



دانشگاه اصفهان

دانشکده مهندسی کامپیوتر

گزارش ارائه داده کاوی

نام و نام خانوادگی: سید عرفان نوربخش

شماره دانشجویی: ۹۸۳۶۱۳۰۵۹

گرایش: نرم افزار

استاد: دکتر کیانی

تاریخ تنظیم گزارش: ۱۴۰۱/۹/۲۹

فهرست مطالب

فصل اول: معرفی سیستم های پیشنهاد دهنده	۵
۱-۱ مقدمه	۵
۲-۱ منظور از سیستم های پیشنهاد دهنده	۶
۳-۱ تاثیر استفاده از سیستم های پیشنهاد دهنده در دستاوردهای شرکت ها	۶
۴-۱ گستردگی سیستم های پیشنهاد دهنده	۷
۵-۱ مزایای استفاده از سیستم های پیشنهاد دهنده	۸
فصل دوم: اصول و انواع سیستم های پیشنهاد دهنده	۱۰
۱-۲ مقدمه	۱۰
۲-۲ اصول سیستم های پیشنهاد دهنده	۱۰
۳-۲ روش های فراهم کردن داده ها	۱۱
۴-۲ معیارهای مشابهت	۱۲
۵-۲ معرفی انواع سیستم های پیشنهاد دهنده	۱۴
۱-۵-۲ فیلتر مشارکتی (Collaborative Filtering)	۱۴
۲-۵-۲ فیلتر دموگرافیک (Demographic Filtering)	۱۹
۳-۵-۲ فیلتر مبتنی بر محتوا (Content-based Filtering)	۱۹
۴-۵-۲ فیلتر مبتنی بر اجتماعی (Social-based Filtering)	۲۳
۵-۵-۲ فیلتر متن آگاه (Context-aware Filtering)	۲۳
۶-۵-۲ سیستم های توصیه آگاه از موقعیت مکانی	۲۴
۷-۵-۲ فیلتر مبتنی بر دانش (Knowledge-based Filtering)	۲۴
۸-۵-۲ روش های الهام گرفته شده از طبیعت (Bio-Inspired)	۲۴

۲-۶	ارزیابی سیستم‌های پیشنهاد دهنده	۲۵
فصل سوم:	پیاده‌سازی سیستم‌های پیشنهاد دهنده	۲۶
۱-۳	مقدمه	۲۶
۲-۳	پروژه سیستم پیشنهاد دهنده براساس فیلتر مبتنی بر محتوا	۲۶
۳-۳	پروژه سیستم پیشنهاد دهنده براساس فیلتر مشارکتی	۲۷
منابع و لینک	پروژه‌های انجام شده	۲۸

چکیده.

اینترنت به عنوان منبعی برای حجم انبوه داده‌ها و اطلاعات محسوب می‌شود. همچنین، کالاها و خدمات متنوعی نیز از طریق اینترنت در دسترس عموم مردم هستند. در این اقیانوس، نیاز به ابزاری برای پالایش، اولویت‌بندی و تحویل موثر اطلاعات مورد نیاز و مرتبط به هر کاربر به او محسوس است. امروزه، سرریز اطلاعات مشکلات متعددی را برای کاربران اینترنت به وجود آورده است. ریکامندر سیستم در صدد است تا این مشکل را با جستجو در میان حجم انبوهی از اطلاعاتی حل کند که همه روزه به صورت پویا تولید می‌شوند و محتوا و خدمات شخصی‌سازی شده برای هر کاربر را در اختیار او قرار دهد.

فصل اول

معرفی سیستم‌های پیشنهاد دهنده

۱-۱ مقدمه

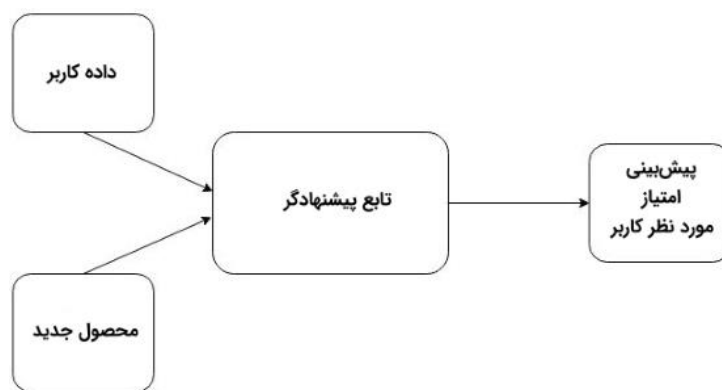
یکی از اهداف اصلی بازاریابی، جذب توجه و علاقه مخاطب به محصول و کسب و کار است. هر چقدر که بتوان مشتری را با محصول خود بیشتر درگیر و علاقه‌مند کرد، احتمال خرید او و تبدیل شدنش به یک مشتری وفادار بالاتر می‌رود. در این سال‌ها که مشتریان با حجم زیادی از تبلیغات و محصولات روبرو می‌شوند، وقت کمتری را برای توجه به تبلیغات برندها صرف می‌کنند. در چنین وضعیتی، این موضوع بسیار اهمیت پیدا می‌کند که در زمان کمی که می‌توانیم توجه مخاطب را به خود جلب کنیم، چه محصول یا تبلیغی به مخاطب پیشنهاد شود تا شانس جلب توجه او بیشتر شود. این توصیه هر چقدر که مناسب ویژگی‌های مشتری باشد احتمال تعامل مشتری بالا می‌رود. سیستم‌های توصیه‌گر با همین هدف ساخته شده‌اند تا با تحلیل اطلاعات ورودی از مشتریان، بتوانند بهترین توصیه را به آن‌ها بکنند. همچنین، از کاربردهای موثر علم داده بخصوص در فروشگاه‌های آنلاین است که سعی می‌کند علایق یک کاربر خاص را پیش‌بینی کند و آیتم‌هایی متناسب با علایقش را به وی نشان دهد.

۲-۱ منظور از سیستم‌های پیشنهاد دهنده

یک سیستم پیشنهاد دهنده یا سامانه پیشنهادگر (Recommender System) نوعی الگوریتم است که هدف آن پیشنهاد دادن اقلام مورد علاقه کاربر به اوست. مواردی مانند فیلم برای تماشا، متن برای خواندن، محصول برای خرید یا هر چیز دیگری که در صنعت یافت می‌شود. هدف سیستم پیشنهاد دهنده ارائه توصیه‌های معنی‌دار به مجموعه‌ای از کاربران برای استفاده از کالا یا خدماتی است که ممکن است مورد علاقه آنها باشد. پیشنهاداتی برای خرید کتاب در آمازون و یا فیلم در نتفلیکس، نمونه‌های واقعی از کاربرد سیستم‌های پیشنهاد دهنده در صنعت می‌باشد.

سیستم پیشنهاد دهنده هوشمند یک زیرکلاس از «سامانه پالایش اطلاعات» (Information Filtering System) است. ریکامندر سیستم در تلاش است که امتیاز احتمالی که یک کاربر ممکن است به یک قلم بدهد را پیش‌بینی کند. در واقع، ریکامندر در تلاش است تا ترجیحات کاربر از میان یک مجموعه از اقلام را پیش‌بینی کند. ریکامندر سیستم‌ها، سیستم‌های یادگیری ماشین هستند که به کاربران کمک می‌کنند تا محصولات و خدمات جدید را کشف کنند.

یک مولفه مهم در تمام سیستم‌های پیشنهاد دهنده، تابع پیشنهادگر (Recommender Function) است که اطلاعات را پیرامون کاربر دریافت و بر اساس آن، برای مثال، امتیاز (رتبه) احتمالی که کاربر ممکن است به یک محصول بدهد را پیش‌بینی می‌کند. پیش‌بینی امتیازهای مورد نظر کاربر حتی پیش از آنکه کاربر واقعا آن رتبه را ثبت کرده باشد، یک ریکامندر سیستم را به ابزاری قدرتمند و بعضا شگفت‌انگیز مبدل می‌کند.



۳-۱ تاثیر استفاده از سیستم‌های پیشنهاد دهنده در دستاوردهای شرکت‌ها

یک کسب و کار موفق ممکن است بتواند بدون داشتن ریکامندر سیستم بقا پیدا کند. اگرچه، در صورتی که کسب و کار بخواهد از قدرت داده‌ها به عنوان اهرمی برای ساخت تجربه کاربری بهتر و افزایش درآمد استفاده کند، باید مسئله پیاده‌سازی ریکامندر سیستم را به طور جدی در نظر داشته باشد.

در ادامه، برخی از آمارهای موجود از شرکت‌های مختلف که از سیستم‌های پیشنهاد دهنده در سیستم‌ها و نرم‌افزارهای خود استفاده کرده‌اند، برای سال‌های قبل از ۲۰۲۱ ارائه شده است که این آمارها برای سال‌های نزدیک به ۲۰۲۲، رشد افزاینده‌ای داشته است.

✓ بنابر آمار ارائه شده توسط مک‌کنزی (McKinsey)، ۳۵ درصد از خریدهای انجام شده از وبسایت خرده‌فروشی آمازون، در نتیجه استفاده از ریکامندر سیستم بوده است.

✓ مطابق با آمار ارائه شده توسط علی‌زیلا (Alizila)، شرکت علی‌بابا با استفاده از صفحات لندینگ شخصی‌سازی شده، شاهد رشد ۲۰ درصدی در نرخ تبدیل‌های (Conversion Rate) خود بوده است.

✓ ۷۰ درصد زمانی که افراد به مشاهده ویدئو در یوتیوب می‌پردازند، با استفاده از سیستم‌های پیشنهاد دهنده در آن است.

✓ مطابق با آمار ارائه شده توسط مک‌کنزی (McKinsey)، ۷۵ درصد از آنچه که افراد در نتفلیکس مشاهده می‌کنند، بر اساس پیشنهادهای ریکامندر سیستم است.

✓ مطابق با مقاله ژورنال ارائه شده توسط کارلوس ای گومز اوریب (Carlos A Gomez-Urbe) و نیل هانت (Neil Hunt)، استفاده از یک سیستم پیشنهاد دهنده، نتفلیکس را قادر به آن می‌سازد که در حدود یک میلیارد دلار در هر سال ذخیره کند.

✓ بر اساس آمار ارائه شده توسط دیجی مگ، طی سال ۲۰۲۱، استفاده از سیستم توصیه‌های هوشمند در دیجی کالا ۳۰ درصد رشد داشته است.

۴-۱ گستردگی سیستم‌های پیشنهاد دهنده

تقریباً هر کسب و کاری می‌تواند از یک سیستم پیشنهاد دهنده بهره‌مند شود. در ادامه دو جنبه مهم که تعیین می‌کند یک کسب‌وکار به چه میزان از یک سیستم توصیه دهنده سود می‌کند، می‌پردازیم.

✓ **وسعت داده‌ها:** کسب‌وکاری که فقط به تعداد انگشت شماری از مشتریانی که به روش‌های مختلف رفتار می‌کنند خدمات ارائه می‌کند، از یک سیستم توصیه خودکار سود چندانی دریافت نخواهد کرد. انسان‌ها هنوز در زمینه یادگیری بسیار بهتر از ماشین‌ها هستند. در چنین مواردی، کارکنان از منطق، درک کمی و کیفی مشتریان برای ارائه توصیه‌های دقیق استفاده می‌کنند.

✓ **عمق داده‌ها:** وجود یک نقطه داده واحد برای هر مشتری نیز برای سیستم‌های توصیه مفید نیست. داده‌های عمیق در مورد فعالیت‌های آنلاین مشتریان و در صورت امکان خریدهای آفلاین می‌تواند پیشنهادهای دقیقی بکنند.

با این چارچوب، می‌توانیم کسب‌وکارهایی را که از سیستم‌های توصیه سود می‌برند را شناسایی کنیم:

✓ **فروشگاه‌های اینترنتی:** یکی از کاربردهای اصلی سیستم‌های ریکامندر در فروشگاه‌های اینترنتی و با ارائه پیشنهاد هدفمند محصول‌هایی به بازدیدکننده سایت است که احتمالاً به آن محصولات سر می‌زنند. فروشگاه‌های بزرگ مانند آمازون دارای پیشرفته‌ترین سیستم‌های ریکامندر هستند که فروش زیادی را از این طریق به دست می‌آورند. در ایران نیز سایت‌های فروشگاهی بزرگ مانند دیجیکالا از این سیستم استفاده می‌کنند. فروشگاه‌های متوسط و کوچک نیز می‌توانند از سطح ساده‌تر این سیستم‌ها به شکل افزونه استفاده کنند. این افزونه می‌تواند از تگ‌های موجود در محصولات و مقالات استفاده کرده و پیشنهاد‌های ساده ولی مرتبط به بازدیدکننده‌ها بدهد.

✓ **سایت‌های UGC:** سایت‌ها و برنامه‌هایی که از محتوای ساخته شده توسط کاربر (UGC) استفاده می‌کنند، یکی از مشتریان اصلی سیستم‌های ریکامندر هستند. سایت‌های ویدیویی مانند یوتیوب و آپارات، اپلیکیشن‌هایی مانند اینستاگرام و پینترست، سایت‌های تعاملی مانند فیسبوک و حتی سایت‌های بیزینسی مثل لینکدین همگی از نمونه‌های این سایت‌ها هستند.

✓ **تبلیغات اینترنتی:** همانند محتوا اگر تبلیغات پیشنهاد شده مناسب علایق کاربران باشد، میزان تعامل و احتمال کلیک و خرید آن‌ها بیشتر می‌شود و در نتیجه، کسب و کار به درآمد بیشتری می‌رسد. به همین دلیل این سایت‌ها در نشان دادن تبلیغات هم از سیستم‌های ریکامندر استفاده می‌کنند تا کاربر تبلیغ‌های نامرتب نبیند و تجربه بهتری داشته باشد.

✓ **سایت‌های محتوایی:** در این سایت‌ها، درگیر کردن بیشتر مخاطب با محتوای سایت به عنوان هدف سیستم در نظر گرفته می‌شود. این سیستم‌ها با الگوریتم‌های خود تلاش می‌کنند تا از بین محتواهای زیاد موجود در سایت، بهترین محتواها را بر اساس علایق کاربر به او پیشنهاد دهند.

۱-۵ مزایای استفاده از سیستم‌های پیشنهاد دهنده

در ادامه به اهمیت استفاده از سیستم‌های توصیه‌گر در کسب و کار پرداخته می‌شود تا دلیل توجه زیاد کسب و کارهای بزرگ به این سیستم‌ها مشخص شود.

✓ شرکت‌ها از ریکامندر سیستم‌ها به منظور افزایش فروش با شخصی‌سازی کردن پیشنهاد‌های ارائه شده به کاربر و بهبود تجربه مشتریان استفاده می‌شوند.

✓ ریکامندر سیستم‌ها به شرکت‌ها کمک می‌کنند تا از رقبای خود جلوتر باشند و درآمد خود را افزایش دهند و امنیت بهتری را برای مشتریان خود تضمین بکنند.

✓ پیشنهاد‌های ارائه شده توسط یک سیستم پیشنهادگر جستجو را سرعت می‌بخشد و به کاربران امکان دسترسی به محتوایی که به آن علاقه دارند را می‌دهد.

- ✓ شرکت‌ها با استفاده از ریکامندر سیستم‌ها می‌توانند به ارسال ایمیل‌هایی بپردازند که حاوی لینک‌های مربوط به اقلام پیشنهادی بر اساس تمایلات کاربر است.
- ✓ هنگامی که کاربران شروع به دریافت پیش‌بینی‌ها بر اساس علائق خود می‌کنند، اعتمادشان به شرکت جلب می‌شود. این امر احتمال اینکه آن‌ها مشتری شرکت باقی بمانند و محتوا، محصول یا خدمات بیشتری از سازمان را مصرف کنند، افزایش می‌دهد. و همین باعث می‌شود که خرید مشتریان استمرار پیدا بکند.

فصل دوم

اصول و انواع سیستم‌های پیشنهاد دهنده

۲-۱ مقدمه

سیستم‌های پیشنهاد دهنده به صورت برنامه‌نویسی شده و خودکار عمل می‌کنند و تنها منبع ورودی اطلاعات و هدف خروجی برای این سیستم‌ها مشخص می‌شود تا شروع به کار کنند. هدف اصلی، بر اساس جایی که این سیستم‌ها قرار است مورد استفاده قرار بگیرند، متفاوت است و می‌تواند پیشنهاد محصول، محتوا یا چیزهای دیگر باشد. ریکامندر سیستم با گردآوری اطلاعات پیرامون کاربر و اقلام و دریافت بازخورد صریح و ضمنی از کاربر یاد می‌گیرد که چگونه پیش‌بینی کند و بر این اساس، پیشنهاداتی را به کاربر ارائه می‌کند.

۲-۲ اصول سیستم‌های پیشنهاد دهنده

برای ایجاد یک سیستم پیشنهاد دهنده کارا مواردی وجود دارند که باید به آنها توجه نمود و در پروسه طراحی و پیاده سازی سیستم آنها را باید در نظر گرفت، که این موارد بدین صورت است:

- ✓ **نوع داده‌های موجود:** بنا به کاربرد سیستم ممکن است انواع مختلفی از منابع اطلاعاتی در سیستم وجود داشته باشد. این اطلاعات می‌توانند امتیازهای کاربران به آیتم‌ها (ratings)، اطلاعات شخصی کاربران، محتوای مربوط به محصولات سیستم، ارتباطات موجود در شبکه‌های اجتماعی و اطلاعات مربوط به موقعیت کاربر (location-aware information) باشند.
- ✓ **الگوریتم فیلترینگ مورد استفاده:** هدف سیستم‌های پیشنهاد دهنده در واقع رتبه‌بندی آیتم‌های سیستم به لحاظ نزدیک بودن به علائق کاربران می‌باشد تا در هنگام ارائه پیشنهاد، آیتم‌هایی با رتبه بالاتر را به کاربر پیشنهاد دهند. از مهم‌ترین الگوریتم‌ها می‌توان به Collaborative Filtering، Social-based Filtering، Knowledge-based Filtering، Context-aware Filtering، Hybrid Filtering و Content-based Filtering اشاره کرد.
- ✓ **مدل انتخابی برای سیستم:** برای پیاده‌سازی سیستم‌های پیشنهاد دهنده دو راهکار استفاده می‌شود. اولین راهکار، استفاده مستقیم از داده‌های موجود سیستم است که به این راهکار، روش مبتنی بر حافظه یا memory-based گفته می‌شود و دیگری که کمی هوشمندتر است استفاده از یک مدل در سیستم است که راهکار مبتنی بر مدل یا model-based گفته می‌شود.
- ✓ **تکنیک مورد استفاده در پیشنهادات:** راهکارها و تکنیک‌های مختلفی برای پیاده‌سازی هسته یک سیستم پیشنهاد دهنده وجود دارد. بعنوان مثال می‌توان از الگوریتم‌های ژنتیک، شبکه‌های عصبی، شبکه‌های Bayesian، روش‌های احتمالی و یا الگوریتم‌های همسایگی برای پیاده‌سازی هسته چنین سیستم‌هایی استفاده می‌شوند.
- ✓ **کارایی مطلوب سیستم:** از نظر میزان حافظه مصرفی و همچنین زمان، باید سیستم مدنظرمان را بررسی کرد که متناسب با آن سیستم پیشنهاد دهنده‌ای انتخاب شود.
- ✓ **کیفیت نتایج قابل ارائه:** نتایجی که یک سیستم پیشنهاد دهنده ارائه می‌دهد، می‌تواند با توجه به کاربرد آن و اینکه در چه حوزه‌ای از بازار کسب و کار در حال استفاده است متفاوت باشد. بطور کلی در سیستم‌های پیشنهاد دهنده هدف فراهم آوردن سه فاکتور تازگی (novelty)، دقت (precision) و پوشش (coverage) در نتایج پیشنهادی است.

۲-۳ روش‌های فراهم کردن داده‌ها

دو دسته از روش‌های فراهم کردن داده برای سیستم‌های پیشنهاد دهنده «رتبه‌بندی صریح» (Explicit Rating) و «رتبه‌بندی ضمنی» (Implicit Rating) هستند.

- ❖ **رتبه‌بندی صریح:** توسط کاربران انجام می‌شود و سیستم پیشنهاد دهنده، نظر کاربر را عیناً استخراج می‌کند. مثال‌هایی از این مورد می‌توان به رتبه‌بندی با دادن ستاره (Star Rating)، نوشتن نقد و بررسی (Review)، بازخورد (Feedback)، پسند (Like) و «دنبال کردن» (Following) اشاره کرد.
- ❖ **رتبه‌بندی ضمنی:** هنگامی انجام می‌شود که کاربران با اقلام در تعامل هستند و سیستم پیشنهاد دهنده، رفتار کاربر را استنتاج می‌کند. مثال‌هایی از این مورد می‌توان به «کلیک‌ها» (Clicks)، «بازدیدها» (Views) و «خریدها» (Purchases) اشاره کرد.

۴-۲ معیارهای مشابهت

معیارهای مشابهت سنجه‌هایی در سنجش فاصله هستند. نزدیک‌ترین نقاط به یکدیگر بیشترین شباهت را به یکدیگر دارند و دورترین نقاط، کم‌ترین مشابهت را به هم دارند. متداول‌ترین معیارهای مشابهت عبارتند از:

- ❖ **فاصله مینکوفسکی (Minkowski Distance):** اگر A و B دو نقطه در فضای p بعدی باشند، فاصله مینکوفسکی برایشان به صورت زیر محاسبه می‌شود. پارامتر فاصله مینکوفسکی در اینجا d در نظر گرفته شده. فاصله مینکوفسکی با پارامتر d را به صورت $D_{mink}(A, B; d)$ نشان می‌دهند.

$$D_{mink}(A, B; d) = \left(\sum_{i=1}^p |x_i - y_i|^d \right)^{\frac{1}{d}}$$

- ❖ **فاصله منهتن (Manhattan Distance):** به تابع فاصله حاصل از مجموع قدرمطلق فاصله بین مولفه‌های نقاط، تابع فاصله منهتن می‌گویند. اگر X و Y دو نقطه با p مولفه باشند، فاصله منهتن بین این دو به صورت زیر قابل محاسبه است:

$$D_{man} = \sum_{i=1}^p |x_i - y_i|$$

- ❖ **فاصله اقلیدسی (Euclidean Distance):** کوتاه‌ترین فاصله بین دو نقطه برطبق رابطه فیثاغورث، محاسبه می‌شود. اگر X و Y دو نقطه با p مولفه باشند، فاصله اقلیدسی بین این دو به صورت زیر قابل محاسبه است:

$$D_{euc} = \left(\sum_{i=1}^p (x_i - y_i)^2 \right)^{\frac{1}{2}}$$

- ❖ **فاصله همینگ (Hamming Distance):** فاصله همینگ یعنی تعداد مقادیری که در بین دو بردار باهم متفاوت هستند. این معیار معمولاً برای مقایسه دو رشته باینری با طول یکسان استفاده می‌شود.

همچنین از این معیار برای بررسی شباهت بین دو رشته میتوان استفاده کرد. اینکه چه تعداد از کاراکترهای دو رشته باهم متفاوت هستند.

A	1	0	1	1	0	0
	↑		↑			
B	1	1	1	0	0	0

- "karolin" and "kerstin" is 3.
- "kathrin" and "kerstin" is 4.
- 1011101 and 1001001 is 2.

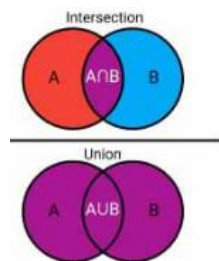
❖ **شباهت کسینوسی (Cosine Similarity):** شباهت کسینوسی یک معیار شباهت است که پایه آن محاسبه‌ی میزان کسینوس زاویه‌ی بین دو بردار است. در صورت انطباق دو بردار (در این معیار نشانه شباهت کامل است) که زاویه‌ی بین دو بردار صفر می‌باشد مقدار آن برابر ۱ خواهد شد و در کمترین میزان شباهت دو بردار یعنی اگر زاویه بین دو بردار ۱۸۰ درجه باشد نتیجه این معیار ۱- خواهد شد.

$$\cos(x, y) = \frac{x \cdot y}{||x|| \cdot ||y||} = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n (x_i)^2} \sqrt{\sum_{i=1}^n (y_i)^2}}$$

❖ **ضریب همبستگی پیرسون (Pearson Coefficient):** اگر X و Y دو متغیر تصادفی بوده که دارای امید-ریاضی $E(X)$ و $E(Y)$ و واریانس $V(X)$ و $V(Y)$ هستند. ضریب همبستگی بین X و Y را با $\rho(X, Y)$ یا $\text{corr}(X < Y)$ نشان داده و به صورت زیر محاسبه می‌کنیم.

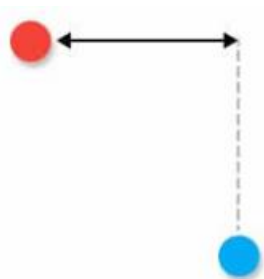
$$\rho(X, Y) = \text{corr}(X, Y) = \frac{E[(X - E(X))(Y - E(Y))]}{[V(X)V(Y)]^{\frac{1}{2}}}$$

❖ **اندیس ژاکارد (Jaccard Index):** شاخص ژاکارد معیاری است که برای محاسبه شباهت و تنوع نمونه‌های مجموعه استفاده می‌شود. فاصله برابر با اندازه اشتراک تقسیم بر اندازه اجتماع نمونه‌های مجموعه است.



$$D(x, y) = 1 - \frac{|x \cap y|}{|y \cup x|}$$

❖ **فاصله چبیشف (Chebyshev Distance):** فاصله چبیشف برابر با ماکزیمم اختلاف بین درایه‌های دو بردار است. به زبان ساده، در این معیار اختلاف تک تک ویژگی‌ها با هم محاسبه شده و سپس ماکزیمم فاصله بین دو ویژگی به عنوان فاصله دو بردار استفاده می‌شود.



$$D(x, y) = \max_i (|x_i - y_i|)$$

۲-۵ معرفی انواع سیستم‌های پیشنهاد دهنده

سیستم‌های پیشنهاد دهنده بر اساس سه رویکرد اصلی زیر ساخته می‌شوند:

✓ روش‌های مشارکت محور (Collaborative based methods) یا پالایش گروهی (Collaborative filtering)

❖ رویکردهای مبتنی بر حافظه (Memory based approaches)

➤ کاربر - کاربر (user-user)

➤ آیتم - آیتم (item-item)

❖ رویکردهای مبتنی بر مدل (Model based approaches)

➤ فاکتورگیری ماتریسی (matrix factorization)

✓ روش‌های محتوا محور (Content based methods)

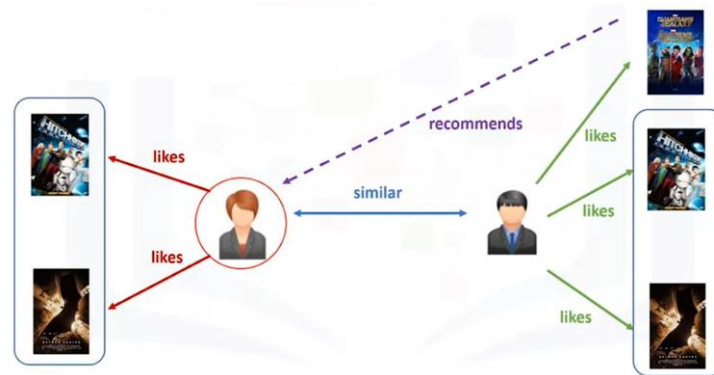
✓ روش‌های ترکیبی (Hybrid methods): در این روش از ترکیب روش‌های مشارکت محور و محتوا محور استفاده می‌شود.

اساسی ترین جزء در سیستم های پیشنهاد دهنده الگوریتم و راهکار فیلترینگ آن می‌باشد. در ادامه به مهمترین راهکارهای مورد استفاده در این زمینه خواهیم پرداخت.

۲-۵-۱ فیلتر مشارکتی (Collaborative Filtering)

یکی از مهمترین و پر استفاده ترین روش های فیلترینگ در سیستم های پیشنهاد دهنده بوده که پایه و مبنای کار در بسیاری از راهکار های دیگر نیز بشمار می رود. در روش Collaborative Filtering آنچه مد نظر قرار داده می شود، بیشتر تجربه دیگران است. در این روش کاربران در سیستم مشارکت می کنند و به آیتم های مختلف

موجود در سیستم امتیاز می‌دهند. البته این امتیاز دادن ها می توان به صورت ضمنی نیز اتفاق بیفتد و توسط سیستم تشخیص داده شود. (آیتم هایی که بیشتر دانلود شده اند احتمالاً از محبوبیت بیشتری برخوردار بوده اند). برای مثال، اگر دو کاربر از نظر فیلم‌های علاقه‌مندشان شبیه هم هستند یا همسایه هستند، می‌توانیم فیلمی را به کاربر فعالی که همسایه‌اش قبلاً دیده است، توصیه کنیم.



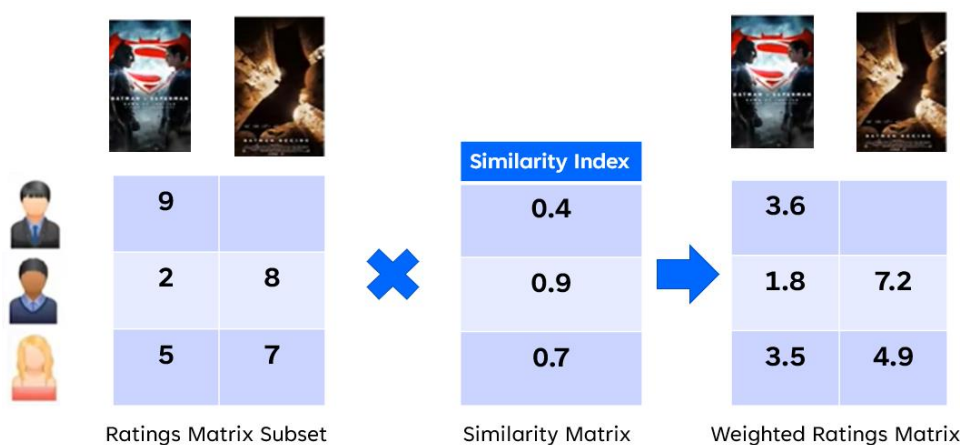
یک ماتریس آیتم کاربر ساده داریم که رتبه بندی چهار کاربر را برای پنج فیلم مختلف نشان می دهد که کاربر فعال سه فیلم از این پنج فیلم را تماشا کرده و به آنها امتیاز داده است. حال می‌خواهیم دریابیم که کدام یک از دو فیلمی که کاربر فعال تماشا نکرده است باید به او توصیه شود. اولین قدم این است که کاربر فعال چقدر شبیه به سایر کاربران است که می توان از طریق چندین تکنیک مختلف آماری و برداری مانند اندازه گیری فاصله یا شباهت از جمله فاصله اقلیدسی، همبستگی پیرسون، تشابه کسینوس و غیره انجام داد. برای محاسبه میزان شباهت بین دو کاربر، از سه فیلمی که هر دو کاربر در گذشته امتیاز داده اند استفاده می کنیم. (صرف نظر از اینکه برای سنجش شباهت از چه چیزی استفاده می کنیم، برای مثال ، شباهت می تواند بین کاربر فعال و سایر کاربران ۰.۷، ۰.۹ و ۰.۴ باشد. این اعداد نشان دهنده وزن شباهت یا نزدیکی کاربر فعال به سایر کاربران در مجموعه داده است).

	9	6	8	4	
	2	10	6		8
	5	9		10	7
	?	10	7	8	?

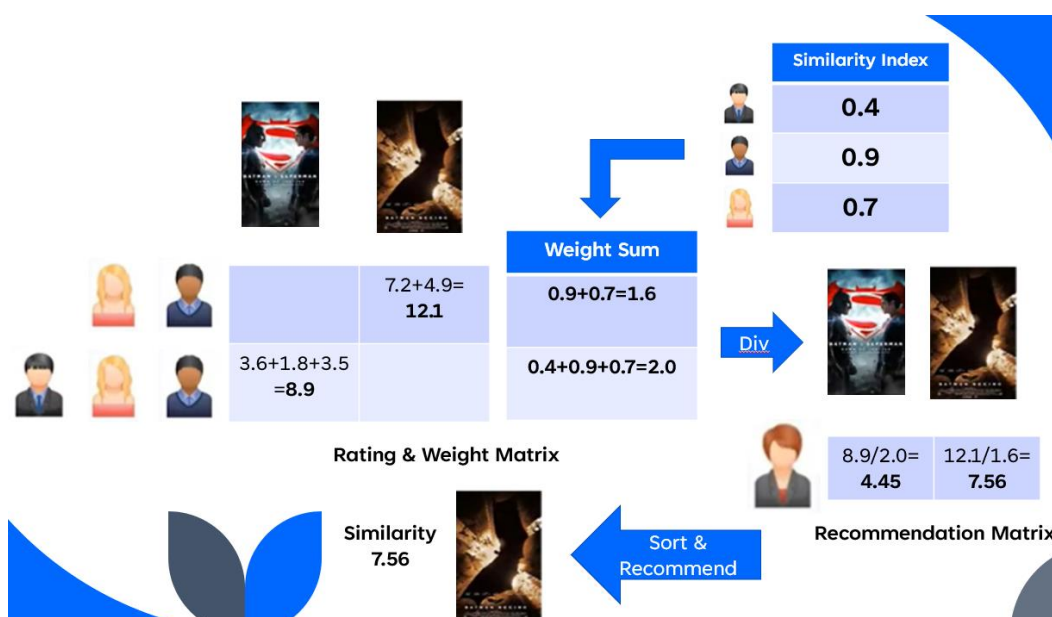
Rating Matrix

Similarity values: 0.4, 0.9, 0.7

مرحله بعدی ایجاد یک ماتریس رتبه بندی وزنی است. اکنون می‌توانیم از شباهت کاربران به کاربر فعال برای محاسبه نظر احتمالی کاربر فعال در مورد دو فیلم هدف استفاده کنیم که با ضرب وزن شباهت در رتبه بندی کاربران به دست می‌آید. این کار منجر به یک ماتریس رتبه بندی وزنی می‌شود که نشان دهنده نظر همسایگان کاربر در مورد دو فیلم کاندید برای توصیه است. در واقع، رفتار سایر کاربران را در بر می‌گیرد و به رتبه بندی کاربرانی که شباهت بیشتری به کاربر فعال دارند، وزن بیشتری می‌بخشد.



اکنون، می‌توانیم ماتریس توصیه را با جمع همه نرخ‌های وزنی ایجاد کنیم. با این حال، از آنجایی که سه کاربر به فیلم اول و دو کاربر به فیلم دوم امتیاز دادند، باید مقادیر رتبه بندی وزنی را نرمال کنیم. این کار را با تقسیم آن بر مجموع شاخص شباهت برای کاربران انجام می‌دهیم. نتیجه امتیاز وزن‌داری است که کاربر فعال بر اساس شباهت خود به سایر کاربران به این فیلم‌ها می‌دهد. بدیهی است که می‌توانیم از آن برای رتبه بندی فیلم‌ها برای ارائه توصیه به کاربر فعال خود استفاده کنیم.



در صورتی که سیستم تازه شروع به کار کرده باشد و یا آیتم جدیدی به سیستم اضافه شود، اطلاعات کافی از محصولات در دسترس نخواهد بود و در نتیجه نمی‌توان به درستی امتیازدهی و رتبه‌بندی را انجام داد که با عنوان شروع سرد (Cold Start) شناخته می‌شود. از مشکلات دیگر این سیستم‌ها می‌توان به پراکندگی داده‌ها (Data Sparsity) اشاره کرد، که اطلاعات در سیستم وجود دارد اما پراکنده هستند و نمی‌توان بدرستی و با قطعیت گفت که چه محصولی مقبولیت بیشتری دارد.

در ادامه دلایل بوجود آمدن مشکل cold start و رفع آن در این سیستم‌ها می‌پردازیم.

✓ **شروع کار سیستم پیشنهاد دهنده:** بهترین راهکار برای آن این است که، با استفاده از روش‌های مناسب کاربران را تشویق به دادن رای به آیتم‌ها نماییم و زمانی اقدام به پیشنهاد به کاربر کنیم که به اندازه کافی اطلاعات جمع‌آوری شده باشد.

✓ **ورود کاربر جدید به سیستم:** بهترین راهکار برای آن این است که، عموماً این روش را با دیگر روش‌های رایج در سیستم‌های پیشنهاد دهنده ترکیب شود و یک سیستم ترکیبی (Hybrid) را ساخته شود. (مثلاً Collaborative Filtering با Content-based Filtering).

✓ **اضافه کردن آیتم جدید در سیستم:** آیتم‌های جدید معمولاً در لیست پیشنهادات هرگز آورده نمی‌شوند و از دیدگاه کاربران نیز پنهان می‌مانند. بهترین راهکار برای آن این است که، بر اساس روش‌ها و ابزارهای دیگری این آیتم‌ها را به کاربران نشان داد تا به آنها رای دهند.

برای رفع مشکل پراکندگی داده‌ها در این نوع از سیستم‌ها، معمولاً از تکنیک dimensionality reduction استفاده می‌شود. در کنار این روش، تکنیک‌های Latent Semantic Index و Singular Value Decomposition نیز وجود دارند که تکنیک SVD علیرغم نتایج بسیار خوب، سربار پردازشی بالایی دارد و از آنها بهتر است در کاربردهای آفلاین استفاده نمود که تغییرات زیادی در اطلاعات آنها وجود ندارند.

از دیگر راهکارهای مقابله با این مشکلات می‌توان به استفاده از تکنیک‌های خوشه‌بندی، پیاده‌سازی Collaborative Tagging در یک سیستم مبتنی بر Collaborative Filtering اشاره کرد. در تکنیک خوشه‌بندی، با خوشه‌بندی کردن آیتم‌ها یا کاربران و یا هر دوی آنها (bi-clustering) باعث بهبود کارایی سیستم پیشنهاد دهنده می‌گردد. در تکنیک Collaborative Tagging، با شناخت صلاحیت کاربران، آیتم‌ها را براساس تمایلات کاربران دسته‌بندی کرد.

الگوریتم K نزدیک‌ترین همسایگی (KNN)

از رایج‌ترین الگوریتم‌ها در روش CF بوده که دو رویکرد کاربر به کاربر (user-to-user) و آیتم به آیتم (item-to-item) به سیستم دارد.

❖ **رویکرد کاربر به کاربر:** الگوریتم‌هایی که بر مبنای این نگاه پیاده سازی می شوند شامل سه گام هستند:

➤ **گام اول:** بر اساس یک معیار شباهت (Mean, Pearson Correlation, cosine)

(square difference) برای کاربر a تعداد k همسایه که بیشترین شباهت را به کاربر a دارند انتخاب می شود.

➤ **گام دوم:** برای تمامی آیتم‌های موجود در سیستم معیاری کمی برای پیشبینی

(prediction) آنکه آیا آیتم جدید، مورد پسند کاربر a قرار خواهد گرفت یا خیر محاسبه می شود. محاسبه این معیار کمی با استفاده از راهکارهای مختلفی (میانگین امتیازات (average)، حاصل جمع وزن دار (weighted sum) و ...) از روی امتیاز‌هایی که همسایگان کاربر a به آیتم i داده اند حاصل می شود.

➤ **گام سوم:** از بین تمامی آیتم‌های گام دوم، N آیتمی که بیشتری مقدار پیش‌بینی را دارند به کاربر پیشنهاد داده می‌شوند.

از جمله مزیت‌های آن به سادگی و دقت نتایج حاصل از آن می‌توان اشاره کرد که در عین حال از مشکلات این روش شامل مقیاس پذیری کم (Scalability) و آسیب پذیری در مقابل پراکندگی داده‌ها (Sparsity) در پایگاه داده می‌شود. در واقع با اضافه شدن کاربر جدید به سیستم معیارهای شباهت و مقادیر پیش‌بینی‌ها باید مجدداً حساب شوند که با افزایش تعداد کاربران و بزرگ شدن سیستم این مسئله سر بار محاسباتی زیادی را به سیستم وارد اضافه خواهد کرد.

❖ **رویکرد آیتم به آیتم:** الگوریتم‌هایی که بر مبنای این نگاه پیاده سازی می شوند شامل سه گام هستند:

➤ **گام اول:** بر اساس معیارهای شباهت برای هر آیتم i تعداد q همسایه را تعیین می کنیم.

➤ **گام دوم:** در صورتی که کاربر a به آیتم i تاکنون امتیازی نداده باشد، بر اساس امتیازاتی که این کاربر به آیتم های همسایه i داده است مقدار پیشبینی (prediction) را محاسبه می کنیم.

➤ **گام سوم:** بر اساس مقادیر پیشبینی ها (recommendations) آیتم هایی را که بیشترین مقدار پیشبینی را دارند به کاربر a پیشنهاد می کنیم.

۲-۵-۲ فیلتر دموگرافیک (Demographic Filtering)

اطلاعاتی مانند سن، جنسیت، ملیت و ... در گروه اطلاعات دموگرافیک (Demographic) قرار می گیرند. سیستم هایی که از این روش استفاده می کنند بر این اساس عمل می کنند که کاربرانی که صفات دموگرافیک مشابهی دارند (مثلا در یک بازه سنی قرار می گیرند) احتمالا سلیق و خواسته های مشابهی نیز دارند.

۲-۵-۳ فیلتر مبتنی بر محتوا (Content-based Filtering)

در این سیستم ها، نیاز به تحلیل و آنالیز اطلاعات و محتویات (contents) مربوط به کاربر و آیتم های موجود در سیستم است تا بتوان میزان شباهت بین کاربر و آیتم های سیستم و همچنین شباهت بین آیتم ها با یکدیگر را تعیین نمود که برای پیشنهاد دادن به کاربر، آیتم هایی آورده می شوند که شباهت بیشتری به آیتم هایی دارند که قبلا کاربر آنها را انتخاب نموده است.

از مشکلات آن می توان به موارد زیر اشاره کرد:

✓ **دشوار بودن استخراج داده ها و اطلاعات در مورد آیتم ها و کاربران:** در برخی از کاربردها و حوزه ها (مانند موزیک، ویدئو وب لاگ ها) پیدا کردن شباهت بین آیتم های مختلف براساس صفات و محتویات آنها، بسیار پیچیده و دشوار است که راهکار آن استخراج صفات بصورت خود کار می باشد.

✓ **مشکل انتزاع بالا (Overspecialization):** در این مشکل، آیتم هایی که ممکن است مورد پسند کاربر باشند ولی شباهتی به آیتم های انتخاب شده در گذشته ندارند، به کاربر هرگز پیشنهاد داده نشوند و از دید وی مخفی بمانند.

✓ **عدم امکان گرفتن بازخورد کاربران:** در چنین سیستم هایی معمولا کاربران به آیتم ها امتیاز نمی دهند که این مسئله باعث می شود که صحت پیشنهاد داده شده به کاربر نامعلوم خواهد بود.

راهکارهای مناسب برای مقابله با این مشکلات، می توان به ایجاد سیستم های hybrid ترکیب Content-based Filtering با شبکه های اجتماعی اشاره کرد. در این سیستم ها، علاوه بر ratings از اطلاعات موجود در شبکه های

اجتماعی مانند کامنت ها، بلاگ ها، ارتباطات بین دوستان، like ها و followers استفاده می شود تا بتوان کیفیت و دقت نتایج را بهبود داد.

سه گام اصلی در روش Content-based Filtering بدین گونه است:

❖ **استخراج صفات (attributes) مربوط به آیتم ها:** استخراج صفاتی که بطور صریح همراه با

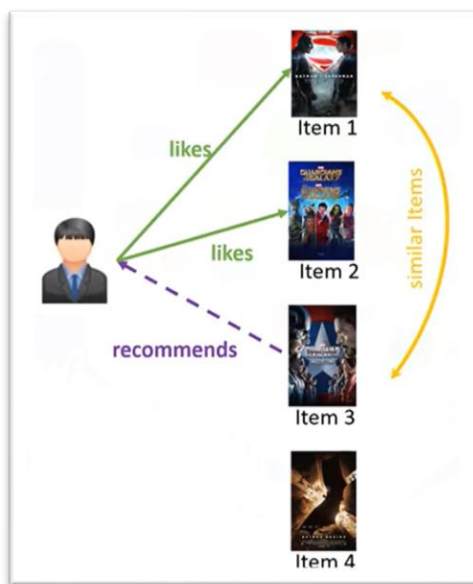
آیتم ها در سیستم درج می شوند، مشکل خاصی نداشته، در حالی که برای صفاتی که بر اساس دامنه سیستم هستند، از تکنیک‌های خاصی استفاده کرد. به عنوان مثال، در سیستم‌هایی با آیتم‌های اسناد متنی، از روش های کلاسیک بازیابی اطلاعات (information retrieval) استفاده می‌شود.

❖ **مقایسه صفات آیتم ها با سلايق کاربر:** برای این کار عموماً با استفاده از روش هایی از قبیل

روش‌های اکتشافی (heuristic) و یا الگوریتم‌های classification انجام می‌شود.

❖ **پیشنهاد دادن آیتم هایی با شباهت زیاد به سلايق کاربر**

به عنوان مثال، اگر چهار فیلم داشته باشیم و اگر کاربر دو مورد اول را دوست داشته باشد یا به آن امتیاز دهد، و اگر آیتم ۳ از نظر ژانر مشابه مورد ۱ باشد، سیستم پیشنهاد دهنده مورد ۳ را نیز به کاربر توصیه می‌کند.



حال اگر ۶ فیلم داشته باشیم که فیلم هایی است که کاربر تماشا کرده و همچنین ژانر هر یک از فیلم ها را نشان می دهد. به عنوان مثال، بتمن در مقابل سوپرمن در ژانر ماجراجویی، سوپر قهرمان و نگرهبانان کهکشانی در ژانرهای کمدی، ماجراجویی، سوپر قهرمان و علمی تخیلی هستند. فرض کنید کاربر تا کنون سه فیلم را تماشا کرده و به

آن امتیاز داده است و به فیلم اول امتیاز دو از ۱۰، به فیلم دوم ۱۰ از ۱۰ و به فیلم سوم هشت از ۱۰ داده است. وظیفه موتور توصیه‌گر این است که یکی از سه فیلم کاندید را به این کاربر توصیه کند، یا به عبارت دیگر می‌خواهیم پیش‌بینی کنیم که در صورت تماشای سه فیلم کاندید، امتیاز احتمالی کاربر از سه فیلم کاندید چقدر خواهد بود.



برای رسیدن به این هدف، باید پروفایل کاربری را بسازیم. ابتدا، ما یک برداری ایجاد می‌کنیم تا رتبه‌بندی کاربر را برای فیلم‌هایی که قبلاً تماشا کرده است نشان دهیم که **Input User Ratings** می‌نامیم. سپس، فیلم‌ها را از طریق رویکرد رمزگذاری تک داغ (**one-hot encoding**) رمزگذاری می‌کنیم. ژانر فیلم‌ها در اینجا به عنوان یک مجموعه ویژگی‌ها استفاده می‌شود. همچنین از سه فیلم اول برای ساخت این ماتریس استفاده می‌کنیم که نشان دهنده ماتریس مجموعه ویژگی‌های فیلم است.




Input User Rating		Movies Matrix			
		Comedy	Adventure	Super-Hero	Sci-Fi
2		0	1	1	0
10		1	1	1	1
8		1	0	1	0

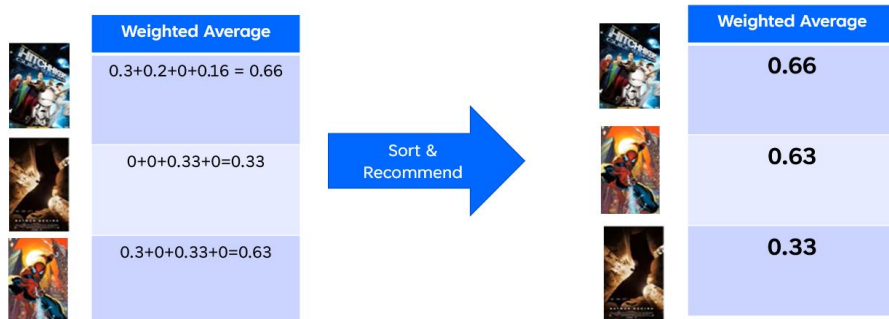
اگر این دو ماتریس را ضرب کنیم می‌توانیم مجموعه ویژگی‌های وزنی فیلم‌ها را بدست آوریم که ماتریس **Weighted Genre** نامیده می‌شود و نشان دهنده علایق کاربر برای هر ژانر بر اساس فیلم‌هایی است که او تماشا کرده است.

$$\text{Input User Rating} * \text{Movies Matrix} = \text{Weighted Genre Matrix}$$

	Comedy	Adventure	Super-Hero	Sci-Fi
	0	2	2	0
	10	10	10	10
	8	0	8	0

اکنون با توجه به **Weighted Genre Matrix**، می‌توانیم پروفایل کاربر را شکل دهیم. اساساً، ما می‌توانیم ژانرهای وزن‌دار را جمع‌آوری کنیم و سپس آنها را نرمال کنیم تا پروفایل کاربر را پیدا کنیم. این به وضوح نشان می‌دهد که او فیلم‌های ابرقهرمانی را بیشتر از ژانرهای دیگر دوست دارد. ما از این نمایه استفاده می‌کنیم تا بفهمیم چه فیلمی برای توصیه به این کاربر مناسب است. به سادگی ماتریس نمایه کاربر را در ماتریس فیلم کاندید ضرب می‌کنیم که نتیجه آن ماتریس فیلم‌های وزن‌دار است. وزن هر ژانر را با توجه به پروفایل کاربر نشان می‌دهد. حال، اگر این رتبه‌بندی‌های وزنی را جمع‌آوری کنیم، سطح علاقه احتمالی کاربر فعال به این سه فیلم را دریافت می‌کنیم. در اصل، این لیست‌های توصیه ما هستند که می‌توانیم آن‌ها را مرتب کنیم تا فیلم‌ها را رتبه‌بندی کنیم و به کاربر توصیه کنیم. به عنوان مثال، می‌توانیم بگوییم که راهنمای سفر به کهکشان، بالاترین امتیاز را در لیست ما دارد و مناسب است که به کاربر توصیه شود.

User Profile									
	Comedy	Adventure	Super-Hero	Sci-Fi	Mul	Weighted Movies Matrix			
	0.3	0.2	0.33	0.16		Comedy	Adventure	Super-Hero	Sci-Fi
	1	1	0	1		0.3	0.2	0	0.16
	0	0	1	0		0	0	0.33	0
	1	0	1	0		0.3	0	0.33	0



۲-۵-۴ فیلتر مبتنی بر اجتماعی (Social-based Filtering)

اطلاعات موجود در این شبکه‌ها شامل Followers, friends, comments, blog و tags می‌شوند که با بکارگیری آنها باعث بهبود نتایج پیشنهادی و همچنین کاهش مشکل sparsity شده است.

مطالعه و تحقیقات در این زمینه شامل دو دسته می‌شود:

❖ **بهبود کارایی سیستم‌های موجود:** نتایج حاصل از کارهای آنها گواه بر تاثیر مثبت این اطلاعات در سیستم های RS دارد.

❖ **ایجاد یک سیستم پیشنهاد دهنده جدید مبتنی بر Social Filtering:** از پتانسیل های موجود در چنین شبکه هایی برای ایجاد یک سیستم مستقل استفاده شود.

در میان اطلاعات موجود در شبکه های اجتماعی، trust و reputation نسبت به بقیه توجهی بیشتری را به خود جذب کرده اند. هرچه میزان trust یک کاربر بیشتر باشد، امتیازهایی که او به آیتم‌ها می دهد از درجه اهمیت و وزن بیشتری نسبت به سایرین برخوردار است. دو راهکار اصلی برای تعیین آن شامل کسب اطلاعات به صورت صریح از کاربر و روابط بین کاربران در شبکه‌های اجتماعی می‌باشد. معیار reputation را از روی تعداد امتیازاتی که کاربران به یک آیتم می دهند (صریح) و یا با بررسی نحوه کار کاربران با آیتم ها (ضمنی) تعیین نمود.

۲-۵-۵ فیلتر متن آگاه (Context-aware Filtering)

با استفاده از دستگاه ها و سنسور های گوناگون، اطلاعاتی از شرایط کاربر (context) را جمع آوری شده و در سیستم های پیشنهاد دهنده مورد استفاده قرار داده می‌شود. تاکید این سیستم ها بر اطلاعاتی از قبیل زمان، مکان، اطلاعات حاصل از دوربین های امنیتی، RFID ها و شبکه های حسگر بیسیم ونیز پارامتر های سلامت، عادات خرید و غذاخوردن فرد می باشد. این اطلاعات را می توان بصورت صریح و یا با استفاده از روش های داده کاوی (Data Mining) کسب نمود.

۲-۵-۶ سیستم های توصیه آگاه از موقعیت مکانی (Location-aware Recommendation Systems)

این گونه از سیستم ها که عموماً در برنامه های تلفن همراه نمود دارند و بر اساس موقعیت فعلی کاربر پیشنهاداتی را در حوزه خاصی به وی می دهند. این مسئله باعث ایجاد پیچیدگی در سیستم می شود. چون با تغییر مکان کاربر آیتمی که قبلاً مطلوب کاربر بوده و باید به وی پیشنهاد داده می شده، ممکن است دیگر مورد پسند کاربر نباشد.

۲-۵-۷ فیلتر مبتنی بر دانش (Knowledge-based Filtering)

نسل جدیدی از سیستم های پیشنهاد دهنده هستند که مبتنی بر دانش موجود در رابطه با کاربران و آیتم ها می باشند. معمولاً پیشنهادات خود را بر پایه تفسیر و استنباط خود از سلايق و نیازهای کاربر ارائه می دهند و از دیدگاه تئوری نسبت به سایر روش های ذکر شده از دقت و کیفیت بیشتری برخوردار هستند که برای پیاده سازی چنین سیستم هایی نیاز به یک بستر و ساختار مبتنی بر دانش وجود دارد. (آنتولوژی ها، case-based reasoning، constraint-based reasoning، knowledge vectors و social knowledge)

دو حوزه تحقیقاتی در این زمینه شامل مدل users-roles-tasks و peer-to-peer می شود که به ترتیب براساس اینکه کاربر در چه نقش هایی چه وظایفی داشته و توزیع دانش سیستم در مورد کاربران و آیتم ها در بین peer ها می باشد.

۲-۵-۸ روش های الهام گرفته شده از طبیعت (Bio-Inspired)

این راهکار ها را به سه گروه راهکار های الگوریتم تکاملی (GA)، راهکار های شبکه های عصبی (NN) و مدل هایی بر پایه شبکه های ایمنی بدن (Artificial Immune Network) تقسیم بندی نمود.

در حوزه الگوریتم های عموماً کارها در جهت ایجاد مدل های ترکیبی (hybrid) و یا خوشه بندی (clustering) بوده است. در خوشه بندی کاربران به گروه هایی دسته بندی شده به طوری که کاربران مشابه در گروه های یکسانی قرار بگیرند که باعث می شود بجای فرد، پیشنهادات به گروه داده شود و بدین ترتیب از زمان و سر بار محاسبات کاسته شود. مدل های ترکیبی کاربران (hybrid user models)، عموماً ترکیبی از Collaborative Filtering با روش هایی از قبیل Demographic Filtering و یا Content-based Filtering هستند. کروموزوم های الگوریتم ژنتیک را می توان بعنوان Demographic Information کاربران در نظر گرفت.

الگوریتم های شبکه های عصبی بر پایه عملکرد و رفتار سیستم عصبی بدن شکل گرفته و در راهکارهای مبتنی بر آنها هدف برخورداری نمودن سیستم از این قوه یادگیری است.

راهکار های مبتنی بر شبکه ایمنی بدن نیز به منظور رفع مشکل پراکندگی داده ها (data sparsity) و افزایش مقیاس پذیری سیستم ارائه شده اند. این روش نیز بصورت ترکیبی با روش های Collaborative Filtering بیشتر مورد استفاده قرار گرفته است.

۲-۶ ارزیابی سیستم های پیشنهاد دهنده

مناسب ترین راهکار برای ارزیابی سیستم ریکامندر وابسته به اهداف توسعه دهنده ریکامندر سیستم است. حالت انتخابی ارزیابی تاثیر مهمی را روی راهکاری دارد که سیستم طراحی می شود. دو نوع از ارزیابی ریکامندر سیستم ها شامل رویکردهای آنلاین و آفلاین می باشند.

❖ **روش آنلاین:** به آن تست A/B نیز گفته می شود. تعامل های کاربر وقتی که پیشنهادهای ریکامندر سیستم به کاربر داده می شود، سنجیده می شود. این رویکرد برای ارزیابی ایده آل است. معمولاً پیاده سازی این روش کاری دشوار است، زیرا که تنها راه برای اجرای آزمایش تعامل با سیستم است. هر شکستی در آزمایش مانند آن است که تاثیر مستقیمی بر درآمد کسب و کار و تجربه کاربر داشته است که استفاده از مشتریان واقعی برای آزمایش ها در ریکامندر سیستم ها بسیار کندتر از آن است که داده ها از پیش موجود باشند.

❖ **روش آفلاین:** برای مراحل آزمایشی ایده آل محسوب می شوند؛ چونکه کاربر به طور مستقیم درگیر نمی شود و برخلاف روش های آنلاین، سیستم برای مورد ارزیابی قرار گرفتن نیازی به مستقر شدن ندارد. بخشی از داده ها برای ساخت سیستم مورد استفاده قرار خواهد گرفت و بخش دیگر برای ارزیابی آن استفاده می شود.

فصل سوم

پیاده‌سازی سیستم‌های پیشنهاد دهنده

۳-۱ مقدمه

در این قسمت قصد داریم چند پروژه برای ساخت سیستم پیشنهاد دهنده براساس فیلتر مبتنی بر محتوا و همچنین براساس فیلتر مشارکتی که دو نوع اصلی از این نوع سیستم‌ها هستند، پیاده‌سازی بکنیم که در هر قسمت به طور مختصر توضیح داده شده‌اند و داخل خود نوت بوک‌ها اطلاعات بیشتر موجود می‌باشد.

۳-۲ پروژه سیستم پیشنهاد دهنده براساس فیلتر مبتنی بر محتوا

در این قسمت قرار است یک سیستم پیشنهاد دهنده فیلم پیاده‌سازی شود که با دو تا دیتاست فیلم‌ها و امتیازهای داده شده به آن فیلم‌ها اینکار انجام می‌شود. این سیستم پیشنهاد دهنده براساس فیلتر مبتنی بر محتوا پیاده‌سازی می‌شود.

۳-۳ پروژه سیستم پیشنهاد دهنده براساس فیلتر مشارکتی

در این قسمت دو سیستم پیشنهاد دهنده براساس فیلتر مشارکتی پیاده سازی شده است که اولین سیستم همان سیستم پیشنهاد دهنده فیلم است که با دو دیتاست فیلمها و امتیازهای داده شده به آن فیلمها انجام می شود و دومین پروژه هم سیستم پیشنهاد دهنده تلفن همراه است که دیتاست تلفن همراهها و نظرات کاربران که از سایت آمازون جمع آوری شده است، انجام می شود. در هر دو پروژه از حالت User-based استفاده می شود که در فصل گذشته به طور مفصل توضیح داده شده است.

منابع و لینک پروژه‌های انجام شده

- [ریکامندر چیست؟ | مبانی سیستم پیشنهاد دهنده — به زبان ساده](#)
- [سیستم پیشنهاد دهنده چیست؟](#)
- [سیستم ریکامندر \(Recommender\) چیست و چه کاربردی در بازاریابی دارد؟](#)
- [مروری بر انواع سیستم های پیشنهاد دهنده](#)
- [Recommendation Systems: Applications and Examples in 2022](#)
- [Introduction to recommender systems](#)
- [Types of Recommendation Systems & Their Use Cases](#)
- ...

لینک گیت‌هاب ارائه (فایل پاورپوینت ارائه — فایل گزارش — پروژه‌ها):

<https://github.com/Erfan2001/DataMining-RecommenderSystems>