Journal of Business Intelligence Management Studies
Vol. 10, No.38, Winter 2022
ims.atu.ac.ir
DOI: 10.22054/IMS.2022.61702.1994

ATU PRESS

Original Research
Accepted: 09/02/2022
Revised: 09/02/2021
Received: 16/06/2021
eISSN: 2821-0816
ISSN: 2821-0964

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.

مقاله پژوهشی
تاریخ ارسال: ۱۴۰۰/۰۳/۲۶
تاریخ بازنگری: ۱۴۰۰/۱۱/۲۰
تاریخ پذیرش: ۱۴۰۰/۱۱/۲۰
ISSN: 2821-0964
eISSN: 2821-0816

مطالعات مدیریت کسب و کار هوشمند
سال دهم، شماره ۳۸، زمستان ۱۴۰۰، ص ۱۷۳ تا ۱۹۹
ims.atu.ac.ir
DOI: 10.22054/IMS.2022.61702.1994

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

نوذر ابراهیمی لامع — دانشجوی دکتری گروه مدیریت سیستم‌های اطلاعاتی، واحد قزوین، دانشگاه آزاد اسلامی، قزوین، ایران.

فاطمه ثقفی * — دانشیار گروه مدیریت صنعتی، دانشکده مدیریت، دانشگاه تهران، تهران، ایران.

مجید قلی پور — استادیار دانشکده برق و رایانه، دانشگاه آزاد اسلامی واحد قزوین، قزوین، ایران.

چکیده

سیستم‌های توصیه‌گر یکی از ضروری ترین ابزارهای هوشمندسازی تجارت الکترونیک است. این سیستم‌ها با انواع مختلف روش‌های فیلترکردن داده‌ها و داده‌کاوی، قادر به انتخاب و ارایه بهترین پیشنهادات از بین انبوه موارد قابل انتخاب برای مشتریان هستند. در بین روش‌های متنوع سیستم‌های توصیه‌گر، فیلترهای اشتراکی پرکاربردترین روش برای ارایه 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 ارایه مدلی بهینه شده از سیستم توصیه‌گر 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

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

کلیدواژه‌ها: سیستم‌های توصیه‌گر، فیلتر اشتراکی، تجزیه مقادیر منفرد، پیش‌بینی امتیازات.

مقدمه

در دهه‌های اخیر، با توسعه وب‌سایت ها و شبکه‌های اینترنتی، تجارت الکترونیک نیز توسعه فراوانی یافته و تبدیل به بخش مهمی از تجارت جهانی شده‌است (محمدی و همکاران، ۱۳۹۹). تجارت الکترونیک فروشندگان را قادر می سازد تا حجم انبوهی از کالا و خدمات را در وب سایت های خود به مشتریان ارایه 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 می نامند. سیستم‌های توصیه‌گر، نرم افزارهایی ضروری برای فیلتر کردن داده‌ها و ارایه پیشنهادات متناسب با سلایق و علائق مشتریان هستند. این سیستم‌ها برای اولین بار در اوایل دهه ۱۹۹۰ میلادی مورد استفاده قرارگرفتند (Guan et al. , 2014؛ Rahul et al., 2021).

وظیفه اصلی سیستم‌های توصیه‌گر، ارایه پیشنهادات متناسب برای هر شخص و بصورت شخصی می‌باشد. به عبارت دیگر، وب سایت ها برای هر مشتری محصولات مشخصی را نمایش می‌دهند (ویترین شخصی). این روش ارائه پیشنهادات را شخصی سازی پیشنهادات می‌نامند. شخصی سازی[1] پیشنهادات باعث افزایش جذب و وفاداری مشتریان و همچنین افزایش فروش وب سایتها می گردد (Alamdari et al., 2020؛ Zhang et al., 2019؛ Lekakos & Caravelas, 2008؛ Rghuwanshi & Pateriya, 2019؛ Guan et al., 2017). سایت های معتبری مانند آمازون[2]، ای بی[3]، نتفلیکس[4]، علی بابا[5] در سطح بین المللی و سایت هایی مانند دیجی کالا، دیوار، باما، نماوا و فیلیمو در سطح کشور از سیستم‌های توصیه‌گر بصورت گسترده‌ای استفاده می نمایند.

---

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

برای تولید سیستم‌های توصیه‌گر، 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[1] انجام است (Aggarwal, 2016).

پرکاربردترین روش پالایش داده‌ها در سیستم‌های توصیه‌گر، روش *فیلتر اشتراکی*[2] است. در این روش از ماتریس امتیازات برای توصیه آیتم‌ها به استفاده می‌شود. فیلترهای اشتراکی دو نوع می باشند. فیلتر اشتراکی حافظه محور[3]، که با استفاده از امتیازات کاربران و تشابه بین کاربران یا آیتم‌ها با یکدیگر، کار می‌کنند و سیستم *مدل محور*[4]، که بر اساس مدل ارایه شده نسبت به پیش‌بینی امتیازات کاربران با استفاده از روش‌های متنوع یادگیری ماشین کار می‌کنند (Aggarwal, 2016؛ Ahuja et al., 2019؛ Guan et al., 2017). یکی دیگر از روش‌های پالایش داده‌ها، روش *محتوا محور* است. در این روش از خصوصیات آیتم‌های مورد علاقه کاربر استفاده می‌شود تا آیتم‌های مشابه به او پیشنهاد شود (Guan et al., 2017). با توجه به اینکه سیستم‌های توصیه‌گر با داده‌های کلان سر و کار دارند و از طرفی حجم امتیازات در مقابل تعداد کاربران و آیتم‌ها بسیار اندک است (ماتریس امتیازات خالی[5])، بنابراین تهیه سیستم توصیه‌گر با دقت مناسب و غلبه بر خالی بودن ماتریس امتیازات، از مشکلات اصلی روش فیلتر اشتراکی است. در سیستم‌های توصیه‌گر فیلتر اشتراکی، استفاده از روش‌های خوشه‌بندی مانند نزدیکترین همسایه ها[6]، روش‌های یادگیری ماشین[7] و تجزیه ماتریس[8]، جهت افزایش دقت پیشنهادات و کاهش حجم محاسبات، از روش های 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[1] نسبت به روش متداول نزدیکترین همسایه‌ها، دارای دقت بیشتری بوده و همچنین با روش‌های کاهش ابعاد ماتریس، امکان کاهش حجم محاسبات نیز وجود دارد (Aggarwal, 2016؛ Falk, 2019). روش سوم پالایش داده‌ها، استفاده از ترکیب دو روش فیلتر اشتراکی و فیلتر محتوا محور است که به آن فیلتر ترکیبی[2] 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[3] با هدف ارتقاء دقت پیشنهادات با روش فیلتر اشتراکی بود (سایت نتفلیکس[4]).

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 رسانه ای، استفاده از سیستم‌های توصیه‌گر جهت توصیه محصولات رسانه‌ای مانند کتاب، خبر، موزیک، سرگرمی، مجله و فیلم بسیار پرکاربرد 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[5] (2020)، بر روی مقالات منتشر شده با جستجوی کلمات کلیدی "سیستم‌های توصیه‌گر[6]" و "تجارت الکترونیک[7]" توسط جستجوگر گوگل اسکولار[8] نشان می‌دهد از بین ۳۵۸ مقاله یافت شده در این موضوعات، ۲۸۹ مورد آن (بیش از ۸۰٪) بین سال های ۲۰۰۸ تا ۲۰۱۹ انتشار یافته‌اند که نشانگر فعال بودن موضوع پژوهش است. البته برخلاف چنین سابقه ای 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

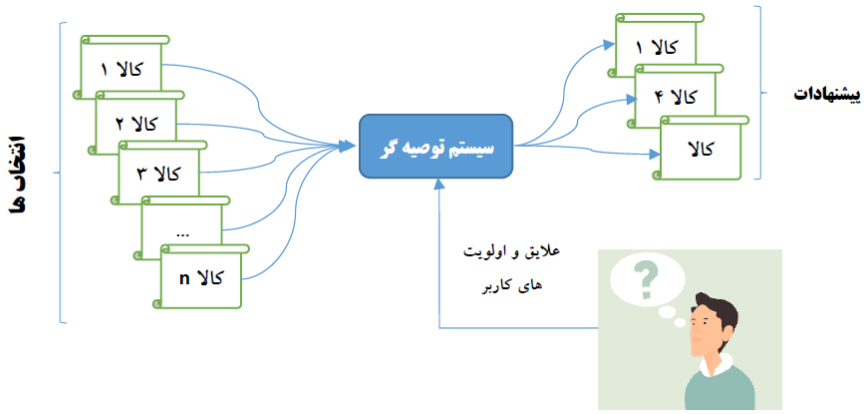
تعداد کارهای پژوهش هایی در زمینه سیستم‌های توصیه‌گر در داخل کشور محدود بوده و در زمینه سیستم‌های توصیه‌گر فیلم، نتایج جستجو در گوگل اسکولار، ایران داک[1] و مگیران[2] و با کلمات کلیدی "سیستم‌های توصیه‌گر فیلم" و "سیستم‌های پیشنهاد دهنده فیلم"، منجر به یافتن تعداد فقط پنج مقاله به زبان فارسی شد که به عنوان نمونه می‌توان به حیدری و پروین نیا (۱۳۹۶) اشاره نمود. البته شرکت‌های داخلی بسیار زیادی مانند دیجی‌کالا، نماوا و فیلیمو از سیستم‌های توصیه‌گر استفاده می‌کنند ولی در خصوص روش‌های بکارگرفته شده و میزان دقت و کارائی آنها مطالب اندکی ارایه شده‌است. در این خصوص، می‌توان به سیستم توصیه‌گر شرکت فیلیمو 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 توصیه‌گر فیلم با روش تجزیه ماتریس به مقادیر منفرد ارایه شده‌است که با استفاده از سه روش مختلف، برای حجم محاسبات و خطای امتیازات پیش بینی شده کاهش قابل ملاحظه‌ای حاصل شده‌است و خطای بدست‌آمده نسبت به موارد مشابه دارای دقت بهتری می‌باشد. در ادامه مقاله، به مرور ادبیات و پیشینه پژوهش، ارایه روش پژوهش، یافته های پژوهش و بحث در مورد نتایج پژوهش خواهیم پرداخت.

پیشینه پژوهش

وظیفه اصلی سیستم‌های توصیه‌گر پالایش حجم انبوهی از داده‌ها و پیشنهاد محصولات متناسب با علائق و سلایق کاربر است. شکل (۱) سیستم‌های توصیه‌گر را بصورت شماتیک نمایش می‌دهد.

1. Irandoc
2. Magiran



شکل ۱. سیستم توصیه‌گر

در مجموع پنج نوع سیستم توصیه‌گر وجود دارد (Alamdari, et.al., 2020؛ Aggarwal, 2016؛ Choi, et. al., 2012) که سه نوع آن متداول و پرکاربرد است. روش اول، سیستم توصیه‌گر *فیلتر محتوا محور*[1] است که از تشابه بین ویژگی‌های آیتم‌ها برای ارایه پیشنهاد استفاده می‌کند. این روش سریع است ولی از چالش هایی مانند کمبود اطلاعات در مورد آیتم‌ها و دقت پایین رنج می‌برد. روش دوم، سیستم توصیه‌گر فیلتر اشتراکی است که با استفاده از ماتریس امتیازات، از تشابه بین کاربران و یا آیتم‌ها برای ارایه پیشنهادات استفاده می‌کند. چالشهای اصلی روشهای فیلتر اشتراکی، شروع سرد کاربر[2] و ماتریس خالی می‌باشد. این روش، پر کاربردترین روش برای ارایه سیستم‌های توصیه‌گر است. روش‌های فیلتر اشتراکی خود به دو دسته فیلتر اشتراکی حافظه محور و فیلتر اشتراکی مدل محور تقسیم می‌شوند. روش سوم، سیستم توصیه‌گر ترکیبی است که برای استفاده از مزایای روش‌های مختلف و کاهش معایب آنها، از ترکیب روش‌های دیگر استفاده می کند. دو روش فیلتر جمعیتی محور[3] و فیلتر دانش محور[4] با توجه به کاربردهای خاص آنها، جزء

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

روش‌های متداول نمی‌باشند (Aggarwal, 2016). جدول (۱) اهداف مفهومی و ورودی های هر یک از روش‌های پایه را بصورت خلاصه نشان می‌دهد. آگاروال (2016) معتقد است اگر چه اصلی ترین هدف از بکارگیری سیستم‌های توصیه‌گر، افزایش فروش محصولات است و این هدف با ارایه پیشنهادات مناسب به کاربران و جلب توجه آنها میسر می‌شود، ولی برای رسیدن به هدف مذکور لازم است در طراحی سیستم‌های توصیه‌گر و ارایه پیشنهادات، موارد و اهداف فنی مهمی دیگری مانند مرتبط بودن[1]، تازگی[2]، پیشنهادات غیر منتظره[3] و تنوع[4] پیشنهادات نیز در نظر گرفته شود (ابراهیمی و همکاران، ۱۳۹۸).

جدول ۱. اهداف مفهومی انواع سیستم‌های توصیه‌گر پایه

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

نکته مهمی که در طراحی سیستم‌های توصیه‌گر باید در نظر گرفته شود اینست که هر یک از روش‌های پالایش داده‌ها، دارای چالش های مخصوص به خود می‌باشند. مهم ترین چالش‌هایی که معمولاً این سیستم‌ها با آنها مواجه می شوند عبارتند از: دشواری استخراج محتوا، شروع سرد، خالی بودن ماتریس امتیازات، مقیاس پذیری، حفظ حریم خصوصی و کلمات مترادف (Rghuwanshi & Pateriya,2019؛ Aggarwal, 2016؛ Choi, et. al., 2012). از این‌رو لازم است روش‌های انتخابی برای ارایه پیشنهادات، قادر به برطرف

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

نمودن چالش های مذکور باشند. همچنین نکته مهمی که در طراحی سیستم های توصیه‌گر باید در نظر داشت اینست که هیچ یک از روش‌ها به تنهایی قادر به برطرف نمودن همه چالش های مذکور نمی‌باشند. لذا برای انتخاب روش مناسب لازم است مواردی مانند مشخصات آیتم‌ها 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 همسایه ها[1] و میانگین امتیازات آنها[2]، استفاده از شبکه های عصبی و روش تجزیه ماتریس به حاصلضرب چند ماتریس نام برد. روش تجزیه ماتریس به عوامل نهانی[3] 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 ماتریس ها، برای سیستم‌های توصیه‌گر کاربردهای فراوانی پیدا نمود. این روش، سابقه کاربرد در موارد دیگری را داشته ولی برای اولین بار توسط سیمون فانک[4] 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

می‌شود. 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[1]، تجزیه و تحلیل عوامل[2] و تجزیه و تحلیل عوامل اصلی[3] جزئ روش‌های پرکاربرد می‌باشند (Andreas, 2017). در این پژوهش به جهت قابلیت‌های 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 شده‌است.

مبانی روش تجزیه ماتریس

اگر R، ماتریس امتیازات کاربران به فیلم‌ها با ابعاد m*n باشد که در آن m تعداد کاربران (سطرها) و n تعداد فیلم‌ها (ستون ها) باشد شکل (۲)، در اینصورت هدف از تجزیه ماتریس R، پیداکردن 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:

$$R_{m*n} \approx U_{m*k.} V^{T}_{k*n} \qquad (۱)$$

بنابراین برای بدست آوردن مقدار امتیاز کاربر i به آیتم j می‌توان از رابطه زیر استفاده نمود:

$$\hat{r_{ij}} = u_i v_j^T = \sum_{k=1}^{k} u_{ik} v_{kj} \qquad (۲)$$

شکل (۲) نشان می‌دهد چگونه ماتریس امتیازات R به دو ماتریس U و V^T تجزیه شده‌است. لازم به ذکر است اگر ماتریس دارای مقادیر نامعلوم باشد، حاصلضرب دو ماتریس تقریبی از ماتریس R را خواهد داد.

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

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

R **U** **V**

شکل ۲. تجزیه ماتریس به دو ماتریس

برای بدست آوردن مقادیر ماتریس های U و V ابتدا مقادیر اولیهٔ تصادفی‌ای برای سلول-های آنها تعیین می‌شود و پس از حاصلضرب دو ماتریس، اختلاف بین مقادیر حاصل ضرب و مقادیر واقعی در ماتریس R محاسبه می‌شود. سپس با روش گرادیان کاهشی[1] نسبت به حداقل نمودن خطا اقدام می‌گردد. در روش گرادیان کاهشی هدف حداقل نمودن مقدار خطا است. برای اینکار، لازم است مجذور تفاضل مقادیر واقعی و پیش بینی حداقل گردد. در اینحالت تابع گرادیان کاهشی بصورت زیر خواهد شد:

$$\min J = \frac{1}{2} \left(\sum_{(i,j)\varepsilon R}^{n} \left(r_{i,j} - u_i \, v_j^T \right)^2 + \lambda \left(\|u_i\|^2 + \|v_j\|^2 \right) \right) \qquad (۳)$$

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 λ مقدار پارامتر تنظیم است.

از طرفی، برای ارزیابی میزان دقت حاصل از تجزیه ماتریس برای پیش‌بینی مقادیر نامعلوم و جلوگیری از بیش برازش 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 می‌شود. در مرحله بعد، با استفاده از مقادیر تقریبی بدست آمده از تجزیه ماتریس امتیازات، مقادیر واقعی داده‌های تست و مقادیر تقریبی بدست آمده با

1. Decent Gradient

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

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

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

در روش تجزیه ماتریس به مقادیر منفرد که دارای کاربردهای فراوانی در کاهش ابعاد[1]، فشرده سازی تصاویر[2] و نادیده گرفتن داده‌های نویز[3] دارد، یک ماتریس بجای دو ماتریس به سه ماتریس U، S و V بصورت زیر تجزیه می‌شود:

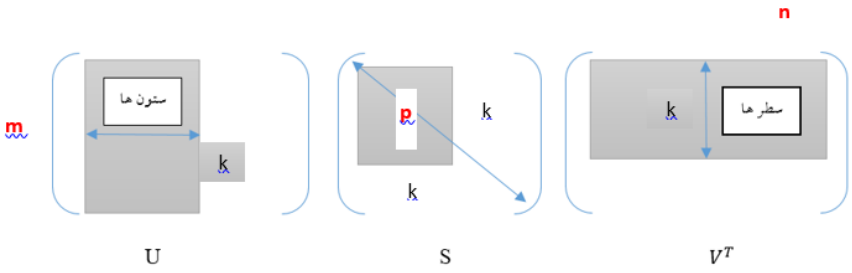
$$R_{m\times n} = U_{m\times p} * S_{p\times p} * V^T_{p\times n} \quad (۶)$$

که در آن ماتریسU، ماتریس متعاملی است که به آن ماتریس عوامل نهانی[4] کاربر می‌گویند و خصوصیات کاربران را بصورت فشرده و ترکیبی نمایش می‌دهد. ماتریس S مقادیر منفرد ماتریس امتیازات R است و ماتریس V نیز ماتریس متعاملی است که به آن ماتریس عوامل نهانی آیتم‌ها می‌گویند و خصوصیات آیتم‌ها (محصولات) را بصورت ترکیبی و فشرده در خود دارد. نهانی بودن عوامل فوق به این معنی است که توسط مدل محاسبه و تعیین می‌شود و در تعیین آنها، انسان نقشی ندارد. همچنین لازم به ذکر است مقادیر عوامل نهانی اغلب قابل تفسیر و توضیح نمی‌باشند.

در این روش، هر دو ماتریس U و V متعامد می‌باشند 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 می‌دهدهر یک از مقادیر حاوی چه مقدار اطلاعات می‌باشند. 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 بهینه ای 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 ماتریس ها بر اساس تعداد مقادیر انتخاب شده از ماتریس S تعیین می‌شود. شکل (۳) این حالت را نمایش می‌دهد. (Aggarwal, 2016 ؛Falk, 2019). در شکل (۳)، ماتریس R_{m*n} به کمک روش تجزیه به مقادیر منفرد به حاصلضرب سه ماتریس U_{m*p}، S_{p*p} و V^T_{p*n} 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 ماتریس های فوق به صورت U_{m*k}، S_{k*k} و V^T_{k*n} 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 پر طرفدار در سیستم‌های توصیه‌گر تبدیل

شد (Falk, 2019). در سال های بعد، پژوهش های متعددی در خصوص ارایه راهکارهایی برای ارایه سیستم‌های توصیه‌گر بر این مبنا توسعه داده شد که در زیر به برخی از آنها اشاره می‌شود. لی و همکاران[1] (2020) در پژوهش خود روشی تحت عنوان "فیلتر کردن مبتنی بر دسته‌های فیلم- تجزیه مقادیر منفرد[2]" را برای پیش‌بینی امتیازات کاربران فعال، به فیلم‌ها بر مبنای داده‌های صریح[3] و ضمنی[4] ناشی از دیتاست فیلم‌ها ارایه می‌دهند. در ابتدا، روش فوق از عوامل وزن دار برای بررسی اثر انواع دسته بندی فیلم بر روی پیش‌بینی امتیازات کاربران و همچنین ارتقا روش پالایش بر مبنای دسته بندی فیلم‌ها استفاده می‌کند. در نهایت، روش فیلتر کردن مبتنی بر دسته‌های فیلم با الگوریتم تجزیه ماتریس به مقادیر منفرد[5] ترکیب می‌شود تا از عملکرد خوب آن در فیلتر اشتراکی استفاده نماید. در مقایسه با انواع الگوریتم‌های مختلف موجود، روش پیشنهادی به میزان قابل توجهی دقت پیش‌بینی امتیازدهی را ارتقاء می بخشد و مقیاس پذیری و اثربخشی سیستم‌های توصیه‌گر شخصی را افزایش می‌دهد. ژو و همکاران[6] (2015) یک الگوریتم تدریجی[7] 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 می‌دهد. سپس آنها اقدام به مقایسه ای از دقت پیش‌بینی و زمان اجرا بین الگوریتم پیشنهادی و الگوریتم تجزیه مقادیر منفرد تدریجی بر روی مجموعه داده‌های مووی لنز و فلیسکتر[8] ارایه داده اند و نتیجه گرفته اند که روش پیشنهادی از سایر روش‌های مشابه خود بهتر است. در مقاله بال و همکاران[9] (2020) یک روش فیلتر ترکیبی از ترکیب فیلتر محتوا محور بر اساس همبستگی بین فیلم‌ها 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 دیتاست های ۱۰۰ هزار مووی لنز نشان از بهبود خطا نسبت به روش‌های دیگر از جمله تجزیه مقادیر منفرد تنها است. همچنین باوانا و همکاران[1] (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 واحد های پردازشگر موازی جهت تسریع در بدست آوردن نتایج را مورد بررسی قرار داده‌اند. پاترا و گانگولی[2] (2019) روش‌های مختلف پیش پردازش برای اعمال به الگوریتم تجزیه مقادیر منفرد را با هم مورد بررسی قرار داده اند. برای بررسی از دیتاست های یک میلیون مووی لنز استفاده شده‌است. در ابتدا با استفاده از روش خوشه بندی نزدیکترین همسایه های کاربران، با روش‌های متفاوت فاصله یابی 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 همسایه های 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[3] (2021) برای سیستم توصیه‌گر فیلم از روش تجزیه مقادیر منفرد برای کاهش ابعاد و از روش متوسط امتیاز همسایگان برای خوشه بندی استفاده کرده‌اند. برای پیاده سازی مدل از دیتاست های 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[4] (2019) در مقاله خود نسبت به ارایه نتایج حاصل از پیاده سازی 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
٤	Bhavana et al., 2019	سرعت پردازش	SVD
٥	Patra& Ganguly,2019	پیش پردازش	SVD+ K_Means
٦	Saadati et al.,2019	خطای کمتر	SVD و سایر روش‌ها
۷	Zhou et al., 2015	مقیاس پذیری و عملکرد	Incremental SVD

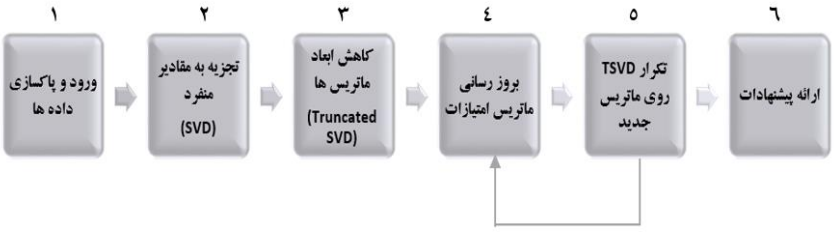
روش پژوهش

در این پژوهش، مدل بهینه شده ای از سیستم توصیه‌گر 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 شده‌است. 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 شده‌است. مراحل مدلسازی و بهینه سازی آن در شکل (۴) نمایش داده شده‌است.

مرحله۱. 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 سلول های بدون مقدار با میانگین امتیاز هر فیلم
- نرمال کردن امتیازات با کم کردن میانگین امتیاز هر فیلم از امتیازات



شکل ٤. 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 می‌آوریم.

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

مراحل ۴و 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 می‌دهیم و با تجزیه مجدد ماتریس امتیازات به مقادیر منفرد، میزان خطای بهینه را بدست می‌آوریم.

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

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 می‌دهند:

'data'

	userId	movieId	rating
0	1	1	4.0
1	1	3	4.0
2	1	6	4.0
3	1	47	5.0
4	1	50	5.0
...	...	...	...
100831	610	166534	4.0
100832	610	168248	5.0
100833	610	168250	5.0
100834	610	168252	5.0
100835	610	170875	3.0

100836 rows × 3 columns

شکل ۵. دیتاست اصلی

'train'

	movieId	rating	userId
2052	27831.0	3.5	18.0
11736	165.0	5.0	71.0
49865	318.0	4.5	319.0
51918	95.0	5.0	337.0
57780	56775.0	4.0	380.0
...	...	...	...
4788	1224.0	4.0	29.0
58227	2085.0	3.0	381.0
38745	1275.0	4.0	266.0
28875	6565.0	4.0	199.0
45346	95441.0	2.0	298.0

63451 rows × 3 columns

شکل ٦. دیتاست آموزش

'test'

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

15863 rows × 3 columns

شکل ۷. دیتاست تست

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[1] و میانگین قدر مطلق خطا[2] 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.

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

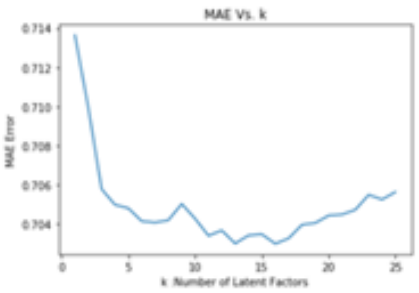
	movieId	rating	userId	rating_hat
50022	1682.0	3.5	322.0	3.83
25564	50872.0	4.0	177.0	3.84
87281	55118.0	4.5	562.0	4.04
2083	47200.0	3.0	18.0	3.11
51265	7438.0	3.0	331.0	3.85
...	...	...	...	...
11962	96588.0	3.5	73.0	3.59
80284	71535.0	4.0	505.0	3.83
55967	5418.0	3.5	369.0	3.85
17707	94777.0	1.0	111.0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

15863 rows × 4 columns

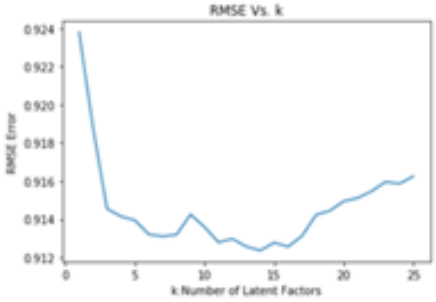
```
iteration= 1 k= 590 mae_error= 0.73212 rmse_error= 0.94482
**************************************************
tims(s)= 504.09
```

شکل ۸ نتیجه اعمال روش SVD

با در نظر گرفتن تعداد مقادیر منفرد (K) از یک تا ۲۵ و اعمال بر روی ماتریس های U، S و V^T، ماتریس های با ابعاد کاهش یافته جدید U_R، S_R و V^T_R بدست می‌آیند که از طریق ضرب آنها، مقادیر پیش‌بینی برای دیتاست تست بدست آمده و مقدار خطاها را اندازه گیری می‌کنیم. شکل (۹) نمودار میزان تغییرات جذر میانگین مربعات خطا داده‌های تست بر مبنای تعداد مقادیر منفرد (K) را نمایش می‌دهد. همانگونه که از نمودار (۹) مشاهده می‌شود، کمترین مقدار خطا در K=۱۴ رخ داده است که مقدار آن برابر با ۰/۹۱۲۳۶۰ است. این مقدار نسبت به روش تجزیه مقادیر منفرد، به میزان ۳/۴۳ ٪خطا را کاهش داده است. شکل (۱۰) نیز نمودار تغییرات خطای میانگین قدر مطلق خطا بر حسب K را نشان می‌دهد.



شکل ۹. نمودار تغییرات خطای RMSE بر حسب k

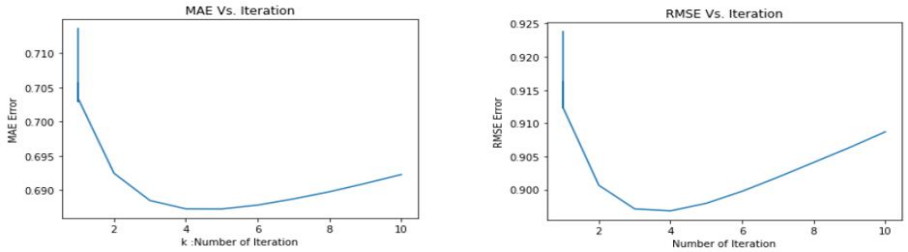


شکل ۱۰ نمودا تغییرات خطای MAE بر حسب k

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 تعدادتکرار حاصل شود. در این مرحله همواره مقدارK برابر با ۱۴ است که از مرحله قبل بدست آمده است. نتایج حاصل در شکل (۱۱) نمایش داده شده‌است. همانگونه که در نمودار مذکور مشاهده می‌شود کمترین مقدار جذر میانگین مربعات خطا در تکرار چهارم و با مقدار ۰/۸۹۶۸۱۴ بدست می آید. در این حالت میزان خطا به مقدار ۱/۷ ٪ نسبت به مرحله قبل بهبود می یابد. بنابراین مجموع میزان کاهش خطا نسبت به روش تجزیه مقادیر منفرد، به میزان ۵/۰۸ درصد کاهش می یابد. همچنین مقدار میانگین قدر مطلق خطا در این‌حالت برابر با ۰/۶۷۷۰۵ خواهد شد. نتایج حاصل در مقایسه با نتایج مندرج در مقالات (Rahul et al., 2021؛ Saadati, et al. 2019؛ Andreas, et. al. ,2017) 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 شده‌است. لازم به ذکر است مقادیر RATING در جداول خروجی، امتیازات پیش‌بینی شده توسط مدل پیشنهادی" تجزیه مقادیر منفرد با ابعاد کاهش یافته تکرار شده[1]" است.



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

user_id= 250

'فیلم های پیشنهادی'

movieId	rating	title
1041	4.69	Secrets & Lies (1996)
89904	4.64	The Artist (2011)
1104	4.61	Streetcar Named Desire, A (1951)
3451	4.52	Guess Who's Coming to Dinner (1967)
1209	4.50	Once Upon a Time in the West (C'era una volta ...
2360	4.46	Celebration, The (Festen) (1998)
1178	4.42	Paths of Glory (1957)
306	4.42	Three Colors: Red (Trois couleurs: Rouge) (1994)
2329	4.40	American History X (1998)
3429	4.38	Creature Comforts (1989)
912	4.37	Casablanca (1942)
527	4.37	Schindler's List (1993)
904	4.35	Rear Window (1954)
1213	4.34	Goodfellas (1990)
58559	4.34	Dark Knight, The (2008)
1221	4.34	Godfather: Part II, The (1974)
475	4.34	In the Name of the Father (1993)
858	4.34	Godfather, The (1972)
593	4.33	Silence of the Lambs, The (1991)
356	4.33	Forrest Gump (1994)
3083	4.33	All About My Mother (Todo sobre mi madre) (1999)

user_id= 10

'فیلم های پیشنهادی'

movieId	rating	title
89904	4.73	The Artist (2011)
1270	4.70	Back to the Future (1985)
1212	4.70	Third Man, The (1949)
457	4.70	Fugitive, The (1993)
1041	4.69	Secrets & Lies (1996)
1207	4.63	To Kill a Mockingbird (1962)
1250	4.59	Bridge on the River Kwai, The (1957)
1204	4.59	Lawrence of Arabia (1962)
1104	4.55	Streetcar Named Desire, A (1951)
1223	4.51	Grand Day Out with Wallace and Gromit, A (1989)
1197	4.49	Princess Bride, The (1987)
1209	4.49	Once Upon a Time in the West (C'era una volta ...
3089	4.48	Bicycle Thieves (a.k.a. The Bicycle Thief) (a...
1284	4.47	Big Sleep, The (1946)
2360	4.45	Celebration, The (Festen) (1998)
1234	4.45	Sting, The (1973)
260	4.44	Star Wars: Episode IV - A New Hope (1977)
2318	4.43	Happiness (1998)
1178	4.43	Paths of Glory (1957)
1262	4.43	Great Escape, The (1963)
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 نتیجه‌گیری

در بین روش‌های 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 روش‌های مختلف برای کاهش خطای مقادیر پیش‌بینی شده‌است. در مدل ارایه شده، با اعمال سه روش: حذف کاربران و فیلم‌های دارای تعداد امتیاز کم، پیدا کردن بهترین میزان کاهش ابعاد ماتریس ها و تکرار جاگذاری مقادیر و تجزیه مجدد ماتریس پیش‌بینی با ابعاد کاهش یافته، خطای مقادیر پیش‌بینی به مقدار مناسبی بهبود یافتند. همانطور که از جدول (۴) مشاهد می‌شود در هر مرحله از اعمال روش‌های بهینه سازی مقدار خطای جذر میانگین مربعات خطا به میزان قابل توجهی کاهش می یابد.

جدول ٤. 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 ۹۷۲۴ فیلم ۶۱۰ کاربر	۹۸/۴
TSVD	۱۴	۰/۹۱۲۳۶۰		
ITSVD	۱۴	۰/۸۹۶۸۱		

هر چند به دلیل تفاوت دیتاست ها، روش‌های پالایش داده‌ها 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 می‌شود. شرکت نتفلیکس در سال ۲۰۰۷ اعلام نمود سیستم توصیه‌گر آن شرکت با دیتاست ۱۰۰ میلیون امتیازی دارای خطای RMSE برابر ۰/۹۵۱۴ است (سعادتی و همکاران، ۲۰۱۹). البته سعادتی 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) مقدارخطاهای بدست آمده از روش‌های مختلف تجزیه مقادیر منفرد با یکدیگر مقایسه شده‌اند. همانطور که دیده می‌شود نتایج حاصل از مدل بهینه ارایه شده دارای خطای کمتری نسبت به سایر موارد می‌باشد.

جدول ۵. مقایسه خطای روش پیشنهادی با سایر مراجع

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

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 مدل ها 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 شبکه های 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 **http://orcid.org/0000-0003-4512-8879**
Fatemeh Saghafi **http://orcid.org/0000-0003-4843-6885**
Majid Gholipour **http://orcid.org/0000-0003-0804-5310**

منابع

ابراهیمی، ل، میرابی، و، رنجبر، م و حسن پور، ا. (۱۳۹۸). مدل وفاداری مشتری برای سیستم‌های توصیه‌گر 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 توصیه‌گر فیلم مبتنی بر رویکرد مشارکت محور، *مجله فناوری اطلاعات در طراحی مهندسی*، ۱۰ (۱)، ۱-۹.

محمدی، ف، یزدانی، ح و ادیب زاده، م. (۱۳۹۹). فرا تحلیل مطالعات خرید آنلاین، بررسی و ترکیب نتایج پژوهش های 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://hellanicus.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

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

Patra, S., & Ganguly, B. (2019). Improvising 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). Improvising 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

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://aparat.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 Desig*n, 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.، قلی‌پور، مجید. (۱۴۰۰). سیستم توصیه‌گر فیلم ابتکاری با استفاده از روش تجزیه ماتریس به مقادیر منفرد با ابعاد کاهش یافته تکرار شده، *مطالعات مدیریت کسب وکار 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..