بسم الله الرحمن الرحيم

پردیس دانشگاهی

پایان‌نامه کارشناسی ارشد

طراحی یک سامانه توصیه‌گر شخصی در پایگاه‌های علمی

نگارش

رامین صفا

استاد راهنما

دکتر سیّد ابوالقاسم میرروشندل

شهریور ماه ۱۳۹۳

تقدیم به

پدر و مادر عزیزم

که همواره با پشتیبانی مادی و معنوی خود صمیمانه مرا در رسیدن به اهدافم یاری نموده‌اند ...

سپاسگزاری

بدین وسیله از زحمات بی‌دریغ و صادقانه جناب آقای دکتر میرروشندل که با حسن نیت و فروتنی در جمع‌آوری این پایان‌نامه مرا راهنمایی فرموده‌اند، تشکر و قدردانی نموده و سعادت و تندرستی ایشان را از خداوند بزرگ مسالت می‌نمایم.

چکیده

طراحی یک سامانه توصیه‌گر شخصی در پایگاه‌های علمی

رامین صفا

مواجهه با مسئله حجم بالای اطلاعات همواره از موضوعات قابل توجه در تحقیقات علمی بوده است و یکی از رویکردهای موجود در این زمینه استفاده از سیستم‌های توصیه‌گر است. سیستم‌های توصیه‌گر رویکردی برای رویارویی با حجم بسیار زیاد اطلاعات بوده و به کاربر این امکان را می‌دهند تا با سرعت بیشتری به مقصود خود برسد. با وجود فراگیر شدن این سیستم‌ها در زمینه‌های گوناگون، مطالعات نیاز به توجه بیشتر به بکارگیری آنها در پایگاه‌های علمی را نشان می‌دهند.

استفاده از سیستم‌های توصیه‌گر در زمینه ارائه پیشنهادهای علمی شامل پیشنهاد مقاله، پیشنهاد متخصص یا همکار، پیشنهاد ارجاعات و پیشنهاد داور جهت مطالعه مقالات است که همگی موضوعاتی نوظهور و در حال توسعه هستند. در این میان با توجه به رشد چشم‌گیر رویدادها و مجلات علمی، یکی از دیگر مسائل پر اهمیت انتخاب مناسب‌ترین محل جهت انتشار مقالات علمی است. به‌طوری که وجود ابزاری جهت تسهیل این روند، برای محقق ضروری به‌نظر می‌رسد. با وجود اهمیت این سیستم‌ها در تسریع فرآیند انتشار و کاهش خطاهای احتمالی، در تحقیقات مرتبط کمتر به این مسئله پرداخته شده است.

از این رو در این پایان‌نامه سعی بر آن است که پس از بررسی رویکردهای موجود در این زمینه، روشی کارآمد جهت توصیه کنفرانس یا مجلات علمی مرتبط با مقاله مورد نظر کاربر پیشنهاد داده شود. سیستم پیشنهادی با بکارگیری مفاهیم تجزیه و تحلیل شبکه اجتماعی و روش محتوا محور، قادر خواهد بود مناسب‌ترین محل‌ها را جهت انتشار مقاله نوشته شده، بر اساس ترجیحات و سابقه نزدیکان محقق به او پیشنهاد دهد. نتایج ارزیابی با استفاده از داده‌های دنیای واقعی نیز عملکرد مناسب آن در ارائه پیشنهادهای موثر نهایی را نشان می‌دهند؛ بگونه‌ای که با پیمایش اولین عمق از شبکه همکاران نویسنده(گان)، قدرت پاسخگویی سیستم برابر ۷۰٫۶۹٪ بوده و دقت آن در پیش‌بینی نتیجه دقیق نهایی تنها در ۲۰ پیشنهاد ابتدایی به ۴۸٫۵۳٪ می‌رسد.

واژه‌های کلیدی: سیستم‌های توصیه‌گر، سیستم‌های توصیه‌گر علمی، بازیابی اطلاعات، پیشنهاد محل انتشار.

فهرست مطالب

فصل ۱: مقدمه ۱

۱-۱- مقدمه ۲

۱-۲- هدف و اهمیت انجام پژوهش ۴

۱-۳- ساختار کلی پایان‌نامه ۵

فصل ۲: مفاهیم پیش زمینه ۷

۲-۱- مقدمه ۸

۲-۲- سیستم‌های توصیه‌گر ۸

۲-۳- انواع رویکردهای موجود در سیستم‌های توصیه‌گر ۹

۲-۳-۱- پالایش مشارکتی یا همکارگونه ۱۰

۲-۳-۲- مسائل عمده در پالایش مشارکتی ۱۱

۲-۳-۳- پالایش مبتنی بر محتوا ۱۱

۲-۳-۴- چگونگی نمایش اقلام ۱۳

۲-۳-۵- مسائل عمده در پالایش مبتنی بر محتوا ۱۴

۲-۳-۶- سایر روش‌ها ۱۵

۲-۳-۷- روش‌های ترکیبی ۱۵

۲-۴- گذری بر داده‌کاوی و تکنیک‌های آن در سیستم‌های توصیه‌گر ۱۵

۲-۴-۱- نزدیکترین همسایگی ۱۷

۲-۴-۲- درخت تصمیم ۱۸

۲-۴-۳- دسته‌بندی بیزین ۱۹

۲-۴-۴- شبکه‌های عصبی مصنوعی ۲۰

۲-۴-۵- قوانین وابستگی یا انجمنی ۲۱

۲-۴-۶- خوشه‌بندی ۲۱

۲-۴-۷- تجزیه و تحلیل لینک ۲۲

۲-۴-۸- رگرسیون ۲۲

۲-۵- کاربرد سیستم‌های توصیه‌گر در زمینه علمی ۲۲
۲-۶- سیستم‌های توصیه‌گر در ترکیب با تجزیه و تحلیل شبکه‌های اجتماعی ۲۳
۲-۷- جمع‌بندی ۲۳

فصل ۳: کارهای مرتبط گذشته **۲۵**

۳-۱- مقدمه ۲۶
۳-۲- گذری بر سیستم‌های توصیه‌گر در زمینه‌های گوناگون ۲۶
۳-۳- انواع سیستم‌های توصیه‌گر علمی ۳۱
۳-۳-۱- پیشنهاد مقالات علمی ۳۱
۳-۳-۲- پیشنهاد محل انتشار مقالات ۳۴
۳-۳-۳- پیشنهاد متخصص یا همکار ۳۵
۳-۳-۴- پیشنهاد ارجاعات ۳۷
۳-۳-۵- سایر موارد ۳۹
۳-۴- جمع‌بندی ۳۹

فصل ۴: روش پیشنهادی **۴۰**

۴-۱- مقدمه ۴۱
۴-۲- مشخصات اولیه ۴۱
۴-۳- معماری کلی سیستم پیشنهادی ۴۲
۴-۳-۱- پیش‌پردازش زبان‌شناسی ۴۳
۴-۳-۲- شباهت‌سنج ۴۵
۴-۳-۳- رتبه‌بندی ۴۵
۴-۴- پیاده‌سازی و ارزیابی ۴۶
۴-۴-۱- مجموعه داده‌ها ۴۶
۴-۴-۲- مثالی واقعی از نحوه انجام عملیات در سیستم پیشنهادی ۴۷
۴-۴-۳- ارزیابی روش پیشنهادی ۵۱
۴-۵- جمع‌بندی ۵۵

فصل ۵: جمع‌بندی و نتیجه‌گیری **۵۶**

۵-۱- مقدمه ۵۷

۵-۲- جمع‌بندی و نتایج حاصل از پژوهش ۵۷

۵-۳- کارهای آینده ۵۹

منابع **۶۰**

ضمائم و پیوست‌ها **۶۵**

پیوست (۱): فهرست کلمات اضافه ۶۶

پیوست (۲): نمایی از واسط کاربری سیستم پیشنهادی ۷۰

واژه‌نامه (فارسی - انگلیسی) **۷۱**

واژه‌نامه (انگلیسی - فارسی) **۷۴**

فهرست جدول‌ها

جدول (۴-۱): نمونه‌ای از کلمات اضافه ۴۴
جدول (۴-۲): انواع موجودیت‌های دی.بی.ال.پی. ۴۶
جدول (۴-۳): نتایج ارزیابی ۵۳
جدول (۴-۴): جزئیات پیاده‌سازی در عمق ۱ ..... ۵۳
جدول (۴-۵): جزئیات پیاده‌سازی در عمق ۲ ..... ۵۴
جدول (۴-۶): توانایی پاسخگویی سیستم در عمق‌های ۱ تا ۴ ..... ۵۴

فهرست شکل‌ها

شکل (۱-۱): آمار انتشار مقالات با موضوع شخصی‌سازی از سال ۲۰۰۰ تا ۲۰۰۷ در پایگاه‌های گوناگون [۵] ۳
شکل (۱-۲): تعداد کنفرانس‌ها به تفکیک موضوعاتی مشخص در سال‌های ۲۰۰۰ تا ۲۰۰۹ [۱۱] ۵
شکل (۱-۳): تعداد رویدادهای موجود در وبسایت دی.بی.ال.پی. از سال ۱۹۹۷ تا ۲۰۰۷ [۱۲] ... ۵
شکل (۲-۱): معماری سطح بالای سیستم‌های توصیه‌گر محتوا محور [۲] ۱۲
شکل (۲-۲): مراحل اصلی در فرآیند داده‌کاوی ۱۶
شکل (۲-۳): مثالی از نزدیکترین همسایگی [۲] ۱۸
شکل (۲-۴): مثالی از یک درخت تصمیم ۱۹
شکل (۲-۵): مدل یک نرون مصنوعی ۲۰
شکل (۳-۱): نمای سیستم ارائه شده جهت پیشنهاد مقاله [۲۷] ۳۳
شکل (۳-۲): نمای کلی سیستم ارائه شده جهت پیشنهاد همکار [۳۵] ۳۶
شکل (۳-۳): تحلیل هم استنادی [۳۸] ۳۸
شکل (۳-۴): اتصال کتاب‌شناختی [۳۸] ۳۸
شکل (۴-۱): معماری کلی سیستم پیشنهادی ۴۳
شکل (۴-۲): استخراج همکاران نویسنده از پایگاه داده ۴۳
شکل (۴-۳): نمونه‌ای از رکورد دی.بی.ال.پی. ۴۷
شکل (۴-۴): نتایج پیاده‌سازی در TOPهای متفاوت عمق ۱ ..... ۵۲

فصل ۱:

مقدمه

۱-۱- مقدمه

بشر امروزه با حجم عظیمی از داده‌ها مواجه است. رشد فناوری، افزایش توانایی جمع‌آوری و ذخیره‌سازی داده‌ها موجب به‌وجود آمدن مسئله اطلاعات بیش از حد[۱] شده است؛ چرا که حجم اطلاعات تولید شده بسیار فراتر از ظرفیت‌های محدود پردازش غیرخودکار است. این حجم بالا همواره از موضوعات مهم و قابل توجه در تحقیقات علمی بوده و سبب پیدایش فرصت‌های جدید در علوم مهندسی و کسب‌وکار شده است. به‌گونه‌ای که توانایی استخراج دانش مفید از این مجموعه داده‌ها[۲] و بکارگیری آن در جهان رقابتی امروز، برای بقای سازمان‌ها بیش از پیش حائز اهمیت است [۱].

تکنیک‌های سنتی بازیابی اطلاعات[۳] به‌دلیل تکیه بر تطابق پرس‌وجو[۴] و عدم توجه به نیاز اطلاعاتی کاربر، ترجیحات[۵] او و همچنین اطلاعات زمینه[۶]، عموماً موجب بازیابی مجموعه عظیمی از نتایج مرتبط و غیرمرتبط در کنار یکدیگر می‌شوند؛ این امر بسیاری از کاربران[۷] را در فرآیند انتخاب و تصمیم‌گیری با مشکل مواجه نموده و سبب صرف زمان نسبتاً زیاد در یافتن نتایج مورد نظر می‌شود. بعلاوه پرس‌وجوهای مشابه، خروجی‌های یکسانی دارند و یا حتی در بسیاری از موارد کاربر تعریف روشن و واضحی از نیاز اطلاعاتی خود ندارد. در این میان نیاز به سیستم‌هایی که بتوانند مناسب‌ترین کالا و یا خدمات را به کاربر پیشنهاد دهند، اساسی به نظر می‌رسد.

سیستم‌های توصیه‌گر[۸] رویکردی برای مواجهه با مشکلات مذکور هستند و به کاربر این امکان را می‌دهند تا در میان حجم بسیار بالای اطلاعات با سرعت بیشتری به مقصود خود برسد. این سیستم‌ها که امروزه از محبوبیت ویژه‌ای نیز برخوردار هستند، شامل ابزارهای نرم‌افزاری و تکنیک‌هایی هستند که با استفاده از انواع متفاوتی از دانش، داده‌های مربوط به کاربران، اقلام[۹] موجود و تراکنش[۱۰]های گذشته، اقدام به پیشنهاد مناسب‌ترین گزینه می‌نمایند. این امر علاوه بر سودآوری بیشتر برای موسسه یا سازمان مورد نظر می‌تواند سبب افزایش میزان رضایتمندی و وفاداری کاربر نیز شود [۲]؛ به‌گونه‌ای که شاید بتوان یکی از دلایل موفقیت وبسایت‌هایی چون آمازون[۱۱]، ای‌بی[۱۲]، یوتیوب[۱۳]، یاهو[۱۴] و غیره را بهره‌گیری از این سیستم‌ها دانست.

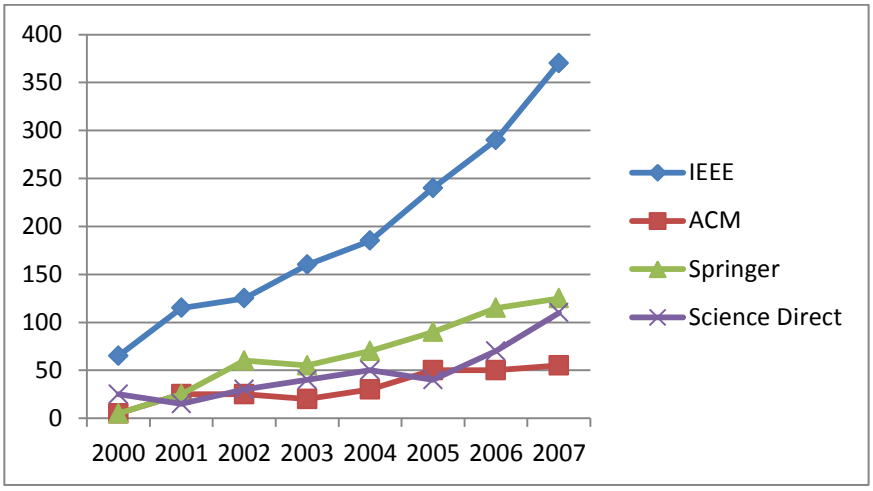
---

1 Information Overload
2 Datasets
3 Information Retrieval
4 Query
5 Preferences
6 Context
7 Users
8 Recommender Systems
9 Items
10 Transaction
11 http://amazon.com/
12 http://ebay.com/
13 http://youtube.com/
14 http://yahoo.com/

پایگاه‌های علمی [1] نیز از این قاعده مستثنی نبوده و با مسئله حجم زیاد اطلاعات مواجه هستند. این پایگاه‌ها شامل اطاعات مربوط به ارتباطات [2] بین محققین [3] و اطلاعات کتابخانه‌ای انتشارات هستند که با فراهم نمودن امکان اشتراک‌گذاری انتشارات علمی به ترویج دانش کمک می‌کنند. در این بین یکی از سودمندترین سناریوهای موجود می‌تواند بهره‌گیری از سیستم‌های توصیه‌گر متفاوتی از قبیل سیستم پیشنهاد مقاله، محل انتشار، همکار و غیره متناسب با علایق محقق باشد. بعلاوه تنوع پایگاه‌های علمی، مقالات و زمینه‌های تخصصی متفاوت افراد، نیاز به شخصی‌سازی [4] محتوا را برای کاربران بیشتر نموده است.

منظور از شخصی‌سازی محتوا ارائه نتایجی است که نه تنها با خواسته کاربر مرتبط هستند بلکه اطلاعات زمینه را نیز در نظر می‌گیرند [۲]. "توانایی فراهم‌آوری محتویات مناسب برای کاربران، بر اساس رفتار و اطلاعات مربوط به ترجیحات آنها" تعریفی است که در اکثر مراجع برای شخصی‌سازی در نظر گرفته شده است [۳,۴]. به عنوان مثال فرض کنید کاربری با وارد نمودن پرس‌وجوی خود در یک موتور جستجو [5] به‌دنبال مطالب مرتبط می‌گردد. در حالت عادی موتور جستجو پرس‌وجوی کاربر را به عنوان ورودی گرفته و پس از عملیات تطبیق [6]، نتایج مربوطه را بازیابی می‌کند. اما در حالت شخصی‌سازی شده و پیشرفته‌تر می‌توان سن، موقعیت، زمان، سابقه جستجو، ترجیحات فرد و اطلاعات دیگر را در هنگام ارائه نتایج در نظر گرفت.

بررسی مقالات منتشر شده در زمینه شخصی‌سازی در پایگاه‌های علمی معتبر، بین سال‌های ۲۰۰۰ تا ۲۰۰۷ که در شکل (۱-۱) نشان داده شده، بیانگر رشد روز افزون تحقیقات در این زمینه است [۵].



شکل (۱-۱): آمار انتشار مقالات با موضوع شخصی‌سازی از سال ۲۰۰۰ تا ۲۰۰۷ در پایگاه‌های گوناگون [۵]

1 Scientific Bases
2 Relationships
3 Researcher
4 Personalization
5 Search Engine
6 Matching

در این پایان‌نامه سعی خواهد شد با بهره‌گیری از مفاهیم سیستم‌های توصیه‌گر و تجزیه و تحلیل شبکه‌های اجتماعی[1]، رویکردی جهت مواجهه با مسائل ذکر شده ارائه شود؛ به‌طوری که متناسب با نیاز اطلاعاتی هر کاربر پیشنهادهایی مرتبط و شخصی‌سازی شده به او ارائه شوند.

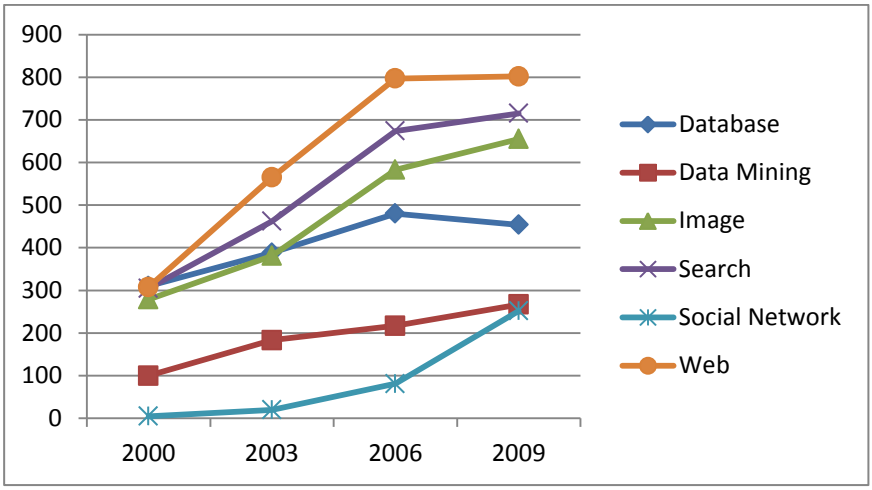
۱-۲- هدف و اهمیت انجام پژوهش

امروزه شاهد استفاده از سیستم‌های توصیه‌گر در زمینه‌های گوناگون مانند تجارت الکترونیک[2]، تبلیغات، اخبار، فیلم و غیره هستیم و مطالعات نشان دهنده افزایش تحقیقات علمی در این زمینه هستند [۲,۶]. بعلاوه بسیاری از موسسات تجاری در حال حاضر مشغول توسعه و ایجاد سیستم‌های توصیه‌گر به عنوان بخشی از خدمات برای مشتریان خود هستند [۲]. اما تحقیقات نشان می‌دهند که استفاده از سیستم‌های توصیه‌گر در زمینه پیشنهادهای علمی، با توجه به پتانسیل بالای جوامع علمی[3] در شخصی‌سازی محتوا، نیازمند توجه بیشتر بوده و از اهمیت ویژه‌ای برخوردار است [۷,۸,۹]. از سوی دیگر بکارگیری اطلاعات مربوط به شبکه اجتماعی افراد در سیستم‌های توصیه‌گر به عنوان موضوعی نوظهور و درحال پیشرفت شناخته می‌شود [۱۰].

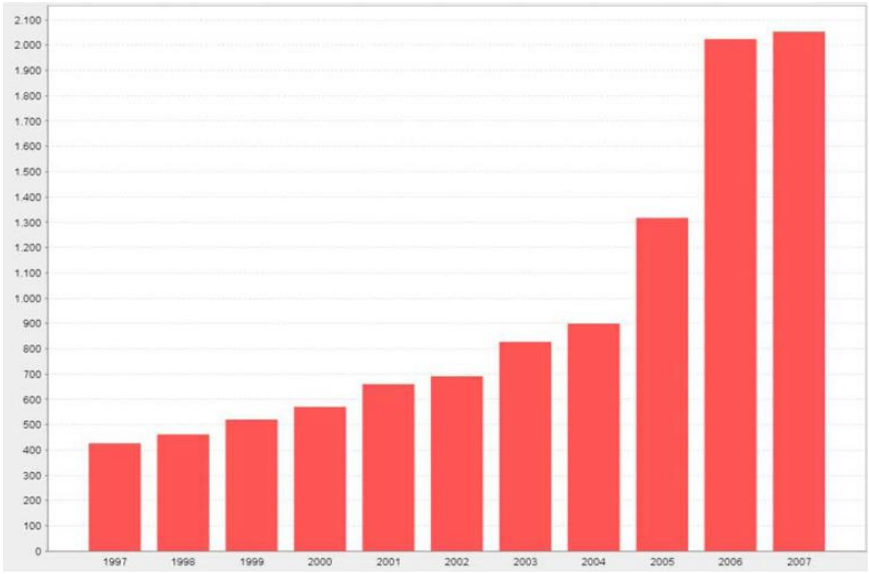
با توجه به تعداد بسیار زیاد و رو به رشد کنفرانس‌ها[4] و مجلات علمی[5] (شکل ۱-۱ و ۱-۲)، انتخاب مناسب‌ترین محل جهت انتشار مقاله برای محققین می‌تواند چالش‌برانگیز باشد. این مسئله که در تحقیقات مرتبط کمتر به آن توجه شده است، می‌تواند انگیزه‌ای برای استفاده از سیستم‌های توصیه‌گر باشد؛ به‌طوری که با شناخت ترجیحات محقق، امکان پیشنهاد کنفرانس یا مجلات علمی مرتبط و مناسب را فراهم نمود.

لذا در این پژوهش پس از بررسی استفاده از سیستم‌های توصیه‌گر در زمینه علمی و تکنیک‌های موجود، یک سیستم توصیه‌گر شخصی جهت پیشنهاد محل انتشار مقالات علمی ارائه خواهد شد. در روش پیشنهادی از مفاهیم تجزیه و تحلیل شبکه‌های اجتماعی جهت بهبود روند توصیه استفاده شده است.

1 Social Networks Analysis
2 E-Commerce
3 Scientific Communities
4 Conferences
5 Journals



شکل (۱-۲): تعداد کنفرانس‌ها به تفکیک موضوعاتی مشخص در سال‌های ۲۰۰۰ تا ۲۰۰۹ [۱۱]



شکل (۱-۳): تعداد رویدادهای موجود در وبسایت دی.بی.ال.پی. از سال ۱۹۹۷ تا ۲۰۰۷ [۱۲]

۱-۳- ساختار کلی پایان‌نامه

این پایان‌نامه شامل پنج فصل است. پس از شناخت موضوع و اهداف پژوهش در این فصل، در فصل دوم تعاریف و مفاهیم اولیه و مورد نیاز تشریح می‌شوند. برخی از سیستم‌های توصیه‌گر ارائه شده در زمینه‌های گوناگون و کارهای مرتبط انجام شده در زمینه استفاده از سیستم‌های توصیه‌گر در پایگاه‌های علمی و به‌ویژه توصیه محل انتشار در فصل سوم مورد بررسی قرار خواهند گرفت. در فصل چهارم ساختار کلی روش پیشنهادی و جزئیات مراحل آن جهت توصیه مناسب‌ترین محل انتشار مقالات علمی به محقق ارائه می‌شود و مسائل مربوط به ارزیابی و نحوه

آماده‌سازی داده‌ها جهت انجام پیاده‌سازی، به همراه نتایج و دستاوردهای پژوهش مورد بحث قرار خواهند گرفت. در نهایت مطالب ارائه شده در قالب فصل پنجم جمع‌بندی شده و زمینه‌های کاری آینده معرفی می‌شوند.

فصل ۲:

مفاهیم پیش‌زمینه

۲-۱- مقدمه

در این فصل مفاهیم و اصطلاحات مورد نیاز جهت درک بهتر موضوع معرفی می‌شوند. ابتدا در خصوص کلیات سیستم‌های توصیه‌گر توضیحاتی ارائه خواهد شد و سپس به تشریح رویکردهای اصلی در آن پرداخته می‌شود. پس از مروری بر داده‌کاوی و تکنیک‌های آن در سیستم‌های توصیه‌گر، موارد استفاده از این سیستم‌ها در زمینه ارائه پیشنهادهای علمی شرح داده شده و در انتها گریزی به ترکیب تجزیه و تحلیل شبکه‌های اجتماعی و سیستم‌های توصیه‌گر خواهیم زد.

۲-۲- سیستم‌های توصیه‌گر

در حالت کلی می‌توان گفت سیستم‌های توصیه‌گر، سیستم‌هایی هستند که با هدف مقابله با مسئله اطلاعات بیش از حد طراحی شده‌اند و با شناخت کاربر، اقلام و همچنین تراکنش‌های قبلی موجود، اقدام به پیشنهاد مناسب‌ترین گزینه‌ها می‌نمایند [۲,۶].

دو سناریو اصلی جهت مدیریت حجم رو به رشد اطلاعات، بازیابی و پالایش اطلاعات[1] است که به‌صورت زیر تعریف می‌شوند [۱۳,۱۴]:

- ✓ **بازیابی اطلاعات:** هرگونه تلاش برای جستجوی اسناد، به منظور بازیابی مرتبط‌ترین داده با پرس‌وجو.
  - در اینجا داده ایستا و پرس‌وجو پویا است.
- ✓ **پالایش اطلاعات:** هرگونه تلاش برای حذف داده‌های افزونه یا ناخواسته از جریان داده.
  - در اینجا داده پویا و پرس‌وجو ایستا است.

سیستم‌های توصیه‌گر در محلی بین این دو مفهوم قرار می‌گیرند، چرا که ماهیت داده در آن به آرامی تغییر کرده و پرس‌وجو نیز به چند پارامتر وابسته است؛ از این رو در چنین سیستم‌هایی از رویکردهای بازیابی و پالایش اطلاعات استفاده می‌شود. با این حال در اکثر پژوهش‌ها، این سیستم‌ها به عنوان نوعی از سیستم‌های پالایش اطلاعات شناخته می‌شوند که دسترسی به اطلاعات سفارشی‌سازی[2]شده احتمالاً مورد علاقه کاربر را در زمینه‌ای مشخص فراهم می‌کنند. در این سیستم‌ها می‌توان به‌وسیله بازخورد[3]های ضمنی[4] و یا آشکار[5] که در ادامه همین بحث تشریح خواهند شد، دقت پیشنهادهای بعدی را افزایش داد.

---

1 Information Filtering

2 Customization

3 Feedback

4 Implicit

5 Explicit

در حالت کلی در یک سیستم توصیه‌گر سه مولفه اصلی وجود دارد که عبارتند از [۲]:

- ✓ **اقلام:** اقلام مواردی هستند که به کاربر پیشنهاد داده می‌شوند و ارزش این پیشنهادها می‌تواند مثبت و یا منفی تلقی شود؛ اگر قلم پیشنهادی برای کاربر سودمند باشد به آن مثبت و اگر انتخاب آن برای کاربر منجر به تصمیم نادرست شود به آن منفی می‌گوییم.
- ✓ **کاربر:** کاربران سیستم‌های توصیه‌گر می‌توانند اهداف متفاوتی داشته باشند. به منظور شخصی‌سازی پیشنهادها، سیستم دامنه‌ای از اطلاعات کاربر را در نظر می‌گیرد که بسته به رویکرد مورد استفاده، متفاوت است. برای مثال در پالایش همکارگونه [۱] که در بخش ۲-۳-۱ توضیح داده خواهد شد، کاربران با استفاده از لیستی از رتبه [۲]هایی که هر کاربر به اقلامی خاص اختصاص داده است، مدل می‌شوند و یا در نوع جمعت شناختی [۳] از ویژگی‌هایی مانند سن، جنسیت، حرفه و تحصیلات کاربر استفاده می‌شود. باید توجه داشت که بدون مدل‌سازی [٤] مناسب کاربر، شخصی‌سازی امکان‌پذیر نیست، بنابراین مدل‌سازی از اهمیت ویژه‌ای برخوردار است.
- ✓ **تراکنش‌ها:** منظور از تراکنش، تعاملات [٥] بین کاربر و سیستم است. تراکنش‌ها داده‌های مربوط به عملیاتی هستند که اطلاعات پر اهمیتی را برای رویکرد مورد استفاده سیستم در بر دارند. مانند رتبه‌دهی که رایج‌ترین نوع تراکنش در سیستم‌های توصیه‌گر است و در بخش مربوط به پالایش همکارگونه توضیح داده خواهد شد.

۲-۳- انواع رویکردهای موجود در سیستم‌های توصیه‌گر

رویکردهای متفاوتی در سیستم‌های توصیه‌گر وجود دارند که با توجه به عملکرد الگوریتم پیشنهاد دهنده و همچنین دانش مورد استفاده جهت پیشنهاددهی، ماهیت متفاوتی دارند. در این بخش دو رویکرد اصلی پالایش مشارکتی یا همکارگونه و پالایش مبتنی بر محتوا [٦] مورد بررسی قرار گرفته و توضیحاتی در خصوص سایر روش‌های موجود و همچنین روش‌های ترکیبی [۷] ارائه می‌شود.

1 Collaborative Filtering
2 Rating
3 Demographic
4 Modeling
5 Interactions
6 Content-based Filtering
7 Hybrid

۲-۳-۱- پالایش مشارکتی یا همکارگونه

ایده اصلی پالایش مشارکتی بهره‌برداری از اطلاعات مربوط به رفتارهای گذشته کاربر، جهت پیش‌بینی اینکه چه اقلامی ممکن است برای وی جالب باشند بوده و سیستم پس از رتبه‌دهی کافی کاربر فعال[۱] به مجموعه‌ای از اقلام، پیشنهادها را بر اساس علایق کاربرانی با ترجیحات مشابه به وی ارائه می‌دهد. در پالایش مشارکتی محتوای اقلام اهمیتی نداشته و پیشنهادها تنها با استفاده از الگو[۲]ی رتبه‌دهی ارائه می‌شوند. مبنای کار در این روش بر اساس یک ماتریس کاربر-قلم است که کاربران سطرها و اقلام ستون‌های آن را تشکیل داده‌اند. درایه‌های ماتریس نیز نشان دهنده میزان علاقه هر کاربر به یک قلم خاص هستند [۲,۶,۱۵].

در حالت کلی پالایش مشارکتی به دو دسته مبتنی بر همسایگی[۳] یا حافظه[۴] و مبتنی بر مدل[۵] تقسیم می‌شود. در دسته اول از رتبه‌های ذخیره شده جهت پیش‌بینی رتبه اقلام جدید استفاده می‌شود و این امر می‌تواند به دو طریق کاربر محور[۶] و قلم محور[۷] انجام شود [۲,۱۶]. به‌صورت خلاصه در نوع کاربر محور با در نظر گرفتن پایگاه‌داده رتبه‌ها و هویت کاربر به عنوان ورودی، کاربران مشابه (همسایگان[۸]) شناخته شده و برای هر قلمی که کاربر فعال تاکنون مشاهده نکرده است پیش‌بینی بر اساس رتبه کاربران دیگر محاسبه می‌شود. اما در نوع قلم محور پیش‌بینی با استفاده از الگوی رتبه اقلام مشابه انجام می‌شود و زمانی می‌توان دو قلم را مشابه دانست که کاربران متفاوت سیستم، به شکل مشابهی آنها را رتبه‌دهی کرده باشند [۶].

در تمایز با روش مبتنی بر همسایگی که از رتبه‌های ذخیره شده به‌صورت مستقیم استفاده می‌کرد، در روش مبتنی بر مدل از این رتبه‌ها به منظور آموزش یک مدل پیش‌بینی استفاده می‌شود. طراحی و ساخت این مدل با استفاده از الگوریتم‌های یادگیری ماشین[۹] و داده‌کاوی[۱۰] انجام می‌شود و ایده اصلی این است که با مدل‌سازی تعاملات بین کاربر و اقلام به‌وسیله عامل‌های ارائه دهنده ویژگی‌های پنهان آنها در سیستم (مانند دسته ترجیحات کاربر و تقسیم‌بندی دسته‌های اقلام)، پس از آموزش مدل به وسیله داده‌های موجود، از آن برای پیش‌بینی رتبه کاربران به اقلام جدید استفاده شود [۲,۱۶].

---

1 Active User
2 Pattern
3 Neighborhood-based
4 Memory-based
5 Model-based
6 User-based
7 Item-based
8 Neighbors
9 Machine Learning
10 Data Mining