دوره دهم، شماره اول، بهار و تابستان ۱۳۹۶

مجله فناوری اطلاعات در طراحی مهندسی
Information Technology in Engineering Design
http://ited.sinaweb.net

ارائه مدلی برای سیستم‌های توصیه‌گر فیلم مبتنی بر رویکرد مشارکت محور

بهنام حیدری [۱] الهام پروین نیا * [۲]

(۱) گروه مهندسی کامپیوتر، واحد شیراز، دانشگاه آزاد اسلامی، شیراز، ایران

(۲) گروه مهندسی کامپیوتر، واحد شیراز، دانشگاه آزاد اسلامی، شیراز، ایران*

چکیده

روش فیلترینگ مشارکتی از دو مشکل عمده رنج می‌برد: مشکل اول شروع سرد کاربران است و مشکل دوم نیز مسئله اعتماد کاربران وفادار است. در این مقاله با ترکیب روش محتوا محور و فیلترینگ مشارکتی به صورت بوستینگ یک سیستم توصیه گر طراحی شده است که علاوه بر حل مشکل شروع سرد، مسئله اعتماد را نیز پوشش می‌دهد. روش پیشنهادی شامل سه مرحله خوشه بندی اولیه کلیه کاربران و تعیین خوشه، تعیین وزن مناسب برای هر کدام از ویژگیها و تعیین همسایگان نزدیک کاربر جدید، تشکیل ماتریس مجاورتی و محاسبه امتیاز کاربر جدید به هر فیلم. روش ارائه شده به منظور خوشه بندی اولیه، از سیستم محتوا محور بر اساس اطلاعات دموگرافیک کاربران استفاده می‌نماید. جهت شناسایی همسایگان نزدیک و ارائه پیشنهادات از فیلترینگ مشارکتی استفاده می‌کند. جهت انجام آزمایش‌ها از بانک اطلاعاتی movielens استفاده شده است. نتایج نشان می‌دهد که میانگین مجذور خطای روش پیشنهادی نسبت به روش‌های پایه مانند نایوبیز، الگوریتم‌های C24.5، CM4.5،RCA،SVD و ApproSVD به ترتیب ۸/۵۹٪، ۸/٦۷٪، ۸/٤٥٪، ۸/۱۵٪، ٤/۵٪ و ٦/۰۵٪ کاهش یافته است.

واژه‌های کلیدی: سیستم پیشنهاد دهنده، چالش شروع سرد، بوستینگ، معیار شباهت ترکیبی

*عهده‌دار مکاتبات:

الهام پروین‌نیا

نشانی: گروه مهندسی کامپیوتر، واحد شیراز، دانشگاه آزاد اسلامی، شیراز، ایران

تلفن: ۰۷۱۳٦۲٦٦۸۹۳ **پست الکترونیکی:** parvinnia@iaushiraz.ac.ir

۱- مقدمه

سیستم‌های پیشنهاد دهنده در دامنه‌های مختلفی همچون فروشگاه فیلم، کتابخانه‌ها، رستوران‌ها، سیستم‌های گردشگری و سایر محیط‌ها برای ارائه آیتم‌های جذاب مورد استفاده قرار می‌گیرند [1]. این سیستم‌ها نقش حیاتی و بسیار مهمی در تجارت الکترونیک و به خصوص وب سایت آنلاین فیلم ایفا می‌کنند [2]. یکی از محبوب‌ترین سیستم‌هایی که در دهه گذشته تا به امروز مورد توجه محققان قرار گرفته است، سیستم‌های پیشنهاد دهنده فیلم است [4][3]. با توجه به حجم عظیم اطلاعات، ارائه جذاب‌ترین آیتم‌ها (فیلم) به کاربران در زمان مناسب و با دقت کافی یکی از مسائل مطرح در این زمینه می‌باشد. سیستم پیشنهاد دهنده فیلم، این امکان را فراهم می‌سازد که فیلم‌ها بوسیله ویژگی‌هایی مثل عنوان فیلم، کارگردان، نویسنده، تاریخ اکران، نوع فیلم و غیره به کاربران ارائه شود. به طور کلی سیستم‌های پیشنهاد دهنده به دو دسته مهم تقسیم می‌شوند [5] که عبارتند از:

- سیستم‌های فیلترینگ محتوا محور[1]
- سیستم‌های فیلترینگ مشارکتی[2]

در سیستم‌های فیلترینگ محتوا محور پیشنهادها بر اساس رتبه‌ها و امتیازی که کاربر به محتوا، متن خبری، لینک و غیره داده است ارائه می‌گردد. بر همین اساس محتوایی که بالاترین امتیاز را داشته است، پیشنهاد داده می‌شود [6]. در سیستم‌های فیلترینگ مشارکتی، پیشنهادها بر اساس انتخاب کاربران مشابه و امتیازات آنها به فیلم‌ها است. مهم‌ترین چالش‌های پیش رو در سیستم‌های پیشنهاد دهنده مبتنی بر فیلترینگ مشارکتی که در سال‌های اخیر مورد توجه محققان مختلف قرار گرفته است عبارتند از:

- مشکل شروع سرد: به کاربرانی گفته می‌شود که برای سیستم، جدید هستند و پروفایلشان خالی است (بعبارتی در محیط عملیاتی فیلم، این کاربران به هیچ فیلمی امتیاز نداده‌اند) و یا رتبه‌های خیلی کمی از آنها در سیستم ثبت شده است. برای کاربرانی که شروع سرد دارند و در سیستم فعالیتی از آنها در دسترس نیست از فیلترینگ محتوا محور بر اساس پروفایلشان استفاده می‌شود. اما برای کاربرانی که سوابق آنها در سیستم در دسترس است از روش فیلترینگ مشارکتی استفاده می‌شود.
- اعتماد: اعتبار کاربرانی که تاریخچه طولانی‌تری در استفاده از سیستم دارند نسبت به کاربرانی که تاریخچه حضور آنها در سیستم کمتر است، بیشتر است. لذا امتیازات آنها به فیلم‌ها از ارزش بالاتری برخوردار است.
- پراکندگی داده‌ها: معمولاً سایت‌های تجارت الکترونیک شامل محصولات زیادی هستند و هر کاربر تنها بخشی از محصولات سایت را ارزیابی می‌کند. اگر دو کاربر به کالای مشابه ای رتبه ندهند، در صورت داشتن علایق مشابه، سیستم نمی‌تواند مشابهت بین آن‌ها را تشخیص دهد و نمی‌تواند هیچ محصولی را برای پیشنهاد انتخاب کند.
- مقیاس پذیری: با گسترش سیستم‌های نرم افزاری تعداد کاربران و اقلام در فضای کاربر هدف برای جستجوی همسایه‌ها، افزایش می‌یابد. در فیلترسازی مشارکتی مبتنی بر کاربر یا فیلم، برای محاسبه مشابهت باید کل پایگاه داده بررسی شود. بنابراین، با افزایش رکوردهای ثبت شده، پیچیدگی محاسباتی به طور چشم گیری افزایش خواهد یافت که منجر به کاهش کارایی سیستم می‌شود.

در این مقاله سیستم پیشنهاد دهنده‌ای طراحی شده است که با استفاده از یک روش ترکیبی و با در نظر گرفتن امتیاز برای کاربران قدیمی مشکلات شروع سرد و اعتماد را تا حد زیادی حل می‌نماید. روش پیشنهادی شامل سه مرحله خوشه بندی اولیه کلیه کاربران و تعیین خوشه، تعیین وزن مناسب برای هر کدام از ویژگی‌ها و تعیین همسایگان نزدیک کاربر جدید، تشکیل ماتریس مجاورتی و محاسبه امتیاز کاربر جدید به هر فیلم. روش ارائه شده به منظور خوشه بندی اولیه، از سیستم محتوا محور بر اساس

[1] Content-Base Filtering
[2] Collaborative Filtering

اطلاعات دموگرافیک کاربران استفاده می‌نماید. جهت شناسایی همسایگان نزدیک و ارائه پیشنهادها از فیلترینگ مشارکتی استفاده می‌کند.

ادامه این مقاله به این صورت تقسیم بندی شد است: در بخش ۲ کارهای انجام شده در گذشته مورد بررسی قرار گرفته، در بخش ۳ مدل پیشنهادی همراه با توصیف معماری مطرح شده ارائه شده است. در بخش ٤ نتایج بدست آمده و در بخش ۵ نیز نتیجه گیری نهایی و پیشنهادهای آینده مطرح شده است.

۲- سوابق پیشین

در این بخش برخی از جدیدترین پژوهش‌های مطرح شده در جهت رفع مشکل شروع سرد کاربران به منظور ارائه پیشنهادهای فیلم تشریح می‌گردد.

هیونگ و همکارانش در سال ۲۰۱۱، مشکلات شروع سرد فیلم‌ها و کاربران را بررسی کردند. آن‌ها یک سیستم بهبود فیلترینگ مشارکتی سنتی ارائه کردند. در این مدل از دو ماتریس شباهت استفاده شده است که یکی شباهت بین کاربران و فیلم‌ها و دیگری شباهت بین کاربران با یکدیگر را نشان می‌دهد. سپس با توجه به مکانیزم پیش بینی مطرح شده پیشنهاداتی را به کاربران ارائه می‌دهند. یکی از معایب قابل تأمل در این تحقیق استفاده از حافظه زیاد با توجه به تعداد کاربران و فیلم‌ها، بعلت ساخت چندین ماتریس شباهت می‌باشد [7].

بوبادیلا و همکارانش در سال ۲۰۱۲، از شبکه عصبی به عنوان یک سیستم پیشنهاد دهنده فیلترینگ مشارکتی جهت کاهش مسئله شروع سرد کاربران جدید الورود استفاده کردند. آن‌ها ارزیابی‌هایشان را بروی مجموعه داده Movielens, Netflix انجام داده و در تحقیق خود به دلیل استفاده از اطلاعات غیره عددی از معیار شباهت گیری جاکارد استفاده کردند [8].

هنس در سال ۲۰۱۳، پیشنهاد فیلم به کارابران را با استفاده از خوشه بندی فیلم‌ها با استفاده از الگوریتم k-means و بر اساس امتیازات ارائه شده به فیلم‌ها توسط کاربران انجام داد. هنس پژوهش خود را بروی دیتاست معروف movielens انجام داده و ارائه فیلمها را بروی مجموعه داده با حجم ۱۰۱۰۹ فیلم که توسط ۲۱۱۳ کاربر امتیاز دهی شده‌اند پیاده سازی نموده است [9].

کامواتسوس و همکارانش در سال ۲۰۱٤، مدلی را مطرح نمودند که در آن از الگوریتم‌های طبقه بندی مثل نایوبیز، درخت تصمیم و الگوریتم طبقه بندی تصادفی با بکارگیری معیارهای شباهت به منظور پیشنهاد فیلم‌ها به کاربران استفاده شده است. آن‌ها نیز در پژوهش خود، ارزیابی‌ها را بروی دیتاست Movielens انجام دادند [10].

لوئیز و همکارانش در سال ۲۰۱۵، جهت افزایش کارایی سیستم و حل مشکل شروع سرد، یک روش ترکیبی مطرح کردند که روشهای فیلترینگ مشارکتی و اطلاعات دموگرافی را با هم ترکیب می‌کرد. آن‌ها در پژوهش خود از الگوریتم‌های ترکیبی Co-Clustering و یادگیری ماشین جهت حل مسئله شروع سرد استفاده کرده و ارزیابی‌ها بروی دیتاست‌های Movielens, Jester, Netflix انجام شده است [11]. به دلیل کم اهمیت بودن چالش‌هایی مثل مقیاسپذیری، پراکندگی و اعتماد کاربران در مقایسه با شروع سرد کاربران و فیلم‌ها، در کلیه پژوهش‌هایی که تا کنون انجام شده است این چالشها نیز با عملیات پیش پردازش، خوشه بندی و دسته بندی نیز مرتفع گردیده‌اند.

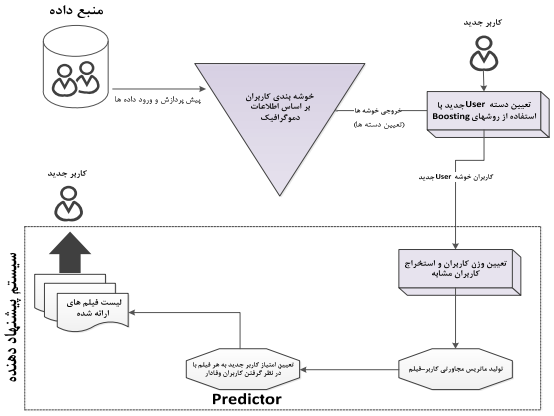
جین و همکارانش در سال ۲۰۱۷، از یک سیستم پیشنهاد دهنده مشارکتی و روش یادگیری عمیق جهت ارائه پیشنهادات فیلم و رفع مشکل شروع سرد در سیستم‌های پیشنهاد دهنده مشارکتی استفاده نمودند. آن‌ها آزمایشات خود را بر روی دیتاست‌های Movie lens و نت فلیکس انجام دادند [12].

لوسیف و همکارانش در سال ۲۰۱۸، از یک روش ترکیبی مبتنی بر دانش و قوانین انجمنی جهت رفع مشکل شروع سرد کاربران در رشته‌های پیشنهاد دهنده متن آگاه استفاده نمودند [13].

در تحقیقی دیگر در سال ۲۰۱۸، کارتیک و همکارانش از ترکیب سیستم‌های پیشنهاددهنده محتوا محور و مشارکتی جهت پیشنهاد فیلم به کاربران جدیدالورود استفاده نمودند [14].

۳- مدل پیشنهادی

طرح مدل پیشنهادی در شکل ۱ ارائه شده است، در ادامه وظایف هر یک از اجزای مدل توضیح داده می‌شود.



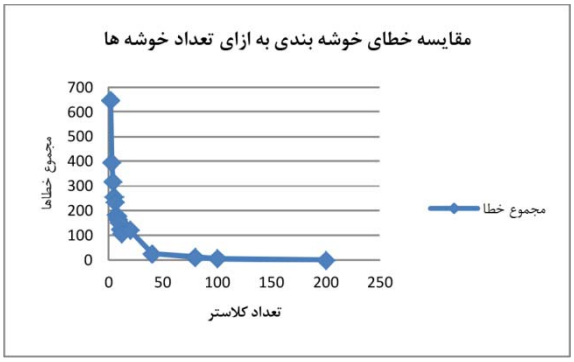
شکل ۱- فلوچارت و معماری روش پیشنهادی

۱- ابتدا بروی داده‌ها عملیات پیش پردازش انجام شده و از پردازش کاربرانی که در سیستم فاقد اطلاعات پروفایلی یا امتیازات به فیلم‌ها است صرف نظر می‌شوند تا سرعت تحلیل و پردازش داده‌ها بهبود حاصل شود. ۲- خوشه بندی کاربران بر اساس اطلاعات دموگرافیک و با استفاده از الگوریتم خوشه‌بندی k-means. ۳- تعیین خوشه مناسب با استفاده از ترکیب خوشه بندی و معیارهای تکنیک boosting به کاربران جدیدالورود به سیستم که با شروع سرد مواجه هستند. ٤- یافتن کاربران مشابه با استفاده از معیارهای شباهت‌گیری ترکیبی مطابق با میزان شباهت سنی، جنسیت و تحصیلات. ٥- تشکیل ماتریس مجاورتی کاربر- آیتم که در آن امتیازات ارائه شده کاربران همسایه (سطرها) به کلیه فیلم‌ها (ستون‌ها) نشان داده شده است. ٦- محاسبه امتیاز کاربر جدید به هر فیلم با در نظر گرفتن کاربران وفادار. ۷- ارائه لیستی از فیلم‌ها به کاربران با استفاده از یک مکانیزم پیش بینی از روی ماتریس مجاورتی.

در قسمت‌های بعدی روش پیشنهادی و فلوچارت ارائه شده به صورت کامل تشریح می‌گردد. قبل از تشریح مراحل تحلیل مدل، لازم است که بروی داده‌های مربوطه عملیات پیش پردازش را انجام داده و داده‌ها را به منظور استفاده‌های بعدی آماده سازی نمائیم.

۳-۱- خوشه بندی کاربران

قبل از اعمال خوشه بندی کاربران بر اساس اطلاعات دموگرافیک آن‌ها لازم است ابتدا نرمال سازی داده‌ها انجام شود. به منظور انتخاب بهینه تعداد خوشه‌ها، با استفاده از نرم افزار داده کاوی weka داده‌ها را به تعداد خوشه‌های (k) مختلف ارزیابی نموده و ضمن محاسبه مجموع مربعات خطا هر خوشه‌بندی، تعداد قابل قبول k را تعیین می‌نمائیم. در شکل ۲ مجموع مربعات خطا عملیات خوشه بندی به ازای k های مختلف رسم شده است.



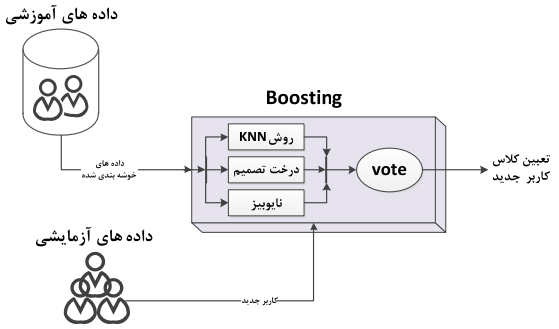
شکل ۲- مقایسه مجموع مربعات خطا خوشه بندی به ازای تعداد k های مختلف

همانطور که در شکل ۲ دیده می‌شود خوشه‌های k=100 دارای دقت ۹۸٪ است. بنابراین تعداد خوشه‌های 100 برای خوشه‌بندی در نظر گرفته می‌شود. در این مقاله خوشه‌بندی کاربران را با استفاده از نرم افزار داده کاوی رپیدماینر انجام داده و محاسبه میزان خطای خوشه بندی با توجه به تعداد

k های متفاوت نیز توسط نرم افزار داده کاوی وکا انجام و مورد ارزیابی قرار می‌گیرد.

۳-۲- انتخاب خوشه مناسب برای کاربر جدید با استفاده از روش boosting

در این مرحله با استفاده از اطلاعات دموگرافیک کاربران جدید و خوشه‌های تعیین شده در مرحله قبل می‌توان خوشه مناسب کاربر جدید را یافت. در شکل ۳ مراحل انتخاب خوشه با استفاده از روش بوستینگ نشان داده شده است.



شکل ۳- تعیین دسته کاربر جدید الورود با استفاده از روش‌های بوستینگ

همانطور که در شکل ۳ دیده می‌شود، داده‌های آموزشی که خروجی مرحله خوشه بندی است به سیستم boosting داده شده و مدل مورد نظر تولید می‌شود. سپس کاربر جدید به عنوان داده‌های آزمایشی به سیستم وارد شده و دسته آن تعیین می‌گردد. پس از اینکه برای کاربر جدید خوشه و یا کلاس تعیین شد، همسایگان آن که شامل کاربران آن خوشه می‌باشند استخراج می‌شوند. نظرات این همسایگان در پیشنهاد فیلم مورد توجه قرار می‌گیرد.

۳-۳- یافتن کاربران مشابه

پس از اینکه برای کاربر جدید خوشه یا کلاس تعیین شد همسایگان آن که شامل کاربران موجود در آن خوشه می‌باشد استخراج می‌شوند. نظرات این همسایگان در پیشنهاد نهایی فیلم به کاربر جدید مؤثر هستند. اما همه این همسایگان به یک اندازه به کاربر جدید شباهت ندارند و در ادامه از یک معیار شباهت برای همسایگان نزدیک‌تر استفاده خواهد شد.

۳-۳-۱- تعریف معیار شباهت

فرض کنید کاربران سیستم به صورت مجموعه $U = \{u_1, u_2, u_3, \dots, u_m,\}$ با خصوصیات دموگرافیک کاربران و مجموعه فیلم‌ها به صورت $D = \{d_1, d_2, d_3, \dots, d_l\}$ تعریف شده باشند.

فرض کنید مجموعه $W = \{w_1, w_2, w_3, \dots, w_i\}$ وزن خصوصیات دموگرافیک هر کاربر است. در بسیاری از سیستم‌های پیشنهاد دهنده وزن همه خصوصیات برابر یک در نظر گرفته می‌شود. نویسندگان بر این باورند که نقش بعضی از خصوصیت‌ها نسبت به بقیه در امتیاز به فیلم مهم‌تر است. برای مثال نقش خصوصیت سن به نسبت خصوصیت شغل یا جنسیت در انتخاب فیلم مهمتر است. بنابراین بصورت تجربی وزن هر یک از خصوصیات را با اعداد بازه [0,1] مقدار دهی می‌نمایند. سپس شباهت یک کاربر جدید (n) و هر یک از همسایگان (u_p) بر اساس رابطه (۱) محاسبه می‌گردد:

$$sim(n,u) = \frac{\sum_{j=1}^{I} SF_j * w_j}{\sum_{j=1}^{I} s(d_j(n,u_p))} \qquad (۱)$$

SF_j مقدار مشابهت j مین ویژگی است، w_j وزن مربوط به ویژگی مورد نظر است. $s(d_j)$ تابعی است که مقدار بین [0,1] دارد. این تابع میزان شباهت یک ویژگی در دو کاربر را محاسبه می‌کند. با توجه به ماهیت ویژگی‌های کاربران مقدار $s(d_j)$ در دو گروه کلی تعریف می‌شود.

برای ویژگی‌های عددی

(۲)

$$s(d_j(n,u_p) = (1 - \frac{|Diff(d_{j,n}, d_{j,u_p})|}{max - Diff_{max}(d_j)})^{\beta}$$

برای مقادیر رشته‌ای و بولین

(۳)

$$s(d_j(n,u_p) = \begin{cases} if\ d_{j,n} = d_{j,u_p} \\ if\ d_{j,n} <> d_{j,u_p} \end{cases}$$

که در آن $d_{j,n}$ مقدار ویژگی در کاربر n و d_{j,u_p} مقدار ویژگی j ام در کاربر u_p و Diff تابع تفاضل و $Diff_{max}$ ماکزیمم تفاضل ویژگی jام کاربر جدید و هر یک از همسایگان است. β پارامتری است که میزان تأثیر تفاضل یک ویژگی را تعیین می‌نماید.

۳-٤- تشکیل ماتریس مجاورتی

پس از اینکه با استفاده از معیارهای شباهت ترکیبی میزان مشابهت کاربر جدید با سایر کاربران همسایه را بدست آوردیم ماتریس مجاورتی مربوط به امتیازات ارائه شده به فیلم‌ها توسط کاربران همسایه تشکیل می‌شود. سپس با استفاده از رابطه (٤) امتیازات فیلم محاسبه می‌گردد. فیلم‌هایی که بیشترین امتیاز را کسب نمایند به عنوان فیلم برتر پیشنهاد می‌گردند. شکل کلی ماتریس مجاورتی مربوط به فیلم و کاربران همسایه به صورت ذیل است.

$$\boldsymbol{R} = \begin{array}{c|cccccccc} & I_1 & I_2 & I_3 & I_4 & I_5 & I_6 & I_7 & I_b \\ u1 & 5 & 4 & 4 & 5 & 2 & 5 & 5 & 3 \\ u2 & 3 & 5 & 5 & 4 & 5 & 4 & 5 & 5 \\ u3 & 4 & 3 & 5 & 4 & 3 & 3 & 4 & 4 \\ u4 & 5 & 5 & 5 & 5 & 2 & 2 & 4 & 3 \\ u5 & 1 & 2 & 4 & 3 & 1 & 3 & 4 & 5 \\ u6 & 2 & 5 & 3 & 2 & 3 & 4 & 5 & 4 \\ u7 & 5 & 4 & 5 & 1 & 5 & 5 & 5 & 5 \\ nj & ? & ? & ? & ? & ? & ? & ? & ? \end{array}$$

همان طور که در ماتریس بالا مشخص است، کاربر ۱ به فیلم ۱ امتیاز ۵ و به فیلم ۲ نیز امتیاز ٤ داده است.

۳-۵- پیش‌بینی امتیازات کاربر جدید الورود به فیلم‌ها

پس از تشکیل ماتریس مجاورتی با استفاده از امتیازات مربوط به هر فیلم که توسط کاربران همسایه امتیاز داده شده‌اند، امتیاز هر فیلم را برای کاربر جدید محاسبه نموده و در قالب پیش بینی ارائه می‌گردد. امتیاز پیش بینی شده برای هر کاربر با استفاده از معادله زیر بدست می‌آید. که در آن p تعداد کاربران همسایه، u_j کاربر همسایه j ام، r_{u_j,i_b} امتیاز کاربر همسایه j ام به فیلم iام و R_{n,i_b} امتیاز کاربر جدید به فیلم i ام می‌باشد.

(٤)

$$R_{n,i_b} = \frac{\sum_{j=1}^{p} Sim(n,u_j).r_{u_j,i_b}}{\sum_{j=1}^{p} Sim(n,u_j)}$$

بنابراین با استفاده از رابطه (٤) امتیازات پیش بینی شده برای هر آیتم توسط کاربر جدید به صورت تقریبی پیش بینی شده و فیلم‌هائی انتخاب می‌شود که دارای بالاترین میزان امتیاز باشند.

٤- ارزیابی نتایج

در این مقاله جهت بررسی موضوع و ارزیابی نتایج، شبیه سازی مورد نیاز بروی دیتاست movielens انجام شده است. جهت دستابی به منبع داده استفاده شده کافیست به [15] مراجعه نموده و از نسخه‌های ارائه شده، مجموعه داده مورد نظر را انتخاب و دانلود نمود. نسخه مورد استفاده در این تحقیق، نسخه ۲۰۱۳ بوده که حجمی معادل با 1 مگابایت دارد. این فایل‌ها شامل 1,000,209 رکورد مربوط به امتیازات ثبت شده توسط کاربران به فیلم‌ها، 3,900 نمونه فیلم و 6,040 نمونه کاربر است که هر کاربر حداقل به ۲۰ فیلم رأی داده است.

از معیارهای ارزیابی میانگین خطای واقعی[1] و مجذور مربع میانگین خطا[2] استفاده شده است.

(۵)

$$MAE = \frac{1}{k}\sum_{u,i} |P_{u,i} - r_{u,i}|$$

(٦)

$$RMSE = \sqrt{\frac{1}{k}\sum_{u,i}(P_{u,i} - r_{u,i})^2}$$

که در آن Pu,i امتیاز پیش بینی شده کاربر u و فیلم i ام و

[1] Mean Absolute Error(MAE)
[2] Root Mean Sequare Error(RMSE)

$r_{u,i}$ امتیاز واقعی کاربر u برای فیلم i ام است.

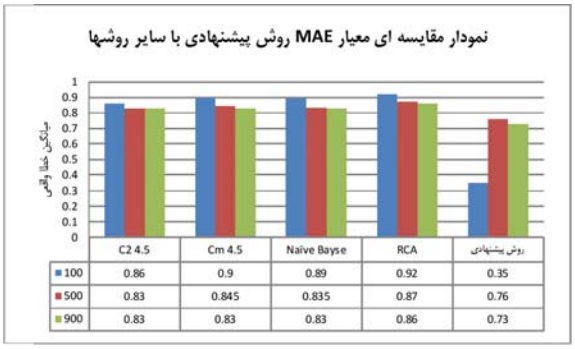
در این قسمت چندین سناریو تعریف شده که در جدول ۲ نشان داده شده است. همانطور که مشاهده می‌شود، به ازای وزن‌های مختلف برای ویژگی‌های سن، جنسیت و شغل نتایج متنوعی بدست می‌آید. سناریوهای بالا با مقادیر مختلف وزن‌ها تعریف شده‌اند. وزن‌هایی که دارای ویژگی‌های بیشتری هستند در واقع تمرکزشان بروی ویژگی مورد نظر نیز بیشتر است و در میزان شباهت تأثیر بیشتری دارد.

۵-۱ ارزیابی نتایج

در این قسمت یک ارزیابی جامع از روش پیشنهادی نسبت به سایر روش‌های بیان شده در بخش قبل ارائه خواهیم نمود. همانطور که در جدول زیر نشان داده شده است کلیه سناریوهای ۱ و ۲ و ۳ که وزن‌های متفاوتی مقدار دهی شده‌اند در قالب یک جدول جامع از کلیه روش‌ها به ازای مقادیر داده‌ای متفاوت ارائه شده‌اند. در شکل ٤ مقادیر مربوط به معیارهای ارزیابی MAE با توجه به تعداد کاربران ۱۰۰, ۵۰۰, ۹۰۰ و سناریوهای ۱،۲،۳ خلاصه شده است.

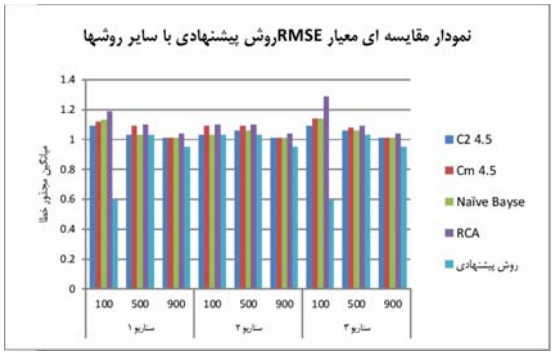
جدول ۲- سناریوهای مختلف

سناریوها	وزون
۱	W_1=0.6, W_2=0.3,W_3=0.1
۲	W_1=0.3, W_2=0.6,W_3=0.1
۳	W_1=0.3, W_2=0.1,W_3=0.6



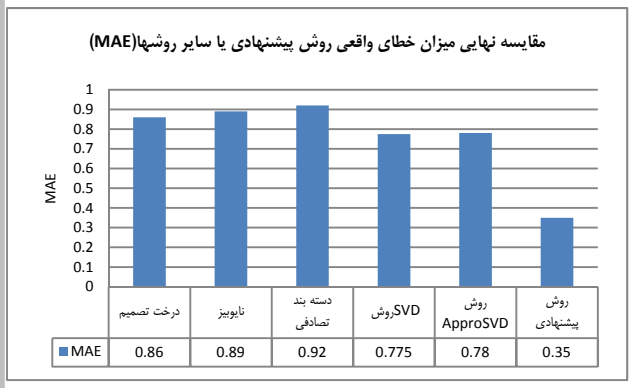
شکل ٤- مقادیر مربوط به MAE برای تعداد کاربران ۱۰۰،۵۰۰،۹۰۰ و سناریوهای ۱،۲،۳

در شکل ۵ مقادیر مربوط به معیارهای ارزیابی RMSE با توجه به تعداد کاربران ۱۰۰, ۵۰۰, ۹۰۰ و سناریوهای ۱،۲،۳ خلاصه شده است.



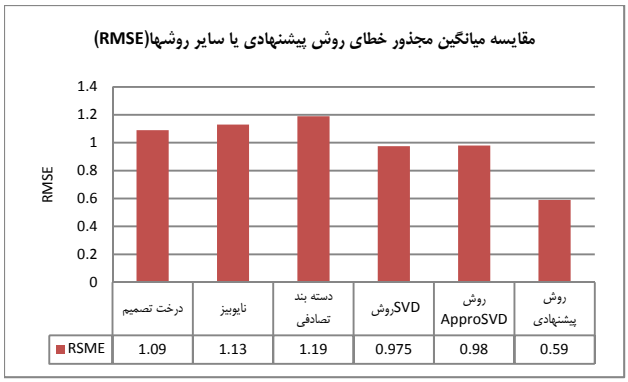
شکل ۵ – مقادیر مربوط به RMSE برای تعداد کاربران ۱۰۰،۵۰۰،۹۰۰ و سناریوهای ۱،۲،۳

لذا با توجه به ارزیابی‌های صورت گرفته بر روی روش پیشنهادی و مقایسه با سایر پژوهش‌های اخیر که در سال‌های 2014,2015 انجام شده است، در نمودار زیر میزان خطای واقعی (MAE) روش پیشنهادی با سایر روش‌های پایه مثل الگوریتم‌های درخت تصمیم، نایوبیز، دسته‌بند تصادفی [10] و همچنین الگوریتم‌هایی مثل SVD, ApproSVD مورد مقایسه قرار گرفته است [16].



شکل ٦- مقایسه نهایی میزان خطای واقعی روش پیشنهادی یا سایر روش‌ها (MAE)

و همچنین میانگین مجذور خطای روش پیشنهادی در مقایسه را سایر روش‌ها در نمودار زیر نشان داده شده است.



شکل ۷– مقایسه نهایی میزان خطای واقعی روش پیشنهادی یا سایر روش‌ها (RMSE)

۵- نتیجه‌گیری و پیشنهادهای آینده

هدف اصلی این مقاله رفع مشکل شروع سرد کاربران در شبکه‌های آنلاین فیلم و ارائه فیلم‌های مناسب به کاربران جدیدالورود با دقت قابل قبول به عنوان پیشنهاد با بکارگیری ترکیب سیستم‌های پیشنهاد دهنده محتوا محور و فیلترینگ مشارکت و همچنین استفاده از تکنیک‌های داده‌کاوی بوده است. لذا در این مقاله، محقق توانسته است که با استفاده از تکنیک‌های خوشه بندی و معیارهای شباهت ترکیبی، فیلم‌هایی را به کاربران جدید الورود ارائه کند که نسبت به سایر روش‌هایی که تا کنون انجام شده است از دقت بالاتری بر خوردار باشد. بنابراین با ورود کاربر جدید به سیستم با توجه به اینکه به هیچ فیلمی امتیاز نداده است و با شروع سرد مواجه است، روش پیشنهادی ابتدا با استفاده از الگوریتم خوشه بندی و به بکارگیری قابلیت‌های روش بوستینگ، بر اساس اطلاعات دموگرفیک کاربر خوشه مورد نظر را برای کاربر تعیین نموده و سپس نزدیک‌ترین کاربران مشابه را با استفاده از یک معیار شباهت ترکیبی استخراج می‌کند. در نهایت با بکارگیری یک استراتژی پیش بینی و بر اساس امتیازات ارائه شده کاربران همسایه به هر فیلم و ترکیب با میزان شباهت کاربران همسایه با کاربر جدید بر اساس وزن دهی اعمال شده در مرحله قبل، امتیاز مورد نظر برای هر فیلم توسط کاربر جدید پیش بینی شده و پس از مرتب سازی، فیلم‌های مرتبط را ارائه می‌نماید. لذا با توجه به شبیه‌سازی انجام شده جهت ارائه فیلم‌های جذاب به کاربران جدیدالورود که با شروع سرد مواجه هستند، روش پیشنهادی با دقت قابل قبولی نسبت به سایر روش‌ها، فیلم‌های مرتبط و جذابی را در زمان مناسب به کاربر جدید ارائه می‌کند. بنابراین مشکل عمده دیگر نیز مسئله اعتماد است که این چالش بر اثر عدم توجه به کاربران قدیمی نسبت به کاربران جدید بوجود می‌آید. روش پیشنهادی شامل سه مرحله است: (۱) خوشه بندی اولیه کلیه کاربران و تعیین خوشه مناسب به کاربر جدید (۲) اعمال وزن مناسب برای هر کدام از ویژگی‌های کاربران خوشه مورد نظر و تعیین همسایگان نزدیک کاربر جدید (۳) تشکیل ماتریس مجاورتی امتیازات کاربران همسایه به فیلم‌های موجود و محاسبه امتیاز کاربر جدید به هر فیلم با توجه به امتیازات کاربران همسایه و میزان مشابهت هر همسایه به کاربر جدید که در این مرحله برای امتیاز به کاربران وفادار سیستم امتیاز مضاعف به فیلم‌ها در نظر گرفته شده است.

در نهایت نیز جهت ارزیابی میزان خطای پیش بینی فیلم در روش پیشنهادی نسبت به سایر روشهای مشابه مثل C24.5, CM4.5, RCA و روش نایوبیز از معیارهای ارزیابی MAE, RMSE استفاده شده است. این معیارهای ارزیابی در سناریوهای 1,2,3 که با توجه به وزن‌های مختلف ارائه شدند که با تعداد کاربران ۱۰۰, ۵۰۰،۹۰۰ مورد ارزیابی قرار گرفتند. میزان خطای روش پیشنهادی از سایر روش‌های مشابه کمتر بوده و دقت قابل قبولی را جهت پیشنهاد فیلم به کاربر جدیدالورود ارائه داده است. برخی از پیشنهادهایی که در آینده سایر محققین می‌توانند ادامه دهند عبارتند از: ۱) در مقاله فعلی از تکنیک خوشه‌بندی k-means جهت خوشه‌بندی کاربران با تعداد k هایی که با تست و خطا بدست آمده استفاده شده است. لذا در مقاله‌های دیگر می‌توان از تکنیک‌هایی مثل الگوریتم خوشه‌بندی Random Walk و یا الگوریتم خوشه‌بندی بهبود یافته عملیات خوشه‌بندی استفاده شود و نتایج به

دست آمده همراه با میزان خطای خوشه بندی با روش فعلی و سایر روش‌ها مورد مقایسه و ارزیابی قرار گیرد. 2) در این مقاله جهت انتقال کاربر جدید به خوشه مورد نظر از ایده بوستینگ استفاده نموده تا بتوان دسته‌هایی را برای کاربران انتخاب نمود که بسیار مناسب هستند. لذا در کارهای آینده می‌توان تکنیک‌های دیگری را جایگزین نموده و نتایج را به روش پیشنهادی مورد مقایسه قرار داد. 3) در مرحله استفاده از تکنیک‌های مشابهت ترکیبی و بکارگیری فرمول‌های متنوع جهت اعمال بروی مقادیر عددی و رشته‌ای، به راحتی می‌توان داده‌ها را نرمال کرده و شباهت‌ها را بر اساس ضرب وزن‌ها در مقادیر هر ویژگی بدست آورد. ٤) در مرحله پیش بینی امتیاز مورد نظر برای یک فیلم توسط کاربر جدیدالورد می‌توان معیارهای دیگری مثل وزن‌دهی فیلم‌ها بر اساس مجموع امتیازات دریافت شده، یا محصولاتی که بیشترین امتیازات را دریافت کرده‌اند و غیره را بجای محاسبه شباهت کاربر جدید با امتیازات اطلاق شده هر کاربر همسایه به فیلم آیتم‌ها بکار برد.

٦- مراجع

[1] Hoang, L. HU-FCF++ "A Novel Hybrid Method for the New User Cold-Start Problem in Recommender", *Systems Engineering Applications of Artificial Intelligence*, 207–222, 2015.

[2] Pera, M., & Ng, Y.-K, "A Group Recommender for Movies Based on Content Similarity and Popularity", *Information Processing and Management*, 673–687, 2013.

[3] Joon Ho Jang, "Movie Recommendation System - based on Matrix Factorization with Trust data",4, 1157-1179, 2017.

[4] RahulKatarya, Om PrakashVerma," An effective collaborative movie recommender system with cuckoo search", *Egyptian Informatics Journal*, Pages 105-112, 2017.

[5] Shardanand, U., & Maes, P. "Social Information Filtering Algorithms for Automating", *Word of Mouth*, 210–217, 1995.

[6] Baeza-Yates, R., & Ribeiro-Neto, B. "Modern Information Retrieval", *New York*, 509-513, 1999.

[7] Heung-Nam, K., El-Saddik, A., & Geun-Sik, J. "Collaborative error-reflected models for cold-start recommender systems", *Decision Support Systems*, 519–531, 2011.

[8] Bobadilla, J., Ortega, F., Hernando, A., & Bernal, J." A collaborative filtering approach to mitigate the new user cold start problem", *Knowledge-Based Systems*, 225–238, 2012.

[9] Byström, H, " Movie Recommendations from User Ratings", *Stanford University*, 1-3, 2013.

[10] Lika, B., Kolomvatsos, K., & Hadjiefthymiades, S," Facing the cold start problem in recommender systems", *Expert Systems with Applications*, 2065–2073, 2014.

[11] Luiz, A., Pereira, V., & Hruschka, E. R, "Simultaneous co-clustering and learning to address the cold start problem in recommender systems", *Knowledge-Based Systems*, 11–19, 2015.

[12] JianWei, JianhuaHe, KaiChen, YiZhou, ZuoyinTang, "Collaborative filtering and deep learning based recommendation system for cold start items", *Expert Systems with Applications*, Volume 69, Pages 29-39, 2017.

[13] Iosif Viktoratosa, Athanasios Tsadirasa, Nick Bassiliadesb, "Combining Community-Based Knowledge with Association Rule Mining to Alleviate the Cold Start Problem in Context-Aware Recommender Systems", *Expert Systems with Applications*, Available online 31 January 2018.

[14] kartik Narendra Jain, Vikrant Kumar, Praveen Kumar, Tanupriya Choudhury, "Movie Recommendation System": *Hybrid Information Filtering System*, pp 677-686, 20 January 2018.

[15] movielens, *movielens*. Retrieved 8 12, 2015, from grouplens: http://grouplens.org/datasets/movielens/2013.

[16] Zhou, X., He, J., & Huang, G. " SVD-based incremental approaches for recommender systems", *Journal of Computer and System Sciences*, 717–733, 2015.