



## سیستم‌های پیشنهاد دهنده

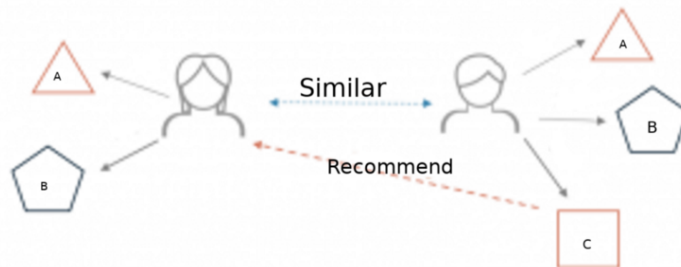
شقایق السادات جلالی  
مهدی لطفی  
دانشکده مهندسی کامپیوتر  
دانشگاه علم و صنعت ایران  
shaghayeghjalali96@gmail.com  
mahdilotfi.j444@gmail.com

### چکیده

در این گزارش می‌خواهیم انواع مختلف از سیستم‌های پیشنهاد دهنده را معرفی و سپس با ایجاد چندین مدل مختلف آن‌ها را با هم مقایسه کنیم.

### ۱ مقدمه

سیستم‌های پیشنهاد کننده با تحلیل رفتار کاربر، اقدام به پیشنهاد مناسب‌ترین اقلام (داده، اطلاعات، کالا و...) میکند. این سیستم کمک زیادی به مشکلات سیستم‌های دارای حجم زیاد اطلاعات می‌کند و به کاربرش کمک می‌کند تا در بین حجم زیاد اطلاعات سریع‌تر به هدفش نزدیک بشود. هدف از یک سیستم پیشنهاد کننده ارائه پیشنهادات شخصی به کاربران، بر اساس حجم زیادی از بازخوردهای زمانی که توصیف ترجیحات کاربران و ویژگی‌های اقلام مورد نظر آنها است میباشد. این به معنی است که می‌توان الگوریتمی آموزش داد که بتواند اقلام را طوری مرتب کند و یا کلیک را طوری پیش بینی کند که به سلیقه فرد نزدیک باشد. فرض کنید یک سایت شبیه GoodReads داریم که کاربران میتوانند در مورد کتاب‌های مختلف در سایت نظر بدهند. ما اطلاعاتی در مورد اینکه هر کاربر به چه کتابی چه امتیازی داده داریم، فرض کنید کاربر الف به کتاب های  $a$ ،  $b$ ،  $c$  رای بالا داده، و ما می‌خوایم پیش‌بینی کنیم این کاربر به کتاب  $d$  احتمالاً چه امتیازی میدهد، برای این کار با تحلیل داده‌های بقیه کاربرهایی که امتیازشون به  $a$ ،  $b$ ،  $c$  شبیه کاربر الف بوده میشود فهمید احتمالاً کاربر الف به کتاب  $d$  چه امتیازی میدهد.



شکل ۱: سیستم پیشنهاد دهنده

## ۱.۱ انواع سیستم پیشنهاد دهنده

شایع ترین نوع سیستم های پیشنهاد دهنده، سیستم های پیشنهاد دهنده مبتنی بر محتوا و همکاری هستند. در فیلترهای همکاری رفتار گروهی از کاربران برای ارائه پیشنهاد به سایر کاربران استفاده می شود. این پیشنهاد بر اساس عملکرد دیگر کاربران است. یک مثال ساده پیشنهاد یک فیلم به یک کاربر بر اساس این که دوستان آن ها فیلم را دوست داشتند. دو نوع مدل همکاری وجود دارد: روش های مبتنی بر حافظه و روش های مبتنی بر مدل. مزیت روش های مبتنی بر حافظه این است که پیاده سازی آن ها ساده است و می توان نتیجه را آسان توضیح داد.

### ۱.۱.۱ فیلترینگ مشارکتی

در روش فیلترینگ مشارکتی بر اساس شباهت رفتاری والگوهای عملکردی کاربرانی که شباهت های رفتاری و الگوهای مشابهی با کاربر فعلی در گذشته داشته اند، پیشنهادات ارائه می شود. شاید تعریف آن کمی پیچیده باشد ولی به طور ساده روش فیلترینگ همکارانه بر این فرض استوار است که کاربرانی که یک سری نظرهای مشابه درباره یک آیت (منظور از آیت، فیلم، عکس، موزیک یا هر چیز دیگری است که پیشنهاد می شود) دارند، درباره آیت های دیگر هم نظرات مشابه دارند.

### ۲.۱.۱ فیلترینگ مشارکتی مبتنی بر کاربر

در روش فیلترینگ مشارکتی مبتنی بر کاربر محصولات ی که به کاربر پیشنهاد می شود بر اساس این واقعیت است که محصولات توسط کاربرانی شبیه به کاربر خریداری شده اند. به عنوان مثال اگر علی و مهدی فیلم های مشابه ای را دوست داشته باشند و فیلم جدیدی می آید که علی آن را دوست دارد، پس می توانیم آن فیلم را به مهدی پیشنهاد کنیم زیرا علی و مهدی به نظر سلیقه یکسانی دارند.

### ۳.۱.۱ فیلترینگ مشارکتی مبتنی بر آیت

در روش فیلترینگ مشارکتی مبتنی بر آیت موارد مشابه را بر اساس رتبه بندی قبلی کاربران مشخص می کند. به عنوان مثال اگر کاربر A، B، C و D را به کتاب X و Y بدهند، زمانی که کاربر D کتاب Y را خریداری می کند سیستم پیشنهاد می کند کتاب X را خریداری کند، زیرا سیستم X و Y را به صورت مشابه بر اساس رتبه بندی کاربران A، B و C در نظر گرفته است.

### ۴.۱.۱ ماتریس عامل بندی

در روش ماتریس عامل بندی درحالی که روش های مبتنی بر کاربر و مبتنی بر آیت ساده هستند، معمولاً روش های ماتریس عامل بندی بیشتر تاثیرگذار هستند. دلیل آن این است که این روش ها به ما این امکان را می دهند که ویژگی های پنهانی که در بین فعل و انفعالات کاربران و آیت ها وجود دارد را کشف کنیم. از این روش برای پیش بینی امتیازها در فیلترینگ همکارانه استفاده می کنند. برای مثال دو کاربر ممکن است به آیت های خاصی امتیاز بالا بدهند و دلیل این کار ممکن است بخاطر بازیگر، کارگردان یا ژانر آن فیلم ها باشد. با تشخیص درست این ویژگی های پنهانی ما می توانیم امتیازها را بر اساس کاربر و آیت های خاص پیش بینی کنیم.

### ۵.۱.۱ مبتنی بر محتوا

در روش مبتنی بر محتوا از متا داده ها استفاده میکند مانند ژانر، تولید کننده بازیگر، موسیقیدان برای توصیف موارد فیلم ها یا موسیقی و ...

### ۶.۱.۱ روش های ترکیبی

در واقع روش های ترکیبی از روش های قبلی است که سعی کرده با ترکیب روش ها از مزیت آن روش ها استفاده کند و محدودیت های آن ها را پوشش دهد.

### ۲.۱ مسائل مربوط به سیستم های فیلترینگ مشارکتی:

- شروع کار: اگر سایت جدیدی را راه اندازی کنیم که خدمات انتخاب یک نمونه را انجام دهد، نمی توانیم پیشنهادات خود را در همان بدو شروع کار توسط کاربر به او ارائه دهیم. چراکه کاربران با تعداد محدودی از موارد ارتباط برقرار کردند.

• اضافه کردن کاربران یا موارد جدید به سیستم: اگر یک کاربر یا آیتم جدید اضافه شود، ما هیچ اطلاعات پیشینی در مورد آن‌ها نداریم، چون تعاملات موجود را نداریم. این مشکلات را می‌توان با درخواست از کاربران برای پاسخ به نوع دیگری از داده‌ها در زمان ثبت نام (جنسیت، سن، منافع و غیره) و استفاده از متا اطلاعات اقلام به منظور ارتباط آن‌ها با سایر موارد موجود در پایگاه داده حل کرد.

## ۲ کارهای مرتبط / پیش‌زمینه

در حال حاضر بسیاری از شرکت‌های مختلف که سایت‌های بزرگی دارند برای پیشروی در کارشان از سیستم‌های پیشنهادگر استفاده می‌کنند. با توجه به تفاوت در سلیقه بین افراد مختلف در سنین مختلف بی‌شک محصولی که یک کاربر انتخاب می‌کند با کاربر دیگر متفاوت بوده و قطعاً عملکرد سیستم‌های پیشنهادگر نیز باید متفاوت باشد. سیستم‌های پیشنهادگر تأثیر بسزایی در آمد شرکت‌های مختلف دارند و اگر درست استفاده شوند می‌توانند موجب سود دهی بسیار بالایی برای شرکت‌ها باشند به طور مثال شرکت Netflix اعلام کرده است که ۶۰ درصد از های DVD که توسط این شرکت به کاربران اجاره داده می‌شود از طریق سیستم‌های پیشنهادگر بوده و این سیستم‌ها توانستند تأثیر بسزایی در انتخاب مشتریان در انتخاب فیلم‌ها بگذارند. در جدول زیر لیست برخی از سایت‌هایی که از سیستم‌های پیشنهادگر استفاده می‌کنند به همراه کالایی که پیشنهاد می‌کنند را مشاهده می‌کنید:

site	recommended is what
Amazon	products Books/other
Facebook	Friends
WeFollow	Friends
Movielen	movies
Nanocrowd	movies
Jinni	movies
Findory	News
Digg	News
Zite	News
Meehive	News
Netflix	DVD
CDNOW	CDs/DVDs
eHarmony	Dates
Chemistry	Dates
True.com	Dates
Perfectmatch	Dates
careerBuilder	Jobs
Monster	Jobs
Pandora	Music
Mufin	Music
StumbleUpon	sites Web

## ۳ مدل پیشنهاد شده

با توجه به تعریف انواع مدل‌های سیستم‌های پیشنهاد دهنده در این پروژه تمامی مدل‌ها پیاده سازی شده‌اند، که در ادامه تماماً شرح می‌دهیم.

### ۱.۳ دادگان

در این قسمت بعد از ایجاد ماتریس عامل بندی چندین مدل بر روی دادگان اعمال می‌کنیم که هر کدام را مفصلاً شرح می‌دهیم اما به خاطر پراکندگی و حجم زیاد دیتاها ممکن است ایجاد ماتریس عامل بندی عملاً غیر ممکن شود که ما برای رفع این مشکل از دو دیتاست یکسان اما با حجم متفاوت استفاده می‌کنیم. در این پروژه از دادگان movielen استفاده شد که این دادگان خود به دو بخش ۲۰ میلیون و ۱۰۰ هزار تقسیم می‌شود. که در زیر شرح مفصلی از دادگان ۲۰ میلیونی آورده شده است مجموعه داده‌ها توصیف رأی‌ها و فعالیت‌های برجسته گذاری رایگان متن از MovieLens، یک سرویس پیشنهاد فیلم است. این مجموعه شامل ۲۰۰۰۲۶۳ رأی و ۲۷۲۷۸ فیلم است. این اطلاعات توسط ۱۳۸۴۹۳ کاربر در بین ژانویه ۰۹، ۱۹۹۵ تا ۳۱ مارس ۲۰۱۵ ایجاد شده است. این مجموعه داده در ۱۷ اکتبر سال ۲۰۱۶ منتشر شد. کاربران به صورت تصادفی انتخاب شدند. تمام

کاربران انتخاب شده در این برنامه حداقل ۲۰ فیلم را رتبه بندی کرده‌اند. اطلاعات جمعیتی موجود نیست. هر کاربر با شناسه نمایش داده می‌شود، و هیچ اطلاعات دیگری ارائه نمی‌شود. داده‌ها در شش فایل قرار دارند.

tag.csv که حاوی برچسب‌هایی است که برای فیلم‌ها توسط کاربر تخصیص می‌شود: movieId, timestamp, tag, userId  
rating.csv که شامل رتبه بندی فیلم‌ها توسط کاربران است: timestamp, rating, movieId, userId  
movie.csv که شامل اطلاعات فیلم است: genres, title, movieId  
link.csv که حاوی شناسه‌هایی است که می‌توانند برای پیوند به منابع دیگر استفاده شوند: tmbdId, imdbId, movieId  
genomescores.csv که حاوی اطلاعات مرتبط با برچسب فیلم است: relevance, tagId, movieId  
genometags.csv که شامل توضیحات برچسب است: tag, tagId

### ۲.۳ روش فیلتر همکاری مبتنی بر حافظه

مدل فیلتر همکاری محبوب‌ترین روش سیستم پیشنهاد دهنده می‌باشد که خود دو مدل مبتنی بر حافظه و مبتنی بر مدل دارد. در بخش‌های زیر مدل‌ها و نتایج را مشاهده می‌کنیم.

#### ۱.۲.۳ مبتنی بر آیتم

در این بخش با توجه به دادگان ماتریس عامل بندی را ساخته و سپس مقادیری که پوچ می‌باشد را صفر می‌کنیم. در ادامه با ساخت سیستم پیشنهاد دهنده بر اساس فیلم می‌توانیم با دادن یک فیلم نزدیک‌ترین فیلم‌ها به آن را دریافت کنیم. این الگوریتم در بین افرادی که فیلم مورد نظر را دیده‌اند به دنبال دیگر فیلم‌های مشابه می‌گردد، و ده تا از بهترین‌ها را نمایش می‌دهد. قابل ذکر است فیلم‌های با رای کمتر از ۱۰ حذف شده‌اند.

#### ۲.۲.۳ مبتنی بر کاربر

در این روش مانند روش قبلی ابتدا ماتریس عامل بندی را ساخته ولی مقادیری که پوچ می‌باشد را به جای صفر با میانگین رای هر کاربر پر می‌کنیم. حال همبستگی بین کاربران را حساب می‌کنیم و با دادن یک کاربر فیلم‌هایی را که تا به حال مشاهده نکرده جدا و برای هر فیلم یک امتیاز پیشبینی می‌کنیم هر چه امتیاز بالاتر احتمال دیده شدن بیشتر است.

### ۳.۳ روش فیلتر همکاری مبتنی بر مدل

در روش فیلتر همکاری مبتنی بر مدل از رویکرد یادگیری عمیق و ابزار کراس استفاده شده است. در حقیقت با استفاده از رویکرد یادگیری عمیق به دنبال پیشبینی کردن بهترین امتیاز که توسط کاربران به فیلم‌های مختلف داده می‌شود هستیم. در اینجا مدل ما خود یاد می‌گیرد که با استفاده از ویژگی‌های مختلف و ساخت ماتریس‌های امبدینگ بهترین پیشبینی را با توجه به انواع ترکیب‌های فیلم-کاربر انجام دهد. در نهایت پس از اینکه تمام اعداد ماتریس عامل بندی پیشبینی شد می‌توان با توجه به امتیازات بالایی که برای کاربران پیشبینی شده است فیلم‌های مناسب را پیشنهاد داد.

### ۴.۳ روش شباهت

این روش مانند روش فیلتر همکاری مبتنی بر آیتم می‌باشد ولی در این روش از الگوریتم نزدیک‌ترین فاصله استفاده می‌کند و ده تا از نزدیک‌ترین فیلم‌ها با استفاده از فاصله کوسینوسی نمایش داده می‌شود

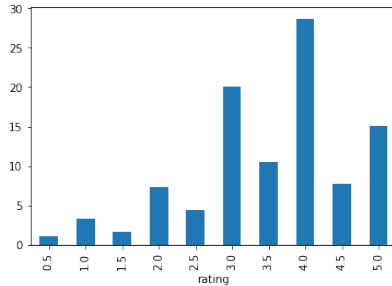
### ۵.۳ مدل مبتنی بر محتوا

در این قسمت اطلاعات ژانر فیلم‌ها را به رشته تبدیل کردیم سپس با استفاده از tf-idf این رشته‌ها را به بردارهایی تبدیل کرده تا بتوان با استفاده از محاسبات ریاضی میزان شباهت این بردارها را با یک دیگر بدست آوریم در نتیجه توانستیم سیستمی مبتنی بر محتوا تولید کنیم که با استفاده از ژانر فیلم‌ها پیشنهاد میدهد.

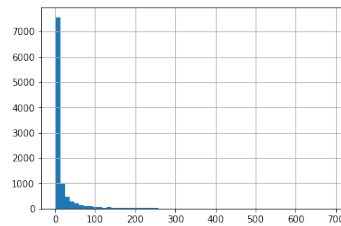
## ۴ تحلیل

### ۱.۴ تحلیل با ابزار داده کاوی

برای هر کار یادگیری ماشین از جمله یادگیری عمیق نیاز به شناخت داده ورودی مسئله می‌باشد که در اینجا چندین الگوریتم اجرا شده، برای مثال در شکل ۲ تعداد رای افراد به کل فیلم‌ها را مشاهده می‌کنید که اگر با دقت به آن نگاه کنیم متوجه می‌شویم بیشتر افراد به کمتر از ۵۰ فیلم رای داده‌اند و افرادی که عاشق فیلم هستند به مراتب کمتر هستند. دیگر رویکرد برای شناخت دادگان



شکل ۳: توزیع رای



شکل ۲: تعداد رای

مسئله که در شکل ۳ مشاهده می‌کنید نمایانگر توزیع رای کاربران می‌باشد همانطور که مشخص است رای ۴ بیشترین امتیاز را دارد.

### ۲.۴ تحلیل با میانگین خطای مربعات

معیار ارزیابی مدل‌های عمیق در این پروژه میانگین خطای مربعات می‌باشد که در جدول ۲ قابل مقایسه است.

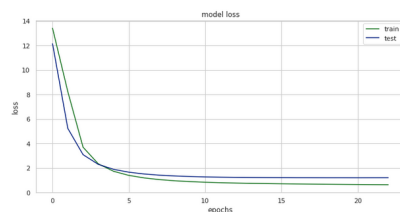
models	loss	val-loss
method deep simple	۰.۵۷۰۸	۱.۲۴۰۰
layer more with method deep	۰.۸۴۴۰	۰.۸۴۲۶
embedding in genres using method deep	۰.۶۲۵۸	۰.۶۸۰۷

همانطور که مشاهده می‌کنید مقدار خطای بر روی دادگان تست در مدل سوم که یک مدل ترکیبی از فیلتر همکاری و مبتنی بر محتوا می‌باشد از همه کمتر می‌باشد و با عمیق‌تر کردن شبکه خطای کمتری گرفتیم. با توجه به جدول بالا مدل پیشنهادی ما که تا حدی یک نوع سیستم هیبرید می‌باشد بهترین عملکرد را در پیشنهادات به ما دارد زیرا این مدل علاوه بر بررسی کاربران و فیلم‌ها ژانر فیلم را هم در ارائه پیشنهادات در نظر می‌گیرد.

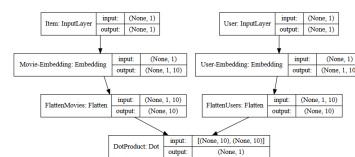
## ۵ نتایج

تمامی پیاده سازی‌ها در گیت هاب به نشانی زیر موجود است.

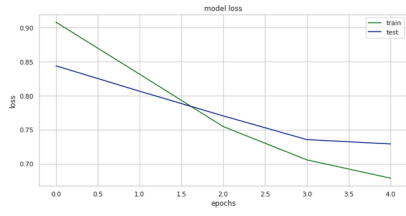
<https://github.com/shaghayeghjalali96/movie-recommender-system>



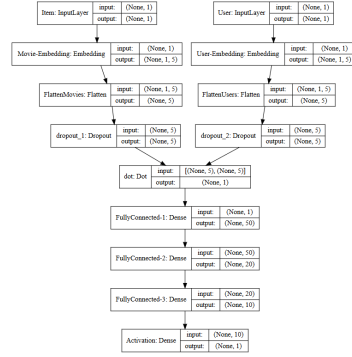
شکل ۵: دقت شبکه



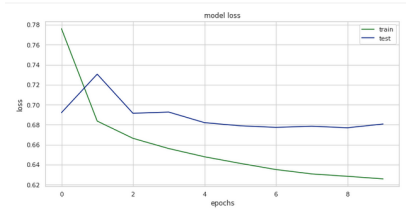
شکل ۴: مدل شبکه



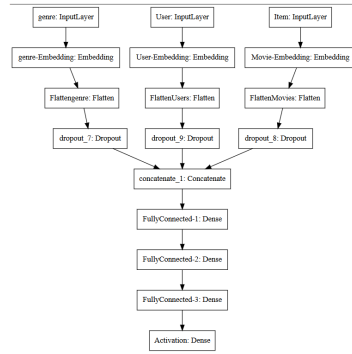
شکل ۷: دقت شبکه



شکل ۶: مدل شبکه



شکل ۹: دقت شبکه



شکل ۸: مدل شبکه

## منابع

- [1] Miguel Campo & J. J. Espinoza & Julie Rieger & Abhinav Taliyan *Collaborative Metric Learning Recommendation System: Application Theatrical Movie Releases*, CoRR, 2018.
- [2] Miller & Bradley N. & Albert, Istvan & Lam, Shyong K. & Konstan, Joseph A. & Riedl, John, *Experiences with an Occasionally Connected Recommender System*, 2003, ACM.
- [3] <http://infolab.stanford.edu/ullman/mmds/ch9.pdf>
- [4] Maral Azizi and Hyunsook Do. 2018. *A collaborative filtering recommender system for test case prioritization in web applications*. In Proceedings of the 33rd Annual ACM Symposium on Applied Computing (SAC '18). ACM, New York, NY, USA, 1560-1567. DOI: <https://doi.org/10.1145/3167132.3167299>