

بررسی میزان دقت سیستم‌های توصیه‌گر در ترکیب فیلترسازی مشارکتی و فیلترسازی مبتنی بر محتوی با روش سوئیچ

مهران صدرائیان فرد جهرمی^{۱*}، ایمان بادروح^۲

۱- دانشجوی کارشناسی ارشد مهندسی کامپیوتر، موسسه آموزش عالی اندیشه جهرم، ایران،

m.sadraian@yahoo.com

۲- عضو هیأت علمی دانشگاه آزاد اسلامی شیراز، خفر، ایران،

iman.badrooh@gmail.com

چکیده

با افزایش نقش اینترنت در زندگی مردم، شرکت‌های اینترنتی و تجارت الکترونیک به وجود آمده است. شرکت‌ها و فروشگاه‌های اینترنتی نسبت به فروشگاه‌های فیزیکی دارای تفاوت‌های زیادی هستند. یکی از این تفاوت‌ها وجود تعداد بسیار زیاد محصول در یک فروشگاه اینترنتی نسبت به شرکت‌های فیزیکی است. از طرف دیگر، این احتمال وجود دارد که بسیاری از کالاها که در یک فروشگاه اینترنتی وجود دارد توسط کاربر دیده نشود. بنابراین وجود سیستم‌های توصیه‌گر برای ارائه پیشنهاد به کاربران اینترنت ضروری است. بر همین اساس تحقیق‌های زیادی برای توسعه سیستم‌های توصیه‌گر صورت گرفته است و روش‌های گوناگونی مانند فیلترسازی مشارکتی مبتنی بر آیتم و کاربر و فیلترسازی مبتنی بر محتوی مورد توجه قرار گرفته است. هر کدام از این روش‌ها مزایا و معایبی دارد. در این مقاله تلاش شده است با ترکیب روش‌های فوق از مزایای آنها بهره برده و میزان دقت به منظور غلبه بر مشکلات شروع سرد و پراکندگی این روش بررسی گردد. در همین راستا یک سیستم توصیه‌گر ترکیبی پیشنهاد شده است که روی مجموعه داده‌های فیلم به کار گرفته شده است و نتایج قابل توجهی نیز بدست آمده و مشخص شده است که میزان دقت در این روش نسبت به روش‌های دیگر بهتر است.

کلمات کلیدی: تجارت الکترونیک، سیستم‌های توصیه‌گر، فیلترسازی ترکیبی، فیلترسازی مبتنی بر محتوی، فیلترسازی مشارکتی

۱- مقدمه

امروزه به سبب پیشرفت فناوری، اینترنت در همه‌جا در دسترس است. تلفن‌های همراه، کامپیوترها، گوشی و ساعت‌های هوشمند با قابلیت دسترسی به اینترنت با قیمت ارزان همه‌جا وجود دارد. افراد از این وسایل استفاده می‌کنند و بیشتر و بیشتر اطلاعات در اینترنت فراهم می‌کنند. بنابراین حجم بسیار زیاد داده وجود دارد که هر روز نیز رشد می‌کند. اطلاعات تنها از گوشی‌های همراه و کامپیوترهای شخصی به دست نمی‌آید، بلکه خریدهای افراد از دستگاه‌های پرداخت الکترونیکی و استفاده از کارت‌های اعتباری نیز حجم بسیار بالایی اطلاعات مفید تولید می‌کند. سایت‌های تجارت الکترونیک از سیستم توصیه‌گر برای دادن پیشنهادها موثر استفاده می‌کنند. برای نمونه تفلیکس و آمازون از داده‌کاوی برای پیشنهاد به مشتریان استفاده می‌کنند آنها محصولات جدید را که احتمالاً مشتریان به آن علاقه‌مند باشند را در کنار دیگر محصولات به کاربران پیشنهاد می‌دهند سیستم‌های توصیه‌گر هم چنین به کاربران کمک می‌کند با توجه به علائق هر فرد داده‌های موجود در اینترنت را برای خود فیلتر کنند. در سیستم‌های توصیه‌گر با مشکلات رایجی مانند شروع سرد و مشکل پراکندگی مواجه هستیم. این مشکلات از کارایی این روش‌ها کاسته و دقت اقلام پیشنهادی به کاربر را کاهش می‌دهند. این مقاله تکنیک‌های داده کاوی را با هدف ایجاد یک سیستم پیشنهاد دهنده ترکیبی و ارزیابی میزان دقت این سیستم با ترکیب فیلترهای مشارکتی و مبتنی بر محتوی مورد بررسی و تحقیق قرار می‌دهد. برای ساختن سیستم‌های توصیه‌گر، لازم است از الگوریتم‌های داده‌کاوی استفاده کنیم. بنابراین مسیر تحقیق شامل سه گام اساسی است:

- ۱- پیدا کردن الگو در داده
- ۲- توسعه یک مدل برای هدف
- ۳- ایجاد یک پیشنهاد دقیق و کارا

۲-پیشینه تحقیق

اولین سیستم که روش فیلترسازی مشارکتی را اجرا کرد، پروژه تپستری در [۱] سال ۱۹۹۲ بود. این پروژه اصطلاح فیلترسازی مشارکتی را ایجاد کرد. یکی از دیگر سیستم‌های اولیه یک سیستم توصیه‌گر موسیقی به نام رینگو [۲،۳] است که در سال ۱۹۹۴ پیشنهاد شده است. یکی دیگر از این سیستم‌ها برای رتبه‌بندی یوزنت در همان سال است. گروه لنز یکی از اولین سیستم‌های پیشنهاد مشارکتی فیلترکردن است که فیلم‌ها را توصیه می‌کند [۴]. مثال‌های دیگر آمازون است که کتاب‌ها را توصیه می‌کند.

شرکت نتفلیکس یک مسابقه را در سال ۲۰۰۶ برگزار کرد. آنها از شرکت‌کنندگان خواستند تا با استفاده از مجموعه داده خود، توصیه‌های بهتر فیلم را ایجاد کنند. مسابقه نتفلیکس منجر به پرش بزرگی در تحقق سیستم‌های توصیه‌گر شد، بیش از ۴۰،۰۰۰ تیم برای ایجاد یک الگوریتم خوب تلاش کردند. در نهایت، یک گروه بین‌المللی [۵،۶،۷] مسابقه را برد، چرا که رویکرد آنها زمان را بر خلاف مطالعات دیگر در نظر گرفت.

یکی دیگر از یافته‌های برجسته متعلق به کریشنان [۸] است که نتایج توصیه‌گر ماشین در برابر انسان را مقایسه می‌کند. با توجه به مطالعه کریشنان، ماشین در بسیاری از مقایسه‌ها برنده است، زیرا ماشین‌ها می‌توانند اطلاعات بیشتری را نسبت به انسان کنترل کنند.

۳-راهکار پیشنهادی

روش‌های ترکیبی، فیلترسازی مشارکتی^۱ و رویکردهای فیلترسازی مبتنی بر محتوی^۲ را برای بدست آوردن توصیه‌ها ترکیب می‌کنند. برای فیلترسازی ترکیبی، ما نیاز به استفاده از یک مجموعه‌داده یکسان برای هر دو الگوریتم‌های فیلترسازی مشارکتی و فیلترسازی مبتنی بر محتوی داریم، بنابراین مجموعه‌داده باید برای هر دوی این روش‌ها مناسب باشد. دو الگوریتم هنگامی که به تنهایی استفاده می‌شوند، مشکلاتی دارند. مشکل شروع سرد برای کاربر به گونه‌ای است که اگر یک آیتم به سیستم معرفی گردد، تا زمانی که کاربری آن را رتبه‌بندی نکند توصیه نمی‌شود. هم‌چنین اگر یک آیتم رتبه‌بندی کافی نداشته باشد، از این که توصیه نمی‌شود رنج می‌برد. [۹،۱۰]. رویکرد فیلترسازی مشارکتی مشکل رتبه‌بندی زود هنگام^۳ دارد. بدین معنا که کاربر اول در سیستم، آیتم‌ها را بدون دریافت هرگونه توصیه‌ای رتبه‌بندی می‌کند. بنابراین باید تصمیم بگیرد کدام آیتم‌ها بدون نگاه به رتبه‌بندی‌های گذشته این کاربر توصیه شوند. رویکردهای مبتنی بر محتوی برخی از معایب فیلترسازی مشارکتی مانند مسئله رتبه‌بندی زود هنگام را حل می‌کند. هنگامی که یک آیتم جدید اضافه می‌شود، شباهت می‌تواند با نگاه کردن به دیگر آیتم‌ها محاسبه شود. فیلترسازی مبتنی بر محتوی نیازمند رتبه‌بندی کاربران روی یک آیتم برای توصیه کردن آن نیست، بنابراین هر آیتمی بدون آن که به وسیله کاربری رتبه‌بندی شده باشد می‌تواند توصیه شود. این کار در روش فیلترسازی مشارکتی غیرممکن است. در فیلترسازی مبتنی بر محتوی مسئله شروع سرد برای آیتم‌ها حل شده است. رویکرد ترکیبی نه تنها تلاش می‌کند مشکلات را حل کند بلکه هم‌چنین تلاش می‌کند که دقت پیش‌بینی را افزایش دهد.

¹ Collaborative filtering (CF)

² Content-Based Filtering (CBF)

³ Early rating

۳-۱- روش ترکیب فیلترسازی مشارکتی و مبتنی بر محتوی

روش های مختلفی از جمله وزن دار^۱، ترکیب^۲، ترکیب ویژگی ها^۳، تقویت ویژگی ها^۴، آبخاری^۵، سوئیچ^۶ برای برای ترکیب فیلترها وجود دارد که در این مقاله از روش سوئیچ استفاده شده است.

سوئیچ: سیستم یکی از کامپوننت های توصیه گر را انتخاب می کند و آن را به کار می گیرد. رویکرد ترکیبی سوئیچ این گونه عمل می کند که ما الگوریتم را بعد از اجرای هر دو الگوریتم انتخاب می کنیم. ابتدا متد مبتنی بر آیتم به تنهایی روی مجموعه داده به کار گرفته می شود. دقت الگوریتم محاسبه می شود. متد مبتنی بر کاربر نیز به تنهایی روی مجموعه داده به کار گرفته می شود. بعد از آن ما می توانیم هر کدام که نتیجه بهتری داشتند انتخاب کنیم. برای یک فیلم، الگوریتم مقدار پیش بینی شده مبتنی بر کاربر (S1) و مقدار پیش بینی شده مبتنی بر آیتم (S2) را پیدا می کند. بعد از نرمال سازی هر دو امتیاز بین صفر تا ۵ است و دو عدد دهمی هستند.

اگر $P(S1) > P(S2)$ آنگاه $ScoreHybrid = S1$

اگر $P(S1) < P(S2)$ آنگاه $ScoreHybrid = S2$

رویکرد ترکیبی به شرح زیر کار می کند:

در هر لیست نتیجه ۱۰۰ فیلم وجود دارد به گونه ای که بیشتر فیلم ها مقدار رتبه بندی از S1 و S2 را دارند. اگر یک فیلم در لیست ۱۰۰ نتیجه روش مبتنی بر کاربر نباشد آنگاه مقدار S1 برابر با صفر است. به همین روش اگر یک فیلم در لیست ۱۰۰ تایی مبتنی بر آیتم نباشد مقدار S2 آن صفر خواهد شد، در روش سوئیچ اگر یک فیلم نتایج S1 و S2 داشته باشد آن را انتخاب می کند که درصد بالاتری دارد. در این روش آیتم های توصیه شده با نگاه کردن به دقت هزار کاربر از مجموعه داده MovieLens که به صورت تصادفی انتخاب شده است انتخاب می گردد. این مجموعه دارای ۶۰۴۰ کاربر و ۳۹۵۲ فیلم مختلف است.

۳-۲- الگوریتم پیشنهادی

خطای مطلق میانگین^۷ یک کمیت به کار گرفته شده برای میزان نزدیکی پیش بینی به خروجی پایانی است. انحراف مربع میانگین ریشه یا خطای ریشه میانگین مربع یک معیار پرتکرار به کار گرفته شده در محاسبه مقدار پیش بینی و مقدار واقعی نهایی است. در این مقاله MAE محاسبه گردید و متدهای پیشنهادی با دیگر متدها با تکنیک ارزیابی K-fold برشی مقایسه گردید.

¹ Weighted

² Mixed

³ Feature combination

⁴ Feature Augmentation

⁵ Cascade

⁶ Switching

⁷ MAE

```

1. Function HybridRecommender (Input: DataSet, Output: Recommendation)
2. {
3.   For Fold  $K = 1$  to 5 do
4.   {
5.     model ← Training data
6.     testModel ← Test Data
7.     Foreach User  $u$  in testModel do
8.     {
9.       ResultUserBased ← UserBasedCF (model,  $u.ID$ )
10.      CalculateMalidationMetrics (ResultUserBased, testModel)
11.      ResultItemBased ← ItemBasedCF (model,  $u.ID$ )
12.      Normalize(ResultItemBased)
13.      CalculateMalidationMetrics (ResultItemBased, testModel)
14.      MergeResult ← merge (ResultItemBased, ResultUserBased)
15.      Foreach movie  $m$  in MergeResult do
16.      {
17.        If precision(ResultItemBased) > Precision(ResultUserBased)
18.        CFResult ← precision(ResultItemBased)
19.        Else
20.        CFResult ← precision(ResultUserBased)
21.      }
22.      Sort(CfResult)
23.      HybridResult ← CBF(CFResult)
24.      Normalize(HybridResult)
25.      Sort(HybridResult)
26.      CalculateValidationMetrics (HybridResult, testModel)
27.    }
28.  }
29. }

```

۴- نتایج آزمایش ها و تفسیر آن ها

۴-۱- روش آزمایش

ابتدا آزمایش هایی برای فیلترسازی مشارکتی مبتنی بر کاربر انجام شده که در آن از معیارهای شباهت همبستگی پیرسون و الگوریتم نزدیک ترین N کاربر برای فیلترسازی مشارکتی مبتنی بر کاربر استفاده شده است. سپس آزمایش هایی با استفاده از فیلترسازی مشارکتی مبتنی بر آیتم انجام شده که در این فیلترسازی از معیارهای شباهت ضریب تانیموتو استفاده شده است. در مرحله سوم از ترکیب مبتنی بر آیتم و مبتنی بر کاربر به عنوان فیلترسازی مشارکتی با روش ترکیب سوئیچ بهره گرفته شده و در پایان میانگین رتبه بندی مبتنی بر ژانر به عنوان روش فیلترسازی مبتنی بر محتوی در راستای تولید لیست مرتبط تر با کاربر در نظر گرفته شده است.

¹ KNN

سومین همایش ملی مهندسی کامپیوتر، داده کاوی و داده های حجیم

الگوریتم ترکیبی مورد نظر روی مجموعه داده های MovieLens ارزیابی شده است و نتایج با بخش بندی K-fold مورد ارزیابی قرار گرفته است. دقت از طریق رابطه ۱ محاسبه محاسبه می شود:

$$Precision = \frac{|relevant\ Movies| \cap |retrieved\ Movies|}{|retrieved\ Movies|} \quad (1)$$

۴-۲- مجموعه داده MovieLens

مجموعه داده ای که برای بررسی روش پیشنهادی انتخاب شده است MovieLens است [۱۱]. این مجموعه از ۱۰۰۰۲۰۹ رتبه بندی بی نام از ۳۹۵۲ فیلم تولید شده توسط ۶۰۴۰ کاربر MovieLens تولید شده است. این داده ها برای سیستم توصیه گر فیلم بسیار مناسب است.

کد کاربری :: کد فیلم :: رتبه بندی :: مهر زمانی

- کد کاربری بین ۱ تا ۶۰۴۰ است.
- MovieId بین ۱ و ۳۹۵۲ است.
- رتبه بندی بر مبنای مقیاس ۵ ستاره است.
- مهر زمانی ۱ بر اساس ثانیه است.
- هر کاربر حداقل ۲۰ رتبه بندی انجام داده است.

۴-۳- نتایج روی مجموعه داده MovieLens

۴-۳-۱- نتایج فیلترسازی مشارکتی مبتنی بر کاربر

۱۰ کاربر با پروفایل های مشابه مشخص شده اند و بررسی شد چه فیلم هایی را دوست دارند. اگر ۸ نفر از ۱۰ نفر فیلم جدیدی را دوست داشته باشند به احتمال زیاد کاربر ۱ نیز این فیلم را دوست خواهد داشت. بر طبق فیلم هایی که کاربر ۱ مشاهده کرده اند، می توان فیلم ها را در ژانرهایشان گروه بندی کرد به گونه ای که امکان محاسبه میانگین امتیاز برای هر ژانر وجود داشته باشد که در جدول ۴-۱ نشان داده شده است.

جدول ۴-۱: میانگین امتیاز کاربر ۱ برای هر ژانر

ژانر	تعداد فیلم ها	میانگین امتیاز
درام	۲۱	۴,۴۲
کودکان	۲۰	۴,۲۵
انیمیشن	۱۸	۴,۱۱
موزیکال	۱۴	۴,۲۸
کمدی	۱۴	۴,۱۴
عاشقانه	۶	۳,۶۶
اکشن	۵	۴,۲۰
ماجراجویی	۵	۴
فانتزی	۳	۴
هیجانی	۳	۳,۶۶
علمی-تخیلی	۳	۴,۳۳
جنایی	۲	۴

¹ TimeStamp

سومین همایش ملی مهندسی کامپیوتر، داده کاوی و داده های حجیم

۵	۲	جنگی
---	---	------

بنابراین می توان گفت که این کاربر ژانرهای درام و انیمیشن را بیشتر از دیگر انواع فیلمها دوست دارد. واضح است که این کاربر توصیه هایی از ژانرهای درام، کودکان، انیمیشن، موزیکال و کمدی دریافت خواهد کرد.

مقدارهای شباهت ضریب پیرسون به اضافه الگوریتم N همسایه نزدیک کاربر رتبه بندی پیش بینی شده را برای مجموعه داده MovieLens بین صفر و ۵ تولید می کند. همان گونه که از جدول ۴-۲ قابل مشاهده است.

جدول ۴-۲ - فیلم های توصیه شده بر اساس ۱۰ همسایه نزدیک

۰: آیتم ۸۵۸ پدرخوانده سال ۱۹۷۲، اکشن، جنایی، درام مقدار: ۵
۱: آیتم ۲۸۵۸ زیبایی آمریکایی، کمدی، درام مقدار: ۵
۲: آیتم ۱۶۱۷: محرمانه لس آنجلس، جنایی، معمایی، هیجانی ۱۹۹۷ مقدار: ۵
آیتم ۵۸۷: روح کمدی، عاشقانه، هیجانی مقدار: ۵
آیتم ۲۸۶۳: حماسه توماس کراون، ۱۹۹۹، اکشن، هیجانی
آیتم ۳۵۷۸: گلا دیاتور، ۲۰۰۰، اکشن، درام مقدار: ۵
آیتم ۱۵۱۷، آستین پاورز ۱۹۹۷، کمدی، مقدار: ۵
آیتم ۱۴۱: قفس پرنده، ۱۹۹۶، کمدی مقدار: ۵
آیتم ۱۲۲۱، پدرخوانده قسمت دوم اکشن، جنایی، درام، ۱۹۷۴، مقدار: ۵
آیتم ۴۵۷: فراری، ۱۹۹۳، اکشن، هیجانی مقدار: ۵

۴-۳-۱ نتایج فیلترسازی مشارکتی مبتنی بر آیتم

تعداد ۱۰ نتیجه از توصیه های مبتنی بر آیتم روی مجموعه داده MovieLens برای کاربر ۱ در جدول ۴-۳ شرح داده شده است.

سومین همایش ملی مهندسی کامپیوتر، داده کاوی و داده های حجیم

جدول ۴-۳- نتایج توصیه مبتنی بر آیتم

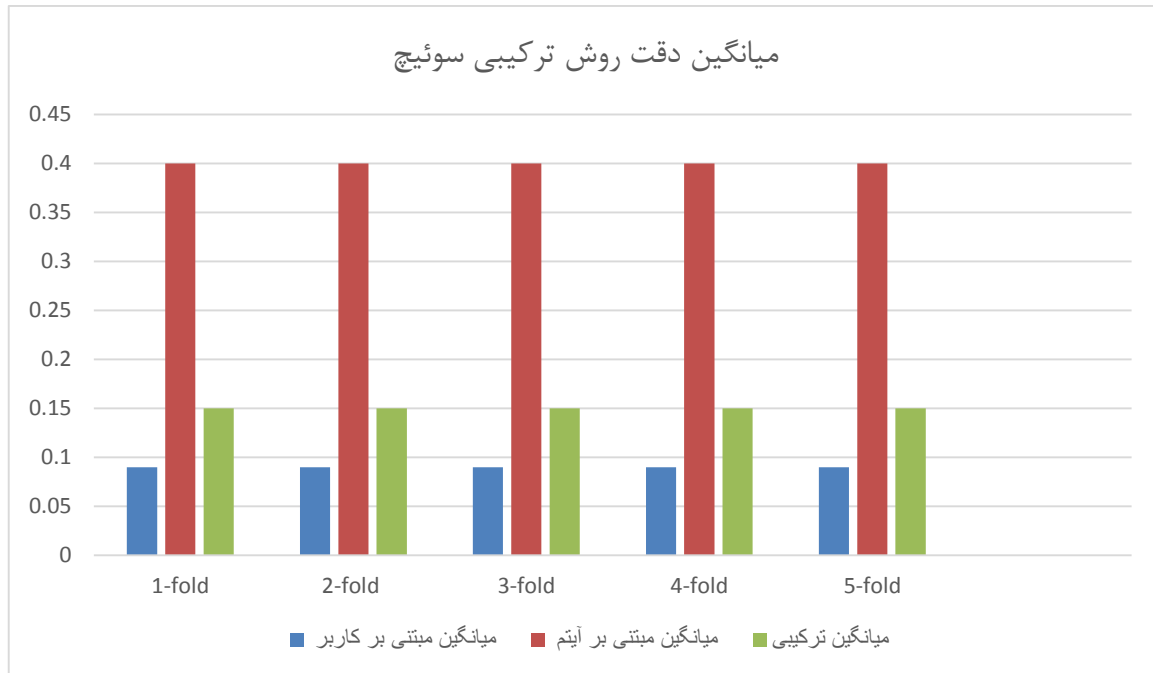
شاه شیر (۱۹۹۴) انیمیشن، کودکان، موزیکال	آیتم: ۳۶۴ مقدار: ۱۲,۹۹	۰
جنگ ستارگان بخش پنجم، ۱۹۸۰، اکشن، هیجانی، درام، جنگی	آیتم: ۱۱۹۶ مقدار: ۱۲,۳۸	۱
روز گراندهاگ، کمدی، عاشقانه ۱۹۹۳	آیتم: ۱۲۶۵ مقدار: ۱۲,۳۷	۲
Beetlejuice، کمدی، فانتزی، ۱۹۹۸	آیتم: ۲۱۷۳ مقدار: ۱۲,۳۱	۳
پری دریایی کوچک، انیمیشن، کودکان، کمدی، موزیکال، عاشقانه ۱۹۸۹	آیتم: ۲۰۸۱ مقدار: ۱۲,۲۸	۴
شکارچیان شباح، کمدی، ترسناک ۱۹۸۴	آیتم: ۲۷۱۶ مقدار: ۱۲,۰۸	۵
ویلی ونکا و چوکو، هیجانی، کودکان، فانتزی، کمدی ۱۹۷۱	آیتم: ۱۰۷۳ مقدار: ۱۲,۰۴	۶
مهاجمان صندوق گمشده، اکشن، هیجانی ۱۹۸۱	آیتم: ۱۱۹۸ مقدار: ۱۲,۰۳	۷
رستگاری در شاونک، درام، ۱۹۹۴	آیتم: ۳۱۸ مقدار: ۱۲,۰۳	۸
وقتی هاری مری را ملاقات کرد، کمدی، عاشقانه ۱۹۸۹	آیتم: ۱۳۰۷ مقدار: ۱۲,۰۲	۹

همان گونه که از نتایج مشاهده می شود ژانر بیشتر فیلم های توصیه شده کمدی، انیمیشن و کودکان است که متناسب با پروفایل کاربر است. در بخش بعد نتایج دو بخش قبلی با یکدیگر ترکیب می شوند.

۴-۴ ارزیابی میزان دقت رویکرد ترکیبی روی مجموعه داده MovieLens

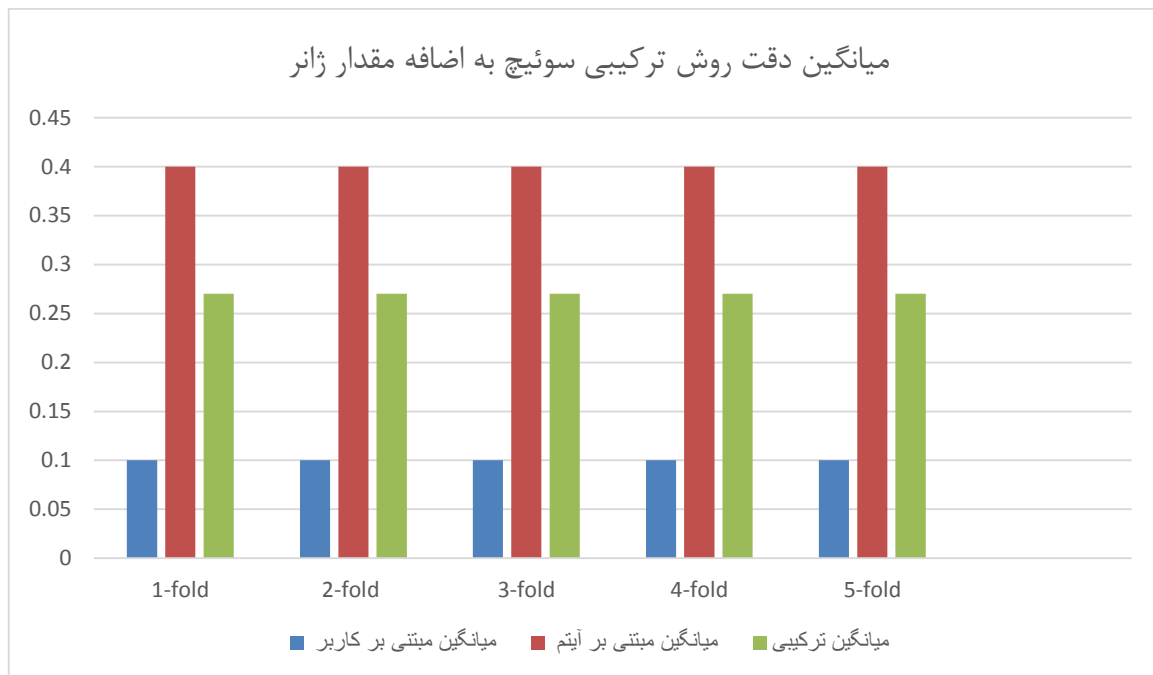
مجموعه داده MovieLens به ۵ مجموعه داده برابر تقسیم شده است. اگر یک کاربر ۱۰۰ رتبه بندی انجام داده باشد آنگاه ۲۰ رتبه بندی تصادفی در مجموعه داده آزمون و ۸۰ رتبه بندی دیگر در مجموعه داده آموزش قرار می گیرد. در هر بخش مجموعه داده آزمون با ۲۰ رتبه بندی تصادفی دیگر جایگزین می شود و در هر فاز مقدار دقت محاسبه می شود. از رویکرد سوئیچ بدون فیلترسازی مبتنی بر محتوی استفاده شده است. الگوریتم با دقت بهتر از بین روش مبتنی بر آیتم و روش مبتنی بر کاربر انتخاب می شود. برنامه برای هر کاربر به طور جداگانه انتخاب می کند. برای روش ترکیبی سوئیچ میانگین دقت، برای هر بخش در شکل ۴-۱ نشان داده شده است.

سومین همایش ملی مهندسی کامپیوتر، داده کاوی و داده های حجیم



شکل ۴- میانگین دقت روش ترکیبی سوئیچ بدون مقدار ژانر

در مرحله بعد آزمون، الگوریتم ترکیبی سوئیچ در فیلترسازی مشارکتی به همراه مقدار ژانر به عنوان فیلترسازی مبتنی بر محتوی مورد ارزیابی قرار گرفته است. در این روش الگوریتم با دقت بهتر از روش بدون استفاده از ژانر و دیگر روش های قبل انتخاب می شود، و نتایج آن در آن در شکل ۴-۲، نشان داده شده است.



شکل ۴-۲ میانگین دقت روش ترکیبی سوئیچ به اضافه مقدار ژانر

۵-پیشنهادهای و کارهای آینده

در این پژوهش تنها مجموعه داده MovieLens مورد ارزیابی قرار گرفته است و می توان همین پژوهش را مجموعه دادهها در زمینه های متفاوت مورد توجه قرار داد. بنابراین یکی از کارهایی که در آینده در ادامه این پژوهش پیشنهاد می شود بررسی الگوریتم های ترکیبی روی مجموعه دادهها در زمینه های به غیر از فیلم و یا مجموعه داده های در زمینه های دیگر است. همچنین به عنوان کارهای آینده می توان از دیگر روش های ترکیب دو فیلتر استفاده کرد و میزان فراخوانی و معیار F را نیز مورد ارزیابی قرار داد.

۶-مراجع

- 1- He, Chen, Denis Parra, and Katrien Verbert. "Interactive recommender systems: A survey of the state of the art and future research challenges and opportunities." *Expert Systems with Applications* 56 (2016): 9-27.
- 2- S. Upendra. Social information filtering for music recommendation. SM Thesis, Program in Media Arts and Sciences, Massachusetts Institute of Technology, 1994.
- 3- U. Shardanand and P. Maes. Social information filtering: algorithms for automating 'word of mouth'. Proceedings of the ACM Conference on Human Factors in Computing Systems, 1:210-217, 1995
- 4- P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, J. G. Riedl, and An. Open Architecture for Collaborative Filtering of Netnews. In in Proceedings of CSCW'94 (Chapel Hill NC, October, pages 175-186, 1994.
- 5- R. M. Bell, M. Jahrer, and A. Töscher. The BigChaos Solution to the Netflix Prize 2008. Netflix Prize, Report, pages 1-17, 2008.
- 6- Y. Koren. The BellKor Solution to the Netflix Grand Prize. Baseline,(August):1-10, 2009.
- 7- M. Piatte and M. Chabbert. The Pragmatic Theory solution to the Netflix Grand Prize. Working Paper, (August):1-92, 2009.
- 8- V. Krishnan, P. K. Narayanashetty, M. Nathan, R. T. Davies, and J. a. Konstan. Who predicts better? Results from an online study comparing humans and an online recommender system. In Proceedings of the 2008 ACM conference on Recommender systems - RecSys '08, volume October, page 211, 2008.
- 9- Agrawal, S., & Agrawal, J. (2015). Survey on anomaly detection using data mining techniques. *Procedia Computer Science*, 60, 708-713.
- 10- Bobadilla, Jesús, et al. "Recommender systems survey." *Knowledge-based systems* 46 (2013): 109-132.
- 11- F. Maxwell Harper and Joseph A. Konstan. 2015. The MovieLens Datasets: History and Context. *ACM Transactions on Interactive Intelligent Systems (TiS)* 5, 4, Article 19 (December 2015), 19 pages.