

## A Novel Movie Recommendation System with Iterated Truncated Singular Value Decomposition (ITSVD)

**Nozar Ebrahimi  
Lame** 

PhD student of Department of Management Information Systems, Qazvin Branch, Islamic Azad University, Qazvin, Iran

**Fatemeh Saghafi** \*

Associate professor of Faculty of Management, University of Tehran, Tehran, Iran.

**Majid Gholipour** 

Assistant Professor of Faculty of Electrical and Computer Engineering, Islamic Azad University, Qazvin Branch, Qazvin, Iran

### Abstract

Recommendation systems are one of the most essential tools for e-commerce intelligence. These systems with different types of data filtering methods are able to offer the best recommendations from a multitude of selectable items. Collaborative Filtering is the most widely used method of filtering data to make recommendations. One of the advanced models for predicting ratings in the Collaborative Filtering is the Singular Value Decomposing (SVD). In this paper, an optimized model of the film recommending system based on the SVD method is developed, which while reducing the dimensions of the matrices and the volume of computations and memory, and with iteration replacement method, has appropriate accuracy compared with other methods. For this research, a set of 100k Movie Lens datasets and Python programming have been used. Evaluation of error rate with root mean square error (RMSE) and mean absolute error (MAE) value shows a good improvement over similar methods in other references.

**Keywords:** Recommendation System, Collaborative Filtering, Singular Value Decomposition, Ratings Prediction.

\* Corresponding Author: : Fatemeh Saghafi, fsaghafi@ut.ac.ir

**How to Cite:** Ebrahimi Lame, N., Saghafi, F., Gholipour, M. (2022). A Novel Movie Recommendation System with Iterated Truncated Singular Value Decomposition (ITSVD), *Journal of Business Intelligence Management Studies*, 10(38), 173-199.



## سیستم توصیه گر فیلم ابتکاری با استفاده از روش تجزیه ماتریس به مقادیر منفرد با ابعاد کاهش یافته تکرار شده

دانشجوی دکتری گروه مدیریت سیستم‌های اطلاعاتی، واحد قزوین، دانشگاه آزاد اسلامی، قزوین، ایران.

نوذز ابراهیمی لامع 

دانشیار گروه مدیریت صنعتی، دانشکده مدیریت، دانشگاه تهران، تهران، ایران.

فاطمه ثقفی  \*

استادیار دانشکده برق و رایانه، دانشگاه آزاد اسلامی واحد قزوین، قزوین، ایران.

مجید قلی پور 

### چکیده

سیستم‌های توصیه گر یکی از ضروری ترین ابزارهای هوشمندسازی تجارت الکترونیک است. این سیستم‌ها با انواع مختلف روش‌های فیلتر کردن داده‌ها و داده‌کاوی، قادر به انتخاب و ارائه بهترین پیشنهادات از بین انبوه موارد قابل انتخاب برای مشتریان هستند. در بین روش‌های متنوع سیستم‌های توصیه گر، فیلترهای اشتراکی پرکاربردترین روش برای ارائه پیشنهادات است. فیلترهای اشتراکی دامنه وسیعی از الگوریتم‌ها را شامل می‌شود و در این بین، روش تجزیه ماتریس به مقادیر منفرد یکی از مدل‌های پیشرفته در فیلتر اشتراکی است. در این مقاله به ارائه مدلی بهینه شده از سیستم توصیه گر فیلم مبتنی بر روش تجزیه مقادیر منفرد پرداخته شده که ضمن کاهش ابعاد ماتریس و کاهش حجم محاسبات و حافظه، با روش تکرار جاگذاری، دارای دقت مناسب نسبت به روش تجزیه ماتریس به مقادیر منفرد ساده و سایر روش‌های دیگر است. برای این پژوهش از مجموعه دیتاست‌های ۱۰۰ هزار امتیازی مووی لنز و از برنامه نویسی پایتون

مقاله حاضر برگرفته از رساله دکتری رشته مدیریت فناوری اطلاعات گرایش کسب و کار هوشمند دانشگاه آزاد واحد قزوین است.

\* نویسنده مسئول: Fatemeh Saghafi, fsaghafi@ut.ac.ir

سیستم توصیه گر فیلم ابتکاری با استفاده از روش تجزیه ماتریس به ...؛ ابراهیمی لامع و همکاران | ۱۷۵

استفاده شده است. ارزیابی میزان خطا با روش های جذر میانگین مربعات خطا و میانگین قدر مطلق خطا، نشان از بهبود مناسب نسبت به روش های مشابه در مراجع دیگر دارد.

**کلیدواژه ها:** سیستم های توصیه گر، فیلتر اشتراکی، تجزیه مقادیر منفرد، پیش بینی امتیازات.



## مقدمه

در دهه‌های اخیر، با توسعه وبسایت‌ها و شبکه‌های اینترنتی، تجارت الکترونیک نیز توسعه فراوانی یافته و تبدیل به بخش مهمی از تجارت جهانی شده است (محمدی و همکاران، ۱۳۹۹). تجارت الکترونیک فروشندگان را قادر می‌سازد تا حجم انبوهی از کالا و خدمات را در وبسایت‌های خود به مشتریان ارائه نمایند، ولی انتخاب کالا و خدمات مناسب از بین حجم زیادی از موارد عرضه شده، نیازمند سیستم‌های نرم‌افزاری است که بر مبنای مشخصات و سلیقه مشتریان، قادر به پیشنهاد بهترین موارد باشد. در اینصورت در وقت مشتریان صرفه جویی شده و میزان فروش نیز افزایش خواهد یافت. این سیستم‌ها را سیستم‌های توصیه‌گر می‌نامند. سیستم‌های توصیه‌گر، نرم‌افزارهایی ضروری برای فیلتر کردن داده‌ها و ارائه پیشنهادات متناسب با سلیقه و علائق مشتریان هستند. این سیستم‌ها برای اولین بار در اوایل دهه ۱۹۹۰ میلادی مورد استفاده قرار گرفتند (Guan et al., 2014; Rahul et al., 2021).

وظیفه اصلی سیستم‌های توصیه‌گر، ارائه پیشنهادات متناسب برای هر شخص و بصورت شخصی می‌باشد. به عبارت دیگر، وبسایت‌ها برای هر مشتری محصولات مشخصی را نمایش می‌دهند (وبترین شخصی). این روش ارائه پیشنهادات را شخصی سازی پیشنهادات می‌نامند. شخصی سازی<sup>۱</sup> پیشنهادات باعث افزایش جذب و وفاداری مشتریان و همچنین افزایش فروش وبسایتها می‌گردد (Zhang et al., 2019; Alamdari et al., 2020; Guan et al., 2019; Lekakos & Caravelas, 2008; Rghuwanshi & Pateriya, 2019; al., 2017). سایت‌های معتبری مانند آمازون<sup>۲</sup>، ای بی<sup>۳</sup>، نتفلیکس<sup>۴</sup>، علی بابا<sup>۵</sup> در سطح بین‌المللی و سایت‌هایی مانند دیجی کالا، دیوار، باما، نماوا و فیلمو در سطح کشور از سیستم‌های توصیه‌گر بصورت گسترده‌ای استفاده می‌نمایند.

- 
1. Personalization
  2. Amazon
  3. eBay
  4. Netflix
  5. Alibaba

برای تولید سیستم‌های توصیه گر، لازم است از مجموعه داده‌های موجود در مورد مشخصات مشتریان، مشخصات کالاها، سابقه کالاهای خریداری شده قبلی و همچنین نظر و امتیازات کاربران به کالاها استفاده نمود. این امر به کمک روش‌های مختلف پالایش داده و داده کاوی روی داده‌های کلان<sup>۱</sup> انجام است (Aggarwal, 2016).

پرکاربردترین روش پالایش داده‌ها در سیستم‌های توصیه گر، روش فیلتر اشتراکی<sup>۲</sup> است. در این روش از ماتریس امتیازات برای توصیه آیت‌ها به استفاده می‌شود. فیلترهای اشتراکی دو نوع می‌باشند. فیلتر اشتراکی حافظه محور<sup>۳</sup>، که با استفاده از امتیازات کاربران و تشابه بین کاربران یا آیت‌ها با یکدیگر، کار می‌کنند و سیستم مدل محور<sup>۴</sup>، که بر اساس مدل ارائه شده نسبت به پیش‌بینی امتیازات کاربران با استفاده از روش‌های متنوع یادگیری ماشین کار می‌کنند (Guan et al., 2017؛ Ahuja et al., 2019؛ Aggarwal, 2016). یکی دیگر از روش‌های پالایش داده‌ها، روش محتوا محور است. در این روش از خصوصیات آیت‌های مورد علاقه کاربر استفاده می‌شود تا آیت‌های مشابه به او پیشنهاد شود (Guan et al., 2017). با توجه به اینکه سیستم‌های توصیه گر با داده‌های کلان سر و کار دارند و از طرفی حجم امتیازات در مقابل تعداد کاربران و آیت‌ها بسیار اندک است (ماتریس امتیازات خالی<sup>۵</sup>)، بنابراین تهیه سیستم توصیه گر با دقت مناسب و غلبه بر خالی بودن ماتریس امتیازات، از مشکلات اصلی روش فیلتر اشتراکی است. در سیستم‌های توصیه گر فیلتر اشتراکی، استفاده از روش‌های خوشه‌بندی مانند نزدیکترین همسایه ها<sup>۶</sup>، روش‌های یادگیری ماشین<sup>۷</sup> و تجزیه ماتریس<sup>۸</sup>، جهت افزایش دقت پیشنهادات و کاهش حجم محاسبات، از روش‌های پرکاربرد می‌باشند و در این بین، روش تجزیه ماتریس به

- 
1. Big Data
  2. Collaborative Filtering
  3. Memory Based
  4. Model Based
  5. Sparse Ratings Matrix
  6. K Nearest Neighbors (KNN)
  7. Machine Learning Methods
  8. Matrix Factorization (MF)

مقادیر منفرد<sup>۱</sup> نسبت به روش متداول نزدیکترین همسایه‌ها، دارای دقت بیشتری بوده و همچنین با روش‌های کاهش ابعاد ماتریس، امکان کاهش حجم محاسبات نیز وجود دارد (Falk, 2019؛ Aggarwal, 2016). روش سوم پالایش داده‌ها، استفاده از ترکیب دو روش فیلتر اشتراکی و فیلتر محتوا محور است که به آن فیلتر ترکیبی<sup>۲</sup> می‌گویند (Guan et al., 2017؛ Rahul et al., 2020).

طی سه دهه گذشته، روش‌های بسیار زیادی برای سیستم‌های توصیه‌گر توسعه یافته‌اند و یکی از اتفاقات تاثیرگذار در این زمینه، مسابقه یک میلیون دلاری شرکت تولید و توزیع فیلم نتفلیکس<sup>۳</sup> با هدف ارتقاء دقت پیشنهادات با روش فیلتر اشتراکی بود (سایت نتفلیکس<sup>۴</sup>).

امروزه سیستم‌های توصیه‌گر تقریباً در تمام شقوق تجارت الکترونیک جایگاه ارزشمندی دارند و در این بین، به جهت گستردگی دامنه مخاطبان و جنبه سرگرمی و تفریحی محصولات رسانه‌ای، استفاده از سیستم‌های توصیه‌گر جهت توصیه محصولات رسانه‌ای مانند کتاب، خبر، موزیک، سرگرمی، مجله و فیلم بسیار پر کاربرد است (Choi, et. al., 2012). به همین دلیل، در این پژوهش بر روی موضوع فیلم که دارای تعدد در فیلم‌ها و کاربران است، کار شده‌است. بررسی علمداری و همکاران<sup>۵</sup> (2020)، بر روی مقالات منتشر شده با جستجوی کلمات کلیدی "سیستم‌های توصیه‌گر"<sup>۶</sup> و "تجارت الکترونیک"<sup>۷</sup> توسط جستجوگر گوگل اسکولار<sup>۸</sup> نشان می‌دهد از بین ۳۵۸ مقاله یافت شده در این موضوعات، ۲۸۹ مورد آن (بیش از ۸۰٪) بین سال‌های ۲۰۰۸ تا ۲۰۱۹ انتشار یافته‌اند که نشانگر فعال بودن موضوع پژوهش است. البته برخلاف چنین سابقه‌ای در دنیا،

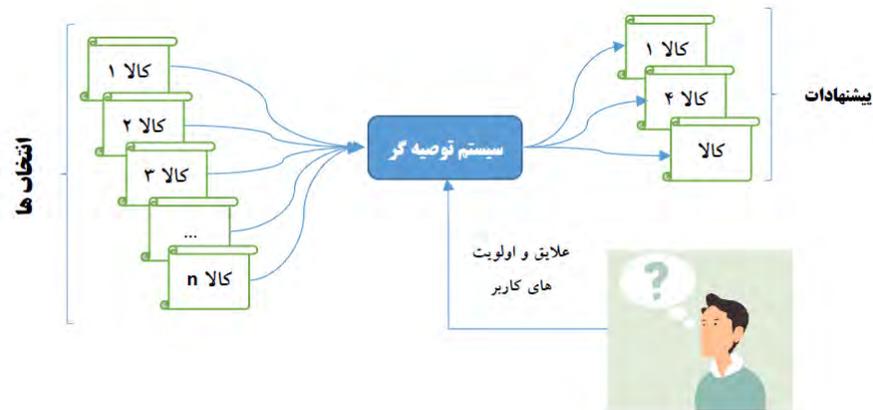
- 
1. Singular Value Decomposition (SVD)
  2. Hybrid Filtering
  3. Netflix
  4. Netflix Prize Site
  5. Alamdari et al.
  6. Recommendation Systems/Recommender System
  7. e-commerce
  8. Google Scholar

تعداد کارهای پژوهش هایی در زمینه سیستم های توصیه گر در داخل کشور محدود بوده و در زمینه سیستم های توصیه گر فیلم، نتایج جستجو در گوگل اسکولار، ایران داک<sup>۱</sup> و مگیران<sup>۲</sup> و با کلمات کلیدی "سیستم های توصیه گر فیلم" و "سیستم های پیشنهاد دهنده فیلم"، منجر به یافتن تعداد فقط پنج مقاله به زبان فارسی شد که به عنوان نمونه می توان به حیدری و پروین نیا (۱۳۹۶) اشاره نمود. البته شرکت های داخلی بسیار زیادی مانند دیجی کالا، نماوا و فیلمو از سیستم های توصیه گر استفاده می کنند ولی در خصوص روش های بکارگرفته شده و میزان دقت و کارائی آنها مطالب اندکی ارائه شده است. در این خصوص، می توان به سیستم توصیه گر شرکت فیلمو به عنوان یکی از کارهای ارزنده و به روز اشاره کرد که حاصل کار تیمی است (واهب، ۱۳۹۹).

در این مقاله مدل بهینه شده ای از سیستم توصیه گر فیلم با روش تجزیه ماتریس به مقادیر منفرد ارائه شده است که با استفاده از سه روش مختلف، برای حجم محاسبات و خطای امتیازات پیش بینی شده کاهش قابل ملاحظه ای حاصل شده است و خطای بدست آمده نسبت به موارد مشابه دارای دقت بهتری می باشد. در ادامه مقاله، به مرور ادبیات و پیشینه پژوهش، ارائه روش پژوهش، یافته های پژوهش و بحث در مورد نتایج پژوهش خواهیم پرداخت.

### پیشینه پژوهش

وظیفه اصلی سیستم های توصیه گر پالایش حجم انبوهی از داده ها و پیشنهاد محصولات متناسب با علائق و سلیقه کاربر است. شکل (۱) سیستم های توصیه گر را بصورت شماتیک نمایش می دهد.



شکل ۱. سیستم توصیه گر

در مجموع پنج نوع سیستم توصیه گر وجود دارد (Aggarwal, Alamdari, et al., 2020؛ Choi, et. al., 2012؛ 2016) که سه نوع آن متداول و پرکاربرد است. روش اول، سیستم توصیه گر فیلتر محتوا محور<sup>۱</sup> است که از تشابه بین ویژگی‌های آیتم‌ها برای ارائه پیشنهاد استفاده می‌کند. این روش سریع است ولی از چالش‌هایی مانند کمبود اطلاعات در مورد آیتم‌ها و دقت پایین رنج می‌برد. روش دوم، سیستم توصیه گر فیلتر اشتراکی است که با استفاده از ماتریس امتیازات، از تشابه بین کاربران و یا آیتم‌ها برای ارائه پیشنهادات استفاده می‌کند. چالش‌های اصلی روش‌های فیلتر اشتراکی، شروع سرد کاربر<sup>۲</sup> و ماتریس خالی می‌باشد. این روش، پرکاربردترین روش برای ارائه سیستم‌های توصیه گر است. روش‌های فیلتر اشتراکی خود به دو دسته فیلتر اشتراکی حافظه محور و فیلتر اشتراکی مدل محور تقسیم می‌شوند. روش سوم، سیستم توصیه گر ترکیبی است که برای استفاده از مزایای روش‌های مختلف و کاهش معایب آنها، از ترکیب روش‌های دیگر استفاده می‌کند. دو روش فیلتر جمعیتی محور<sup>۳</sup> و فیلتر دانش محور<sup>۴</sup> با توجه به کاربردهای خاص آنها، جزء

1. Content Based Filtering
2. Cold Start
3. Demographic Based Filtering
4. Knowledge Based Filtering

روش های متداول نمی باشند (Aggarwal, 2016). جدول (۱) اهداف مفهومی و ورودی های هر یک از روش های پایه را بصورت خلاصه نشان می دهد. آگاروال (2016) معتقد است اگر چه اصلی ترین هدف از بکارگیری سیستم های توصیه گر، افزایش فروش محصولات است و این هدف با ارایه پیشنهادات مناسب به کاربران و جلب توجه آنها میسر می شود، ولی برای رسیدن به هدف مذکور لازم است در طراحی سیستم های توصیه گر و ارایه پیشنهادات، موارد و اهداف فنی مهمی دیگری مانند مرتبط بودن<sup>۱</sup>، تازگی<sup>۲</sup>، پیشنهادات غیر منتظره<sup>۳</sup> و تنوع<sup>۴</sup> پیشنهادات نیز در نظر گرفته شود (ابراهیمی و همکاران، ۱۳۹۸).

جدول ۱. اهداف مفهومی انواع سیستم های توصیه گر پایه

نوع فیلتر	روش کار	ورودی اصلی	چالش های اصلی	مزیت ها
محتوا محور	تشابه بین آیتم ها	امتیاز کاربران و مشخصات آیتم ها	دقت پایین	سرعت و سادگی پیاده سازی
اشتراکی	تشابه بین کاربران	امتیازات کاربران	خالی بودن امتیازات و شروع سرد	دقت بالای پیشنهادات
ترکیبی	ترکیب روش های مختلف	اطلاعات روش های مختلف	زمان زیاد، هزینه بالا، امنیت	دقت بالا و امکان رفع چالش های گوناگون

نکته مهمی که در طراحی سیستم های توصیه گر باید در نظر گرفته شود اینست که هر یک از روش های پالایش داده ها، دارای چالش های مخصوص به خود می باشند. مهم ترین چالش هایی که معمولاً این سیستم ها با آنها مواجه می شوند عبارتند از: دشواری استخراج محتوا، شروع سرد، خالی بودن ماتریس امتیازات، مقیاس پذیری، حفظ حریم خصوصی و کلمات مترادف (Choi, et. al., Aggarwal, 2016؛ Rghuwanshi & Pateriya, 2019). از این رو لازم است روش های انتخابی برای ارایه پیشنهادات، قادر به برطرف

1. Relevance
2. Novelty
3. Serendipity
4. Diversity

نمودن چالش‌های مذکور باشند. همچنین نکته مهمی که در طراحی سیستم‌های توصیه‌گر باید در نظر داشت اینست که هیچ‌یک از روش‌ها به تنهایی قادر به برطرف نمودن همه چالش‌های مذکور نمی‌باشند. لذا برای انتخاب روش مناسب لازم است مواردی مانند مشخصات آیت‌ها و کاربران و اطلاعات موجود از آنها در نظر گرفته شود و با بررسی روش‌های مختلف نسبت به انتخاب روش بهینه با بیشترین دقت و کمترین هزینه از نظر حجم عملیات، حجم حافظه و میزان سخت افزار مورد نیاز اقدام شود (Alamdari, et al., 2020؛ Aggarwal, 2016).

### روش تجزیه ماتریس امتیازات به مقادیر منفرد

روش‌های بسیار زیادی در زمینه سیستم‌های توصیه‌گر فیلتر اشتراکی وجود دارد و همچنان در حال تکامل است. از روش‌های مرسوم در این زمینه می‌توان به روش محاسبه همبستگی بین کاربران با روش‌هایی مانند ضریب همبستگی پیرسون یا کسینوسی، استفاده از خوشه‌بندی همسایه‌ها<sup>۱</sup> و میانگین امتیازات آنها<sup>۲</sup>، استفاده از شبکه‌های عصبی و روش تجزیه ماتریس به حاصلضرب چند ماتریس نام برد. روش تجزیه ماتریس به عوامل نهانی<sup>۳</sup> یکی از الگوریتم‌های فیلتر اشتراکی است که کاربرد آن در سیستم‌های توصیه‌گر پس از مسابقه نتفلیکس مورد توجه قرار گرفت و با توجه به دقت مناسب آن در پیش‌بینی امتیازات و همچنین امکان کاهش ابعاد ماتریس‌ها، برای سیستم‌های توصیه‌گر کاربردهای فراوانی پیدا نمود. این روش، سابقه کاربرد در موارد دیگری را داشته ولی برای اولین بار توسط سیمون فانک<sup>۴</sup> برای سیستم‌های توصیه‌گر استفاده شد و مبنایی برای روش‌های بهبود یافته دیگر گردید (Falk, 2019).

بر مبنای آگاروال (2016)، فالک (2019) و آندریاس (2017)، در روش تجزیه ماتریس، یک ماتریس با ابعاد بزرگ به حاصلضرب چند ماتریس با ابعاد کوچکتر تبدیل

- 
1. K Nearest Neighbors (KNN)
  2. K\_Means
  3. Latent Factors
  4. Simon Funk

می شود. روشهای مختلفی برای تجزیه یک ماتریس وجود دارد که از بین آنها روش تجزیه ماتریس به سه ماتریس (تجزیه به مقادیر منفرد)، تجزیه به دو ماتریس<sup>۱</sup>، تجزیه و تحلیل عوامل<sup>۲</sup> و تجزیه و تحلیل عوامل اصلی<sup>۳</sup> جزئی روش های پر کاربرد می باشند (Andreas, 2017). در این پژوهش به جهت قابلیت های مهم روش تجزیه به مقادیر منفرد از جهت امکان کاهش ابعاد ماتریس های سه گانه و تقریب مناسب و جدید بودن، از این روش استفاده شده است.

### مبانی روش تجزیه ماتریس

اگر  $R$ ، ماتریس امتیازات کاربران به فیلم ها با ابعاد  $m \times n$  باشد که در آن  $m$  تعداد کاربران (سطرها) و  $n$  تعداد فیلم ها (ستون ها) باشد شکل (۲)، در اینصورت هدف از تجزیه ماتریس  $R$ ، پیدا کردن دو ماتریس  $U$  و  $V$  به گونه ای است که حاصل ضرب دو ماتریس در هم ماتریس  $R$  را با تقریب مناسبی بازسازی نماید. یعنی:

$$R_{m \times n} \approx U_{m \times k} \cdot V_{k \times n}^T \quad (1)$$

بنابراین برای بدست آوردن مقدار امتیاز کاربر  $i$  به آیت  $j$  می توان از رابطه زیر استفاده نمود:

$$\hat{r}_{ij} = u_i v_j^T = \sum_{k=1}^k u_{ik} v_{kj} \quad (2)$$

شکل (۲) نشان می دهد چگونه ماتریس امتیازات  $R$  به دو ماتریس  $U$  و  $V^T$  تجزیه شده است. لازم به ذکر است اگر ماتریس دارای مقادیر نامعلوم باشد، حاصل ضرب دو ماتریس تقریبی از ماتریس  $R$  را خواهد داد.

1. QR Decomposition
2. Factor Analysis (FA)
3. Principal Component Analysis (PCA)

$$\begin{matrix}
 \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{bmatrix} & = & \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1k} \\ u_{21} & u_{22} & \cdots & u_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ u_{m1} & u_{m2} & \cdots & u_{mk} \end{bmatrix} & \times & \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{k1} & v_{k2} & \cdots & v_{kn} \end{bmatrix}^T \\
 \mathbf{R} & & \mathbf{U} & & \mathbf{V}
 \end{matrix}$$

شکل ۲. تجزیه ماتریس به دو ماتریس

برای بدست آوردن مقادیر ماتریس های  $U$  و  $V$  ابتدا مقادیر اولیه تصادفی ای برای سلول-های آنها تعیین می شود و پس از حاصلضرب دو ماتریس، اختلاف بین مقادیر حاصل ضرب و مقادیر واقعی در ماتریس  $R$  محاسبه می شود. سپس با روش گرادیان کاهش<sup>۱</sup> نسبت به حداقل نمودن خطا اقدام می گردد. در روش گرادیان کاهش<sup>۱</sup> هدف حداقل نمودن مقدار خطا است. برای اینکار، لازم است مجذور تفاضل مقادیر واقعی و پیش بینی حداقل گردد. در اینحالت تابع گرادیان کاهش<sup>۱</sup> بصورت زیر خواهد شد:

$$\min J = \frac{1}{2} \left( \sum_{(i,j) \in R}^n (r_{i,j} - u_i v_j^T)^2 + \lambda (\|u_i\|^2 + \|v_j\|^2) \right) \quad (3)$$

در رابطه فوق:

$r_{i,j}$ : مقدار واقعی،  $u_i v_j^T$ : مقدار تقریبی،  $\|u_i\|^2$  و  $\|v_j\|^2$  مقدار اندازه بردارهای  $u$  و  $v$  و  $\lambda$  مقدار پارامتر تنظیم است.

از طرفی، برای ارزیابی میزان دقت حاصل از تجزیه ماتریس برای پیش بینی مقادیر نامعلوم و جلوگیری از بیش برآزش مدل، لازم است ماتریس امتیازات به دو ماتریس آموزش و تست با نسبت مناسب تقسیم شود. سپس، نسبت به تجزیه ماتریس آموزش و تعیین مقادیر نامعلوم اقدام می شود. در مرحله بعد، با استفاده از مقادیر تقریبی بدست آمده از تجزیه ماتریس امتیازات، مقادیر واقعی داده های تست و مقادیر تقریبی بدست آمده با

## 1. Decent Gradient

هم مقایسه می شوند و با استفاده از روش های ارزیابی RMSE و MAE، میزان خطای مدل محاسبه می شود. برای محاسبه خطاهای مذکور از روابط زیر استفاده می شود:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (r_i - p_i)^2}{n}} \quad (۴)$$

$$MAE = \frac{\sum_{i=1}^n |r_i - p_i|}{n} \quad (۵)$$

در روش تجزیه ماتریس به مقادیر منفرد که دارای کاربردهای فراوانی در کاهش ابعاد<sup>۱</sup>، فشرده سازی تصاویر<sup>۲</sup> و نادیده گرفتن داده های نویز<sup>۳</sup> دارد، یک ماتریس بجای دو ماتریس به سه ماتریس U، S و V بصورت زیر تجزیه می شود:

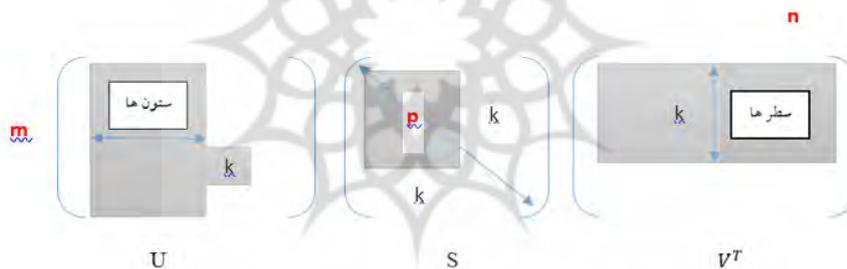
$$R_{m \times n} = U_{m \times p} * S_{p \times p} * V_{p \times n}^T \quad (۶)$$

که در آن ماتریس U، ماتریس متعاملی است که به آن ماتریس عوامل نهانی<sup>۴</sup> کاربر می گویند و خصوصیات کاربران را بصورت فشرده و ترکیبی نمایش می دهد. ماتریس S مقادیر منفرد ماتریس امتیازات R است و ماتریس V نیز ماتریس متعاملی است که به آن ماتریس عوامل نهانی آیتم ها می گویند و خصوصیات آیتم ها (محصولات) را بصورت ترکیبی و فشرده در خود دارد. نهانی بودن عوامل فوق به این معنی است که توسط مدل محاسبه و تعیین می شود و در تعیین آنها، انسان نقشی ندارد. همچنین لازم به ذکر است مقادیر عوامل نهانی اغلب قابل تفسیر و توضیح نمی باشند.

در این روش، هر دو ماتریس U و V متعامد می باشند و ماتریس S یا سیگما، ماتریس قطری است که مقادیر منفرد آن بصورت نزولی بر روی قطر اصلی ماتریس چیده شده و اندازه این مقادیر نشان می دهد هر یک از مقادیر حاوی چه مقدار اطلاعات می باشند. برای

- 
1. Dimensionality Reduction
  2. Image Compression
  3. Noising Data
  4. Latent Factors

کاهش همانطور که در معادله (۶) دیده می‌شود، مقدار بعد  $p$ ، هایپر پارامتر است و قابل تعیین است. لذا در تعیین این مقدار می‌توان مقدار بهینه‌ای را طوری انتخاب کرد که ابعاد سه ماتریس  $U$ ،  $S$  و  $V$  کاهش یافته و مقدار خطا نیز حداقل شود. برای اینکار لازم است میزان خطا به ازای مقادیر مختلف  $p$  محاسبه شود و مقدار بهینه انتخاب شود. در حقیقت میزان کاهش ابعاد ماتریس‌ها بر اساس تعداد مقادیر انتخاب شده از ماتریس  $S$  تعیین می‌شود. شکل (۳) این حالت را نمایش می‌دهد. (Falk, 2019؛ Aggarwal, 2016). در شکل (۳)، ماتریس  $R_{m \times n}$  به کمک روش تجزیه به مقادیر منفرد به حاصلضرب سه ماتریس  $U_{m \times p}$ ،  $S_{p \times p}$  و  $V_{p \times n}^T$  تبدیل می‌شود که در آن  $p$  تعداد مقادیر منفرد غیر صفر ماتریس  $R$  است. حال اگر تعداد  $k$  مقدار منفرد حفظ و بقیه برابر صفر قرار داده شود، در اینصورت ماتریس‌های فوق به صورت  $U_{m \times k}$ ،  $S_{k \times k}$  و  $V_{k \times n}^T$  کاهش ابعاد خواهند یافت که در آنها  $k \leq p$  است.



شکل ۳. نمایش نحوه کاهش ابعاد ماتریس با استفاده از روش SVD

### پیشینه پژوهش

هر چند مبانی ریاضی و کاربردهای روش تقریب ماتریس از طریق تجزیه به مقادیر منفرد از سالیان دور مورد استفاده بوده است، ولی استفاده از روش تجزیه مقادیر منفرد برای سیستم‌های توصیه‌گر، اولین بار توسط سیمون فانک و در جریان مسابقه شرکت نتفلیکس مورد توجه قرار گرفت. این روش در مسابقه نتفلیکس رتبه سوم را کسب کرد، ولی به جهت مزایای آن به سرعت به یکی از روش‌های طرفدار در سیستم‌های توصیه‌گر تبدیل

شد (Falk, 2019). در سال های بعد، پژوهش های متعددی در خصوص رایه راهکارهایی برای رایه سیستم های توصیه گر بر این مبنا توسعه داده شد که در زیر به برخی از آنها اشاره می شود. لی و همکاران<sup>۱</sup> (2020) در پژوهش خود روشی تحت عنوان "فیلتر کردن مبتنی بر دسته های فیلم- تجزیه مقادیر منفرد"<sup>۲</sup> را برای پیش بینی امتیازات کاربران فعال، به فیلم ها بر مبنای داده های صریح<sup>۳</sup> و ضمنی<sup>۴</sup> ناشی از دیتاست فیلم ها رایه می دهند. در ابتدا، روش فوق از عوامل وزن دار برای بررسی اثر انواع دسته بندی فیلم بر روی پیش بینی امتیازات کاربران و همچنین ارتقا روش پالایش بر مبنای دسته بندی فیلم ها استفاده می کند. در نهایت، روش فیلتر کردن مبتنی بر دسته های فیلم با الگوریتم تجزیه ماتریس به مقادیر منفرد<sup>۵</sup> ترکیب می شود تا از عملکرد خوب آن در فیلتر اشتراکی استفاده نماید. در مقایسه با انواع الگوریتم های مختلف موجود، روش پیشنهادی به میزان قابل توجهی دقت پیش بینی امتیازدهی را ارتقاء می بخشد و مقیاس پذیری و اثربخشی سیستم های توصیه گر شخصی را افزایش می دهد. ژو و همکاران<sup>۶</sup> (2015) یک الگوریتم تدریجی<sup>۷</sup> مبتنی بر تجزیه مقادیر منفرد با مقیاس پذیری مناسبی را پیشنهاد داده اند که ترکیبی از الگوریتم تجزیه مقادیر منفرد تدریجی و الگوریتم تقریب مقدار تجزیه مقادیر منفرد است. تجزیه و تحلیل دقیق خطا، اثربخشی عملکرد این روش را نشان می دهد. سپس آنها اقدام به مقایسه ای از دقت پیش بینی و زمان اجرا بین الگوریتم پیشنهادی و الگوریتم تجزیه مقادیر منفرد تدریجی بر روی مجموعه داده های مووی لنز و فلیکستر<sup>۸</sup> رایه داده اند و نتیجه گرفته اند که روش پیشنهادی از سایر روش های مشابه خود بهتر است. در مقاله بال و همکاران<sup>۹</sup> (2020) یک روش فیلتر ترکیبی از ترکیب فیلتر محتوا محور بر اساس همبستگی بین فیلم ها و روش

1. Li et al.
2. MCBF-SVD
3. Explicit
4. Implicit
5. Singular Value Decomposition
6. Zhou et al.
7. Incremental Algorithm
8. Flixster
9. Bahl et al.

اشتراکی مبتنی بر تجزیه به مقادیر منفرد را ارایه داده اند. هدف از ارایه روش مذکور بدست آوردن حداقل خطای ممکن برای پیش‌بینی امتیازات کاربران به فیلم‌ها است. نتایج بر روی دیتاست‌های ۱۰۰ هزار مووی لنز نشان از بهبود خطا نسبت به روش‌های دیگر از جمله تجزیه مقادیر منفرد تنها است. همچنین باوانا و همکاران<sup>۱</sup> (2019) با اعتقاد به اینکه روش تجزیه مقادیر منفرد یکی از روش‌های قدرتمند در تجزیه ماتریس است، در مقاله خود موضوع استفاده از این الگوریتم برای سیستم‌های توصیه‌گر بر مبنای توزیع محاسبات در واحد‌های پردازشگر موازی جهت تسریع در بدست آوردن نتایج را مورد بررسی قرار داده‌اند. پاترا و گانگولی<sup>۲</sup> (2019) روش‌های مختلف پیش‌پردازش برای اعمال به الگوریتم تجزیه مقادیر منفرد را با هم مورد بررسی قرار داده‌اند. برای بررسی از دیتاست‌های یک میلیون مووی لنز استفاده شده‌است. در ابتدا با استفاده از روش خوشه‌بندی نزدیکترین همسایه‌های کاربران، با روش‌های متفاوت فاصله‌یابی مانند ژاکارد و اقلیدسی مشخص شده و امتیازات بدست آمده از این روش را جایگزین مقادیر نامعلوم در ماتریس امتیازات نموده‌اند. نتایج بدست آمده نشان می‌دهد، وقتی که همسایه‌های کاربران با استفاده از روش فاصله اقلیدسی بدست آید و سپس با استفاده از روش میانگین امتیازات، متوسط امتیاز آنها در ماتریس امتیازات اعمال و از الگوریتم تجزیه مقادیر منفرد استفاده شود، نتایج بهتری حاصل خواهد شد. در مقاله راهول و همکاران<sup>۳</sup> (2021) برای سیستم توصیه‌گر فیلم از روش تجزیه مقادیر منفرد برای کاهش ابعاد و از روش متوسط امتیاز همسایگان برای خوشه‌بندی استفاده کرده‌اند. برای پیاده‌سازی مدل از دیتاست‌های مووی لنز و فلیکستر استفاده شده‌است. نتایج بدست آمده از روش مذکور را با روش‌های مختلف ارزیابی نموده و به این نتیجه رسیده‌اند که روش پیشنهادی آنها دارای عملکرد بهتری نسبت به سایر روش‌ها است. سعادت‌ی و همکاران<sup>۴</sup> (2019) در مقاله خود نسبت به ارایه نتایج حاصل از پیاده‌سازی چند مدل سیستم توصیه‌گر فیلم، فیلتر اشتراکی کاربر محور، فیلتر اشتراکی

1. Bhavana et al.
2. Patra & Ganguly
3. Rahul et al.
4. Saadati et al.

آیتم محور، سیستم محتوا محور، تجزیه مقادیر منفرد و شبکه عصبی برای پیش‌بینی امتیازات کاربران با استفاده از زبان برنامه نویسی پایتون اقدام کرده و نتایج حاصل را با یکدیگر مقایسه کرده‌اند. بررسی نتایج نشان می‌دهد، روش تجزیه مقادیر منفرد نسبت به سایر روش‌های فیلتر اشتراکی دارای خطای کمتری است.

جدول ۳. جمع بندی مقالات بررسی شده فیلتر اشتراکی با روش SVD

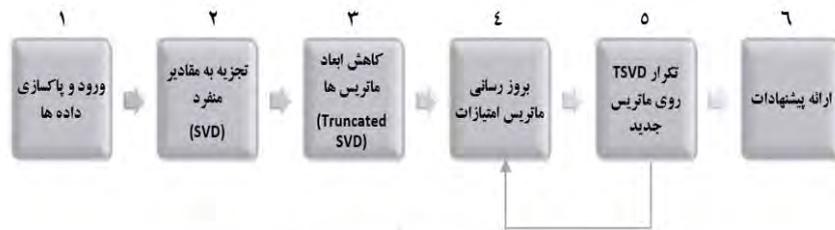
ردیف	نویسنده ها	چالش مورد بررسی	روش بکار رفته
۱	Rahul, et.al., 2021	عملکرد	Truncated SVD+ K_Means
۲	Bahl et al.,2020	خطای کمتر	CBF+SVD
۳	Li et al., 2020	کاهش خطا+ مقیاس پذیری	MCBF-SVD
۴	Bhavana et al., 2019	سرعت پردازش	SVD
۵	Patra& Ganguly,2019	پیش پردازش	SVD+ K_Means
۶	Saadati et al.,2019	خطای کمتر	SVD و سایر روش‌ها
۷	Zhou et al., 2015	مقیاس پذیری و عملکرد	Incremental SVD

### روش پژوهش

در این پژوهش، مدل بهینه شده ای از سیستم توصیه گر فیلم با استفاده از روش تجزیه به مقادیر منفرد و بر اساس مدل های ارایه شده توسط راهول و همکاران (2021) و آندریاس (2017) توسعه داده شده است. به عنوان داده‌های ورودی از نسخه ۱۰۰ هزار امتیازی دیتاست مووی لنز دانشگاه مینه سوتا با ۹۷۲۴ فیلم، ۶۱۰ کاربر و حدود ۱۰۰ هزار امتیاز کاربر به فیلم‌ها استفاده شده و با استفاده از زبان برنامه نویسی پایتون ویرایش ۳ مدل پیاده سازی شده است. مراحل مدل‌سازی و بهینه سازی آن در شکل (۴) نمایش داده شده است. مرحله ۱. ورود و پاکسازی داده‌ها: با توجه به اینکه بیش از ۹۸ درصد دیتاست امتیازات مورد استفاده خالی است، لذا برای اطمینان از صحت و دقت مدل لازم است ابتدا نسبت به پاکسازی و مرتب سازی داده‌ها اقدام شود. اهم اقدامات به شرح ذیل است:

- حذف سطر و ستون های اضافی
- حذف سطر و ستون های تکراری و خالی

- حذف کاربران با کمتر از ۲۰ امتیاز و فیلم‌های با کمتر از ۱۰ امتیاز
- تقسیم امتیازات به ماتریس آموزش و تست با نسبت ۸۰ به ۲۰
- پر کردن سلول‌های بدون مقدار با میانگین امتیاز هر فیلم
- نرمال کردن امتیازات با کم کردن میانگین امتیاز هر فیلم از امتیازات



شکل ۴. فرایند پیاده سازی مدل بهینه شده

مرحله ۲. تجزیه به مقادیر منفرد: نسبت به تجزیه ماتریس امتیازات آموزش اقدام و مقادیر ماتریس های  $U$ ،  $S$  و  $V$  را بدست می آوریم.

مرحله ۳. کاهش ابعاد ماتریس‌ها: در این مرحله برای مقادیر مختلف  $k$ ، میزان دقت مقادیر پیش‌بینی شده برای داده‌های تست را بدست می آوریم و بهترین مقدار آن را انتخاب می کنیم.

مراحل ۴ و ۵. مرحله بهینه سازی: در این مرحله با توجه به مقدار کاهش ابعاد بدست آمده در مرحله قبل، مقادیر بدست آمده برای داده‌های تست را در ماتریس کاهش ابعاد بدست آمده قرار می دهیم و با تجزیه مجدد ماتریس امتیازات به مقادیر منفرد، میزان خطای بهینه را بدست می آوریم.

مرحله ۶. ارائه پیشنهادات: در این مرحله با مرتب کردن امتیازات پیش‌بینی شده برای فیلم‌های دیده نشده برای کاربر فعال، نسبت به پیشنهاد تعدادی فیلم برتر اقدام می شود.

#### یافته‌ها

شکل (۵) دیتاست اصلی امتیازات، شکل (۶) و (۷) به ترتیب دیتاست آموزش و دیتاست

تست را پس از تفکیک به نسبت ۸۰ به ۲۰ نمایش می دهند:

'test'				'train'				'data'			
movieid	rating	userid		movieid	rating	userid		userid	movieid	rating	
50022	1682.0	3.5	322.0	2052	27831.0	3.5	18.0	0	1	1	4.0
25564	50872.0	4.0	177.0	11736	165.0	5.0	71.0	1	1	3	4.0
87281	55118.0	4.5	562.0	49865	318.0	4.5	319.0	2	1	6	4.0
2083	47200.0	3.0	18.0	51918	95.0	5.0	337.0	3	1	47	5.0
51265	7438.0	3.0	331.0	57780	56775.0	4.0	380.0	4	1	50	5.0
...	...	...	...	...	...	...	...	...	...	...	...
11962	96588.0	3.5	73.0	4788	1224.0	4.0	29.0	100831	610	166534	4.0
80284	71535.0	4.0	505.0	58227	2085.0	3.0	381.0	100832	610	168248	5.0
55967	5418.0	3.5	369.0	38745	1275.0	4.0	266.0	100833	610	168250	5.0
17707	94777.0	1.0	111.0	28875	6565.0	4.0	199.0	100834	610	168252	5.0
38039	2692.0	4.5	260.0	45346	95441.0	2.0	298.0	100835	610	170875	3.0

15863 rows × 3 columns      63451 rows × 3 columns      100836 rows × 3 columns

شکل ۷. دیتاست تست

شکل ۶. دیتاست آموزش

شکل ۵. دیتاست اصلی

با اعمال روش تجزیه مقادیر منفرد بر روی دیتاست آموزش، مقادیر پیش‌بینی شده برای دیتاست تست بصورت شکل (۸) بدست می آید. مقدار جذر میانگین مربعات خطا<sup>۱</sup> و میانگین قدر مطلق خطا<sup>۲</sup> به ترتیب عبارتند از ۰/۹۴۴۸۲ و ۰/۷۳۲۱۲. در شکل (۸) مقادیر پیش‌بینی با نماد RATING\_HAT و مقادیر واقعی با نماد RATING نمایش داده شده‌اند.

پژوهشگاه علوم انسانی و مطالعات فرهنگی  
پرتال جامع علوم انسانی

1. RMSE (Root Mean Square Error)
2. MAE (Mean Absolute Error)

movieId	rating	userId	rating_hat	
50022	1682.0	3.5	322.0	3.83
25564	50872.0	4.0	177.0	3.84
87281	55118.0	4.5	562.0	4.04
2083	47200.0	3.0	18.0	3.11
51265	7438.0	3.0	331.0	3.85
...	...	...	...	...
11962	96588.0	3.5	73.0	3.59
80284	71535.0	4.0	505.0	3.83
55967	5418.0	3.5	369.0	3.85
17707	94777.0	1.0	111.0	3.55
38039	2692.0	4.5	260.0	3.94

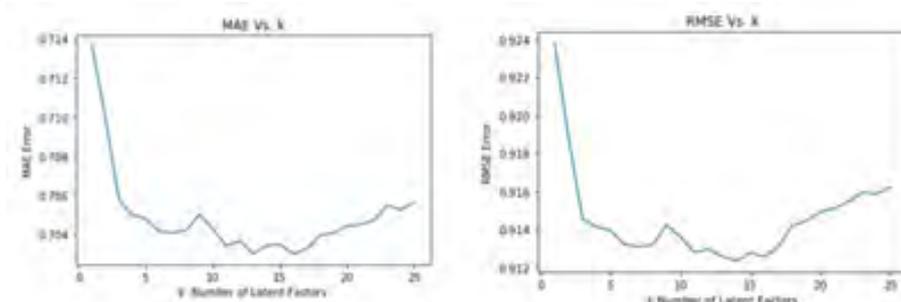
15863 rows × 4 columns

iteration= 1 k= 590 mae\_error= 0.73212 rmse\_error= 0.94482

\*\*\*\*\*  
time(s)= 504.09

#### شکل ۸. نتیجه اعمال روش SVD

با در نظر گرفتن تعداد مقادیر منفرد ( $K$ ) از یک تا ۲۵ و اعمال بر روی ماتریس های  $S$ ،  $U$  و  $V^T$ ، ماتریس های با ابعاد کاهش یافته جدید  $U_R$ ،  $S_R$  و  $V^T_R$  بدست می آیند که از طریق ضرب آنها، مقادیر پیش بینی برای دیتاست تست بدست آمده و مقدار خطاها را اندازه گیری می کنیم. شکل (۹) نمودار میزان تغییرات جذر میانگین مربعات خطا داده های تست بر مبنای تعداد مقادیر منفرد ( $K$ ) را نمایش می دهد. همانگونه که از نمودار (۹) مشاهده می شود، کمترین مقدار خطا در  $K=14$  رخ داده است که مقدار آن برابر با  $0.91236$  است. این مقدار نسبت به روش تجزیه مقادیر منفرد، به میزان  $3/43\%$  خطا را کاهش داده است. شکل (۱۰) نیز نمودار تغییرات خطای میانگین قدر مطلق خطا بر حسب  $K$  را نشان می دهد.

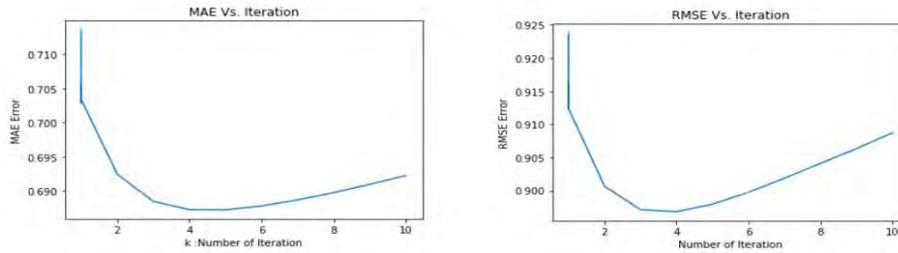


شکل ۹ نمودار تغییرات خطای RMSE بر حسب k | شکل ۱۰ نمودار تغییرات خطای MAE بر حسب k

در مرحله بعد برای کاهش بیشتر میزان خطا، از جاگذاری مقادیر پیش‌بینی در ماتریس امتیازات اولیه کاهش ابعاد یافته و انجام تجزیه مقادیر منفرد روی آن استفاده می‌کنیم. مرحله جاگذاری و تجزیه مقادیر منفرد را تا ۱۰ مرتبه تکرار می‌کنیم تا نمودار میزان خطا بر حسب تعداد تکرار حاصل شود. در این مرحله همواره مقدار  $K$  برابر با ۱۴ است که از مرحله قبل بدست آمده است. نتایج حاصل در شکل (۱۱) نمایش داده شده است. همانگونه که در نمودار مذکور مشاهده می‌شود کمترین مقدار جذر میانگین مربعات خطا در تکرار چهارم و با مقدار  $0/896814$  بدست می‌آید. در این حالت میزان خطا به مقدار  $1/7\%$  نسبت به مرحله قبل بهبود می‌یابد. بنابراین مجموع میزان کاهش خطا نسبت به روش تجزیه مقادیر منفرد، به میزان  $5/08$  درصد کاهش می‌یابد. همچنین مقدار میانگین قدر مطلق خطا در این حالت برابر با  $0/67705$  خواهد شد. نتایج حاصل در مقایسه با نتایج مندرج در مقالات (Andreas, et. al., 2017؛ Saadati, et al. 2019؛ Rahul et al., 2021) دارای خطای کمتری است.

برای مشاهده خروجی مدل پیشنهادی و با توجه به ماتریس پیش‌بینی بدست آمده می‌توان پیشنهادات را برای کاربرهای مختلف استخراج کرد. در این مرحله لازم است پس از استخراج فیلم‌های مشاهده توسط کاربر فعال، لیست فیلم‌های مذکور از لیست کلیه فیلم‌ها حذف شده، فیلم‌های باقیمانده بر اساس امتیازات پیش‌بینی شده بصورت نزولی مرتب گردند و ۲۰ فیلم اول به عنوان پیشنهادات انتخاب و نمایش داده شود. به عنوان نمونه

خروجی مدل برای کاربرهای با شناسه ۱۰ و ۲۵۰ در شکل (۱۲) نمایش داده شده است. لازم به ذکر است مقادیر RATING در جداول خروجی، امتیازات پیش‌بینی شده توسط مدل پیشنهادی "تجزیه مقادیر منفرد با ابعاد کاهش یافته تکرار شده" است.



شکل ۱۱. نمودار تغییرات RMSE و MAE نسبت به تعداد تکرار

user_id= 250 فیلم های پیشنهادی			user_id= 10 فیلم های پیشنهادی		
movielid	rating	title	movielid	rating	title
1041	4.69	Secrets & Lies (1996)	89904	4.73	The Artist (2011)
89904	4.64	The Artist (2011)	1270	4.70	Back to the Future (1985)
1104	4.61	Streetcar Named Desire, A (1951)	1212	4.70	Third Man, The (1949)
3451	4.52	Guess Who's Coming to Dinner (1967)	457	4.70	Fugitive, The (1993)
1209	4.50	Once Upon a Time in the West (C'era una volta ...)	1041	4.69	Secrets & Lies (1996)
2360	4.46	Celebration, The (Festen) (1998)	1207	4.63	To Kill a Mockingbird (1962)
1178	4.42	Paths of Glory (1967)	1250	4.59	Bridge on the River Kwai, The (1957)
306	4.42	Three Colors: Red (Trois couleurs: Rouge) (1994)	1204	4.59	Lawrence of Arabia (1962)
2329	4.40	American History X (1998)	1104	4.55	Streetcar Named Desire, A (1951)
3429	4.38	Creature Comforts (1989)	1223	4.51	Grand Day Out with Wallace and Gromit, A (1989)
912	4.37	Casablanca (1942)	1197	4.49	Princess Bride, The (1987)
527	4.37	Schindler's List (1993)	1209	4.48	Once Upon a Time in the West (C'era una volta ...)
904	4.35	Rear Window (1954)	3089	4.48	Bicycle Thieves (a.k.a. The Bicycle Thief) (a...)
1213	4.34	Goodfellas (1990)	1284	4.47	Big Sleep, The (1946)
8859	4.34	Dark Knight, The (2008)	2360	4.45	Celebration, The (Festen) (1998)
1221	4.34	Godfather Part II, The (1974)	1234	4.45	Sting, The (1973)
475	4.34	In the Name of the Father (1993)	250	4.44	Star Wars: Episode IV - A New Hope (1977)
858	4.34	Godfather, The (1972)	2318	4.43	Happiness (1998)
893	4.33	Silence of the Lambs, The (1991)	1178	4.43	Paths of Glory (1967)
356	4.33	Forrest Gump (1994)	1262	4.43	Great Escape, The (1963)
3083	4.33	All About My Mother (Todo sobre mi madre) (1999)	3429	4.41	Creature Comforts (1989)

شکل ۱۲. خروجی مدل برای کاربرهای با شناسه ۱۰ و ۲۵۰

## 1. Iterated Truncated Singular Value Decomposition (ITSVD)

## بحث و نتیجه گیری

در بین روش های مختلف فیلتر اشتراکی، بکارگیری روش تجزیه ماتریس به مقادیر منفرد دارای قابلیت انعطاف مناسبی در کاهش حجم محاسبات، کاهش حافظه مورد نیاز و همچنین اعمال روش های مختلف برای کاهش خطای مقادیر پیش بینی شده است. در مدل ارایه شده، با اعمال سه روش: حذف کاربران و فیلم های دارای تعداد امتیاز کم، پیدا کردن بهترین میزان کاهش ابعاد ماتریس ها و تکرار جاگذاری مقادیر و تجزیه مجدد ماتریس پیش بینی با ابعاد کاهش یافته، خطای مقادیر پیش بینی به مقدار مناسبی بهبود یافتند. همانطور که از جدول (۴) مشاهده می شود در هر مرحله از اعمال روش های بهینه سازی مقدار خطای جذر میانگین مربعات خطا به میزان قابل توجهی کاهش می یابد.

### جدول ۴. کاهش خطای جذر میانگین مربعات خطا با بهینه سازی روش تجزیه مقادیر

#### منفرد

روش	مقدار k	RMSE	Dataset	خالی بودن دیتاست (درصد)
SVD	۵۹۰	۰/۹۴۴۸۲	۱۰۰ هزار امتیازی مووی لنز با	۹۸/۴
TSVD	۱۴	۰/۹۱۲۳۶۰	۹۷۲۴ فیلم	
ITSVD	۱۴	۰/۸۹۶۸۱	۶۱۰ کاربر	

هر چند به دلیل تفاوت دیتاست ها، روش های پالایش داده ها و سرعت اجرای برنامه ها مقایسه نتایج بدست آمده با مراجع دیگر، از اعتبار کافی برخوردار ناست ولی جهت روشن شدن حدود خطا در مراجع مختلف به برخی موارد اشاره می شود. شرکت نتفلیکس در سال ۲۰۰۷ اعلام نمود سیستم توصیه گر آن شرکت با دیتاست ۱۰۰ میلیون امتیازی دارای خطای RMSE برابر ۰/۹۵۱۴ است (سعادت و همکاران، ۲۰۱۹). البته سعادت و همکاران (۲۰۱۷) با استفاده از روش تجزیه به مقادیر منفرد، مقدار خطا را ۳/۵۴۲ بدست آورده اند که مقدار بسیار بزرگی است. جدول (۵) مقدار خطاهای بدست آمده از روش های مختلف تجزیه مقادیر منفرد با یکدیگر مقایسه شده اند. همانطور که دیده می شود نتایج حاصل از مدل بهینه ارایه شده دارای خطای کمتری نسبت به سایر موارد می باشد.

جدول ۵. مقایسه خطای روش پیشنهادی با سایر مراجع

خالی بودن (درصد)	Dataset	RMSE	مقدار k	روش	مراجع
۹۸/۴	۱۰۰ هزار مووی لنز	۰/۹۴۴۸۲	۵۹۰	SVD	پژوهش حاضر
۹۸/۴	۱۰۰ هزار مووی لنز	۰/۹۱۲۳۶۰	۱۴	TSVD	پژوهش حاضر
۹۸/۴	۱۰۰ هزار مووی لنز	۰/۸۹۶۸۱	۱۴	ITSVD	پژوهش حاضر
۹۳/۷	۱۰۰ هزار مووی لنز	۰/۹۲۹۳۴	-	SVD+K_means	Rahul et al., 2021
نامعلوم	Epinions.com	۱/۱۴۷۴	۱۰	TSVD	Andreas, 2017
۹۵/۸	یک میلیون مووی لنز	۳/۵۴۲	۷۵	SVD	Saadati, et. al., 2019
نامعلوم	۱۰۰ میلیون نتفلیکس	۰/۹۵۱۴	-	فیلتر اشتراکی	Saadati, et. al., 2019

بنا بر خروجی مدل نهایی، می‌توان نتیجه گرفت مدل تجزیه ماتریس امتیازات به مقادیر منفرد، بر روی مجموعه داده‌های موجود که دارای بیش از ۹۸ درصد مقادیر نامعلوم است، بخوبی عمل کرده و تمام مقادیر نامعلوم را با دقت مناسبی می‌تواند پیش‌بینی نماید. البته بکارگیری دو روش بهینه سازی کاهش ابعاد و همچنین تکرار جاگذاری و تجزیه ماتریس، دقت مقادیر پیش‌بینی شده را تا حد بسیار مناسبی ارتقاء داده و نتایج حاصل در مقایسه با سایر مدل‌ها دارای دقت مناسبی است. برای کارهای آتی پیشنهاد می‌شود ترکیب مدل مذکور با سایر روش‌ها مانند روش‌های فیلتر محتوا محور و سایر روش‌های فیلتر اشتراکی خصوصاً شبکه‌های عصبی برای کاهش بیشتر میزان خطا بررسی شود. همچنین استفاده از روش‌های مقدار دهی اولیه برای امتیازات ناموجود از طریق روش‌های خوشه بندی همسایگی، استفاده از میانگین امتیازات کاربران و یا میانگین کل می‌تواند روش‌هایی برای مقایسه با روش پیشنهادی باشد.

### تعارض منافع

تعارض منافع ندارد.

### ORCID

Nozar Ebrahimi Lame

Fatemeh Saghafi

Majid Gholipour



<http://orcid.org/0000-0003-4512-8879>



<http://orcid.org/0000-0003-4843-6885>



<http://orcid.org/0000-0003-0804-5310>

## منابع

- ابراهیمی، ل، میرابی، و، رنجبر، م و حسن پور، ا. (۱۳۹۸). مدل وفاداری مشتری برای سیستم‌های توصیه گر در تجارت الکترونیک، نشریه علمی مطالعات مدیریت کسب و کار هوشمند، ۸ (۲۹)، ۱۴۳-۱۷۰. <https://dx.doi.org/10.22054/ims.2019.10379>
- حیدری، ب، پروین نیا، ا. (۱۳۹۶). ارایه مدلی برای سیستم‌های توصیه گر فیلم مبتنی بر رویکرد مشارکت محور، مجله فناوری اطلاعات در طراحی مهندسی، ۱۰ (۱)، ۹-۱.
- محمدی، ف، یزدانی، ح و ادیب زاده، م. (۱۳۹۹). فرا تحلیل مطالعات خرید آنلاین، بررسی و ترکیب نتایج پژوهش‌های انجام شده در زمینه خرید آنلاین، نشریه علمی مطالعات مدیریت کسب و کار هوشمند، ۹ (۳۳)، ۱۰۱-۱۴۲. <https://dx.doi.org/10.22054/ims.2020.12038>

## References

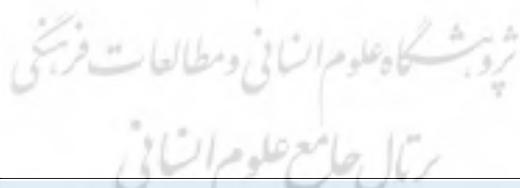
- Aggarwal, C. C., & Textbook, T. (2016). *Recommender Systems : Text Book*.
- Ahuja, R., Solanki, A., & Nayyar, A. (2019). Movie Recommender System Using K-Means Clustering AND K-Nearest Neighbor. *Proceedings of the 9th International Conference On Cloud Computing, Data Science and Engineering, Confluence 2019*, 263–268. <https://doi.org/10.1109/CONFLUENCE.2019.8776969>
- Alamdari, P. M., Navimipour, N. J., Hosseinzadeh, M., Safaei, A. A., & Darwesh, A. (2020). A Systematic Study on the Recommender Systems in the E-Commerce. *IEEE Access*, 8, 115694–115716. <https://doi.org/10.1109/ACCESS.2020.3002803>
- Andreas, M. (2017). *Matrix Factorization Techniques for Recommender Systems* [University of The Aegean, School of Engineering]/thesis [https://hellenicus.lib.aegean.gr/bitstream/handle/11610/18038/Matrix\\_Factorization\\_techniques\\_for\\_Recommender\\_Systems.pdf?sequence=1](https://hellenicus.lib.aegean.gr/bitstream/handle/11610/18038/Matrix_Factorization_techniques_for_Recommender_Systems.pdf?sequence=1)
- Bahl, D., Kain, V., Sharma, A., & Sharma, M. (2020). A novel hybrid approach towards movie recommender systems. *Journal of Statistics and Management Systems*, 23 (6), 1049–1058. <https://doi.org/10.1080/09720510.2020.1799503>
- Bhavana, P., Kumar, V., & Padmanabhan, V. (2019). *Block based Singular Value Decomposition approach to matrix factorization for recommender systems*. <http://arxiv.org/abs/1907.07410>
- Choi, S. M., Ko, S. K., & Han, Y. S. (2012). A movie recommendation algorithm based on genre correlations. *Expert Systems with*

- Applications*, 39 (9) , 8079–8085. <https://doi.org/10.1016/j.eswa.2012.01.132>
- Falk, K. (2019). *Practical Recommender Systems*. [https://www.mendeley.com/catalogue/4db84cb8-e911-3167-a969-98d4c25707f3/?utm\\_source=desktop&utm\\_medium=1.19.4&utm\\_campaign=open\\_catalog&userDocumentId=%7B09985284-1648-45ae-9586-8265d1bb4ad2%7D](https://www.mendeley.com/catalogue/4db84cb8-e911-3167-a969-98d4c25707f3/?utm_source=desktop&utm_medium=1.19.4&utm_campaign=open_catalog&userDocumentId=%7B09985284-1648-45ae-9586-8265d1bb4ad2%7D)
- Guan, X., Li, C. T., & Guan, Y. (2017). Matrix Factorization with Rating Completion: An Enhanced SVD Model for Collaborative Filtering Recommender Systems. *IEEE Access*, 5, 27668–27678. <https://doi.org/10.1109/ACCESS.2017.2772226>
- Lekakos, G., & Caravelas, P. (2008). A hybrid approach for movie recommendation. *Multimedia Tools and Applications*, 36 (1–2) , 55–70. <https://doi.org/10.1007/s11042-006-0082-7>
- Li, X., Zhao, H., Wang, Z., & Yu, Z. (2020). Research on Movie Rating Prediction Algorithms. *2020 5th IEEE International Conference on Big Data Analytics, ICBDA 2020*, 121–125. <https://doi.org/10.1109/ICBDA49040.2020.9101282>
- Netflix Prize*. [www.netflixprize.com](http://www.netflixprize.com)
- Patra, S., & Ganguly, B. (2019). Improving Singular Value Decomposition by KNN for Use in Movie Recommender Systems. *Journal of Operations and Strategic Planning*, 2 (1) , 22–34. <https://doi.org/10.1177/2516600x19848956>
- Patra, S., & Ganguly, B. (2019). Improving Singular Value Decomposition by KNN for Use in Movie Recommender Systems. *Journal of Operations and Strategic Planning*, 2 (1) , 22–34. <https://doi.org/10.1177/2516600x19848956>
- Raghuwanshi, S. K., & Pateriya, R. K. (2019). Recommendation systems: Techniques, challenges, application, and evaluation. In *Advances in Intelligent Systems and Computing* (Vol. 817, pp. 151–164). [https://doi.org/10.1007/978-981-13-1595-4\\_12](https://doi.org/10.1007/978-981-13-1595-4_12)
- Rahul, M., Kumar, V., Yadav, V., & Rishabh. (2021). Movie recommender system using single value decomposition and K-means clustering. *IOP Conference Series: Materials Science and Engineering*, 1022 (1). <https://doi.org/10.1088/1757-899X/1022/1/012100>
- Saadati, M., & Shihab, S. (n.d.). *Movie Recommender Systems: Implementation and Performance Evaluation/2019*
- Vaheb, A.. <https://aparart.design/recommendation-systems-in-filimo-cqzge4gfsevi>
- Zhang, Y., Abbas, H., & Sun, Y. (2019). Smart e-commerce integration with recommender systems. In *Electronic Markets* (Vol. 29, Issue 2, pp. 219–220). <https://doi.org/10.1007/s12525-019-00346-x>

Zhou, X., He, J., Huang, G., & Zhang, Y. (2015). SVD-based incremental approaches for recommender systems. *Journal of Computer and System Sciences*, 81 (4), 717–733. <https://doi.org/10.1016/j.jcss.2014.11.016>

#### References [In Persian]

- Ebrahimi, L., Mirabi, and, Ranjbar, M. and Hassanpour, A. (1398). Customer Loyalty Model for Advisory Systems in E-Commerce, *Scientific Journal of Intelligent Business Management Studies*, 8 (29), 143-170. <https://dx.doi.org/10.22054/ims.2019.10379> [ In Persian]
- Heydari, B, Parvin Nia, A. (1396). Providing a Model for Participatory Oriented Film Recommending Systems, *Journal of Information Technology in Engineering Design*, 10 (1), 1-9. [ In Persian]
- Mohammadi, F. Yazdani, H. and Adibzadeh, M. (1399). Meta-analysis of online shopping studies, review and synthesis of research results in the field of online shopping, *Scientific Journal of Smart Business Management Studies*, 9 (33), 101-142. <https://dx.doi.org/10.22054/ims.2020.12038>[ In Persian]



**استناد به این مقاله:** ابراهیمی لامع، نوذر، ثقفی، فاطمه، قلی پور، مجید. (۱۴۰۰). سیستم توصیه گر فیلم ابتکاری با استفاده از روش تجزیه ماتریس به مقادیر منفرد با ابعاد کاهش یافته تکرار شده، *مطالعات مدیریت کسب و کار هوشمند*، ۱۰(۳۸)، ۱۷۳-۱۹۹.

DOI: 10.22054/IMS.2022.61702.1994



Journal of Business Intelligence Management Studies is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License..



پروہشگاہ علوم انسانی و مطالعات فرہنگی  
پرتال جامع علوم انسانی