

$FNHSM_{HRS}$: سیستم توصیه‌گر ترکیبی مبتنی بر خوشه‌بندی فازی و معیار شباهت اکتشافی

مصطفی خلجی، چیترا دادخواه

دانشکده مهندسی کامپیوتر، دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران، ایران

چکیده

امروزه سیستم‌های توصیه‌گر به یک سیستم فراگیر برای راهنمایی و هدایت کاربران در حجم عظیمی از داده در اینترنت، تبدیل شده است. پالایش همکارانه که پیشنهادات به کاربر فعال را براساس امتیازدهی مجموعه‌ای از کاربران ارائه می‌دهد، یکی از مدل‌های ساده و قابل درک و موفق برای پیدا کردن افراد هم‌سلیقه در سیستم‌های توصیه‌گر است. در این مدل، با افزایش تعداد کاربران و اقلام، سیستم دچار مشکل مقیاس پذیری می‌شود. از طرفی دیگر بهبود عملکرد سیستم در مواقعی که اطلاعات کمی از امتیازات در دسترس داریم، امری مهم است. در این مقاله یک سیستم توصیه‌گر ترکیبی به نام HRS_FNHSM که مبتنی بر معیار شباهت اکتشافی (NHSM) به همراه خوشه‌بندی فازی است، ارائه شده است. استفاده از روش خوشه‌بندی فازی در سیستم پیشنهادی باعث بهبود پذیری مساله مقیاس پذیری گذشته و دقت پیشنهادات سیستم را افزایش می‌دهد. سیستم پیشنهادی مبتنی بر مدل پالایش همکارانه بوده و با استفاده از معیار شباهت اکتشافی، عملکرد و صحت سیستم را ارتقا می‌دهد. ارزیابی نتایج سیستم پیشنهادی بر روی مجموعه داده MovieLens صورت گرفته، نتایج ارزیابی با استفاده از معیارهای MAE ، $Recall$ و $Precision$ ، $Accuracy$ بیانگر بهبود کارایی سیستم و افزایش دقت پیشنهادات نسبت به روش‌های پالایش همکارانه‌ای که از معیارهای دیگری برای پیدا کردن شباهت استفاده می‌نمایند، می‌باشد.

Mathematics Subject Classification (2010): 62A86, Email: Khalaji@email.kntu.ac.ir.

عبارات و کلمات کلیدی: سیستم‌های توصیه‌گر، پالایش همکارانه، خوشه‌بندی فازی، معیار شباهت اکتشافی.

۱۳۹۸ (انجمن سیستم‌های فازی ایران)

۱ مقدمه

با افزایش اطلاعات در اینترنت و فضای مجازی و خریدهای اینترنتی و تعاملات کاربران با یکدیگر، سیستم‌های توصیه‌گر^۱ جهت هدایت کاربران به سمت سلیق یا نیازهایی که دارند، در بیست سال اخیر و به ویژه در دهه‌ی اول قرن بیست و یکم مورد مطالعه قرار گرفته‌اند و پژوهش‌های بسیاری در این زمینه انجام شده است [۱]. سیستم‌های توصیه‌گر براساس نحوه پیشنهاد دهی به مدل‌های مختلفی دسته بندی می‌شوند که یکی از مهم‌ترین آنها مبتنی بر پالایش همکارانه^۲ است. عملکرد این روش به این طریق است که براساس شباهت بین کاربران یا اقلام، پیش بینی و پیشنهاد دهی به کاربر فعال را انجام می‌دهد. از این رو شامل تکنیک‌های مبتنی بر حافظه و مبتنی بر مدل است [۶]. مدل مبتنی بر حافظه در ابتدا شباهت بین کاربران را محاسبه و پس از آن کاربران مشابه به کاربر فعال را به عنوان کاربران همسایه انتخاب کرده و در پایان پیشنهادات را براساس این کاربران به کاربر فعال ارائه می‌کند. پالایش همکارانه مبتنی بر مدل در ابتدا یک مدلی از رفتار کاربران را ایجاد می‌کند و پس از آن امتیازات اقلام مشاهده نشده را براساس آن مدل پیش بینی می‌نماید. شروع سرد^۳ و مقیاس پذیری^۴ مشکلاتی در سیستم‌های توصیه‌گر هستند. هنگامی که سیستم توصیه‌گر با کمبود اطلاعات (امتیازات) از سوی سوابق کاربر مواجه می‌شود، مشکل شروع سرد از نوع کاربر جدید در سیستم بوجود می‌آید و در صورتیکه قلم جدیدی وارد سیستم شود از آنجایی که کاربران سیستم این قلم را قبلاً مشاهده ننموده‌اند، مشکل شروع سرد برای قلم جدید رخ خواهد داد. از سوی دیگر با افزایش چشمگیر تعداد کاربران در فضای مجازی، مشکل مقیاس پذیری برای سیستم‌های توصیه‌گر رخ می‌دهد که می‌توان گفت سیستم عملکرد و دقت خود را تا حدی از دست خواهد دهد. سیستم توصیه‌گر از نوع پالایش همکارانه مدل‌های بسیار ساده و قابل فهم و به راحتی قابل پیاده سازی می‌باشند. در این مقاله تمرکز اصلی سیستم توصیه‌گر پیشنهادی بر روی عملکرد سیستم با استفاده از ترکیبی از روش‌های مبتنی بر حافظه و مبتنی بر مدل جهت رفع مشکل مقیاس پذیری است. هسته اصلی این سیستم

¹Recommender Systems

²Collaborative Filtering

³Cold-Start

⁴Scalability

در ابتدا از روش خوشه‌بندی فازی و الگوریتم‌های پالایش همکارانه مشتق شده است که از معیار شباهت اکتشافی که در مرجع [۱۲] با نام FNHSM^۵ معرفی شده است، جهت تعیین کاربران همسایه با کاربر فعال استفاده می‌نماید. استفاده از روش‌های خوشه‌بندی در این نوع سیستم‌های توصیه‌گر باعث می‌شود تا کاربرانی که از نظر سلیقه به یکدیگر دارای شباهت‌هایی هستند در گروه‌های خاصی خوشه‌بندی شوند و پس از آن عملیات پیش بینی و پیشنهاد دهی اقلام در هر خوشه به صورت مجزا صورت گیرد و همین امر باعث بهبود عملکرد سیستم توصیه‌گر از نظر زمان و دقت می‌شود. ساختار مقاله به این صورت می‌باشد که در بخش دوم به مروری بر کارهای محققین می‌پردازیم. در بخش سوم سیستم توصیه‌گر پیشنهادی FNHSM_HRS^۶ معرفی می‌شود. بخش چهارم، بخش ارزیابی و نتایج پیاده سازی می‌باشد که در آن به نتایج آزمایشات و مقایسه با روش‌های دیگر پرداخته می‌شود. در نهایت در بخش آخر نیز نتیجه‌گیری ارائه می‌گردد.

۲ مروری بر کارهای محققین

سیستم‌های توصیه‌گر در ابتدا توسط گلدبرگ^۷ و همکارانش معرفی شدند [۷]. پس از آن چندین روش براساس نحوه پیشنهاد دهی ارائه شدند که یکی از آنها روش پالایش همکارانه است. پالایش همکارانه به عنوان یکی از محبوب‌ترین روش پیشنهاد دهی خصوصی سازی شده برای کاربران است که در بسیاری از حوزه‌ها مورد استفاده قرار می‌گیرد. این روش براساس یک سری معیارهای شباهت و یکسری مدل‌های از پیش تعریف شده، عملیات خود را انجام می‌دهند. اگرچه این روش از مشکلاتی از قبیل شروع سرد، پراکندگی داده و مقیاس پذیری رنج می‌برد اما فهم و پیاده سازی آنها بسیار راحت است و از مدل‌های پایه در سیستم‌های توصیه‌گر می‌باشند. برای بهبود عملکرد سیستم، بسیاری از پژوهشگران انواع روش‌های معیار شباهت را معرفی کرده‌اند. این معیارها شامل معیار شباهت، Pearson، PIP Cosine، و غیره [۲] و [۳] است. بلاگین^۸ و همکارانش روش‌هایی را برای بهبود عملکرد سیستم‌های توصیه‌گر معرفی کردند که از روش‌های

^۵New Heuristic Similarity Measure

^۶Fuzzy NHSM_Hybrid Recommender System

^۷Goldberg

^۸Bellogin

وزن دار هرلوکر^۹ و مک لافلین^{۱۰} جهت تعیین کاربرانی که از نظر سلیقه به کاربر فعال نزدیک بودند را انتخاب می‌کرد [۴]. همچنین هایفنگ^{۱۱} و همکارانش یک معیار شباهت اکتشافی با نام NHSM ارائه کردند که سه جنبه مجاورت، اثر شدید و محبوبیت امتیازات کاربران را در هنگام انتخاب کاربران همسایه برای کاربر فعال مد نظر داشت. از سوی دیگر یک معیار ترکیب شده از Jaccard و Difference Squared Mean توسط بوبادیل^{۱۲} و همکارانش و همکارانش ارائه گردید [۵]. بکارگیری روش‌های خوشه‌بندی برای گروه بندی کاربران مشابه از نظر سلیقه کمک شایانی به حل مسائل مقیاس پذیری می‌کند. از این رو کوهی^{۱۳} و همکارش با بکارگیری خوشه‌بندی فازی و روش فازی‌زدایی، Max به صورت تخصیص کاربران به تمام خوشه‌ها با درجه عضویت متفاوت و استفاده از معیار شباهت Pearson برای پیدا کردن نزدیک ترین همسایه، نشان دادند که عملکرد سیستم آنها نسبت به استفاده از روش‌های K-Means و SOM بهبود بخشیده شده است [۱۰].

۳ سیستم FNHSM_HRS

در شکل شماره ۱ ساختار سیستم پیشنهادی FNHSM_HRS نشان داده شده است. سیستم FNHSM_HRS دارای دو بخش برون خط و برخط است که در بخش برون خط مدل سیستم براساس اطلاعات داخل ماتریس امتیاز دهی کاربر-قلم آموزش داده می‌شود. در بخش بر خط پیشنهادات سیستم براساس مدل بدست آمده در مرحله برون خط برای کاربر فعال ارائه می‌گردد

۱.۳ بخش برون خط

روش‌های سنتی سیستم‌های توصیه‌گر، تنها به پیدا کردن کاربرانی که با کاربر فعال هم سلیقه هستند به انواع روش‌های معیار شباهت، اکتفا می‌کردند که برای مواقعی که ماتریس امتیازات کاربران،

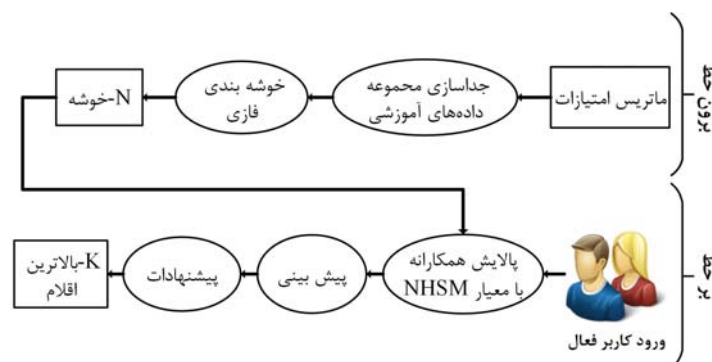
⁹ Herlocker's weighting

¹⁰ McLaughlin's weighting

¹¹ Haifeng

¹² Bobadila

¹³ Koohi



شکل ۱: سیستم پیشنهادی FNHSM_HRS

تنک است، کارا نمی‌باشد. از سوی دیگر اکثر این معیارها از مشکل پیچیدگی زمانی رنج می‌برند. برای حل این مشکلات، با خوشه‌بندی کردن کاربران در خوشه‌های متفاوت در سیستم پیشنهادی، باعث تسریع در روند عملکرد سیستم شده و بیانگر این مطلب می‌باشد که کاربران موجود در هر خوشه از نظر سلیقه باهم دارای شباهت‌هایی هستند. در این بخش ماتریس امتیازات به عنوان ورودی به سیستم توصیه‌گر داده شده است؛ ۸۰ درصد ماتریس امتیازات به عنوان مجموعه داده آموزشی انتخاب و با استفاده از روش خوشه‌بندی فازی کاربران سیستم به تعدادی خوشه تقسیم می‌شوند. در سیستم FNHSM_HRS با توجه به پیشنهادات محققین [۱۰]، ۳ خوشه در نظر گرفته شده است. الگوریتم خوشه‌بندی فازی با تخصیص درجه عضویت متفاوت به کاربران در هر خوشه، تاثیر کاربران در هر خوشه را مشخص می‌نماید. برای رتبه بندی درجه کاربران و انتخاب کاربران در مرحله پیش بینی سیستم از روش فازی زدایی درجات مبتنی بر مرکز ثقل (COG)^{۱۴} استفاده شد.

۲.۳ بخش بر خط

در این بخش، کاربر فعال که قرار است سیستم پیشنهاداتی در رابطه با فیلم‌های مشاهده نشده به او ارائه نماید، وارد سیستم شده و براساس سوابق خود کاربر فعال، عملیات محاسبات نزدیک ترین همسایه در خوشه کاربر فعال توسط معیار شباهت NHSM صورت می‌گیرد. این معیار دارای دو

¹⁴Center of Gravity

ضریب اصلی است که در رابطه ۱ مشخص شده است.

$$NHSM_Sim(u, v) = JPSS_Sim(u, v).URP_Sim(u, v) \quad (۱)$$

برای محاسبه معیار شباهت NHSM_Sim باید در ابتدا معیار شباهت JPSS_Sim محاسبه شود که خود نیز از دو معیار شباهت مشتق شده است که در روابط ۲ و ۳ ذکر شده است.

$$JPSS_Sim(u, v) = PSS_Sim(u, v).Jaccard_Sim(u, v)' \quad (۲)$$

$$Jaccard_Sim(u, v)' = \frac{|I_u \cap I_v|}{|I_u| \times |I_v|} \quad (۳)$$

$|I_u \cap I_v|$ بیانگر تعداد اقلام مشابه ای است که کاربر u و v مشاهده نموده اند و $|I_u|$ بیانگر تعداد اقلامی که کاربر فعال (u) امتیاز داده است و $|I_v|$ بیانگر تعداد اقلامی که کاربر دیگر (v) امتیاز داده است. معیار PSS_Sim از طریق رابطه ۴ بدست می آید.

(۴)

$$PSS_Sim(r_{u,p}, r_{v,p}) = Proximity(r_{u,p}, r_{v,p}).Significance(r_{u,p}, r_{v,p}).Singularity(r_{u,p}, r_{v,p})$$

Significance، Proximity و Singularity از طریق روابط ۵ و ۶ و ۷ بدست می آید.

$$Proximity(r_{u,p}, r_{v,p}) = 1 - \frac{1}{1 + \exp(-|r_{u,p} - r_{v,p}|)} \quad (۵)$$

$$Significance(r_{u,p}, r_{v,p}) = \frac{1}{1 + \exp(-|r_{u,p} - r_{med}| \cdot |r_{v,p} - r_{med}|)} \quad (6)$$

$$Singularity(r_{u,p}, r_{v,p}) = 1 - \frac{1}{1 + \exp(-|\frac{r_{u,p} + r_{v,p}}{2} - \mu_p|)} \quad (7)$$

$r_{u,p}$ امتیاز قلم p توسط کاربر فعال u و $r_{v,p}$ امتیاز قلم p توسط کاربر v است. r_{med} هم میانگین امتیاز بین امتیازاتی که کاربر موظف است برای امتیازدهی به یک قلم وارد کند، می‌باشد. ماتریس امتیازدهی در سیستم توصیه‌گر FNHSM_HRS دارای محدوده امتیازات بین عدد ۱ تا ۵ است که متوسط آن ۳ می‌باشد. μ_p هم میانگین امتیازات قلم p توسط کاربران است. آخرین مرحله از رابطه ۱، محاسبه معیار شباهت URP_Sim است که با استفاده از رابطه ۸ بدست می‌آید.

$$URP_Sim(r_{u,p}, r_{v,p}) = 1 - \frac{1}{1 + \exp(-|\mu_u - \mu_v| \cdot |\sigma_u - \sigma_v|)} \quad (8)$$

μ_u میانگین امتیازات و σ_u انحراف معیار کاربر فعال است که طبق رابطه ۹ بدست می‌آید.

$$\sigma_u = \sqrt{\sum_{p \in I_u} \frac{(r_{u,p} - \bar{r})^2}{I_u}} \quad (9)$$

در مرحله بعدی تعداد کاربران همسایه براساس بالاترین درجه شباهت کاربران همسایه به کاربر فعال در خوشه مربوطه، مشخص و انتخاب می‌شود. از این رو با توجه به رابطه ۱۰ عملیات پیش بینی امتیازات اقلام مشاهده نشده براساس پالایش همکارانه در سیستم، FNHSM_HRS، محاسبه می‌شود.

$$Predict(u, i) = \mu_u + \frac{\sum_{j=1}^m (r_{v_j, i} - \mu_v) \cdot NHSM_Sim(u, v_j)}{\sum_{j=1}^m |NHSM_Sim(u, v_j)|} \quad (10)$$

در اینجا، u کاربر فعال و i فیلمی است که سیستم FNHSM_HRS قرار است برای آن، امتیازی را پیش بینی نماید. سیستم FNHSM_HRS برای کلیه اقلام مشاهده نشده توسط کاربر فعال این امتیازات را محاسبه نموده و k قلم برتر را پیشنهاد می دهد. در رابطه 10 ، μ_u ، میانگین امتیازات کاربر فعال و m تعداد کاربران همسایه داخل خوشه‌ی کاربر فعال است. $NHSM_Sim(u, v_j)$ ، میزان شباهت بدست آمده کاربر فعال u با کاربر v_j است. $r_{v, i}$ ، امتیاز کاربر v به فیلم i است و μ_v ، میانگین امتیازات کاربر v است.

۴ ارزیابی سیستم FNHSM_HRS

داده‌های ورودی سیستم FNHSM_HRS مجموعه داده MovieLens شامل ۹۴۳ کاربر و ۱۶۸۲ فیلم با ۱۰۰ هزار تا امتیاز کاربر به فیلم‌ها، در نظر گرفته شده است [۱۱]. بازه امتیاز دهی در این مجموعه داده از ۱ تا ۵ است که به ترتیب یک نشان دهنده مورد پسند نبودن و عدد ۵ نشان دهنده مورد پسند بودن کاربر به یک فیلم خاص می باشد. برای ارزیابی عملکرد سیستم، از ۵-fold مجموعه داده‌ها به روش Cross-validation استفاده شده است که ۸۰ درصد داده‌ها برای آموزش و ایجاد مدل سیستم پیشنهادی و ۲۰ درصد داده‌ها برای آزمایش سیستم در نظر گرفته شده است. ارزیابی سیستم براساس معیارهای MAE، Precision Accuracy و Recall طبق رابطه‌های ۱۱ تا ۱۴ بر روی داده‌های آزمایش محاسبه شده است که در جدول شماره ۱ ماتریس اغتشاش مربوط را مشاهده می‌نمایید [۸].

جدول ۱: ماتریس اغتشاش [۸]

Predicted / Actual	Negative	Positive
Negative	A	B
Positive	C	D

$$MAE = \frac{\sum_{i=1}^n |\hat{r}_{u,i} - r_{u,i}|}{n} \quad (11)$$

$$Accuracy = \frac{CorrectRecommendation}{TotalPossibleRecommendation} = \frac{A + D}{A + B + C + D} \quad (12)$$

$$Precision = \frac{CorrectlyRecommendedItems}{TotalRecommendedItems} = \frac{D}{B + D} \quad (13)$$

$$Recall = \frac{CorrectlyRecommendedItems}{TotalUsefulRecommendations} = \frac{C}{C + D} \quad (14)$$

همانطور که قبلاً اشاره نموده‌ایم در سیستم FNHSM_HRS تعداد خوشه‌ها را ۳ و تعداد کاربران همسایه را ۵۰ در نظر گرفتیم. سیستم ۵ بار مستقل برای هر Top-N جداگانه اجرا و نتایج ذخیره گردیده است. در سیستم FNHSM_HRS از معیار شباهت اکتشافی استفاده شده است که در نتیجه با ۵ روش دیگری: روش Pearson، روش Cosine، روش وزن دارمک لافلین [۴]، روش وزن دار هرلوکر [۴] و روش RA^{۱۵} [۹] مقایسه و ارزیابی شده است که نتایج را در جداول شماره ۲ و شماره ۳ مشاهده می‌نمایید.

¹⁵Resource Allocation

جدول ۲: نتایج ارزیابی سیستم *FNHSM_HRS*

نام روش	معیار ارزیابی	۵ بالاترین اقلام	۱۰ بالاترین اقلام	۱۵ بالاترین اقلام	۲۰ بالاترین اقلام	۳۰ بالاترین اقلام
<i>FNHSM_HRS</i>	Accuracy	۲۰۳.۶۰	۳۷۳.۵۹	۵۶۲.۰۰	۵۵۷.۵۵	۱۵۵.۵۵
	Precision	۳۹۱.۹۱	۱۴۸.۹۱	۹۶۲.۹۰	۹۰۲.۹۰	۸۴۴.۹۰
	Recall	۶۴۳.۵۸	۰۲۹.۵۱	۴۷۵.۶۰	۴۸۶.۴۵	۳۷۹.۴۴
	MAE	۷۵۶.۰	۷۶۶.۰	۰.۷۷۸	۷۸۲.۰	۷۸۹.۰
<i>F_CF with Pearson</i>	Accuracy	۵۹.۵۳۷	۵۶.۷۵۶	۵۵.۷۸۴	۵۵.۱۶۰	۵۴.۷۶۲
	Precision	۹۰.۶۷۴	۹۰.۳۵۲	۹۰.۲۳۰	۹۰.۱۴۷	۹۰.۰۸۲
	Recall	۵۴.۵۳۷	۵۰.۰۹۰	۴۶.۷۹۰	۴۴.۸۳۴	۴۲.۷۶۸
	MAE	۰.۷۶۸	۰.۷۷۵	۰.۷۸۲	۰.۷۸۷	۰.۷۹۴
<i>F_CF with Cosine</i>	Accuracy	۶۰.۰۱۰	۵۷.۲۷۵	۵۶.۰۳۵	۵۵.۰۷۳	۵۳.۶۱۰
	Precision	۸۲.۰۲۰	۸۱.۳۷۱	۸۰.۹۳۵	۸۰.۶۸۳	۸۰.۴۴۱
	Recall	۵۰.۲۷۸	۵۰.۲۰۶	۴۷.۰۴۱	۴۵.۴۲۱	۴۴.۰۰۳
	MAE	۰.۸۲۷	۰.۸۲۸	۰.۸۲۹	۰.۸۳۴	۰.۸۳۹
<i>F_MW</i>	Accuracy	۵۹.۵۰۷	۵۷.۷۱۷	۵۶.۱۱۳	۵۵.۷۳۴	۵۵.۰۵۶
	Precision	۹۱.۰۹۸	۹۰.۸۵۶	۹۰.۶۹۱	۹۰.۶۱۷	۹۰.۵۵۰
	Recall	۵۸.۰۳۹	۵۱.۰۱۵	۴۷.۵۷۶	۴۵.۴۳۳	۴۳.۳۸۳
	MAE	۰.۷۵۸	۰.۷۶۸	۰.۷۷۸	۰.۷۸۳	۰.۷۹۱
<i>F_HW</i>	Accuracy	۵۸.۱۷۲	۵۷.۰۰۲	۵۵.۹۲۹	۵۵.۴۰۶	۵۴.۹۰۴
	Precision	۹۰.۸۸۱	۹۰.۵۶۲	۹۰.۵۰۰	۹۰.۴۲۹	۹۰.۳۶۴
	Recall	۵۸.۱۲۰	۵۰.۶۰۴	۴۷.۰۷۴	۴۵.۱۳۶	۴۳.۰۴۲
	MAE	۰.۷۷۷	۰.۷۸۵	۰.۷۹۱	۰.۷۹۶	۰.۸۰۳
<i>F_RA</i>	Accuracy	۵۹.۸۸۷	۵۸.۶۷۴	۴۸.۹۵۶	۵۴.۶۰۱	۵۴.۰۳۵
	Precision	۸۹.۶۵۰	۸۹.۳۵۵	۸۹.۲۰۹	۸۹.۱۵۰	۸۹.۰۸۸
	Recall	۵۷.۵۱۳	۵۰.۲۹۷	۵۵.۴۴۹	۴۴.۳۸۵	۴۴.۰۶۷
	MAE	۰.۷۶۰	۰.۷۶۸	۰.۷۷۰	۰.۷۸۳	۰.۷۹۱

جدول ۳: میانگین Top-N برای هر روش

نام روش	معیار ارزیابی	میانگین بالاترین n ها	نام روش	معیار ارزیابی	میانگین بالاترین n ها
<i>FNHSM_HRS</i>	Accuracy	۲۹۷۶.۵۷	<i>F_MW</i>	Accuracy	۵۶۸۲۵۴
	Precision	۰۴۹۴.۹۱		Precision	۹۰.۷۶۲۴
	Recall	۴۱۹۴.۴۹		Recall	۴۹.۰۸۹۲
	MAE	۷۷۴۲.۰		MAE	۰.۷۷۵۶
<i>F_CF with Pearson</i>	Accuracy	۵۶.۳۹۹۸	<i>F_HW</i>	Accuracy	۵۶.۲۸۲۶
	Precision	۹۰.۲۹۷		Precision	۹۰.۵۴۷۲
	Recall	۴۷.۸۰۳۸		Recall	۴۸.۸۱۵۲
	MAE	۰.۷۸۱۲		MAE	۰.۷۹۰۴
<i>F_CF with Cosine</i>	Accuracy	۵۶.۴۰۰۶	<i>F_RA</i>	Accuracy	۵۶.۶۹۷۲
	Precision	۸۱.۰۹		Precision	۸۹.۲۹۰۴
	Recall	۴۷.۳۸۹۸		Recall	۴۹.۱۶۳۲
	MAE	۰.۸۳۱۲		MAE	۰.۷۷۵۸

۵ نتیجه گیری

هدف اصلی سیستم‌های توصیه‌گر ارائه یک سری پیشنهادات براساس سلیقه کاربر و پیدا کردن کاربرانی که از نظر سلیقه با کاربر فعال دارای شباهت‌های زیادی است، می‌باشد. از این رو یکی از چالش‌های اصلی این سیستم‌ها دقت و عملکرد آنها با توجه به حجم زیاد اطلاعات (افزایش تعداد کاربران و اقلام سیستم) در کمترین زمان ممکن است. روش ارائه شده در سیستم FNHSM_HRS که حاصل ترکیب روش‌های خوشه‌بندی فازی و بکارگیری معیار شباهت NHSM است، کمک به سزایی به بهبود، Accuracy، Precision، Recall و MAE کرده است. نتایج آزمایشات سیستم پیشنهادی FNHSM_HRS به صورت میانگین، MAE برابر با ۰/۷۷۴۲، Accuracy برابر با ۵۷/۲۹۷۶، Precision برابر با ۹۱/۰۴۹۴ و Recall برابر با ۴۹/۴۱۹۴ شده است که در مقایسه با سایر معیارهای شباهت، بیانگر بهبود عملکرد سیستم و افزایش دقت می‌باشد. سیستم FNHSM_HRS مشکل مقیاس پذیری در سیستم‌های توصیه‌گر را به کمک روش خوشه‌بندی حل نموده است.

مراجع

- [1] Aggarwal, C. (2016) *Recommender systems*. Springer International Publishing.
- [2] Ahn, H. (2008) A new similarity measure for collaborative filtering to alleviate the new user cold-starting Problem, *Information Sciences*, 178.1, 37-51.
- [3] Adomavicius G. and Tuzhilin A. (2005) Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions, *IEEE Transactions on Knowledge Data Engineering*, 6, 734-749.
- [4] Bellogín, A., Castells, P. and Cantador, I. (2013) Improving memory-based collaborative filtering by neighbour selection based on user preference overlap, *Proceedings of the 10th Conference on Open Research Areas in Information Retrieval*.

- [5] Bobadilla, J., Ortega, F., Hernando, A. and Bernal, J. (2012) A collaborative filtering a roach to mitigate the new user cold start problem, *Knowledge-Based Systems*, 26, 225-238.
- [6] Cacheda, F., Carneiro, V., Fernández, D. and Formoso, V. (2011) Comparison of collaborative filtering algorithms: Limitations of current techniques and proposals for scalable, high-performance recommender systems, *ACM Transactions on the Web (TWEB)*, 5.1.
- [7] Goldberg, D., Nichols, D., Oki, B. and Terry, D. (1992) Using collaborative filtering to weave an information tapestry, *Communications of the ACM*, 35.12, 61-70.
- [8] Hahsler, M. (2011) Developing and testing top-n recommendation algorithms for 0-1 data using recommenderlab, NSF Industry University Cooperative Research Center for Net-Centric Software and System.
- [9] Javari, A., Gharibshah, J. and Jalili, M. (2014) Recommender systems based on collaborative filtering and resource allocation, *Social Network Analysis and Mining*, 4.1.
- [10] Koochi, H. and Kiani, K. (2016) User based Collaborative Filtering using fuzzy C-means, *Measurement*, 91, 134-139.
- [11] Konstan, J., Miller, B., Maltz, D., Herlocker, J., Gordon, L. and Riedl, J. (1997) GroupLens: a lying collaborative filtering to Usenet news, *Communications of the ACM*, 40.3, 77-87.
- [12] Liu, H., Hu, Z., Mian, A., Tian, H. and Zhu, X. (2014) A new user similarity model to improve the accuracy of collaborative filtering, *Knowledge-Based Systems*, 56, 156-166.