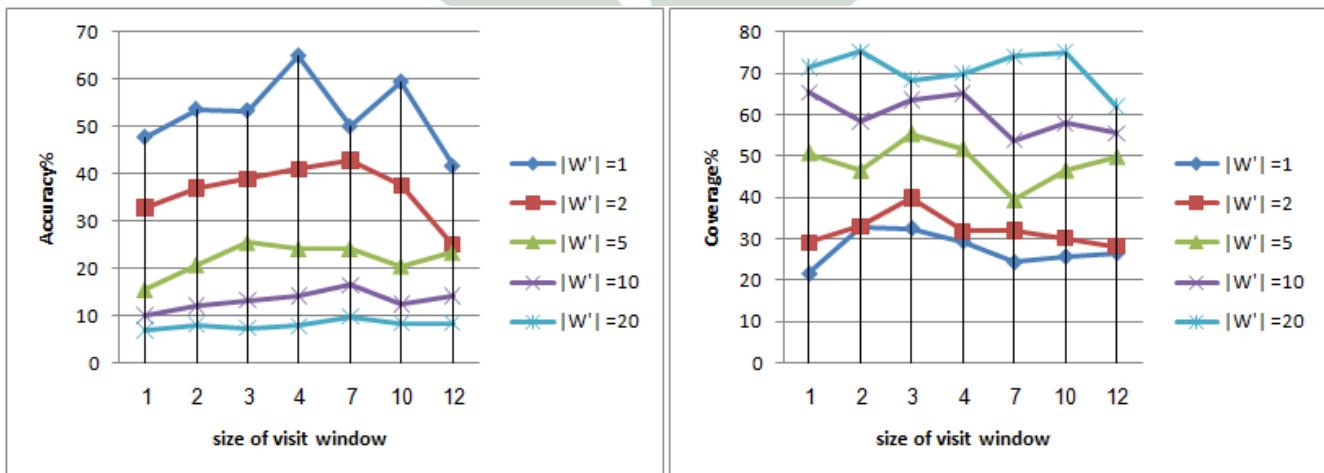
برای بررسی تاثیر اندازه پنجره بازدید، ما اندازه پنجره بازدید را از ۱ تا ۱۲ و اندازه پنجره پیشنهاد را از ۱ تا ۲۰ تغییر می‌دهیم. (مشابه آزمایشاتی که در [۱۳] انجام شده است). نتایج شبیه سازی در اشکال ۶-۶ نشان داده شده است. این اشکال دقت و پوشش رویکردهای پیشنهادی ما، یعنی رویکرد مبتنی بر محتوا، پالایش گروهی و روش ترکیبی را نشان می‌دهند. برای نشان دادن تاثیر اندازه پنجره بازدید و اندازه پنجره پیشنهاد ما این پارامترها را همانگونه که در اشکال ۶-۶ نشان داده شده است، تغییر می‌دهیم. نتایج نشان می‌دهد که بهترین دقت زمانی است که اندازه پنجره بازدید ۱ و بهترین پوشش زمانی است که اندازه پنجره بازدید ۲۰ است.



شکل (۴): میانگین دقت و پوشش در رویکرد مبتنی بر محتوا برای اندازه های مختلف پنجره بازدید

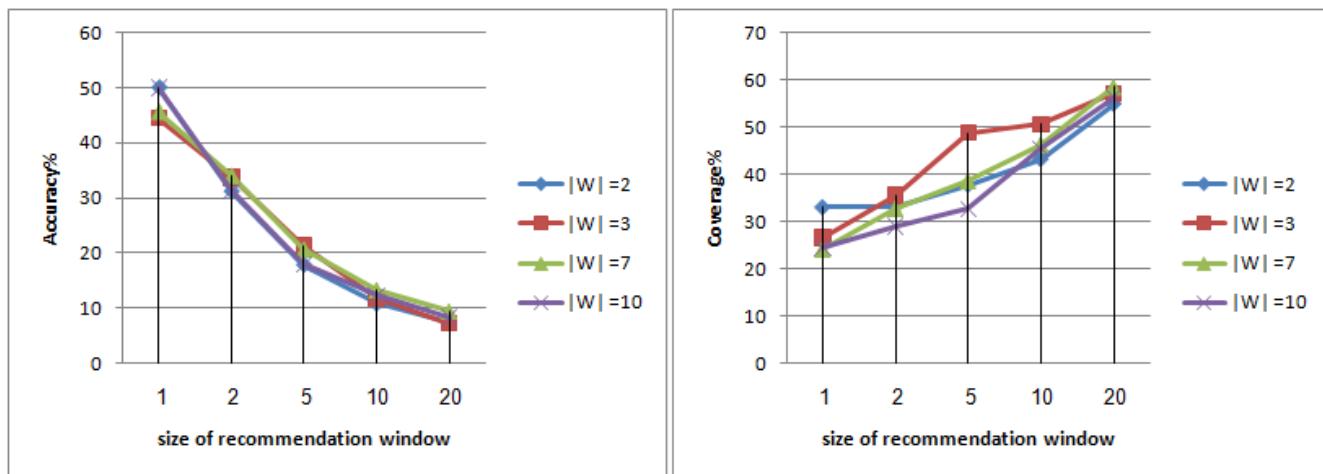


شکل (۵): میانگین دقت و پوشش در رویکرد پالایش گروهی برای اندازه های مختلف پنجره بازدید

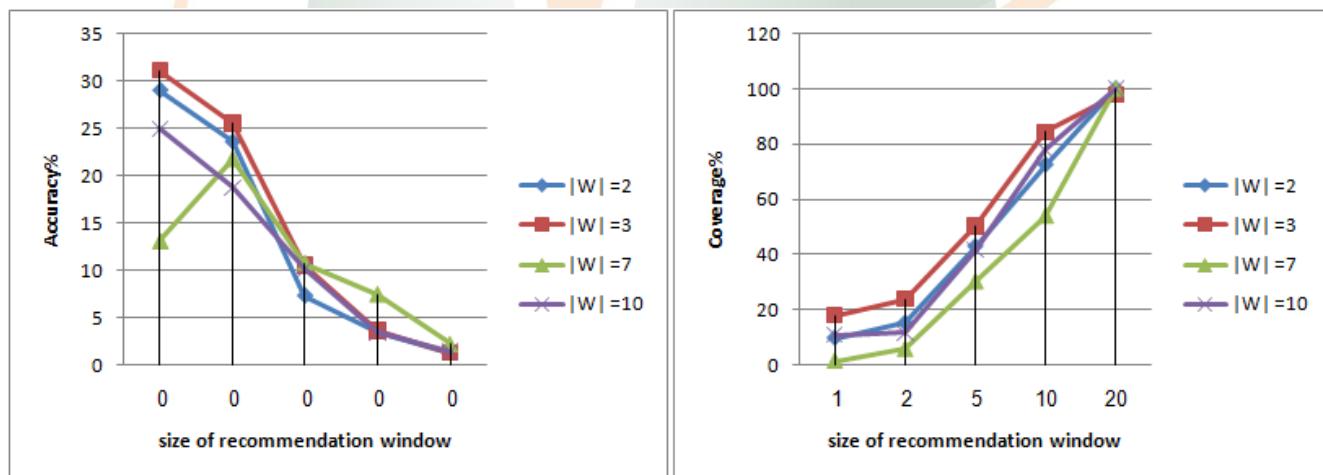


شکل (۶): میانگین دقت و پوشش در رویکرد ترکیبی برای اندازه های مختلف پنجره بازدید

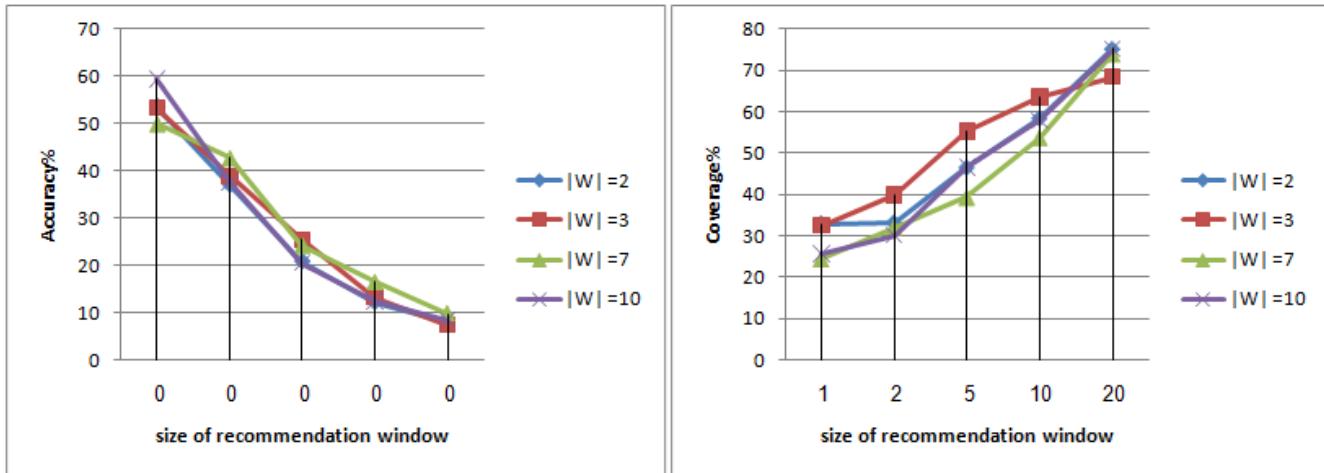
حال، برای بررسی تاثیر اندازه پنجره پیشنهاد، آزمایشات جدیدی انجام دادیم. اشکال ۹-۷ نتایج این آزمایشات را نشان می‌دهند. این آزمایشات نشان می‌دهند که با کاهش اندازه پنجره پیشنهاد دقیق‌تر از افزایش می‌یابد. بهترین کارایی زمانی است که اندازه پنجره پیشنهاد ۱ است و هنگامی که از اندازه پنجره پیشنهاد بزرگ‌تر از ۵ استفاده می‌کنیم میزان دقیق‌تر از این میزان می‌یابیم؛ این بدین معناست که، مقدار بزرگ  $W$  پیشنهادات نامطلوب تری را به کاربر هدف ارائه می‌نماید. همچنین در این اشکال با افزایش اندازه پنجره پیشنهاد، پوشش افزایش می‌یابد. هنگامیکه فقط یک صفحه وب پیشنهاد می‌شود پوشش در حداقل مقدار است.



شکل (۷): میانگین دقیق و پوشش در رویکرد مبتنی بر محتوا برای اندازه‌های مختلف پنجره پیشنهاد

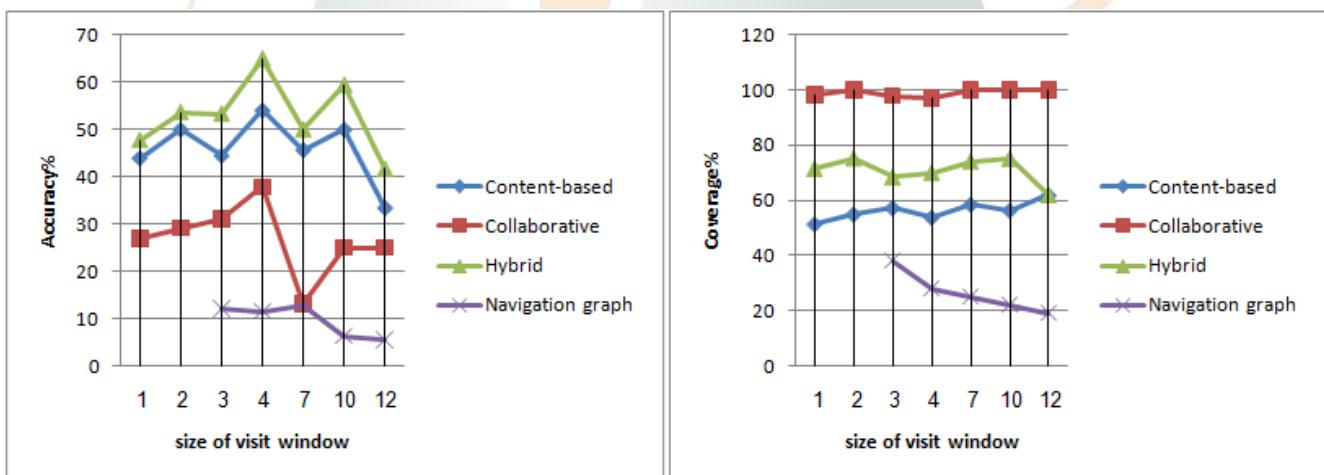


شکل (۸): میانگین دقیق و پوشش در رویکرد پالایش گروهی برای اندازه‌های مختلف پنجره پیشنهاد



شکل (۹): میانگین دقت و پوشش در رویکرد ترکیبی برای اندازه های مختلف پنجره پیشنهاد

مقایسه رویکردهای پیشنهادی نشان داد که رویکرد ترکیبی در مقایسه با رویکردهای مبتنی بر محتوا و پالایش گروهی از کارایی بالاتری برخوردار است. در آخرین مراحل آزمایشات ما کارایی سیستم های پیشنهادیمان را با کارایی یک سیستم پیشنهاددهنده مبتنی بر گراف مقایسه می کنیم.<sup>[۱۳]</sup> شکل (۱۰) مقایسه سیستم های پیشنهادی ما را با این سیستم نشان می دهد. نتایج آزمایشات نشان داد که، رویکرد ترکیبی کارایی قابل توجهی نسبت به رویکرد مبتنی بر محتوا، پالایش گروهی و رویکرد مبتنی بر گراف در [۱۳] دارد.



شکل (۱۰): مقایسه دقت و پوشش در رویکردهای مبتنی بر محتوا، پالایش گروهی، ترکیبی و روش [۱۳]

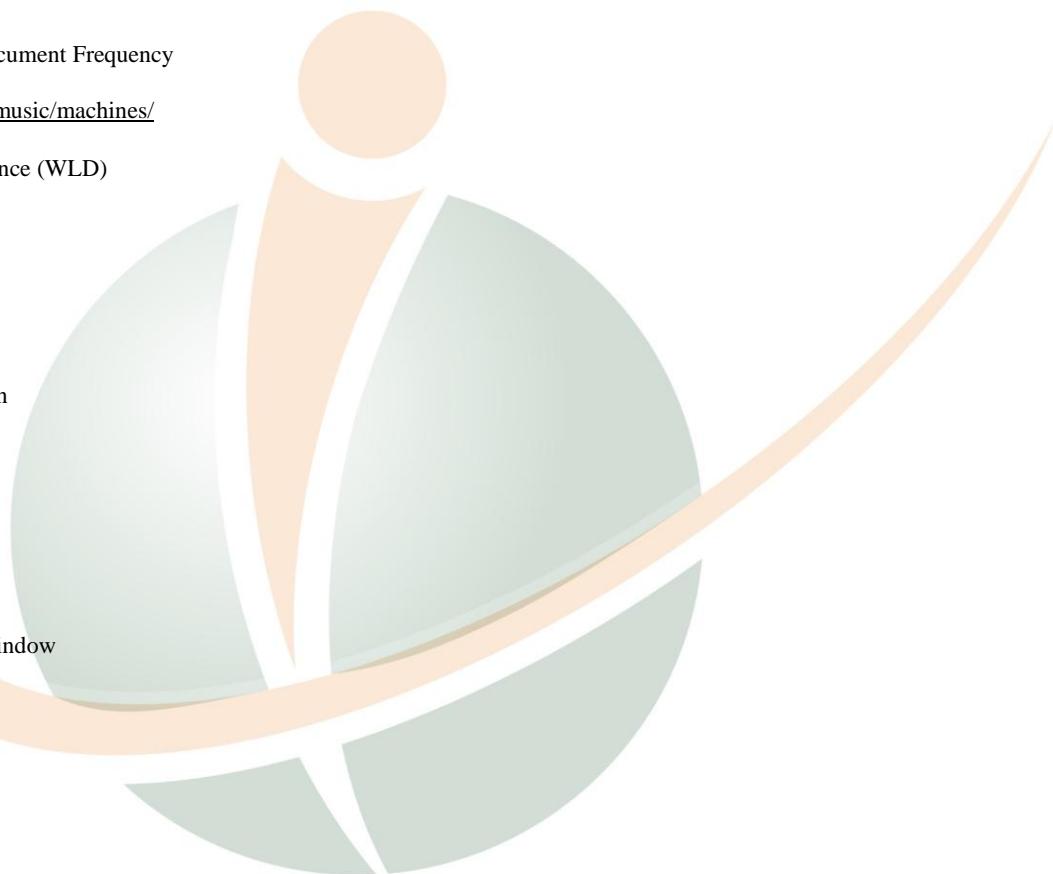
#### ۴. نتیجه گیری

در این مقاله، ما یک رویکرد گراف دو لایه را شرح دادیم که از ترکیب دو روش معمول سیستم های پیشنهاددهنده یعنی مبتنی بر محتوا و پالایش گروهی استفاده می کند. استفاده همزمان از اطلاعات صفحات وب و اطلاعات کاربران سهم به سزاپی در بهبود نتایج روش ترکیبی داشت. روش های نوبنی جهت یافتن شباهت میان صفحات وب و کاربران ارائه شد سپس بر اساس این شباهت ها، عملیات پارتبیشن بندی روی هر دو سطح گراف صورت گرفت. بر مبنای این پارتبیشن بندی صفحاتی که شبیه به صفحات مورد علاقه کاربر هدف هستند به وی پیشنهاد شد و همچنین صفحاتی که مورد علاقه کاربرانی شبیه به این کاربر هدف هستند نیز به او پیشنهاد گردید و باعث بالا رفتن کارایی سیستم پیشنهاددهنده شد. در این رویکرد از سه شاخه وب کاوی، یعنی کاوش محتوای وب، کاوش ساختار وب و کاوش استفاده از وب استفاده شده است.

## ۵. مراجع

- [۱] معتمدی مهر، شهرزاد، تاران، مجید، برادران هاشمی، علی، میبدی، محمدرضا؛ "سیستم پیشنهادهندۀ وب با استفاده از اتوماتای یادگیر توزیع شده و پارتبیشن بندی گراف"، چهارمین کنفرانس داده کاوی ایران، تهران، ایران، ۱۳۸۹.
- [۲] R. Kosalay, and H. Blockeel; "Web Mining Research: A Survey", SIGKDD Explorations Newsletter of the ACM Special Interest Group on Knowledge Discovery and Data Mining, vol. ۲ (۱), pp. ۱-۱۰, ۲۰۰۰.
- [۳] F. Heylighen, and J. Bollen; "Hebbian Algorithms for a Digital Library Recommendation System", International Conference on Parallel Processing Workshops (ICPPW'۲), ۲۰۰۲.
- [۴] J. Zhu, J. Hong, and J. Hughes; "Mining Conceptual Link Hierarchies from Web Log Files for Adaptive Web Site Navigation", ACM Transactions on Internet Technology (TOIT), ۲۰۰۳.
- [۵] P. Resnick, and H.R. Varian; "Recommender Systems", Communications of the ACM, vol. ۴۰ (۳), pp. ۵۶-۵۸, ۱۹۹۷.
- [۶] B. Sarwar, G. Karypis, J. Konstan, J. Riedl; "Item-based Collaborative Filtering Recommendation Algorithms", Proceedings of the ۱۰th international conference on World Wide Web, ACM New York, ۲۰۱۰.
- [۷] D. Lemire; "Scale and Translation Invariant Collaborative Filtering Systems", Journal of Information Retrieval, pp. ۱۲۹-۱۵۰, ۲۰۰۳.
- [۸] C. Basu, H. Hirsh, and W. Cohen; "Recommendation as Classification: Using Social and Content-based Information in Recommendation", Proceedings of the ۱۰th National Conference on Artificial Intelligence, Madison, ۱۹۹۸.
- [۹] M. Bogdanov, M. Haro, F. Fuhrmann, E. Gomez, and P. Herrera; "Content-based Music Recommendation Based on Use Preference Example", The 4th ACM Conference on Recommender Systems. Workshop on Music Recommendation and Discovery, ۲۰۱۰.
- [۱۰] R. Brucke; "Hybrid Recommender Systems", School of Computer Science, Telecommunications and Information Systems, Springer Berlin Heidelberg, ۲۰۰۸, pp. ۳۷۷-۴۰۸.
- [۱۱] M. Goksedef, and S. Gunduz-Oguducu; "Combination of Web Page Recommender Systems", Journal of Expert Systems with Applications, USA, ۲۰۱۰.
- [۱۲] Z. Huang, W. Chung, T.H Ong, and H. Chen; "A Graph-based Recommender System for Digital Library", ACM/IEEE Joint Conference on Digital Libraries, ۲۰۰۲, pp. ۶۵-۷۳.
- [۱۳] Y. Wang, W. Dai, and Y. Yuan; "Website Browsing Aid: A Navigation Graph-based Recommendation System", Journal Decision Support systems, Elsevier Science Publishers B. V. Amsterdam, Netherlands, ۲۰۰۸.
- [۱۴] Z. Huang, W. Chung, and H. Chen; "A Graph Model for E-Commerce Recommender Systems", Journal of the American society for information science and technology, ۲۰۰۴, pp. ۲۰۹-۲۱۴.
- [۱۵] R. Forsati, M.R. Meybodi, M. Mahdavi, and A.G. Neiat; "Hybridization of K-means and Harmony Search Methods for Web Page Clustering", IEEE International Conference on Web Intelligence and Intelligent Agent Technology, ۲۰۰۸.
- [۱۶] W. Wang, and O. R. Zaiane; "Clustering Web Sessions by Sequence Alignment", ۱۸th International Workshop on Database and Expert Systems Applications, University of Alberta, ۲۰۰۲.
- [۱۷] R. Forsati, and M.R. Meybodi; "Effective Web Page Recommendation Algorithms Based on Distributed Learning Automata and Weighted Association Rules", Journal of Expert Systems with Applications, ۲۰۱۰, pp. ۱۳۱۶-۱۳۳۰.
- [۱۸] M. Talabeigi, R. Forsati, and M. R. Meybodi; "A Hybrid Web Recommender System Based on Cellular Learning Automata", grc, ۲۰۱۰ IEEE International Conference on Granular Computing, ۲۰۱۰, pp. ۴۰۳-۴۰۸.
- [۱۹] J. Liu, S. Zhang, and J. Yang; "Characterizing Web Usage Regularities with Information Foraging Agents", IEEE Transactions on Knowledge and Data Engineering, vol. ۱۶ (۴), pp. ۵۶۶-۵۸۴, ۲۰۰۴.
- [۲۰] H. M. Doustdar, R. Forsati, M. R. Meybodi, and M. Shamsfard; "A Bi-section Graph Approach for Hybrid Recommender System", grc, ۲۰۱۱ IEEE International Conference on Granular Computing, ۲۰۱۱.
- [۲۱] B. W. Kernighan, and S. Lin; "An Efficient Heuristic Procedure for Partitioning Graphs", The Bell System Technical journal, vol. ۴۹ (۲), p. ۲۹۱-۳۰۸, ۱۹۷۰.
- [۲۲] C. K. Cheng, and Y. C. A Wei; "An Improved two-way Partitioning Algorithm with Stable Performance", IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. ۱۰ (۱۲), p. ۱۵۰۲-۱۵۱۱, ۱۹۹۱.
- [۲۳] L. Hagen, and A. B. Kahng; "A New Approach to Effective Circuit Clustering", In Proceedings of the IEEE/ACM International Conference on Computer-Aided Design, Santa Clara, CA, USA, ۱۹۹۲.
- [۲۴] T. N. Bui, and C. A. Jones; "Heuristic for Reducing Fill in Sparse Matrix Factorization", In Proceedings of the Sixth SIAM Conference on Parallel Processing for Scientific Computing, Norfolk, Virginia, USA, ۱۹۹۳.
- [۲۵] B. Mobasher, R. Cooley, and J. Srivastava; "Automatic Personalization based on Web Usage Mining", Communications of the ACM, vol. 43 (4), pp. 142-151, ۲۰۰۰.
- [۲۶] J. Li, and O. R. Zaiane; "Combining Usage, Content and Structure Data to Improve Web Site Recommendation", ۵th International Conference on Electronic Commerce and Web, ۲۰۰۴.
- [۲۷] G. Karypis, and V. Kumar; "METIS: A Software Package for Partitioning Unstructured Graphs, Partitioning Meshes, and Computing Fill-Reducing Orderings of Sparse Matrices", Minneapolis, ۲۰۰۷.

- 
- <sup>۱</sup> Recommender system
  - <sup>۲</sup> Cold Start Problem
  - <sup>۳</sup> Sparsity Problem
  - <sup>۴</sup> Gray Sheep Problem
  - <sup>۵</sup> Over-Specification Problem
  - <sup>۶</sup> flexible
  - <sup>۷</sup> comprehensive
  - <sup>۸</sup> modular
  - <sup>۹</sup> Term Frequency Inverse Document Frequency
  - <sup>۱۰</sup> directory
  - <sup>۱۱</sup> <http://www.hyperreal.org/music/machines/>
  - <sup>۱۲</sup> Home Page
  - <sup>۱۳</sup> weighted Levenshtein distance (WLD)
  - <sup>۱۴</sup> minimum edge-cut
  - <sup>۱۵</sup> collapsing
  - <sup>۱۶</sup> coarsening Phase
  - <sup>۱۷</sup> initial Partitioning Phase
  - <sup>۱۸</sup> finite element methods
  - <sup>۱۹</sup> linear programming
  - <sup>۲۰</sup> coarsening Phase
  - <sup>۲۱</sup> Very large scale integration
  - <sup>۲۲</sup> user label
  - <sup>۲۳</sup> data transmission protocol
  - <sup>۲۴</sup> coverage
  - <sup>۲۵</sup> accuracy
  - <sup>۲۶</sup> recall
  - <sup>۲۷</sup> precision
  - <sup>۲۸</sup> size of visit window
  - <sup>۲۹</sup> size of recommendation window



# کنفرانس داده کاوی ایران