بررسی و تحلیل فاکتورهای اساسی در طراحی سیستم‌های توصیه‌گر

زهرا مرادی‌منش
عضو گروه پژوهشی ITBM پژوهشکده ICT جهاددانشگاهی
Moradimanesh.zahra@gmail.com

محمددرزی
عضو گروه پژوهشی ITBM پژوهشکده ICT جهاددانشگاهی
Modarzi@yahoo.com

حبیب اله اصغری
عضو هیات علمی جهاددانشگاهی
تهران،ایران
Asghari@itincubator.com

چکیده

در دنیای جدید اگرچه اینترنت حجم فراوانی از داده‌ها را به عنوان فرصتی مناسب پیش روی کاربران قرارداده است، اما در صورت نبود مدیریتی کارآمد بر روی انبوه داده‌های در دسترس، این امتیاز خود مانعی برای پیشرفت خواهد بود. به طوری که امروزه با توجه به حجم روزافزون داده و اطلاعات، نیاز به سیستم‌هایی که توانایی هدایت کاربران به سمت کالا و سرویس مورد نظر را داشته باشند بیش از پیش احساس می‌شود. سیستم‌های توصیه‌گر سیستم‌های هوشمندی هستند که در فضای اینترنت با شناسایی علایق و اولویت‌های کاربر، اطلاعات موجود را پالایش کرده و پیشنهادات مناسب و مرتبط را به تک تک کاربران ارائه می‌کنند. سیستم‌های توصیه گر ابزاری برای هرچه توانمند کردن شهروندان در بهره برداری از فضای وب محسوب می شوند. با استفاده از سیستمهای توصیه گر، امکان جستجو به دنبال مفاهیمی وجود دارد که در جستجوی عادی داده‌ها دسترسی به آنها میسر نیست.

در این مقاله پس از شرح مقدمه‌ای بر مفاهیم بنیادی سیستم‌های توصیه‌گر، به تعریف و توصیف ویژگی‌های آن‌ها خواهیم پرداخت. در ادامه و طی یک طبقه‌بندی، هشت فاکتور اساسی در طراحی، ساخت و به روز رسانی این سیستم‌ها از دیدگاه ساختاری مورد بررسی قرار گرفته و حدود ده سیستم پیاده‌سازی شده نیز از منظر این طبقه بندی مورد بررسی و تحلیل قرار خواهد گرفت.

واژگان کلیدی

سیستم‌های توصیه گر، پالایش اطلاعات، انباشت اطلاعات، پروفایل کاربر، ساخت پروفایل، نگهداری پروفایل، استخراج اطلاعات پروفایل

۱- مقدمه

دنیای الکترونیکی در حال حرکت به سمت اشباع اطلاعاتی است. در طول دهه گذشته حجم زیادی از داده‌ها بر روی سرورهای اطلاعاتی و پایگاه‌های داده ذخیره شده اند . امروزه میزان داده‌های در دسترس در هر ۵ سال دو برابر می‌شود و با توجه به تنوع زیاد اطلاعات، بازارها، خدمات و محیط‌های کسب و کار، دسترسی به داده‌های مناسب برای تصمیم‌گیری صحیح امری ضروری تلقی می‌شود [1] . این حجم بالای داده روی اینترنت و افزایش روز افزون آن کاربران را با مشکل یافتن اطلاعات و کالاهای مناسب در زمان مناسب مواجه کرده است و یافتن داده و اطلاعات نهایی متناسب با نیاز کاربران تبدیل به فرآیندی پیچیده و زمانگیر شده است. در واقع همگام با افزایش انتخابها، حجم اطلاعاتی که باید برای رسیدن به هدف پردازش شوند و به دنبال آن میزان زمان و انرژی مصرفی برای رسیدن به اطلاعات نهایی به طور چشمگیری بالا می رود. در چنین محیط هایی وجود سیستمی با قابلیت شناخت و به روز رسانی علایق و اولویت‌های کاربران از یک سو و توانایی شاخص گذاری و

Archive of SID
www.SID.ir

ذخیره اطلاعات به روشی قابل جستجو[1] از سوی دیگر، برای پیش‌بینی و تشخیص نیازهای کاربر وهدایت او در جهت یافتن کالاهای متناسب با آن شدیداً احساس می‌شود.

سیستم‌های توصیه‌گر از جمله ابزارهایی هستند که می‌توانند کاربران را در محیطهای الکترونیکی به سمت یافتن اطلاعات، خدمات و کالاهای مورد علاقه هدایت کنند. سیستم‌های توصیه‌گر با توانایی شناخت کاربر و پیش‌بینی اولویت‌های او، اطلاعاتی که احتمال می‌رود مورد توجه کاربر باشد را از بین حجم بالای داده‌ها غربال کرده و با پیشنهاد آنها به کاربر در وقت و انرژی او صرفجویی می‌کنند. از طرف دیگر این سیستم‌ها با توانایی تحلیل رفتارهای گذشته کاربر، خدمات و اطلاعات موجودی را که وی توجهی به آن نداشته ولی احتمالا علاقه‌مند به آن است نیز استنتاج کرده و نتایج جالب توجهی به کاربران پیشنهاد می‌دهد [5].

در واقع سیستم‌های توصیه‌گر یکی از ابزارهای اصلی برای غلبه بر مشکل انباشت اطلاعات[2] بوده [4] و با داشتن قدرت تحلیل رفتارهای کاربر، مکمل هوشمندی برای مفاهیم بازیابی و غربال اطلاعات هستند.

در سال‌های اخیر تعداد زیادی سیستم توصیه گر برای شناسایی کاربران، پیش‌بینی علایق آنها و ارائه‌ی پیشنهادات، طراحی و پیاده‌سازی شده اند. در هر یک از این سیستم‌ها با توجه به دامنه کاری و اهداف، مجموعه ای از تکنیک‌های ساخت، به روز رسانی و استخراج داده‌ها به کار گرفته شده اند ولی محور اساسی در تمامی این سیستم‌ها پروفایل کاربر[3] است. چگونگی ساخت پروفایل [4]و منبع داده‌ای [5]که در ساخت پیشنهاد استفاده خواهد شد، پروفایل پیش فرض سیستم [6]برای کاربران، چگونگی به روز رسانی اطلاعات پروفایل [7]و منبع این به روز رسانی[8] و در نهایت تکنیک‌های استخراج و استفاده از اطلاعات پروفایل [9]فاکتورهایی هستند که در طراحی یک سیستم توصیه گر جایگاه مهمی دارند [6] در طبقه‌بندی ارائه شده در این مقاله هر یک از این مباحث به عنوان یکی از ستونهای جدول نهایی در نظر گرفته شده‌اند. طراح و تحلیل گر سیستم هنگام طراحی، ساخت و به‌هنگام سازی سیستم، باید با داشتن دانش کافی در مورد تکنیک‌های موجود در هر یک از مطالب، تصمیمات مناسبی در راستای انتخاب آن اتخاذ نماید.

امروزه سیستم‌های توصیه‌گر در زمینه‌های متنوعی از غربال کردن اخبار موجود در وب متناسب با نیازها و خواسته‌های کاربر تا تجارت الکترونیکی، پیشنهاد فیلم، موسیقی، کتاب، مقاله و اطلاعات کاربرد دارند. با توجه به جایگاه این ابزار در دنیای الکترونیکی امروز و با در نظر گرفتن نیازهای کاربران برای دریافت اطلاعات مناسب در مکان مناسب و در زمان مناسب ([5])، مقاله‌ی حاضر تلاش دارد گامی در جهت معرفی سیستم‌های توصیه‌گر به عنوان ابزاری برای توانمندسازی کاربران و در نگاه کلان شهروندان الکترونیکی در یافتن خدمات، کالا، اسناد و اطلاعات مورد نظر خود از میان اطلاعات موجود به صورت بهینه، بردارد.

۲- تاریخچه

ساخت سیستمی هوشمند برای کمک به کاربران در جستجو[10]، مرتب‌سازی[11]، کلاسه بندی[12]، غربال کردن [13]و اشتراک اطلاعات [14]جزء اولین ایده‌ها برای ایجاد و توسعه‌ی سیستم‌های توصیه‌گر بوده است. از اواسط دهه ۹۰ تاکنون تعداد زیادی سیستم توصیه گر برای کمک به کاربران در رسیدن به اطلاعات مورد علاقه‌شان پیاده‌سازی شده‌اند [5] . Tapestry [12] که یک سیستم نامه ی آزمایشی در مرکز تحقیقات Xerox Palo Alto بود بر اساس این ایده شکل گرفت که اگر افراد در فرآیند پالایش شرکت کنند، پالایش اطلاعات کارایی بالاتری خواهد داشت. سیستم Group Lens [7,8]، توصیه گر مقالات و اخبار یوزنت و Ringo [9]، سیستم پیشنهاد موسیقی- نیز در ادامه طراحی و پیاده‌سازی شدند. علاوه بر این سیستم‌های عملیاتی تحقیقاتی نیز در زمینه ی ترکیب سیستم‌های توصیه گر و ساخت کاراترین سیستم ممکن توسط Balabanovic، Bruke و Nguyen انجام شده است [10].

۳- مروری بر مفهوم سیستم‌های توصیه گر

در این قسمت ابتدا به شرح مفاهیم بنیادی و تئوریک در حوزه ی سیستم‌های توصیه گر می پردازیم تا زمینه ای برای تحلیل بهتر و ساده تر مفاهیم مرتبط با طراحی و پیاده سازی فراهم شود.

1 - Searchable
2 - Information overload
3 - User profile
4 - Profile learning techniques
5 - User profile representation
6 - Initial profile generation
7 - Profile adaptation techniques
8 - Relevance feedback
9 - Profile Exploitation

10 - Searching
11 - Sorting
12 - Classifying
13 - Filtering
14 - Sharing

Archive of SID
www.SID.ir

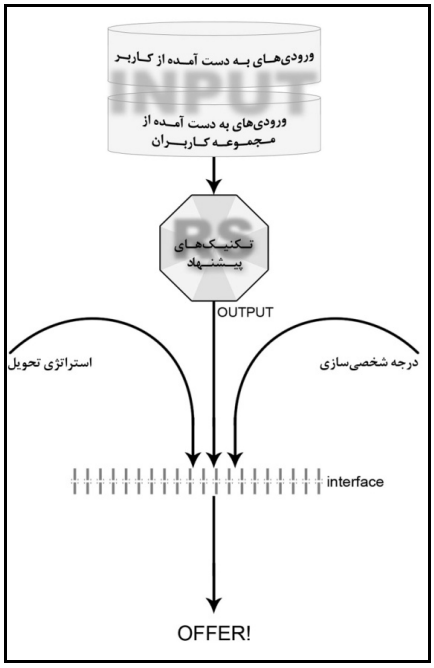
۳-۱- تعریف

سیستم توصیه‌گر یا سامانه پیشنهادگر، با تحلیل رفتار کاربر خود، اقدام به پیشنهاد مناسب‌ترین اقلام (داده، اطلاعات، کالا و...) می‌نماید. این سیستم رویکردی است که برای مواجهه با مشکلات ناشی از حجم فراوان و رو به رشد اطلاعات ارائه شده‌است و به کاربر خود کمک می‌کند تا در میان حجم عظیم اطلاعات، سریع‌تر به هدف خود نزدیک شود [22] . سیستم‌های توصیه‌گر به کاربرانی که از بین حجم بالای اطلاعات به دنبال نوعی خاص از اطلاعات مرتبط با اولویت‌هایشان هستند، پیشنهادات شخصی شده‌ای را ارائه می‌دهد. این سیستم با توانایی که در جمع آوری اطلاعات از رفتار و حرکات کاربران، دسته‌بندی و تفسیرآنها دارند، امکانی فراهم آورده اندکه کاربران با صرف زمان و انرژی کمتر به اطلاعات مناسب‌تری دسترسی پیدا کنند.

ایده‌ی سیستم‌های توصیه‌گر می تواند به کاربرانی که در جستجو، مرتب‌سازی، دسته‌بندی، پالایش اطلاعات و به اشتراک گذاری نیازمند مساعدت هستند، کمک نماید.

۳-۲- شمای کلی یک سیستم توصیه گر

از دیدگاه سیستمی، توصیه‌گر ماشینی است که با گرفتن ورودی از منابع تعریف شده - کاربر نهایی [1] و دیگر کاربران جامعه اطلاعاتی [2] و پردازش آنها با استفاده از تکنیک‌های پیشنهاد [3]، خروجی [4] منحصر به فردی برای کاربر خاص تولید می‌کند. این خروجی می‌تواند یک پیشنهاد [5]، پیش‌بینی [6] و یا ارزیابی [7] از طرف سیستم باشد. در ادامه متغیرهایی چون درجه شخصی‌سازی [8] که تعیین‌کننده میزان هماهنگی خروجی با تک تک کاربران است و می‌تواند طیف وسیعی از خروجی‌های کاملا شخصی شده [9] برای کاربر تا خروجی‌های نیمه شخصی [10] و خروجی‌های عام [11] را در بر گیرد، نحوه ی تحویل پیشنهاد [12] که می‌تواند به صورت ایمیل یا لیست‌های Top-N باشد نیز با توجه به سیاست‌های سیستم، روی خروجی اعمال شده و پیشنهاد نهایی به کاربر ارائه می‌شود. فرآیند فوق به صورت شماتیک در شکل (۱) نشان داده شده است.



شکل (۲) :شمای کلی یک توصیه گر

۴- انواع سیستم‌های توصیه گر

در زمانی نسبتاً کوتاه تعداد زیادی عامل‌های توصیه‌گر ایجاد شدند و در حال حاضر تنوع وسیعی از این سیستم‌ها موجود است که همه‌ی آنها از مزایای مجموعه‌ی خاصی از تکنیک‌های هوش مصنوعی استفاده می‌کنند. در این مطالعه دو روش کلی دنبال شده است: Spatial و Functional [6]. روش Spatial دسته‌بندی عامل‌ها بر اساس دامنه‌های کاری است و روش Functional بر ساختار و چگونگی عملکرد سیستم‌ها تمرکز دارد و دسته‌بندیی بر اساس تکنیک‌هایی که در اجزاء مختلف سیستم‌های توصیه گر استفاده می‌شوند ارائه می‌کند. روش اخیر امکان مطالعه‌ی مفهومی سیستم‌ها را فراهم می‌کند و در نتیجه در این تحقیق زمان بیشتری برای بررسی آن صرف شده است.

۴-۱- مطالعه Spatial

امروزه سیستم‌های توصیه‌گر در زمینه‌های متنوعی از پالایش اخبار موجود در وب متناسب با نیازها و خواسته‌های کاربران تا تجارت الکترونیکی، پیشنهاد فیلم، موسیقی، کتاب، مقاله و اطلاعات کاربرد دارند. در این مقاله، سیستم‌های توصیه گر از دیدگاه دامنه کاری در ۱۰ رده طبقه‌بندی شده‌اند. جدول (۱) این رده‌بندی را به همراه مثالی در

1 - Target customer input
2 - Community inputs
3 - Recommendation methods
4 - Output
5 - Suggestion
6 - Prediction
7 - Rating
8 - Degree of personalization
9 - Full-personalized
10 - Half-personalized
11 - None-personalized
12 - Delivery

Archive of SID
www.SID.ir

هر دامنه نشان می‌دهد. هر یک از این رده‌ها می‌توانند فضای کاری برای پژوهش در حوزه سیستم‌های توصیه گر باشند.

جدول (۱) : رده بندی سیستم‌های توصیه گر از دیدگاه دامنه‌های کاری موجود

دامنه کاری[1]	مثالی از سیستم فعال در این دامنه
غربال کردن اخبار[2]	ACR News
تجارت الکترونیکی[3]	www.Amazon.com
توصیه گر وب[4]	www.webmate.com.au [13]
روزنامه شخصی شده[5]	Anatogonomy
به اشتراک گذاری اخبار[6]	www.thebeehive.org
توصیه گر فیلم[7]	http://movielens.umn.edu
توصیه‌گر اطلاعات[8]	www.infofinder.no [16]
پالایش نامه الکترونیکی[9]	Tapestry
توصیه‌گر موسیقی[10]	Ringo [9]
توصیه‌گر خرید[11]	www.websell.net.au [19]

۲-۴- مطالعه Functional

در این بخش سیستم‌های توصیه گر از دیدگاه کارهایی که باید انجام دهند، مورد مطالعه قرار می گیرند. هدف نهایی یک توصیه گر ساخت پیشنهاد است، بدون شک هنگام تحلیل چگونگی ساخت پیشنهادات، بحث کلیدی پروفایل کاربر۵ است. با این دیدگاه دو گام اصلی در هر سیستم توصیه گر انجام می شود:

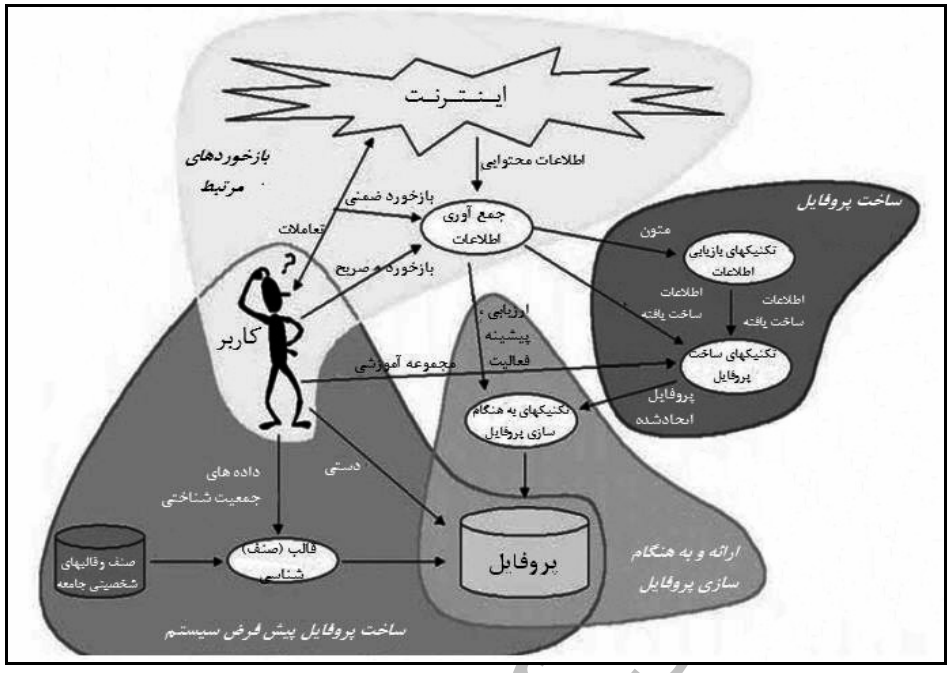
- ساخت و نگهداری پروفایل کاربر[12]
- استخراج اطلاعات از پروفایل برای ساخت پیشنهاد[13]

هر یک از این گامها خود به زیر وظیفه‌هایی تقسیم می‌شوند. نحوه ارائه پروفایل کاربر[14]، ساخت پروفایل اولیه[15] در لحظه آغاز به کار سیستم، استفاده از اطلاعات مدل‌های ارائه و به کار گیری تکنیک‌های یادگیری پروفایل برای کاربران[16]، کشف بازخورد مرتبط[17] و بهنگام کردن پروفایل[18] از مهم‌ترین اقداماتی است که باید در گام اول انجام شوند. پالایش اطلاعات[19]، منطبق‌سازی اطلاعات کاربران با کالاها[20] و با یکدیگر[21]، و ساخت پیشنهاد نیز در گام دوم دنبال می‌شوند. برای درک بهتر این دو فرآیند به شکل‌های (۲) و (۳) توجه کرده و روند فلش‌ها را در هر یک دنبال کنید [6].

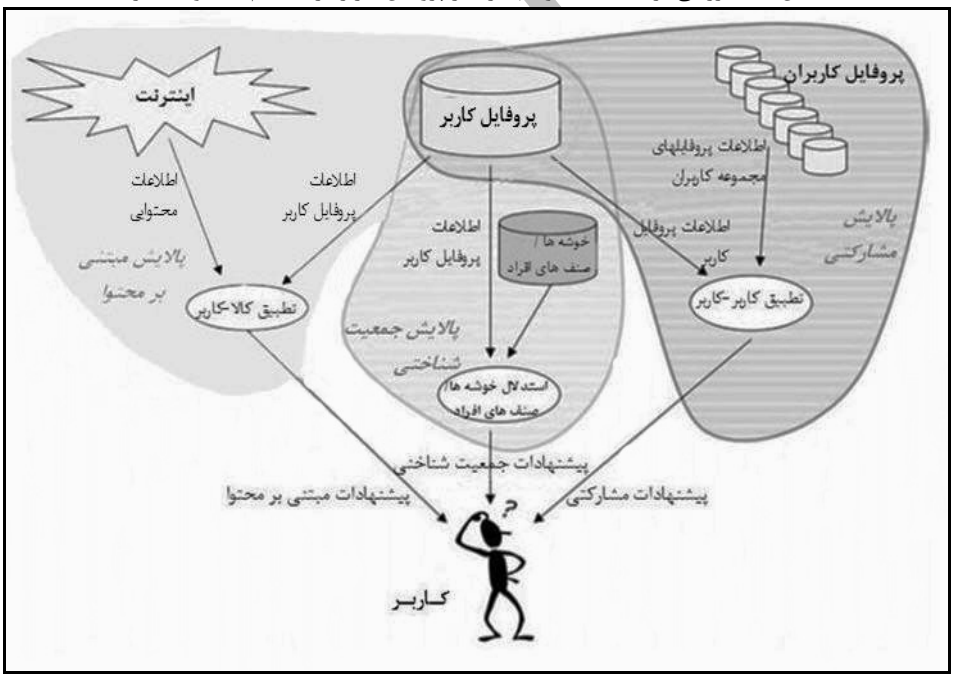
1 - Domain
2 - Netnews Filtering
3 - Ecommerce
4 -Web recommender
5 - Personalized newspaper
6 - Sharing news
7 - Movie recommender
8 - Information recommender
38 - Email filtering
10 - Music recommender
11 - Purchase recommender
12 - Profile generation and maintenance
13 - Profile exploitation for recommendation
14 - User profile representation
15 - Initial profile generation
16 - Profile learning techniques
17 - Relevance feedback
18 - Profile adaptation techniques Information

19 - Information filtering method
20 - User profile-item matching
21 - User profile matching

Archive of SID
www.SID.ir



شکل (۲) : بررسی فرآیند ساخت و نگهداری از پروفایل کاربر در سیستم‌های توصیه گر

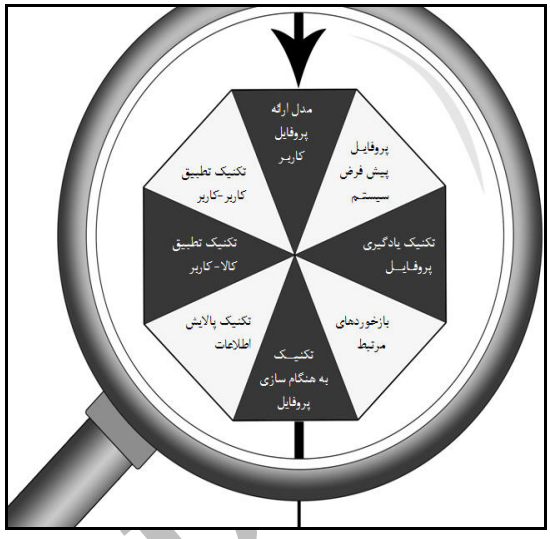


شکل (۳): روند استخراج اطلاعات از پروفایل در فرآیند تولید پیشنهاد در سیستم‌های توصیه گر به تفکیک تکنیک پالایش اطلاعات

۵- معرفی ابعاد طبقه بندی

در بخش قبل ۸ فعالیت اساسی که در هر سیستم توصیه گر باید انجام شود بیان شد. برای انجام هر یک از اقدامات فوق با توجه به اهداف خاص هر سیستم، گروهی از تکنیکهای ریاضی و هوش مصنوعی به کار می‌رود. در این بخش با در نظر گرفتن هر یک از آنها به عنوان یک فاکتور موثر در طراحی و ساخت سیستم نهایی، به معرفی آنها و تکنیک‌های موجود در هریک پرداخته می شود. در ادامه، در بخش ۶ هر یک از این فاکتورها را به منزله‌ی یکی از ابعاد طبقه‌بندی در نظر خواهیم گرفت. در شکل (۴) شمایی نزدیک از یک توصیه‌گر با ذکر این ابعاد آورده شده است.

www.SID.ir



شکل (۴) : نمایش ابعاد داخلی یک توصیه‌گر از شمای نزدیک

۵-۱- نحوه ارائه پروفایل کاربر۴۲

یکی از مهمترین تصمیماتی که هنگام طراحی و ساخت یک سیستم توصیه‌گر یابد اتخاذ شود، انتخاب منبعی است که اطلاعات مربوط به کاربران از آن استنتاج شود.

در این راستا تکنیک‌های هوش مصنوعی با نمایش داده‌ها به صورت مدل، منبعی برای ادامه‌ی کار سیستم ایجاد می کنند. پرکاربردترین منبعی که بسیاری از سیستم‌های توصیه‌گر از اطلاعات آن به عنوان ورودی استفاده می‌کنند، "مدل مبتنی بر گذشته"[1] است که در آن سیستم از لیست خریدهای قبلی کاربران، حرکات و کلیک‌های موس و محتویات Emailها به عنوان منبع اطلاعات برای ساخت پروفایل کاربران استفاده می‌کنند. "ماتریس کاربر- کالا"[2] از دیگر منابع پرکاربرد است که در آن در هر درایه میزان علاقه‌ی کاربر I به کالای j ثبت شده است. از "ویژگیهای جمعیت شناختی"[3] نیز می‌توان به عنوان منبعی برای به دست آوردن اطلاعات کاربران استفاده کرد. "مدل فضای برداری" [4]که در آن اطلاعات به وسیله برداری از ویژگیها ارائه می‌شوند،"Weighted n-grams"، "Weighted associative networks "،"semantic networks" [14] "و "مدل‌های مبتنی بر دسته کننده‌ها"[5] نیز هر یک می‌توانند به عنوان مدلی برای ارائه اطلاعات و منبعی برای ساخت پروفایل استفاده شوند.

1 - History based model
2 - User-item ratings matrix
3 - Demographic features
4 - Vector space model
5 - Classifier-based models

۵-۲- پروفایل پیش فرض سیستم۴۳

از دیگر فاکتورهای پایه ،در طراحی سیستم‌های توصیه‌گر مشخص کردن پروفایل کاربر در ابتدای اجرای سیستم است. از آنجا که سیستم‌های توصیه‌گر در طی زمان به شناخت از کاربران می‌رسند لذا لازم است برای لحظه t=o که سیستم داده‌ای ندارد پیش فرض مناسبی برای شروع کار در نظر گرفته شود. این پیش فرض می‌تواند تکنیک استفاده از یک "پروفایل خالی"[6]، یا یک "مجموعه آموزشی"[7] که با توجه به پاسخ کاربر به کلیه سوالات سیستم انتخاب می‌شود، در نظر گرفته شود و یا می توان برای این منظور از "روش دستی"[8] استفاده نمود که طی آن سیستم از کاربر ثبت نام به عمل آورد و با طرح سوالاتی اقدام به ایجاد پروفایل ابتدایی نماید. به علاوه می‌توان از "تکنیک جمعیت شناختی"[9] نیز استفاده کرد.

۵-۳- تکنیک‌های یادگیری پروفایل۴۴

از جمله فاکتورهایی که بررسی و تعیین آن در یک سیستم توصیه‌گر اهمیت بسیاری دارد انتخاب تکنیکی است که سیستم با استفاده از آن منابع داده‌ی پروفایل را ساختاردهی کرده و پروفایل نهایی کاربر را ایجاد می‌کند . در این میان، سیستم‌هایی که اطلاعات کاربران را از یک پایگاه داده استخراج می‌کنند، سیستم‌های پالایش مشارکتی که از ماتریس کالا به کالا به عنوان پروفایل استفاده می‌کنند و سیستم‌هایی که از اطلاعات جمعیت شناختی به عنوان منبع داده استفاده می‌کنند نیازی به تکنیک‌های جدیدی برای اعمال روی داده‌ها نداشته و هر یک از اطلاعات مذکور می‌توانند به عنوان پروفایل نهایی در نظر گرفته شوند [6]، اما لازم است در بقیه موارد از تکنیک مناسبی برای ساختاردهی به اطلاعات منبع استفاده کرد. تکنیک‌های "بازیابی اطلاعات ساخت یافته"[10] [20] از جمله روش‌های متداول در ساختاردهی به اطلاعات است. این فرایند در دو گام" انتخاب ویژگیها"[11] که طی آن تعداد واژه‌ها کاهش می‌یابد و شاخص گذاری اطلاعات که در آن با توجه به تعداد رخداد واژه‌ها ارتباط بالقوه ی کالا محاسبه می شود، اجرا می گردد. "TF-IDF" یکی از موفق‌ترین تکنیک‌ها در این زمینه است."

6 - Empty
7 - Training set
8 - Manual
9 - Stereotyping
10 - Structured information retrieval techniques
11 - Feature selection

Archive of SID
www.SID.ir

خوشه‌بندی"[1] [2,3] از دیگر تکنیک‌هایی است که می‌تواند اطلاعات را بر اساس شباهت به خوشه‌هایی تقسیم نماید. در این روش اطلاعات منبع خوشه‌بندی شده و اطلاعات جدید پس از مقایسه با خوشه‌های موجود ارزیابی و ساختاردهی می‌شوند.

"دسته‌کننده‌ها" نیز مدل‌های محاسباتی هستند که به مجموعه‌ای از داده‌های ساخت نیافته رده‌ی خاصی را منتسب می‌کنند. تکنیک‌های "شبکه‌های عصبی"[2]، "درخت تصمیم"[3]، "قواعد وابستگی" [4] [17] و "شبکه‌های بیزین"[5] [15] از جمله تکنیک‌های دسته کننده‌ها هستند.

۵-۴- بازخوردهای مرتبط۴۵

اولویت‌های کاربران متغیری است که در طول زمان تغییر می‌کند، در نتیجه، سیستم‌های توصیه‌گر برای ارائه‌ی پیشنهادات مناسب نیازمند منبع و تکنیکی جهت به هنگام سازی پروفایل کاربران هستند. به جز تعداد اندکی از سیستم‌های توصیه‌گر مانند سیستم SIFT Netnews [11] که این بهنگام سازی را به صورت «دستی» انجام می‌دهند، اکثر سیستم‌های موجود از طریق اعمال تکنیک‌های نامبرده شده در بخش ۵-۳ روی اطلاعات به دست آمده از طریق "بازخورد صریح"[6] یا "بازخورد ضمنی" [7] طی زمان‌های مشخصی (با توجه به نوع تکنیک انتخاب شده) پروفایل کاربر را به هنگام می‌کنند. در این میان سیستم‌هایی نیز هستند که در راستای افزایش دقت و کارایی از هر دو نوع اطلاعات صریح و ضمنی به صورت ترکیبی استفاده می‌کنند.

۵-۵- تکنیک‌های به هنگام سازی پروفایل۴۶

برای انجام عمل به هنگام سازی و معتبرسازی پروفایل با توجه به اهداف سیستم توصیه‌گر، باید از میان تکنیک‌های موجود مناسبترین را انتخاب نمود. به هنگام سازی به روش "دستی" از ساده‌ترین تکنیکها است که در آن کاربران به هنگام تغییر در علایق، پروفایل خود را به صورت دستی تغییر می‌دهند. دو مشکل اساسی این روش افزایش فعالیت در سمت کاربر و عدم توانایی کاربران در مشخص نمودن اولویت‌هایشان به طور دقیق است. در نتیجه این روش در محیط‌هایی با آهنگ تغییرات سریع کارایی چندانی ندارد.

در "تکنیک افزایشی" [8]، سیستم اطلاعات جدید استنتاج شده را به اطلاعات پیشین افزوده و پروفایل را به روز می‌رساند. مشکل این روش در نگهداری اطلاعاتی است که ارزشی ندارند ولی همچنان با صرف توان سیستم نگهداری می‌شوند. برای حل این مسئله در سیستم‌هایی که به دقت بالاتری نیاز است با استفاده از تابعی[9] از "تکنیک وزن دادن به نسبت اهمیت اطلاعات" استفاده می‌شود [21] در نتیجه می‌توان با انتساب وزن پایین به اطلاعات قدیمی یا کم ارزش، ارزش آنها را در عملیات تولید پیشنهاد کاهش داد.

۵-۶- تکنیک‌های پالایش اطلاعات۴۷

پس از اینکه منبع و تکنیک فرآیند ساخت و به هنگام سازی پروفایل از طرف طراحان سیستم در گام اول مشخص شد، نیاز به آن است که در گام دوم تصمیمات مناسبی در راستای چگونگی پالایش و استخراج اطلاعات کاربران اتخاذ شود. در یک سیستم توصیه‌گر برای ارائه‌ی پیشنهاد به کاربران، توانایی انتخاب مناسب‌ترین اطلاعات از بین اطلاعات موجود عاملی است که با میزان کارایی و دقت سیستم ارتباط مستقیم دارد. سه متد اصلی در پالایش اطلاعات، "پالایش مبتنی بر جمعیت شناختی"[10]، که در آن تکنیک کلی استناد بر ویژگی‌های آماری است [18]، "پالایش مبتنی بر محتوا"[11]،که با مسئله ی پیشنهاد از زاویه ی جستجو برای پیدا کردن اطلاعات با محتوای مشابه برخورد می کند [17] و "پالایش مشارکتی"[12]، که برای پیشنهاد اطلاعات به یک کاربر خاص از نظرات گروهی کاربران استفاده می کند، هستند.

۵-۷- تکنیک‌های تطبیق کالا-کاربر۴۸

به طور معمول برای پیشنهاد کالاهای مرتبط به کاربر از پروفایل او استفاده می‌شود. در سیستم‌هایی که از متد پالایش مبتنی بر محتوا استفاده می‌کنند ارتباط مستقیم کالا با کاربر بررسی می‌شود. در نتیجه به تکنیکی برای یافتن میزان ارتباط یک کالا با پروفایل خاص نیاز است. یکی از تکنیک‌های پرکاربرد در این زمینه "تطبیق

1 - Clustering
2 - Neural networks
3 - Decision tree
4 - Association rules
5 - Bayesian network
6 - Explicit feedback
7 - Implicit feedback

8 - Add new information
9 - Gradual forgetting function
10 - Demographic filtering
11 - Content based filtering
12 - Collaborative filtering

Archive of SID
www.SID.ir

کلید واژه استاندارد"[1] است که شامل شمارش ساده‌ی اصطلاحاتی است که متناوباً در توصیف کالا و کاربر خاص آمده‌اند. روش دیگر تبدیل پروفایل کاربر و جستجو[2]ی او به بردار و محاسبه اندازه "کسینوس[3] بین دو بردار" است. هرقدر این مقدار به ۱ نزدیک تر باشد کالا و کاربر مورد بررسی به هم شباهت بیشتری دارند. به علاوه می‌توان میزان شباهت را با استفاده از" الگوریتم نزدیک‌ترین همسایه" [4] نیز محاسبه کرد، که در آن فاصله‌ی بین کالایی که مورد علاقه‌ی کاربر است با مجموعه دسته کالاهای دیگر محاسبه شده و کالاهای همسایه پیشنهاد می‌شوند.

از دیگر روشهای تطبیق کالا با کاربر نگاه به این مسئله از زاویه‌ای "دسته‌بندی" است به این صورت که سیستم تلاش می‌کند با استفاده از یکی از دسته کننده‌ها و بر اساس مجموعه‌ای از ویژگی‌های کالا مدلی از رفتارهای کاربر استنتاج کند که در ادامه بتوان با استفاده از این مدل کالاهای جدید را به کلاس‌هایی مثل دسته کالای علاقه‌مند یا دسته کالای بی‌علاقه دسته‌بندی کرد.

۵-۸- تکنیک‌های تطبیق کاربر-کاربر۴۹

در سیستم‌های مبتنی بر پالایش مشارکتی، نیاز به روش هایی است که مستقیم بتواند کاربرانی با علایق مشابه را یافته، همسایگی آنها را شکل داده و بر اساس آن پیش‌بینی نماید. برای هر یک از این ۳ مرحله روشهایی به شرح زیر بررسی شده است.

با استفاده از یکی از سه تکنیک "دسته کننده‌ها"، "خوشه بندی" و "نزدیک ترین همسایه "می‌توان کاربرانی با علایق مشابه را یافت. سپس با به کار گیری یکی از روشهای" آستانه‌ی همبستگی" [5]یا "بهترین n همسایگی"[6] همسایگی کاربر نهایی را ساخته و در نهایت پس از اینکه همسایگی‌ها انتخاب شدند، ارزیابی‌های آنها را برای محاسبه‌ی پیش‌بینی ترکیب کرد. برای محاسبه پیش‌بینی می‌توان از تکنیک "پیشنهاد کالا با بالاترین تکرار"[7]، "پیشنهادات مبتنی بر قواعد وابستگی "[8]و "میانگین ارزیابی‌ها" [9] استفاده کرد.

۶- طبقه بندی

آنچه تا این بخش بیان شد، فاکتورهایی است که هنگام طراحی یک سیستم توصیه گر، طراح سیستم باید با توجه به اهداف و نوع خروجی مورد نظر، تصمیم مناسبی در راستای انتخاب تکنیک‌های موثر اتخاذ کند. در این قسمت برای نمایش کلی حالت‌های ممکن و مرتبط در ۱۰ نمونه از سیستم‌های پیاده سازی شده، هر یک از این فاکتورها به عنوان یک بعد در نظر گرفته شده اند و بر این اساس طبقه بندی به صورت جدول (۲) در ۸ ستون شامل نام سیستم و ابعاد بررسی شده در آنها و ۱۰ سطر که هر سطر به یکی از توصیه گرها اختصاص دارد، آورده شده است.

1 - Standard keyword matching
2 - Query
3 - Cosine similarity
4 - Nearest neighbor
5 - Correlation-thresholding
6 - Best n-neighbor technique
7 - Most frequent item
8 - Association rule-based recommendation
9 - Weighted average of ratings

Archive of SID
www.SID.ir

جدول ۲ : طبقه بندی سیستم‌های توصیه گر (۱۰ نمونه از آنها) از دیدگاه ابعاد بررسی شده در مطالعه Functional [6]

	نحوه ارائه پروفایل کاربر	پروفایل پیش فرض	تکنیک یادگیری پروفایل	بازخوردهای مرتبط	تکنیک یه هنگام سازی پروفایل	تکنیک پالایش اطلاعات	تکنیک تطبیق کالا-کاربر یا کالا-کالا
ACR News	خوشه های URL (Clusters)	مجموعه آموزشی (Training set)	استنتاج قواعد یادگیری - خوشه بندی (Rule learning – clustering)	بازخورد ضمنی (Implicit)	افزایش اطلاعات جدید به اطلاعات قبل (Add new)	مبتنی بر محتوا (Content)	خوشه بندی (Cluster matching)
Amazon	پیشینه حرکات کاربر (History, Ratings)	تکنیک به کارگیری پروفایل خالی (Empty)	نیازی به یادگیری ندارد (Not necessary)	بازخورد صریح (Explicit)	افزایش اطلاعات جدید به اطلاعات قبل(Add new)	ترکیبی (Hybrid)	Unknown
Webmate	بردار ویژگی وزن دار شده (Weighted Feature vector)	تکنیک به کارگیری پروفایل خالی (Empty)	تکنیک TF-IDF	بازخورد صریح (Explicit)	افزایش اطلاعات جدید به اطلاعات قبل(Add new)	مبتنی بر محتوا (Content)	تعیین تشابه با استفاده از کسینوس زاویه (Cosine similarity)
Anatogonomy	بردار ویژگی وزن دار شده (Weighted Feature vector)	تکنیک به کارگیری پروفایل خالی (Empty)	تکنیک TF-IDF	بازخورد صریح (Explicit)	افزایش اطلاعات جدید به اطلاعات قبل(Add new)	ترکیبی (Hybrid)	تعیین تشابه با استفاده از کسینوس زاویه (Cosine similarity)
BeeHive	خوشه ها (Clusters)	تکنیک به کارگیری پروفایل خالی (Empty)	خوشه بندی (Clustering)	بازخورد ضمنی (Implicit)	افزایش اطلاعات جدید به اطلاعات قبل(Add new)	مشارکتی (Collaborative)	خوشه بندی (Cluster matching)
MovieLens	بردار ویژگی وزن دار شده (Weighted Feature vector)	مجموعه آموزشی (Training set)	تکنیک TF-IDF - استنتاج قواعد یادگیری (Rule learning)	بازخورد صریح (Explicit)	افزایش اطلاعات جدید به اطلاعات قبل(Add new)	ترکیبی (Hybrid)	تعیین تشابه با استفاده از کسینوس زاویه (Cosine similarity)
InfoFinder	درخت تصمیم (Decision tree)	مجموعه آموزشی (Training set)	انتخاب ویژگی ها (Feature selection)	بازخورد صریح (Explicit)	افزایش اطلاعات جدید به اطلاعات قبل(Add new)	مبتنی بر محتوا (Content)	Boolean search query string
Tapestry	Indexed messages	تکنیک به کارگیری پروفایل خالی (Empty)	نیازی به یادگیری ندارد (Not necessary)	بازخورد صریح (Explicit)	افزایش اطلاعات جدید به اطلاعات قبل(Add new)	مشارکتی (Collaborative)	Tapestry Query Language
Ringo	ماتریس ارزیابی ها (Ratings matrix)	مجموعه آموزشی (Training set)	نیازی به یادگیری ندارد (Not necessary)	بازخورد صریح (Explicit)	افزایش اطلاعات جدید به اطلاعات قبل(Add new)	ترکیبی (Hybrid)	نزدیک ترین همسایه (Nearest neighbor)
Websell	پیشینه حرکات کاربر (History)	تکنیک به کارگیری پروفایل خالی (Empty)	نیازی به یادگیری ندارد (Not necessary)	بازخورد صریح (Explicit)	افزایش اطلاعات جدید به اطلاعات قبل(Add new)	ترکیبی (Hybrid)	Case-base reasoning

۷- نتیجه‌گیری

رشد توقف‌ناپذیر اینترنت و محیط‌های الکترونیکی، نیاز به ابزارهایی برای کمک به کاربران جهت یافتن اطلاعات و سرویس‌های مورد نظر را ایجاد کرده است. سیستم‌های توصیه گر از جمله ابزارهای مناسب در این راستا هستند که در این مقاله به تعریف، بررسی انواع و عملکرد آنها پرداختیم و نقش آنها را در کمک به شهروندان برای جستجو، مرتب‌سازی، دسته‌بندی، پالایش و به اشتراک‌گذاری اطلاعات مورد بررسی قرارگرفت. در ادامه فرآیندی که هر توصیه‌گـــر برای تولید پیشنهاد طی می‌کند را به همراه تکنیک‌های موجود در هر گام بررسی کرده و یک طبقه‌بندی روی ۱۰ سیستم از ۸ بعد ارائه شد.

به طور کلی می‌توان نتیجه گرفت که با توجه به روند رو به رشد شکل‌گیری تجارت‌های مبتنی بر وب و افزایش میزان کارکرد اینترنت در کشورمان، نیاز به تجهیز کســب و کارهـــا به سیستم‌های توصـــیه‌گر بیش از پیش احساس می‌شود، در حال حاضر مدل‌های بسیار ساده آنها بعضا یا پیاده شده و یا در حال پیاده‌سازی است که شناسائی آن می‌تواند موضوع پژوهش دیگری در این حوزه از دانش باشد.

مراجع

۱- شهرابی، جمال، داده کاوی، تهران، جهاد دانشگاهی واحد صنعتی امیر کبیر، ۱۳۸۶.

2- Gan, Guojun, Chaoqun Ma, and JianhongWu , Data Clustering: Theory, Algorithms, and Applications, ASA-SIAM Series on Statistics and AppliedProbability, SIAM, Philadelphia, ASA, Alexandria, VA, 2007.

3- Jiawei Han and Micheline Kamber Data Mining: Concepts and Techniques, Second Edition, TheMorgan Kaufmann Series in DataManagement Systems, Series Editor: Jim Gray, Microsoft Research

۴- محمد درزی، حبیب اله اصغری، آندره شلتز . "تحلیل الگوریتم‌های پالایش مشارکتی مبتنی بر کالا در ارائه خدمات به شهروند الکترونیک"، اولین کنفرانس شهر الکترونیک، تهران، ۱۳۸۶

۵- مرادی منش، زهرا، تحلیل فنی سیستم‌های توصیه گر، برای اخذ مدرک کارشناسی در رشته مهندسی نرم افزار، دانشگاه علم و فرهنگ.

6- Miquel Montaner , Beatriz López and Josep Lluís de la Rosa , " A taxonomy of recommender Agents on the Internet", Artificial Intelligence Review PublisherSpringer Netherlands, Volume 19, Number 4 / June, 2003

7- Konstan, J., Miller, B., Maltz, D., Herlocker, J., Gordon, L., and Riedl, J. (1997). ,"GroupLens: Applying Collaborative Filtering to Usenet News", Communications of the ACM, 40(3), pp. 77-87

8- Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., and Riedl, J. (1994). " GroupLens: An Open Architecture for Collaborative Filtering of Netnews". In Proceedings of CSCW '94, Chapel Hill, NC.

Archive of SID

www.SID.ir

9- Shardanand, U., and Maes, P. (1995).” Social Information Filtering: Algorithms for Automating ’Word of Mouth’”, In Proceedings of CHI ’95. Denver, CO.
10- Robin Bruke. (November 2002) “Hybrid Recommender Systems” , Pages: 331 – 37
11- Tak W. Yan, Hector Garcia Molina, “SIFT- A tool for Wide-Area Information Dissemination”, in proceedings of the 1995 USENIX Technical Conference, pages 177-86, 1995
12- Goldberg David, Nichols David, M. Oki Brian and Terry Douglas, “Using collaborative filtering to weave an information Tapestry”. Communications of the ACM, Dec 1992 v35 n12 p61(10)
13- Chen, L. & Sycara, K. (1998). “Webmate: A Personal Agent for Browsing and Searching”. InProceedings of AGENTS ’98, 132–139. ACM.
14- Potter, G. & Trueblood, R. (1988). “Traditional, Semantic, and Hyper-Semantic Approaches to Data Modeling”. IEEE Computer 21(6): 53–63.
15- Jensen, F. V. (1996). “An Introduction to Bayesian Networks” . New York: Springer.
16- B.Krulwich,, C.Burkey, “ The InfoFinder agent: learning user interests through heuristicphrase extraction” IEEE Expert, 1997
17- Basu, C., Hirsh, H. & Cohen,W. (1998). Recommendation as Classification: Using Social and Content-Based Information in Recommendation. In Proceedings of AAAI’98, 714–720.
18- Rich, E. (1979). “ User Modeling via Stereotypes”. Cognitive Science 3: 329–354.
19- Cunningham, P., Bergmann, R., Schmitt, S., Traphoner, R., Breen, S. & Smyth, B. (2001). “WebSell: Intelligent Sales Assistants for the World Wide Web”. In E-2001.
20- Salton, G. & Buckley, C. (1988). “Term-Weighting Approaches in Automatic Text Retrieval.” Information Processing and Management 24(5): 513–523.
21- Koychev, I. (2000). “Gradual Forgetting for Adaptation to Concept Drift”. In Proceedings of ECAI 2000 Workshop Current Issues in Spatio-Temporal Reasoning.
22- Sonja Kangas, “Collaborative Filtering and Recommendation Systems”. In Proceedings of LOUHI-project.2000

Archive of SID

www.SID.ir