بسم الله الرحمن الرحیم

دانشکده آموزشهای الکترونیکی

پایان نامه کارشناسی ارشد در رشته مهندسی فناوری اطلاعات (تجارت الکترونیک)

ترکیب سیستم‌های توصیه‌گر مبتنی بر محتوا با استفاده از بهینه‌سازی خطی

توسط

الهه افشار

استادراهنما:

دکتر منصور ذوالقدری جهرمی

اردیبهشت ۱۳۹۲

به نام خدا

اظهارنامه

اینجانب الهه افشار (۸۸۰۹۰۶) دانشجوی رشته‌ی مهندسی فناوری اطلاعات گرایش تجارت الکترونیک دانشکده‌ی آموزش های الکترونیکی اظهار می کنم که این پایان نامه حاصل پژوهش خودم بوده و در جاهایی که از منابع دیگران استفاده کرده ام، نشانی دقیق و مشخصات کامل آن را نوشته ام. همچنین اظهار می کنم که تحقیق و موضوع پایان نامه ام تکراری نیست و تعهد می نمایم که بدون مجوز دانشگاه دستاوردهای آن را منتشر ننموده و یا در اختیار غیر قرار ندهم. کلیه حقوق این اثر مطابق با آیین نامه مالکیت فکری و معنوی متعلق به دانشگاه شیراز است.

نام و نام خانوادگی:

تاریخ و امضاء:

الهه افشار ۹۲/۲/۲۸

تقدیم به:

پدر و مادرم, مهربانان همیشه ماندنی

سپاسگذاری

با سپاس بی کران از زحمات بی دریغ استاد بزرگوارم جناب آقای دکتر ذوالقدری، که از ابتدا تا انتهای کار با راهنمایی های خود موجب پیشرفت رساله شدند و در این راه زحمات فراوانی را متقبل شدند،

همچنین اساتید گرانقدر، جناب آقایان دکتر صدرالدینی و دکتر جعفری که بر من منت نهاده و اساتید مشاور من در این رساله بوده اند.

و نیز جناب دکتر طاهری که در تمام طول مسیر رساله مرا مورد مرحمت و راهنمایی خود قرار دادند.

چکیده

ترکیب سیستم‌های توصیه‌گر مبتنی بر محتوا با استفاده از بهینه‌سازی خطی

به کوشش

الهه افشار

حجم فراوان و رو به رشد اطلاعات بر روی اینترنت و وب فرایند تصمیم‌گیری و انتخاب را برای بسیاری از کاربران دشوار کرده است به گونه‌ای که عدم وجود راهنمایی و ناوبری درست ممکن است منجر به انتخابهایی غط یا غیربهینه شود. سیستم‌های توصیه‌گر[1] سیستم‌هایی تاثیرگذار در راهنمایی و هدایت کاربر در میان حجم عظیمی از انتخابهای ممکن برای رسیدن به گزینه‌ی مفید و مورد علاقه‌ی وی هستند به گونه‌ای که این فرایند برای همان کاربر شخصی‌سازی شده باشد.

ما در این پایان نامه به بررسی سیستم‌های توصیه گرا مبتنی بر محتوا[2] و ترکیب برخی از روش‌های شناخته شده در این زمینه جهت بهبود عملکرد سیستم امتیاز دهی به اقلام مورد نیاز کاربران خواهیم پرداخت. بالطبع امتیاز دهی مناسب در این سیستم ها موجب ارائه مجموعه اقلام مناسب تر به مشتریان و در نتیجه افزایش فروش و سودآوری بیشتر خواهد شد.

[1]- Recommender System

[2]- Content-Based

فهرست مطالب

عنوان **صفحه**

فصل اول:مقدمه..... ۱

۱-۱. گذشته تحقیق ۲

۱-۱-۱. سیستم های توصیه گر مبتنی بر محتوا..... ۴

۱-۲. بیان مساله ۷

۱-۲-۱. تعریف راهبرد حل مساله ۹

۱-۳. اهداف تحقیق و سوالات ۱۰

۱-۴. اهمیت موضوع ۱۱

۱-۵ . طرح کلی پایان نامه ۱۳

فصل دوم:مبانی نظری تحقیق..... ۱۴

۲-۱. مبانی, مفاهیم و معرفی سیستم های توصیه گر ۱۶

۲-۱-۱. مزایای سیستم های توصیه گر ۱۹

۲-۱-۲. تقسیم بندی سیستم های توصیه گر ۲۱

۲-۱-۲-۱. فیلترینگ مبتنی بر محتوا ۲۱

۲-۱-۲-۲. فیلترینگ همکارگونه ۲۲

۲-۱-۲-۳.سیستم های مبتنی بر دانش ۲۳

۲-۱-۲-۴.سیستم های مبتنی بر آمار ۲۳
۲-۱-۳. مدل اولویت ترکیبی ۲۴
۲-۱-۴. توصیف پیاده سازی سیستم های توصیه گر ۲۵
۲-۱-۵.بررسی سیستم های توصیه گر مبتنی بر محتوا ۲۸
۲-۱-۵-۱الگوریتم فیلترینگ مبتنی بر محتوا ۲۹
۲-۲. داده کاوی ۳۲
۲-۲-۱. داده کاوی در وب ۳۳
۲-۲-۲. شکل های مختلف وب کاوی ۳۴
۲-۲-۳. متدهای مورد استفاده در داده کاوی ۳۴
۲-۲-۳-۱. متد پیش بینی ۳۵
۲-۲-۳-۲. متد توصیفی ۳۵
۲-۲-۴. روش های داده کاوی ۳۵
۲-۲-۴-۱.یادگیری ماشین ۳۶
۲-۲-۴-۲. شبکه های عصبی ۳۷
۲-۲-۴-۲-۱. مدل سازی شبکه عصبی ۴۰
۲-۲-۴-۲-۲. امتیاز دهی به هدف ۴۰
۲-۲-۴-۲-۳. پیکر بندی شبکه ۴۱
۲-۲-۴-۲-۴. تحلیل و ارزیابی مدل ۴۲
فصل سوم:مروری بر تحقیقات انجام شده ۴۵
فصل چهارم:روش پیشنهادی ۵۱
۴-۱.بیان کلی ایجاد مدل ترکیبی ۵۲

۴-۲.مدل سازی ۵۴
۴-۲-۱. انتخاب مدل های پایه ۵۴
۴-۲-۲. امتیازدهی به هدف ۵۵
۴-۲-۳. پیکربندی شبکه ۵۶
۴-۲-۴. تحلیل و ارزیابی مدل ۵۶
۴-۳. بیان فنی ایجاد مدل ترکیبی و حل مساله ۵۷
فصل پنجم:بررسی و تحلیل یافته های تحقیق ۶۴
فصل ششم:جمع بندی و نتیجه گیری ۸۰
۶-۱. جمع بندی کلی ۸۱
۶-۲. نتایج بدست آمده ۸۲
۶-۳. پیشنهادات آینده ۸۳
منابع ۸۴
چکیده به زبان انگلیسی ۹۰

فهرست جداول

عنوان **صفحه**

جدول شماره ۱: قسمتی از ماتریس رتبه دهی برای یک سیستم توصیه گر ۲۷

جدول شماره ۲: مدل های پایه انتخاب شده به همراه امتیاز آنها در مرحله آموزش..... ۷۵

فهرست شکل‌ها

عنوان **صفـحه**

شکل شماره ۱: یک چهارچوب عمومی برای سیستم های توصیه گر ۱۹

شکل شماره ۲: شبکه عصبی با یک لایه نهان..... ۳۷

شکل شماره ۳: شبکه عصبی وزن دار ۳۸

شکل شماره ۴: نمودار gain..... ۴۳

شکل شماره ۵: نمودار hit probability..... ۴۳

شکل شماره ۶: چهار نمونه معادله بر روی نمودار ۵۹

شکل شماره ۷: انتخاب بهترین بازه برای وزن دهی با چهار نمونه فرضی ۶۰

شکل شماره ۸: نمودار های حاصل از ۱۲ روش پایه ۶۶

شکل شماره ۹: نمودارها و مساحت زیر آنها براساس تابع هدف ۰ و ۱ ..... ۶۷

شکل شماره ۱۰: نمودارهاو مساحت زیر آنها با افزودن ویژگی میانگین و براساس تابع هدف ۱ تا ۵ ..... ۶۸

شکل شماره ۱۱: نمودارهاو مساحت زیر آنها با افزودن ویژگی میانگین و براساس تابع هدف ۰ و ۱..... ۶۸

شکل شماره ۱۲: نمودارها و مساحت زیر منحنی ها با لایه خروجی purelin..... ۶۹

شکل شماره ۱۳: نمودار و مساحت زیر منحنی های مدلها با الگوریتم های یادگیری مختلف و برچسب ۱ تا ۵. ۷۰

شکل شماره ۱۴: نمودارها و مساحت زیر منحنی های مدل ها با الگوریتم های یادگیری مختلف و برچسب ۰ و ۱ ..... ۷۰

شکل شماره ۱۵: نمودارها و مساحت زیر منحنی های مدل ها با ایجاد تغییر در شیوه انتخاب داده های آموزش و آزمایش ۷۱

شکل شماره ۱۶: **نمودارها و مساحت زیر منحنی روش های پایه برای کاربر شماره ۹۳۲**..... ۷۳

شکل شماره ۱۷: **جمع امتیازات هر روش پایه به ازای تعدادی از ۱۴۹ کاربر انتخابی** ۷۳

شکل شماره ۱۸: **وزن های بدست آمده برای روش های پایه بر اساس الگوریتم پیشنهادی**..... ۷۵

شکل شماره ۱۹: **منحنی و امتیازات مدل های ترکیبی و روش های پایه بر اساس مجموعه آموزشی**.... ۷۶

شکل شماره ۲۰: **منحنی و امتیازات مدل های ترکیبی و روش های پایه بر اساس مجموعه آموزشی**.....۷۶

شکل شماره ۲۱: **مجموع امتیاز مدل های ترکیبی و روش های پایه به ازای تعداد نمونه های انتخاب شده** ۷۸

فصل اول

مقدمه

۱-۱-گذشته تحقیق

معمول ترین سناریو برای سیستم‌های توصیه گر مدرن، یک برنامه وب است که با کاربر تعامل داشته باشد. سیستم لیست خلاصه‌ای از موارد را به کاربر نشان می‌دهد و کاربر از میان این لیست اقلامی را انتخاب و جزئیات مربوط به آن ها را مشاهده می‌کند. این سیستم ها ممکن است در شکل های متفاوتی از صفحات وب گرفته تا مقاله‌ها، اخبار، معرفی رستوران‌ها، برنامه‌های تلویزیونی و... مورد استفاده قرار گیرد.

از آنجا که اغلب, تعداد خیلی زیادی از اقلام در پایگاه‌داده وجود دارد که عملا نمی‌توان تمام آنها را روی یک صفحه وب نشان داد،لازم است که زیر مجموعه‌ای از این اقلام را برای نمایش به کاربر انتخاب نمود. در سیستم‌های توصیه‌گر مبتنی بر محتوا، اقلامی به کاربر توصیه می‌شوند که با اقلام دیگری که کاربر موردنظر، قبلا نسبت به آنها ابراز علاقه کرده است، شباهت دارد. بنابراین ارتباط میان اقلام و کشف قوانین مربوط به وابستگی اقلام در این رویکرد بسیار مورد توجه است. از این رو واضح است در این سیستم تمرکز بر روی یافتن شباهت بین اقلام با توجه به ویژگی‌های آن‌ها می‌باشد.

در این سیستم‌ها، در ابتدا مشخصات تمام محصولات موجود در فروشگاه در نظر گرفته شده و آنها بر اساس این مشخصات به دسته‌هایی تقسیم می‌شوند. سپس با استفاده از پایگاه داده‌ای که در مورد مشتری وجود دارد و آنچه از قبل مشتری خریده است، مشخصات مورد علاقه او را استخراج و کالاهای دارای این خصوصیات به او معرفی میگردد. در این روش فقط

مشتری وعلایق او مدنظر بوده و بر اساس اولویت‌های درک شده از مشتری دسته کالاها و سپس کالاهای مورد نظر او توصیه می‌شوند. وب سایت‌هایی که از این روش استفاده نموده‌اند در ابتدا با دریافت نکات و خصوصیات از مشتری و آنچه مشتری می‌پسندد، مرحله به مرحله کالاها را برای او فیلتر کرده و انتخاب را برای او ساده می‌نمایند. به عنوان مثال سایت NewsWeeder [۱] یکی از این سیستم‌ها را برای ارائه خبرهای خود به کاربران مورد استفاده قرار داده است.

ریشه سیستم‌های توصیه‌گر به فعالیت‌های گسترده در زمینه‌های علوم شناختی، تئوری تخمین، بازیابی اطلاعات، تئوری پیش‌بینی، علم مدیریت و مدل‌سازی انتخاب برمی‌گردد [۲]. اهمیت خدماتی اینچنین در این است که اولویت و ترجیحات خوانندگان اغلب پیچیده است و به راحتی با استفاده از کلمات کلیدی و دسته‌بندی‌های استاندارد موضوعی نمی‌توان این پیچیدگی را کاهش داد. در اواسط دهه‌ی ۱۹۹۰ زمانی که محققان تحقیقاتشان را در زمینه‌ی سیستم‌های توصیه‌گر آغاز نمودند[۲]، این تحقیقات به روشنی روی ساختارهای رتبه‌بندی متمرکز بود. در اغلب فرمول‌ها مسایل پیشنهادی با تخمین امتیاز که معمولا توسط کاربران داده میشد, انجام می‌گرفت. تخمین‌ها معمولا بر اساس امتیاز داده شده به اقلام توسط کاربر و اطلاعات دیگر که به صورت قراردادی وجود دارد زده می‌شد. در اینصورت اقلامی به کاربر پیشنهاد می‌شود که بالاترین امتیاز را دارند. این سیستم‌ها امروزه در زمینه‌های متنوعی مثل پیشنهاد محصولات یا خدماتی نظیر اخبار و ایمیل، صفحات وب، کتاب، موزیک، فیلم و ... به کار رفته‌اند.

تعدادی از شرکت هایی که در این رابطه شروع موفقی داشته‌اند عبارتند از: Net perception, Like Mind , Fireefly و نیز کتابفروشی‌های آنلاین نظیر , Barnes And Amazon [۳].

برای توضیح نحوه پیاده‌سازی سیستم‌های توصیه‌گر و به منظور ارائه این سیستم‌ها به صورت یک مدل ریاضی باید با مفهومی به نام تابع سودمندی آشنا شد. به طور کلی یک سیستم توصیه‌گر را می‌توان با نگاشت U همسان دانست:

$$U: C^* S \rightarrow R \qquad (۱)$$

که در آن C مجموعه تمام مشتریان و S مجموعه تمام محصولات و اقلام در دسترس می‌باشد. تابعی که میزان مفید و متناسب بودن کالای $s \in S$ را برای مشتری $c \in C$ محاسبه می‌کند، با $U(c,s)$ نشان داده می‌شود که با انتساب عددی حقیقی به هر جفت (c,s)، اقلام را برای هر مشتری امتیازدهی می‌کند. هر کدام از عناصر S یا C را می‌توان با مجموعه‌ای از خصوصیات خود، تحت عنوان پروفایل[1] توصیف نمود.

از آنجا که این پایان‌نامه بر روی ترکیب سیستم‌های توصیه‌گر مبتنی بر محتوا متمرکز است، در ادامه به توضیحات عمیق‌تر در رابطه با این سیستم‌ها خواهیم پرداخت.

۱-۱- ۱ سیستم‌های توصیه‌گر مبتنی بر محتوا

در رویکرد مبتنی بر محتوا $U(c,s)$ (سودمندی قلم s برای کاربر c)، از روی اقلام 's و $U(c,s')$ به ازاء $s' \in S$ که شبیه s هستند، تخمین زده می‌شود. برای مثال در برنامه توصیه‌گر فیلم به کاربر c ,این سیستم سعی می‌کند شباهت بین فیلم‌ها مانند نام کارگردان، نوع فیلم، موضوع فیلم، بازیگران خاص را که کاربر c به آنها نرخ بالایی داده است، تشخیص داده و فیلم-هایی که درجه تشابه بالاتری با اولویت‌های مشتری دارد را توصیه کند.

استفاده از روش فوق در بسیاری از موارد نظیر اسناد ویدئویی و صوتی بسیار مشکل و نیاز به تکنیک‌های پیشرفته به منظور شناسایی این اسناد دارد. همچنین، اغلب صفحات وب شامل متون توضیحی می‌باشند که به صورت فیلد اضافی در پایگاه داده ذخیره می‌شوند. این فیلدهای توضیحی برخلاف داده‌های متنی ساخت یافته مشکلاتی را برای یادگیری پروفایل کاربر ایجاد می‌نمایند. به طور کلی متون نامحدود مانند مقالات، اخبار، نظرات کاربران و ... از داده‌های غیر ساخت‌یافته به شمار می‌روند و بر خلاف داده‌های ساخت‌یافته، هیچ‌نام، ویژگی خاص و هیچ مقادیر دقیق تعریف شده‌ای برای آنها وجود ندارد. علاوه بر آن پیچیدگی کامل زبان طبیعی ممکن است در متن وجود داشته باشد، نظیر استفاده از کلماتی که دارای معانی مختلفی می‌باشند، یا استفاده از کلمات مترادف و ...

[1] Profile

یک رویکرد مشترک برای رویایی با این فیلدهای متنی نامحدود استفاده از مدل فضای بردار است[۴]. اسناد مورد آزمایش به یک بردار از فضای n بعدی تبدیل می‌شود به گونه‌ای که هر بردار مشخصه‌ای برای آن سند باشد ابعاد بردار مشخصه تعداد واژه‌های متمایز در سند را مشخص می‌کند. به منظور کاهش مشخصه‌ها و افزایش کارایی می‌توان کلمات پرتکراری نظیر "و" و"یا" و "از" و ... را از متن حذف نمود. راه دیگر برای کاهش ابعاد بررسی ریشه‌ی واژه‌های کلیدی است.

بنابراین محتوا در این سیستم‌ها با کلمات کلیدی مشخص می‌شود. f_{ij} و v_{ij}، به ترتیب، تعداد حضور و سودمندی کلمه k_i در سند d_j را نشان می‌دهند. تعیین مقدار v_{ij}به عنوان یک مقدار ویژگی برای سند d_j یکی از مهمترین تفاوت‌های سیستم‌های توصیه‌گر مبتنی بر محتوا است که توجه بسیاری از تحقیقات را به خود جلب نموده است. به عنوان یکی از ساده‌ترین روش‌ها، می‌توان تعداد حضور کلید واژه k_i در متن d_j را به عنوان اهمیت آن نیز استفاده کرد ($v_{ij}=f_{ij}$). اما تعداد حضور یک کلمه در یک متن کوتاه علمی مرتبط ممکن است با تعداد حضور همان کلمه در یک رمان بلند نامرتبط برابری کند. بنابراین، معمولاً تعداد حضور یک کلمه را با توجه به بیشترین تعداد حضور یک کلمه در آن متن مطابق رابطه (۲) نرمال کرده و آن را با فرکانس کلمهTF_{ij} نشان می‌دهند. ماکزیمم تعداد حضور پرتکرارترین کلمه در یک سند جهت نرمال کردن تعداد حضور تمام کلمات در آن سند مورد استفاده قرار می گیرد.[4]

$$TF_{ij} = f_{ij} / \max_z f_{Zj} \qquad (۲)$$

یکی از بهترین روش‌های مقدار دهی $v_{i,j}$، استفاده از معیار فرکانس کلمه تقسیم برمعکوس فرکانس سند (TF-IDF)[1] است که به صورت زیر تعریف می‌شود:

هنگامی که یک کلمه در سندهای زیادی تکرار می‌شود، دیگر آن کلمه برای نمایش تفاوت بین سندها نمی‌تواند زیاد مفید باشد و با کمک آن کلمه کلیدی نمی‌توان تشخیص داد کدام سند مناسب وکدام نامناسب است. به همین دلیل اغلب از معیار معکوس فرکانس سند برای کلمه (IDF_i) در ترکیب با فرکانس کلمه(TF_{ij})استفاده می‌کنند. فرض کنیم که N تعداد کل متن هایی باشد که می توان به کاربر پیشنهاد داد و همچنین کلمه کلیدی K_i در n_i تا از

[1] . Vector Space

آن ها وجود دارد . بنابراین معکوس فرکانس سند برای کلمه کلیدی K_i معمولا به صورت زیر تعریف می‌شود:

$$IDF_i = LogN / n_i \qquad (۳)$$

سپس وزن TF-IDFبرای کلمه کلیدی k_iدر متن d_j به صورت زیر محاسبه می‌شود:

$$v_{ij} = TF_{ij} * IDF_i \qquad (۴)$$

اگر k کلمه کلیدی استخراج شده باشد، محتوای سندd_j نیز با مقادیر ویژگی‌های آن مانند زیر نشان داده می‌شود.

$$\text{Content } (d_j)=(v_{1j},\ldots,v_{kj})$$

همانگونه که قبلا توضیح دادیم، سیستم‌های مبتنی بر محتوا اقلام مشابه با آنچه که مشتری قبلا انتخاب کرده است به وی پیشنهاد می‌دهد. اقلام کاندیدای گوناگون با مجموعه اقلامی که قبلا توسط کاربر امتیازدهی شده مقایسه می‌شوند و شبیه‌ترین آنها به اقلام مورد دلخواه وی، پیشنهاد داده می‌شود. یک پروفایل مربوط به کاربر c با عنوان پروفایل محتوا گرا Profile(c) در نظر می‌گیریم. این پروفایل با استفاده از تحلیل محتوا اقلامی که قبلا مشتری مشاهده کرده یا امتیازدهی کرده است به دست می‌آید. این امر معمولا با کمک تکنیک‌های تحلیل کلمات کلیدی ارزیابی اطلاعات بدست می‌آید. به عنوان مثال پروفایل c شامل یک بردار از وزن‌های ($p_{c1}, p_{c2},\ldots p_{ck}$)که هر وزن p_{ci} نشان‌دهنده اهمیت کلمه کلیدی k_i برای کاربر c است. این که این بردار برای یک کاربر چگونه محاسبه شود، از نقاط اساسی تفاوت در سیستم-های توصیه‌گر است. به طور کلی، در سیستم‌های مبتنی بر محتوا، تابع سودمندی معمولا به صورت زیر تعریف می‌شود:

$$U(c,s)=\text{Score (Profile(c), Content (s))} \qquad (۵)$$

برای محاسبه دقیق‌تر تابع سودمندی، هم باید از پروفایل مربوط به اقلام و هم پروفایل مربوط به مشتری استفاده کرد و بردارهای v_c و v_i را محاسبه کرد و برای تعیین میزان سودمندی آیتم s برای کاربر c باید میزان تشابه پروفایل آیتم s و پروفایل کاربر c را اندازه

بگیریم. چگونگی اندازه گیری تشابه پروفایل یک کاربر به یک سند نیز از جمله مسایل مورد تحقیق در این عرصه است. برای محاسبه این تشابه، می‌توان از فرمول اندازه‌گیری تشابه کسینوسی به صورت زیر استفاده نماییم:

$$U(c,s) = \cos(\Pr ofile(c), Content(s)) = \frac{\Pr ofile(c).Content(s)}{|\Pr ofile(c)| \times |Content(s)|} = \frac{\sum_{i=1}^{k} p_{ci} \times v_{is}}{\sqrt{\left[\sum_{i=1}^{k} {p_{ci}}^2\right] \times \left[\sum_{i=1}^{k} {v_{is}}^2\right]}}$$

(۶)

که k در اینجا نشان‌دهنده تعداد کلمات کلیدی در کل سیستم است. برای مثال اگر کاربرc مقالات آنلاین زیادی در مورد تجارت الکترونیک می‌خواند، در پروفایل خود، کلمات مرتبط با این موضوع از وزن بالاتری نسبت به باقی کلمات برخوردارند. در معیار شباهت کسینوسی، مقالاتی که بیشترین مقدار را به این کلمات اختصاص داده‌اند، از بالاترین میزان شباهت با کاربر c برخوردارند. بعد از مقایسه محتوای سند با محتوای پروفایل کاربر، بهترین سند که مطابق با علاقه‌ی کاربر است به وی نشان داده می‌شود.

۱-۲- بیان مساله

به طور کلی در این پایان نامه، سعی بر ترکیب مناسب روش های امتیاز دهی در سیستم-های توصیه گر است. در هر سیستم توصیه‌گر، با توجه به اطلاعات کسب شده از پیشینه انتخاب و نظرسنجی هریک از کاربران در مورد اقلام متفاوت، یک مدل ساخته می‌شود. سپس بر اساس این مدل، با در نظر گرفتن کاربری که به سیستم وارد می‌شود، امتیازی به هر یک از اقلام موجود در سیستم (که تا به حال توسط آن کاربر دیده نشده‌اند) اختصاص می‌دهد. سپس

مبتنی بر امتیازی که به هر قلم داده شده است آنها را مرتب کرده و به تعداد دلخواه، اقلام ابتدای لیست را به کاربر مورد نظر پیشنهاد می‌کند.

تا کنون روش‌های متفاوتی برای ترکیب این سیستم‌ها ارائه شده است، اما در این پایان-نامه مدل ترکیب خطی این روش‌ها مد نظر قرار می‌گیرد. در این مدل سعی می‌شود با استفاده از هر یک از سیستم‌های اولیه، به ازاء هر کاربر، به هر قلم خاص یک امتیاز داده می‌شود. سپس با ترکیب خطی این امتیازات، یک امتیاز کلی برای هر ترکیب کاربر-قلم محاسبه می‌شود. در نهایت به ازاء هر کاربر، یک لیست مرتب شده از اقلام (بر اساس این امتیاز کلی) آماده شده و اقلام با بیشترین امتیاز به وی توصیه می‌شود.

مسئله امتیازدهی در سیستم‌های پیشنهاد دهنده، از شباهت بسیاری با سیستم‌های بازاریابی مستقیم برخوردارند. در این سیستم‌ها هدف امتیازدهی به تعدادی از مشتریان بالقوه بر اساس احتمال پاسخگویی به درخواست ارسالی است. بدون در نظر گرفتن اینکه در این سیستم‌ها، این امتیازدهی به چه نحوی صورت می‌پذیرد، می‌توان بر این شباهت پافشاری کرد که هم در سیستم‌های پیشنهاد دهنده و هم در سیستمهای بازاریابی مستقیم، به تعداد مورد نیاز، به ترتیب، اقلام یا مشتریان بالقوه با بیشترین امتیاز انتخاب می‌شوند. در این راستا، در [۵] یک ترکیب خطی امتیاز های به دست آمده از چند روش پایه در بازاریابی مستقیم پیشنهاد می شود. رویکرد اصلی این پایان نامه نیز الهام گرفتن از روش پیشنهادی در [۵] و ترکیب تعدادی از سیستم‌های پیشنهاددهنده با استفاده از وزن دهی این روش‌ها و ترکیب خطی امتیازات می-باشد. فرمول زیر، ترکیب خطی امتیازات هر روش پایه را نشان می‌دهد:

$$U(c,j)=\sum_{1\le p\le n} w^p.U^p(c,j)$$

(۷)

در شرایطی که n روش پایه داریم و $U^p(c,j)$ عبارت است از امتیاز داده شده توسط روش پایه pاُم به قلم شماره j برای مشتری c؛ در حالیکه $U(c,j)$ امتیاز محاسبه شده از ترکیب امتیازات تمام روش‌های پایه می‌باشد. ضرایب این ترکیب خطی توسط بردار $W=[w^1,w^2,...,w^n]$ مشخص می‌شوند به گونه‌ای که، w^p بیانگر ارزش و تأثیر گذاری روش pاُم در امتیاز نهایی است.