

جمهوری اسلامی ایران
وزارت علوم تحقیقات و فناوری



دانشگاه اراک

دانشکده فنی و مهندسی

کارشناسی ارشد مهندسی کامپیوتر-نرم افزار

پیاده سازی یک سیستم توصیه گر ترکیبی برای تجارت الکترونیکی

پژوهشگر

حسین حسن رئیسی

استاد راهنما

دکتر رضا رافع

زمستان ۱۳۹۱



سپاسگزاری

لازم می‌دانم از زحمات و تلاش‌های استاد ارجمند آقای دکتر رضا رافع که در انجام این پایان نامه بنده را راهنمایی کردند، صمیمانه تشکر کنم و از خداوند منان توفیقات روزافزون برای ایشان خواستارم.

چکیده

رشد روزافزون تجارت الکترونیکی در فضای مجازی باعث معرفی محصولات و خدمات متنوع و زیادی از سوی شرکت‌های فعال در این حوزه شده است. در چنین شرایطی انتخاب مناسب و بهینه از میان حجم انبوهی از انتخاب‌های ممکن برای کاربران مشکل می‌شود. سیستم‌های توصیه‌گر با توجه به ویژگی‌ها، علایق و رفتار کاربران سعی می‌کنند بهترین و مناسب‌ترین آیتم‌ها را به کاربران خود پیشنهاد دهند.

در این پایان‌نامه یک سیستم توصیه‌گر ترکیبی از دو روش متداول پالایش گروهی و محدودیت محور، طراحی و پیاده‌سازی شده است. در طراحی این سیستم توصیه‌گر ترکیبی از مزایای هر دو روش استفاده شده است تا دقت پیشنهادات به کاربر نهایی بهبود یافته و خطاهای احتمالی ناشی از معایب دو روش کاهش یابد.

کلمات کلیدی: سیستم توصیه‌گر ، پالایش گروهی ، محدودیت ، پایگاه دانش

فهرست مطالب

ب	فهرست مطالب
د	فهرست جداول
ه	فهرست اشکال
۱	فصل اول مقدمه
۳	۱-۱ پیشنهاد گروهی
۴	۲-۱ پیشنهاد مبتنی بر محتوا
۵	۳-۱ پیشنهاد دانش محور
۶	۴-۱ رویکرد های ترکیبی
۷	۵-۱ کاربرد سیستم های توصیه گر
۸	۶-۱ اهداف پایان نامه
۹	۲ فصل دوم معرفی سیستم های توصیه گر و کارهای مرتبط
۹	۱-۲ سیستم توصیه گر پالایش گروهی
۱۰	۱-۱-۲ مفاهیم پایه
۱۱	۲-۱-۲ روش پالایش گروهی
۱۱	۳-۱-۲ چالش های پالایش گروهی
۱۲	۴-۱-۲ روش پالایش گروهی حساس به زمان
۱۶	۲-۲ سیستم توصیه گر محدودیت محور
۱۸	۱-۲-۲ نمایش دانش و استدلال کردن
۲۰	۳-۲ سیستم توصیه گر ترکیبی
۲۱	۱-۳-۲ سیستم توصیه گر ترکیبی محدودیت محور گروهی در سطح متا
۳۰	۳ فصل سوم سیستم توصیه گر ترکیبی پیشنهادی
۳۱	۱-۳ پیاده سازی سیستم ترکیبی پیشنهادی

۳۱	۱-۱-۳ معرفی حوزه فیلم
۳۲	۲-۱-۳ معرفی وب سایت movielens.org
۳۳	۳-۱-۳ تعریف الگوی محدودیت‌ها و استخراج پایگاه دانش محدودیت‌ها
۳۵	۴-۱-۳ اعمال روش پالایش گروهی
۳۶	۵-۱-۳ اعمال روش محدودیت محور
۳۹	۴ فصل چهارم ارزیابی سیستم توصیه‌گر ترکیبی پیشنهادی
۳۹	۱-۴ معیارهای ارزیابی
۴۰	۱-۱-۴ معیار دقت آماری
۴۰	۲-۱-۴ معیار دقت پشتیبانی تصمیم
۴۱	۲-۴ مجموعه داده
۴۲	۳-۴ نتایج ارزیابی
۴۳	۴-۴ تفسیر نتایج
۴۵	۵ فصل پنجم نتیجه‌گیری و پیشنهاد
۴۶	۱-۵ پیشنهادات برای کارهای آتی
۴۷	فهرست منابع
۴۹	واژه‌نامه
۴۹	واژه‌نامه فارسی به انگلیسی
۵۲	واژه‌نامه انگلیسی به فارسی

فهرست جداول

- جدول ۱-۲ : نمونه ای از کاتالوگ محصول (دوربین دیجیتال) ۱۸
- جدول ۲-۲ : مثالی از یک مسئله پیشنهاد و نتیجه متناظر با آن ۱۹
- جدول ۳-۲ : لیست محصولات فروشگاه اینترنتی سیگار [۵] ۲۴
- جدول ۴-۲ : نمونه‌ای از تعاملات قبلی کاربران با سیستم فروش اینترنتی سیگار [۵] ۲۴
- جدول ۵-۲ : پایگاه دانش محدودیت های استخراج شده [۵] ۲۵
- جدول ۶-۲ : شباهت کاربر جدید با سایر کاربران ۲۷
- جدول ۷-۲ : پایگاه دانش محدودیت شخصی سازی شده برای کاربر جدید ۲۸
- جدول ۱-۳ : ویژگی‌های برخی از کاربران سایت MovieLens ۳۴
- جدول ۲-۳ : ویژگی‌های برخی از فیلم‌های سایت MovieLens ۳۴
- جدول ۳-۳ : نمونه ای از اطلاعات تعاملات کاربران با سیستم MovieLens ۳۴
- جدول ۴-۳ : برخی از محدودیت‌های استخراج شده از تعاملات کاربران سایت MovieLens ۳۵
- جدول ۵-۳ : محاسبه شباهت کاربر هدف با سایر کاربران ۳۶

فهرست اشکال

- شکل ۱-۳ : طرح کلی سیستم ترکیبی از روش پالایش گروهی و محدودیت محور ۲۲
- شکل ۱-۴ : نتایج ارزیابی روش‌های ترکیبی و پالایش گروهی با استفاده از معیار MAE ۴۳
- شکل ۲-۴ : نتایج ارزیابی روش‌های ترکیبی و پالایش گروهی با استفاده از معیار ROC ۴۳

فصل اول

مقدمه

می‌توان گفت که ما در میان حجم عظیمی از اطلاعات قرار گرفته‌ایم که بدون راهنمایی درست ممکن است انتخاب‌هایی غلط و یا غیر بهینه از میان آن‌ها داشته باشیم. سیستم‌های توصیه‌گر¹ سیستم‌های تأثیرگذار در راهنمایی و هدایت کاربر، در میان حجم عظیمی از انتخاب‌های ممکن، برای رسیدن به گزینه مفید و مورد علاقه وی هستند به گونه‌ای که این فرآیند، برای همان کاربر، شخصی‌سازی شده باشد.

سیستم‌های توصیه‌گر با تحلیل رفتار کاربر خود، اقدام به پیشنهاد مناسب‌ترین اقلام (اطلاعات، کالا و یا خدمات) می‌نماید. این گونه سیستم‌ها برای مواجهه با مشکلات ناشی از حجم فراوان و رو به رشد اطلاعات ارائه شده‌اند و به کاربران خود کمک می‌کنند تا در میان حجم عظیم اطلاعات سریع‌تر به هدف خود نزدیک شوند.

یک سیستم توصیه‌گر سیستمی است که توانایی تحلیل رفتارهای گذشته و ارائه توصیه‌هایی برای مسائل جاری را دارا باشد. به زبان ساده‌تر در سیستم‌های توصیه‌گر تلاش بر این است تا با حدس زدن شیوه تفکر کاربر (به کمک اطلاعاتی که از نحوه رفتار وی یا کاربران مشابه وی و نظرات آن‌ها داریم) مناسب‌ترین و نزدیک‌ترین کالا به سلیقه او را شناسایی و به وی پیشنهاد کنیم. توصیه‌هایی که از سوی سیستم‌های توصیه‌گر ارائه می‌شوند به طور کلی می‌توانند دو نتیجه در برداشته باشند :

¹ Recommender Systems

- کاربر را در تصمیم‌گیری یاری می‌کنند (مثلاً از میان چندین گزینه پیش رو کدام بهتر است).

- موجب افزایش آگاهی کاربر در زمینه مورد علاقه وی می‌شود (مثلاً در حین ارائه توصیه به کاربر موجب می‌شود تا وی با اقلام و اشیاء جدیدی که قبلاً آن‌ها را نمی‌شناخته، آشنا شود).

بسیاری از کاربران اینترنتی از طرق مختلفی به صورت خودآگاه یا ناخودآگاه با یک سیستم توصیه‌گر در ارتباط بوده‌اند. برای مثال فرض کنید که یکی از دوستان شما پیشنهاد خواندن یک کتاب جدید را به شما دهد. شما برای سفارش این کتاب به یک کتابفروشی آنلاین مراجعه می‌کنید. عنوان کتاب مورد نظر خود را وارد کرده و لیستی از کتاب‌های یافت شده به شما نشان داده می‌شود. در قسمت دیگری از وب سایت کتابفروشی آنلاین با عنوان «مشتریانی که این کتاب را خریده‌اند همچنین علاقه‌مند هستند به» لیستی از کتب نمایش داده می‌شود که با توجه به خریدهای قبلی مشتریان احتمالاً شما نیز به آن‌ها علاقه‌مند هستید. در صورتی که یکی از کاربران قدیمی‌تر این کتاب فروشی آنلاین باشید به محض ورود به وب سایت آن لیستی از کتاب‌های پیشنهادی به شما نشان داده می‌شود. [۱]

سیستم نرم افزاری که لیست کتاب‌های پیشنهادی به یک بازدید کننده را معین می‌کند یک «سیستم توصیه گر» است. اولین نکته حائز اهمیت در این سیستم‌ها این است که پیشنهادات ارائه شده توسط یک سیستم توصیه گر شخصی است، به عبارت دیگر متناسب با ذائقه بازدید کننده می‌باشد و این مورد بر خلاف بسیاری از فروشگاه‌ها و وب سایت‌های دیگر است که تنها آیتم‌هایی که بیشتر مورد توجه هستند را به عنوان لیست برترین‌ها نمایش می‌دهند.

برای ایجاد کردن لیست پیشنهادات مخصوص هر کاربر به اطلاعاتی درباره تمام کاربران نیاز است. هر سیستم توصیه گر باید یک مدل کاربر^۱ یا حساب کاربری^۲ را در ابتدا طراحی و پیاده سازی و در ادامه نگه داری و بروز رسانی کند. این حساب کاربری می‌تواند شامل اطلاعات فردی کاربر و همچنین ترجیحات^۳ او باشد. در مثال کتاب فروشی آنلاین سیستم می‌بایست کتاب‌های مشاهده شده و یا خریده شده توسط کاربر در گذشته را ذخیره کند تا بتواند از آن‌ها برای تعیین لیست کتاب‌های پیشنهادی مورد علاقه وی استفاده کند. [۱]

¹ User model

² User profile

³ Preferences

هر چند حساب کاربری در یک سیستم توصیه گر نقش محوری ایفا می‌کند، نحوه جمع آوری و استفاده از این اطلاعات به تکنیک‌های استفاده شده توسط سیستم توصیه گر بستگی دارد. ترجیحات کاربران می‌تواند به صورت ضمنی با پایش رفتار وی جمع آوری گردد و یا اینکه در سیستمی دیگر در مورد ترجیحات کاربر به صورت صریح از وی سؤال شود.

دومین نکته ای که در سیستم‌های توصیه گر مورد توجه قرار دارد انواع اطلاعات اضافی است که سیستم جهت تولید لیست پیشنهادات باید استفاده کند. یکی از مهم‌ترین رویکردها در این زمینه که بسیاری از کتاب‌فروشی‌های آنلاین حقیقی از آن بهره می‌برند، در نظر گرفتن رفتار، نظرات و سلیق سایر کاربران سیستم است. این گونه سیستم‌ها را با رویکرد مبتنی بر انجمن^۱ یا گروهی^۲ معرفی می‌کنند. [۱]

۱-۱ پیشنهاد گروهی

ایده اصلی در سیستم‌های توصیه گر گروهی این است که اگر کاربرانی در گذشته علایق یکسانی داشتند (برای مثال کتاب‌های مشترکی را دیده و یا خریداری کرده‌اند) در نتیجه در آینده نیز علایق یکسانی خواهند داشت. برای مثال اگر کاربر A و کاربر B دارای خریدهای بسیار نزدیکی هستند و اخیراً کاربر A اقدام به خرید کتابی کرده که هنوز کاربر B آن را مشاهده نکرده است، بر اساس ایده توضیح داده شده به احتمال زیاد کاربر B به این کتاب علاقه‌مند خواهد بود.

در این روش مجموعه آیتم‌های مورد علاقه یک کاربر از میان تعداد زیادی آیتم که احتمالاً کاربر به آن‌ها علاقه بیشتری دارد فیلتر می‌شود و از طرف دیگر سایر کاربران به صورت ضمنی در این عمل شرکت می‌کنند، این روش پالایش گروهی^۳ نیز نامیده می‌شود.

امروزه این گونه سیستم‌ها در حوزه‌های مختلفی مورد استفاده قرار می‌گیرند و در پانزده سال اخیر مورد مطالعه و توسعه زیادی قرار گرفته‌اند. جنبه‌های مختلف این روش را در فصل بعدی مورد بررسی بیشتری قرار خواهیم داد. سؤالاتی که در زمینه رویکرد گروهی مطرح می‌شوند و باید پاسخی برای آن‌ها یافت به صورت زیر است:

¹ community

² collaborative

³ collaborative filtering

• چگونه باید کاربرانی مشابه به کاربر مورد نظر که می‌خواهیم به او پیشنهاد دهیم، بیابیم؟

- چگونه باید شباهت کاربران را اندازه‌گیری کرد؟
- در مورد کاربران جدید که تاریخچه خرید آن‌ها خالی است چه باید کرد؟
- در مورد کالاهای جدیدی که هنوز هیچ کاربری آن‌ها را نخریده چه باید کرد؟
- تکنیک‌های دیگری که در کنار یافتن شباهت کاربران می‌توان استفاده کرد تا میزان علاقه یک کاربر به یک آیتم را پیش‌بینی کرد کدامند؟

رویکرد های محض پالایش گروهی به اطلاعاتی در مورد خود آیتم‌ها احتیاجی ندارد. در مثال کتاب فروشی آنلاین، سیستم توصیه گر نیازی به دانستن موضوع کتاب یا نویسندگان آن ندارد. مزیت این استراتژی این است که نیازی به در نظر گرفتن اطلاعات آیتم‌ها در امر تولید پیشنهاد نیست. از طرف دیگر استفاده از برخی ویژگی‌های کتاب‌ها برای نشان دادن شباهت کاربران به یکدیگر با توجه به خرید های قبلی شان می‌تواند نتایج بهتری را تولید کند.[۱]

۱-۲ پیشنهاد مبتنی بر محتوا

عموماً سیستم‌های توصیه گر برای تحقق دو هدف متفاوت ایجاد می‌شوند. از یک طرف می‌توانند برای تحریک کاربران جهت انجام کاری استفاده شوند مانند خرید یک کتاب خاص و یا تماشای یک فیلم خاص، از طرف دیگر سیستم‌های توصیه گر می‌توانند به عنوان ابزارهایی جهت برآورد «سربار اطلاعات»^۱ استفاده شوند به عبارت دیگر آن‌ها به انتخاب آیتم‌های مورد علاقه گروه زیادی از کاربران کمک می‌کنند.[۲]

بنابراین تحقیقات در حوزه سیستم‌های توصیه گر کاملاً به رشته های «بازیابی اطلاعات»^۲ و «پالایش اطلاعات»^۳ وابسته است. بسیاری از تکنیک‌های توسعه داده شده در این حوزه‌ها از اطلاعات مشتق شده از محتوای آیتم‌ها برای رتبه دهی به آن‌ها استفاده می‌کنند. اساس پیشنهاد مبتنی بر محتوا^۴ بر مبنای توصیف ویژگی‌های آیتم‌ها و یک حساب کاربری که ارجحیت‌های کاربران را به این ویژگی‌ها منتسب می‌کند، است. اگر مجدداً به مثال کتاب فروشی آنلاین توجه کنیم می‌توان ویژگی‌های احتمالی کتاب‌ها از جمله طبقه، موضوع یا

¹ information overload

² information retrieval

³ information filtering

⁴ content based recommendation

نویسنده را شناسایی کرد. همانند توصیف آیتم‌ها، حساب کاربری کاربران نیز می‌تواند به صورت خودکار با تجزیه و تحلیل رفتار و بازخوردهای آن‌ها مشتق شود و یا به صورت صریح در مورد ارجحیت‌ها و علایق کاربران از آن‌ها سؤال شود.

در ارائه پیشنهاد مبتنی بر محتوا به این سؤال‌ها باید پاسخ داد:

- سیستم چگونه به صورت خودکار می‌تواند اطلاعات حساب کاربری کاربران را جمع‌آوری کرده و بروز رسانی کند؟
- چگونه می‌توان تطابق و یا حداقل شباهت آیتم‌ها با علایق کاربران را مشخص کرد؟
- چه تکنیک‌هایی را می‌توان برای استخراج توصیف ویژگی‌های آیتم‌ها به صورت خودکار به کار برد؟

سیستم‌های توصیه‌گر مبتنی بر محتوا نسبت سایر روش‌ها دارای دو مزیت است. اول اینکه نیازی به داشتن گروه کاربران زیاد برای ارائه پیشنهادات دقیق نیست. دوم اینکه آیتم‌های جدید را می‌توان بلافاصله بعد از مشخص شدن ویژگی‌هایشان پیشنهاد داد. [۱]

۳-۱ پیشنهاد دانش محور

اگر به سایر کاربردها مانند مصرف کنندگان وسایل الکترونیکی توجه کنیم مشاهده می‌کنیم که بسیاری از افراد فقط یک بار اقدام به خرید می‌کنند. بدین ترتیب نمی‌توان به تاریخچه خرید کاربران دسترسی داشت که پیش‌نیازی برای ارائه پیشنهاد در روش‌های مبتنی بر محتوا و پالایش گروهی محسوب می‌شود. از طرف دیگر ممکن است جزئیات بیشتر و ساختمان تری از آیتم‌ها مانند ویژگی‌های کیفی و تکنیکی در دسترس باشد. [۳]

برای مثال یک سیستم توصیه‌گر دوربین‌های دیجیتال را در نظر بگیرید که به کاربران نهایی در یافتن دوربینی متناسب با نیازهایش کمک می‌کند. مسلماً هر مشتری در هر چند سال اقدام به خرید یک دوربین می‌کند، بنابراین سیستم توصیه‌گر نمی‌تواند حساب کاربری ایجاد کند تا با استفاده از اطلاعات آن دوربین‌هایی که سایر کاربران پسندیده‌اند را پیشنهاد دهد.

بنابراین این سیستم به دانشی کاربردی و اضافه برای تولید پیشنهادات کارآمد نیاز دارد. در رویکرد های دانش محور^۱، سیستم توصیه گر معمولاً از اطلاعات اضافی و تهیه شده به صورت دستی در مورد کاربر جاری و آیتم‌های موجود استفاده می‌کند. یکی از این سیستم‌ها، سیستم‌های توصیه گر محدودیت محور^۲ است که جنبه های مختلف آن را در فصول بعدی بررسی خواهیم کرد.

در حوزه دوربین دیجیتال، یک سیستم محدودیت محور می‌تواند از دانش جزئی درباره ویژگی‌های دوربین‌ها شامل وضوح تصویر، وزن و قیمت استفاده کند. علاوه بر این، ممکن است از محدودیت‌های صریح برای توصیف نوع ارتباط ویژگی‌های خاص با مشتری‌ها استفاده شود. برای مثال یک دوربین با وضوح بالا برای مشتری که قصد چاپ عکس‌های بزرگ را داشته باشد مناسب است.

ارائه محصولاتی که تمامی نیاز های مشتری را برآورده می‌کنند به تنهایی نمی‌تواند یک پیشنهاد خوب محسوب شود. در این حالت جنبه های شخصی هر کاربر نادیده گرفته می‌شود و برای تمام کاربران با نیاز های یکسان پیشنهادات یکسانی ارائه می‌شود. بنابراین سیستم‌های توصیه گر محدودیت محور نیز به حساب کاربری نیاز دارند. در مثال دوربین دیجیتال سیستم می‌تواند در مورد نسبت اهمیت ویژگی‌ها از کاربر سؤالاتی بپرسد. برای مثال می‌تواند از کاربر بپرسد: آیا برای شما وضوح تصویر از وزن اهمیت بیشتری دارد؟

سؤالاتی که در خصوص سیستم‌های توصیه گر دانش محور باید پاسخ داد عبارتند از:

- چه نوع حوزه دانشی را می‌توان در پایگاه دانش نمایش داد؟
- چه نوع مکانیزمی را می‌توان برای انتخاب و رتبه دهی آیتم‌ها بر مبنای ویژگی‌های کاربران استفاده کرد؟
- چگونه می‌توان اطلاعات حساب کاربری را در حوزه‌هایی که تاریخچه تعامل کاربر با سیستم وجود ندارد، جمع آوری کرد و ارجحیت‌های صریح کاربران را در نظر گرفت؟ [۱]

۴-۱ رویکرد های ترکیبی^۳

¹ knowledge-based approach

² constraint-based recommender system

³ Hybrid approaches

با توجه به روش‌های بحث شده مشخص است با توجه به حوزه مورد نظر هر روش دارای مزایا و همچنین معایبی است. یک راه حل بدیهی ادغام کردن تکنیک‌های مختلف در جهت تولید پیشنهادات دقیق‌تر و بهینه‌تر است. برای نمونه اگر اطلاعات کاربران و همچنین اطلاعات جزئی در خصوص آیتم‌ها وجود داشته باشد می‌توان یک سیستم توصیه‌گر بهینه‌تر با ترکیب کردن روش پالایش گروهی با تکنیک‌های مبتنی بر محتوا ایجاد کرد. در این نمونه این سیستم می‌تواند بر مشکل شروع سرد^۱ روش پالایش گروهی غلبه کند و بر اساس تجزیه و تحلیل محتوای آیتم‌ها و کاربران جدید اقدام به ارائه پیشنهاد مناسب نماید.[۴]

وقتی روش‌های مختلف را در یک سیستم توصیه‌گر با یکدیگر ترکیب می‌کنیم، باید به سؤال‌ها زیر پاسخ دهیم:

- کدام تکنیک‌ها را می‌توان با هم ترکیب کرد و پیش‌نیاز این ترکیب چیست؟
- چگونه باید نتیجه‌ی تکنیک‌های مختلف را وزن داد و چگونه می‌توان آن‌ها را به صورت پویا تعیین کرد؟

۱-۵ کاربرد سیستم‌های توصیه‌گر

سیستم‌های توصیه‌گر کاربردهای فراوانی دارند که برخی از زمینه‌های کاربردی آن به شرح زیر است:

- تجارت الکترونیک: برای توصیه محصولات و خدمات مختلف.
- اینترانت بنگاهی: برای پیدا کردن افراد خبره در یک زمینه خاص و یا افرادی که در رویارویی با شرایط مشابه، تجاربی کسب کرده و راه‌حلهایی یافته‌اند. (درون سازمانی)
- کتابخانه دیجیتال: برای پیدا کردن کتاب، مقاله و اقلام دیجیتالی دیگر.
- کاربردهای پزشکی: برای انتخاب پزشک متناسب با شرایط (مکان، نوع بیماری و زمان) بیمار، انتخاب دارو.
- مدیریت ارتباط با مشتری: برای ارائه راهکارهایی جهت حل مشکلات تولیدکننده و مصرف‌کننده در زنجیره تأمین.

¹ ramp-up problem

۱-۶ اهداف پایان نامه

در این پایان نامه قصد داریم روش پالایش گروهی و روش محدودیت محور را با یکدیگر ترکیب کنیم. در روش محدودیت محور قبل از ارائه پیشنهاد باید پایگاه دانش آن توسط خبره‌های حوزه مورد نظر ایجاد شود. ایجاد کردن و نگهداری از پایگاه دانش به تلاش زیادی نیاز دارد. از این رو با استفاده از روش پالایش گروهی به صورت خودکار از اطلاعات کاربران موجود پایگاه دانش مورد نیاز روش محدودیت محور را تولید می‌کنیم و از آن برای ایجاد پیشنهادات استفاده می‌کنیم.

در فصل دوم به صورت مختصر سیستم‌های توصیه گر پالایش گروهی و محدودیت محور را بررسی می‌کنیم و همچنین مزایا و معایب روش پالایش گروهی مبتنی بر زمان که نوع خاصی از روش‌های پالایش گروهی است مشخص می‌شود.

در فصل سوم طراحی و پیاده سازی یک سیستم توصیه گر ترکیبی پیشنهادی مورد توجه قرار می‌گیرد و جنبه های مختلف آن مورد بررسی کامل قرار می‌گیرد.

در فصل چهارم نتایج اجرا و ارزیابی روش پیشنهادی با روش پالایش گروهی مبتنی بر زمان ذکر شده است.

در نهایت در فصل پنجم مطالب مطرح شده در پایان نامه جمع بندی شده و پیشنهادهایی برای انجام کارهای آتی آمده است.

فصل دوم

معرفی سیستم‌های توصیه‌گر و کارهای مرتبط

در این فصل به صورت مختصر سیستم‌های توصیه‌گر پالایش گروهی، محدودیت محور و ترکیبی را مورد بررسی قرار می‌دهیم. همچنین سیستم توصیه‌گر پالایش گروهی مبتنی بر زمان که قبلاً در [۸] طراحی و پیاده‌سازی شده است و مبنای ارزیابی سیستم ترکیبی که در این پایان‌نامه توسعه داده شده است، توصیف می‌شود.

در ادامه سیستم توصیه‌گر ترکیبی توصیف شده در [۵] که روش پالایش گروهی و محدودیت محور را با یکدیگر ترکیب کرده و در حوزه فروشگاه اینترنتی سیگار به کار برده است، بررسی می‌شود.

۱-۲ سیستم توصیه‌گر پالایش گروهی

به طور کلی عمده‌ترین دلیل استفاده از سیستم‌های پیشنهادگر این است که برای یک کاربر تعداد زیادی آیت‌م برای انتخاب در یک زمان کوتاه وجود دارد و او نمی‌تواند تمام آیت‌ها را از نظر کاربرد ارزیابی کند. سیستم‌های توصیه‌گر با پالایش کردن اطلاعات غیر مفید، کاربر مورد نظر را در یافتن آیت‌م مورد علاقه‌اش یاری می‌کنند.

این سیستم‌ها ارائه پیشنهادات را با در نظر گرفتن قانون وجود سلايق مشابه بين کاربران، به یک فرآیند خودکار تبدیل می‌کند. به عبارت دیگر آن دسته از کاربرانی که در

گذشته رفتارهایی مشابه یکدیگر داشته‌اند به احتمال زیاد در آینده نیز رفتار مشابهی در مورد آیتم‌های دیگر خواهند داشت. مدل کاربر بر اساس خصوصیات شخصی و همچنین نوع تعاملات او با سیستم ایجاد می‌شود. [۶]

۱-۱-۲ مفاهیم پایه

• کاربر نهایی

هدف نهایی سیستم توصیه‌گر ارائه پیشنهاد به کاربران است. کاربر نهایی^۱ یا کاربر جاری^۲ یا کاربر هدف^۳، کاربری است که سیستم قصد ارائه پیشنهاد آیتم‌هایی را به او دارد.

• آیتم

آیتم یک واژه عمومی است که به محتوای چیزهای استفاده شده توسط سیستم گفته می‌شود که می‌تواند موسیقی، فیلم، صفحه وب، کتاب یا کالاهای دیگری باشد.

• نرخ‌گذاری

یکی از مهم‌ترین عناصر در سیستم‌های توصیه‌گر نرخ‌گذاری آیتم‌ها توسط کاربران است زیرا سیستم بر اساس این اطلاعات عملیات خود را انجام می‌دهد. در این سیستم‌ها معمولاً مفید بودن یک آیتم به وسیله نرخ‌گذاری مشخص می‌شود و میزان علاقه یک کاربر به آن آیتم را نشان می‌دهد.

• نمایه کاربر

مدل کاربر یا نمایه^۴ کاربر شامل اطلاعات شخصی کاربر و همچنین تعاملات او با سیستم می‌باشد. با کنکاش نمایه کاربر می‌توان علاقه او در طول زمان به آیتم‌های مختلف را ارزیابی کرد و همچنین کاربران مشابه به او را یافت.

• ماتریس کاربر-آیتم

نرخ‌گذاری‌های انجام شده توسط کاربران مختلف را می‌توان در قالب یک ماتریس ارائه کرد. در این ماتریس ستون‌ها متناظر با آیتم‌ها و سطرهای آن کاربران می‌باشند. هر عنصر ماتریس برابر با نرخ (مقدار عددی) است که کاربر به آیتم مربوطه داده است. البته در بعضی

¹ end user

² current user

³ target user

⁴ profile

از موارد این ماتریس به صورت دودویی نمایش داده می‌شود. در ماتریس دودویی اگر عنصری برابر با یک بود یعنی کاربر آیتم مورد نظر را خریده است و در صورتی که کاربر آیتم را خریداری نکرده باشد آن مقدار برابر با صفر خواهد بود. [۶]

۲-۱-۲ روش پالایش گروهی

روش پالایش گروهی میزان علاقه کاربر به یک آیتم را بر اساس تعاملات قبلی کاربران با آیتم‌ها، پیش بینی می‌کنند. در ابتدا کاربران بازخوردهایی را به سیستم می‌دهد سپس سیستم می‌تواند تخمین‌هایی بر اساس بازخوردهای جمع آوری شده و مقایسه آن‌ها با بازخوردهای کاربر هدف تولید کند.

سیستم‌های پالایش گروهی محاسبات خود را بر پایه ماتریس تعاملات کاربر-آیتم انجام می‌دهند که هر ردیف آن بیانگر نمایه کاربر است و ستون‌ها آیتم‌های موجود در سیستم را معرفی می‌کنند. [۶]

۳-۱-۲ چالش‌های پالایش گروهی

اگر چه پالایش گروهی یکی از پر کاربردترین روش‌های سیستم‌های توصیه‌گر هستند همچنان با چالش‌های زیر روبرو هستند.

(۱) پراکندگی داده^۱

یکی از مهم‌ترین مشکلات در وضعیتی که حجم وسیعی از کاربران و آیتم‌ها در مجموعه داده وجود داشته باشد رخ می‌دهد که سطح پایین پوشش نرخ گذاری کاربران بین آیتم‌هاست. در بسیاری از موارد ماتریس کاربر-آیتم دارای پوشش ۱ درصد است. در نتیجه پیدا کردن همسایه‌های قابل اطمینان کاربر هدف کار دشواری است.

موضوع دیگری که در این زمینه وجود دارد کاربران با علائق غیر معمول است. این امر باعث شده که همسایگان اندکی توسط سیستم برای آن‌ها یافت شود در نتیجه پیشنهادات ارائه شده از اعتبار کمی برخوردار هستند.

(۲) مشکل شروع سرد^۲

این مشکل برای هر دو مؤلفه کاربران و آیتم‌ها رخ می‌دهد. از آنجایی که الگوریتم‌های پالایش گروهی بر اساس نرخ گذاری‌های کاربران محاسبات خود را انجام می‌دهند. ارائه

¹ Data dispersion

² ramp-up (cold start)