



# یک سامانه توصیه‌گر ترکیبی با استفاده از اعتماد و خوشه‌بندی دوجته به منظور افزایش کارایی پالایش گروهی

منیره حسینی<sup>\*</sup>، مقصود نصراللهی و علی باغی

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

## چکیده

موفقیت سامانه‌های تجارت الکترونیکی و تراکنش‌های کسب‌وکار بر خط، به‌طور قابل توجهی وابسته به طراحی مؤثر سازوکار توصیه محصولات است. فراهم کردن توصیه‌های باکیفیت برای سامانه‌های تجارت الکترونیکی بسیار مهم است تا بدین ترتیب به کاربران در تصمیم‌گیری مؤثر میان انتخاب‌های متعدد کمک کند. پالایش گروهی یک روش برای تولید توصیه‌ها بر اساس رتبه‌های کاربران مشابه است که به‌صورت وسیعی مورد قبول واقع شده است. این روش دارای چندین مشکل ذاتی مانند کم‌پشتی داده‌ها، شروع سرد و مقیاس‌پذیری است. حل این مشکلات و بهبود کارایی پالایش گروهی از چالش‌های مطرح در این زمینه است. در این مقاله یک سامانه ترکیبی جدید که شبکه اعتماد و خوشه‌بندی دوجته را برای افزایش کارایی پالایش گروهی به کار می‌بندد، پیشنهاد شده است. نتایج تجربی بر روی زیرمجموعه‌ای از مجموعه داده‌های epinions، اثربخشی و کارایی سامانه پیشنهادی در مقابل روش‌های پالایش گروهی مبتنی بر کاربر و پالایش گروهی ترکیبی با اعتماد را تأیید می‌کند.

واژگان کلیدی: سامانه‌های توصیه‌گر، پالایش گروهی، اعتماد، خوشه‌بندی دوجته، سامانه توصیه‌گر ترکیبی

## A hybrid recommender system using trust and bi-clustering in order to increase the efficiency of collaborative filtering

Monireh Hosseini\*, Maghsoud Nasrollahi & Ali Baghaei

IT Department, Industrial Engineering Faculty of K. N. Toosi University of Technology, Tehran, Iran

### Abstract

In the present era, the amount of information grows exponentially. So, finding the required information among the mass of information has become a major challenge. The success of e-commerce systems and online business transactions depend greatly on the effective design of products recommender mechanism. Providing high quality recommendations is important for e-commerce systems to assist users in making effective selection decisions from a plethora of choices. Recommender systems have been developed in order to respond this problem in order to customize the required information for users.

So far, several types of recommender systems have been developed such as collaborative filtering recommender systems, content-based recommender systems and knowledge-based recommender systems. Each of these systems has advantages and disadvantages. Most of the recommender systems are based on collaborative filtering; Collaborative filtering is a widely accepted technique to generate recommendations based on the ratings of like-minded users. In fact, the main idea of this technique is to benefit from the past behavior or existing beliefs of the user community to predict products that are likely to be liked by the current user of the system. In collaborative filtering, we use the similarity between users or items to recommend products. However, this technique has several inherent problems such as cold start, sparsity and scalability.

Since the collaborative filtering system is considered to be the most widely used recommender system, solving these problems and improving the effectiveness of collaborative filtering is one of the challenges raised in this context.

<sup>\*</sup> Corresponding author

\* نویسنده عهده‌دار مکاتبات

فصلنامه



سال ۱۳۹۷ شماره ۲ پیاپی ۳۶

None of the proposed hybrid systems have ever been able to resolve all of the collaborative filtering problems in a single and desirable manner; in this paper, we proposed a new hybrid recommender system that applies trust network as well as bi-clustering to improve the effectiveness of collaborative filtering. Therefore, the objectives of this research can be summarized as follows: sparsity reduction, increasing the speed of producing recommendations and increasing the accuracy of recommendations.

In the proposed system, the trust between users is used to fill the user-item matrix which is a sparse matrix to solve the existing problem of sparsity. Then using bi-clustering, the user-item matrix is subdivided into matrices to solve the problem of scalability of the collaborative filtering and then the collaborative filtering is implemented for each sub matrix and the results from the implementation of the collaborative filtering for the sub-matrices are combined and recommendations are made for the users.

The experimental results on a subset of the extended Epinions dataset verify the effectiveness and efficiency of our proposed system over user-based collaborative filtering and hybrid collaborative filtering with trust techniques.

#### Improve sparsity problem

Experimental results showed that our proposed system solves some of the sparsity problems which is due to the using the trust in the hybrid recommender system. By using trust, we can predict many uncertain ratings. Thus, transforming the user-item sparsity matrix into a half-full matrix.

#### Improve scalability problem

The results show that the proposed system has a higher speed compared with the user-based collaborative filtering algorithm and hybrid collaborative filtering with trust, and increasing the volume of data has little effect on increase online computing time. The reason can be summarized as a using of bi-clustering. Bi-directional clusters are made offline and break down the matrix of rankings into smaller subsets. Implementing the collaborative filtering on these smaller sets has led to increased computing speed.

#### Improve the new user problem

This system can provide accurate results for the new users due to the use of trust, because product collections viewed by new user can increase with the trust between the users. This system can predict the similarity between the new user and other users. So, the results are more accurate than the results of the user-based collaborative filtering and hybrid collaborative filtering with trust.

**Keywords:** Recommender systems, Collaborative filtering, Trust, Bi-clustering, Hybrid recommender systems.

## ۱- مقدمه

حرکت به سوی تجارت الکترونیکی، سبب شده است که شرکت‌ها برای مشتریان خود گزینه‌های بیشتری فراهم کنند. این افزایش گزینه‌ها موجب شده است که مشتریان وقت زیادی را صرف پیدا کردن محصولات مورد نظر خود کنند [31]. برای مقابله با این مسأله که از آن به عنوان سربار اطلاعاتی<sup>۱</sup> یاد می‌شود، سامانه‌های توصیه‌گر<sup>۲</sup> طراحی و تولید شدند.

سامانه‌های توصیه‌گر، سامانه‌هایی هستند که به کمک اطلاعات موجود و تحلیل رفتار و خصوصیات کاربران، پیشنهاد‌های خودکار ارائه می‌دهند. این سامانه‌ها هوشمند بوده و برای تسهیل تراکنش‌هایی با سربار زیاد اطلاعات و مشکلات اطلاعاتی تعریف می‌شوند [15]. سامانه‌های توصیه‌گر به کاربران کمک می‌کنند تا سریع‌تر به هدف خود نزدیک شوند و گزینه مورد علاقه خود را پیدا کنند. درواقع این سامانه‌ها به مانند فیلتری عمل می‌کنند که فقط آنچه مطلوب کاربر است به وی نمایش می‌دهند. به این عمل شخصی سازی کردن اطلاعات نیز می‌گویند [39]. این سامانه‌ها

علاوه بر این که برای کاربران فهرستی از محصولاتی که ممکن است مورد پسند وی باشد، فراهم می‌کنند، می‌توانند میزان علاقه مندی یک کاربر را به محصولی خاص پیش‌بینی کنند [40].

سامانه‌های توصیه‌گر درحقیقت همان فرآیندی را به کار می‌برند که ما در زندگی روزمره خود به کار می‌بریم و طی آن تلاش می‌کنیم تا افرادی با سلیقه نزدیک به خود را پیدا کرده و از آن‌ها در مورد انتخاب‌هایمان نظرخواهی کنیم [7]. به کارگیری این سامانه‌ها منجر به جذب و جلب رضایت مشتریان می‌شود و فروش را در سایت‌های تجارت الکترونیکی از طریق تبدیل بازدیدکنندگان به خریداران، دگر فروشی و وفاداری افزایش می‌دهند [31]. به این ترتیب با افزایش روزافزون تمرکز کسب و کارهای الکترونیکی بر روی ارزش بلندمدت مشتریان، بنگاه‌های اقتصادی باید از سامانه‌های توصیه‌گر قدرتمند بهره گیرند.

راه کارهای متفاوتی برای پالایش اطلاعات به وجود آمده‌اند. از پرکاربردترین راه کارهای پالایش، پالایش گروهی (CF)<sup>۳</sup> است [27]. پالایش گروهی به صورت وسیعی برای

<sup>1</sup> Information Overload

<sup>2</sup> Recommender System (R.S)

<sup>3</sup> Collaborative Filtering

از سویی به فرموله‌سازی مسئله توصیه، خوشه‌بندی دوجته<sup>۱</sup> و اعتماد خواهیم پرداخت.

## ۱-۲- مفاهیم اصلی و مرور بر ادبیات

بدون شک اساسی‌ترین جزء در سامانه‌های توصیه‌گر، راه‌کار پالایش آن است. بوادیلا<sup>۲</sup> و همکارانش، راه‌کارهای پالایش مختلفی برای توصیه معرفی کرده‌اند. مهم‌ترین راه‌کارها عبارت از: توصیه مبتنی بر همکاری، توصیه مبتنی بر محتوا، توصیه مبتنی بر ویژگی‌های جمعیت‌شناختی، توصیه مبتنی بر دانش، توصیه مبتنی بر موقعیت، توصیه مبتنی بر زمینه و توصیه مبتنی بر جامعه می‌باشند. هریک از این راه‌کارها مزایا و مشکلاتی دارند [4].

پالایش گروهی از پرکاربردترین راه‌کارهای پالایش است [27]. این نوع از سامانه‌ها از سال‌ها پیش موردتوجه قرار گرفته‌اند و مزیت‌ها، کارایی‌ها و محدودیت‌های آن‌ها امروزه به‌خوبی فهمیده شده است. در طول سال‌ها، الگوریتم‌ها و روش‌های متنوعی پیشنهاد و با موفقیت در جهان واقعی و داده‌های ساختگی ارزیابی شده‌اند. پالایش گروهی، مبتنی بر این حقیقت است که عقاید دیگران به‌طور قابل‌توجهی بر تصمیم‌گیری خریداران مؤثر است [39]. الگوریتم‌های پالایش گروهی بر شباهت میان کاربران یا شباهت میان موردها با استفاده از رتبه‌های کاربران تمرکز می‌کند. برای سامانه‌های توصیه‌گر مبتنی بر پالایش گروهی، کیفیت توصیه‌ها وابسته به کیفیت بازخوردها (رتبه‌های کاربران) می‌باشد [21].

پالایش گروهی در زمینه پژوهش‌ها، عمل، پالایش اطلاعات و کاربردهای تجارت الکترونیکی بسیار موفق بوده است. درواقع هدف الگوریتم پالایش گروهی، پیشنهاد موردهای جدید یا پیش‌بینی سودمندی یک مورد مشخص برای یک کاربر خاص براساس پسندهای پیشین و نظرات دیگر کاربران که سلیق مشابه دارند، می‌باشد [30]. در ادامه برخی از چالش‌ها و مشکلات سامانه‌های توصیه‌گر مبتنی بر پالایش گروهی شرح داده می‌شوند.

شروع سرد<sup>۳</sup> یا خلوتی (کم‌پشتی داده‌ها)<sup>۴</sup>: در صورتی که سامانه تازه شروع به کار کرده باشد و یا مورد جدیدی به سامانه اضافه شود، اطلاعات کافی از موردها در دسترس نخواهد بود و در نتیجه به درستی پیش‌بینی را نمی‌توان انجام داد. این مشکل زمانی ایجاد می‌شود که به‌منظور ارائه پیشنهاد، اطلاعات لازم و کافی (رتبه‌ها) در سامانه وجود نداشته باشد.

ساخت سامانه‌های توصیه‌گر در بخش تجارت الکترونیکی استفاده شده است. این روش، سودمندی یک محصول برای کاربر خاص را بر اساس پسندهای پیشین و عقاید دیگران که مشابه کاربر خاص هستند، پیش‌بینی می‌کند [30]. پالایش گروهی رتبه یک مورد فرض شده را، بر اساس رتبه‌بندی‌های پیشین کاربر محاسبه می‌کند. زمانی که اطلاعات رتبه‌بندی وجود ندارد یا رضایت‌بخش نیست، پالایش گروهی توصیه‌های باکیفیتی تولید نمی‌کند. همچنین پیدا کردن کاربران مشابه به‌دلیل کم‌بودن موردهایی که رتبه‌بندی شده‌اند، امکان‌پذیر نیست [18]. بهبود کارایی این راه‌کار، از جمله چالش‌های مطرح در این زمینه است.

برای بهبود عملکرد سامانه‌های توصیه‌گر مبتنی بر پالایش گروهی و تولید سامانه‌های توصیه‌گر جدید بهبودیافته، از اطلاعات حاصل از شبکه‌های اجتماعی که اشتراک‌گذاری دیدگاه‌ها را برای کاربران آسان کرده‌اند، می‌توان استفاده کرد. یک سامانه توصیه‌گر اجتماعی، صحت سامانه‌های توصیه‌گر سنتی را با استفاده از اطلاعات اعتماد اجتماعی و علاقه‌مندی اجتماعی بین کاربران در یک شبکه اجتماعی برخط به‌عنوان ورودی‌های اضافی، بهبود می‌بخشد [40].

بر این اساس در این مقاله یک راه‌کار ترکیبی به‌منظور افزایش کارایی پالایش گروهی با استفاده از اعتماد و خوشه‌بندی پیشنهاد شده است. سامانه پیشنهادی در این مطالعه، با رفع محدودیت‌های سامانه‌های پیشین که در ادامه اشاره خواهد شد، به‌منظور تولید توصیه‌های بهتر و دقیق‌تر برای مشتریان یک تارنمای تجارت الکترونیکی می‌تواند به‌کار گرفته شود. درواقع نوآوری این مقاله ترکیب دو راه‌کار خوشه‌بندی و اعتماد برای حل تمامی مشکلات و چالش‌های سامانه‌های توصیه‌گر مبتنی بر پالایش گروهی است.

در ادامه این مقاله و در بخش دوم، مفاهیم اصلی مطرح شده و مروری بر ادبیات موضوع خواهیم داشت. در بخش سوم مدل پیشنهادی بیان شده و در بخش چهارم مدل بر روی داده‌های واقعی پیاده‌سازی و آزمایش می‌شود. در انتها نتایج ارزیابی مدل پیشنهادی بیان و نتیجه‌گیری و پیشنهادهای آینده مطرح می‌شود.

## ۲- مروری بر ادبیات موضوع

در این بخش مفاهیم اصلی در حوزه سامانه‌های توصیه‌گر بیان شده و همچنین مروری بر ادبیات این موضوع خواهیم داشت.

<sup>1</sup> Biclustering

<sup>2</sup> Bobadilla

<sup>3</sup> Cold Start

<sup>4</sup> Sparsity

این حالت (شروع سرد) ممکن است به یکی از دلایل شروع کار سامانه پیشنهاددهنده، ورود کاربر جدید به سامانه و یا درج مورد جدید در سامانه رخ دهد [4,21].

مقیاس پذیری<sup>۱</sup>: الگوریتم نیازمند محاسباتی است که با شمار کاربران و شمار محصولات رشد می کند. با میلیون ها کاربر و مورد، یک سامانه توصیه معمولی مبتنی بر وب در

اجرای الگوریتم های موجود با مشکل جدی مقیاس پذیری روبه رو خواهد [30].

مقالات در حوزه سامانه های توصیه گر و ویژگی ها و کارهای انجام شده هر یک از مقالات به طور خلاصه در جدول (۱) آورده شده است.

(جدول ۱-): مروری بر ادبیات موضوع  
(Table-1): Literature review

منبع	عنوان مقاله	کلیات پژوهش های انجام شده	تمرکز مقاله
[14]	توصیه پالایش گروهی مبتنی بر SOM و خوشه نمایه سازی CBR	افزایش صحت و حل مشکل مقیاس پذیری با استفاده از فرآیندهای یادگیری ماشین	افزایش صحت و مقیاس پذیری
[9]	یک چارچوب پالایش گروهی مقیاس پذیر با استفاده از خوشه بندی دوجهته	حل مشکل مقیاس پذیری با خوشه بندی دوجهته	افزایش مقیاس پذیری
[28]	سامانه توصیه مبتنی بر مکان، با استفاده از مدل بیزی اولویت های کاربر، برای دستگاه های تلفن همراه	سامانه توصیه مبتنی بر کاربر با استفاده از شبکه های بیزی	افزایش دقت و صحت توصیه ها
[37]	سامانه توصیه بهبود یافته با استفاده از زیر گروه های مورد-کاربر	بهبود سامانه پالایش گروهی با استفاده از خوشه بندی هم زمان کاربرها و مورد ها	افزایش مقیاس پذیری و صحت
[42]	یک سامانه توصیه مبتنی بر اطلاعات برجسب و زمان برای سامانه های برجسب اجتماعی	بهبود توصیه با استفاده از اطلاعات برجسب و زمان	افزایش صحت و دقت
[8]	مدل توصیه شخصی سازی شده بر اساس الگوریتم ژنتیک	افزایش صحت توصیه ها با استفاده از الگوریتم ژنتیک	افزایش صحت
[11]	ترکیب اعتماد در پالایش گروهی برای کاهش خلوتی داده ها و شروع سرد	حل مشکل خلوتی داده ها و شروع سرد با استفاده از شبکه اعتماد	حل مشکل خلوتی داده ها و شروع سرد
[30]	الگوریتم توصیه پالایش گروهی مبتنی بر مورد	معرفی پالایش گروهی مبتنی بر مورد	حل مشکل کاربر جدید
[34]	توسعه یک سامانه توصیه گر همکاری مبتنی بر مکان برای برنامه های کاربردی GIS	تمرکز بر توصیه مکان های محبوب بر اساس رفتار GPS کاربران متعدد. استفاده از الگوریتم ژنتیک برای پیش بینی علاقه کاربران برای مکان های بازدید نشده بر اساس رتبه بندی ضمنی و صریح	افزایش صحت توصیه ها
[35]	پالایش گروهی نزدیکترین خوشه، بر اساس مقدارهای ثابت و منسجم	استفاده از خوشه بندی دوجهته برای افزایش صحت و مقیاس پذیری پالایش گروهی	افزایش صحت و مقیاس پذیری
[33]	ترکیب روابط اعتماد در سامانه توصیه گر پالایش گروهی	ارائه دو مدل اعتماد عمومی و مدل اعتماد سطح-منبع جهت پالایه کردن محتوای ناخواسته بر اساس روابط اعتماد بین کاربران	افزایش صحت
[16]	یک سامانه توصیه گر همکاری بهبود یافته توسط روش بهینه سازی ازدحام ذرات	پیشنهاد یک مدل ترکیبی جدید براساس رویکرد توصیه گر پالایش گروهی برای تولید توصیه های فیلم با بهره گیری از سامانه تقسیم نوع با بهینه سازی ازدحام ذرات (به منظور کاهش فضای داده های چندبعدی متراکم) و الگوریتم های خوشه بندی FCM و K-means	افزایش مقیاس پذیری و افزایش صحت
[19]	یک الگوریتم خوشه بندی اصلاح شده DBSCAN استفاده شده در سامانه توصیه گر پالایش گروهی برای پیشنهاد موسیقی	ارائه رویکرد جدید خوشه بندی (DBSCAN) اصلاح شده برای توصیه های پالایش گروهی برای حل مشکل مقیاس پذیری در یک توصیه گر مبتنی بر مورد برای پیشنهاد موسیقی	افزایش مقیاس پذیری

<sup>1</sup> Scalability

[17]	یک سامانه توصیه با استفاده از خوشه‌بندی برای فروشگاه‌های برخط	افزایش صحت سامانه توصیه با استفاده از خوشه‌بندی k-means و الگوریتم ژنتیک	افزایش صحت و مقیاس‌پذیری
[20]	پالایش گروهی مبتنی بر اعتماد	استفاده از کاربران مورد اعتماد به جای کاربران مشابه برای غلبه بر مشکلات پالایش گروهی	حل مشکل شروع سرد
[23]	پالایش گروهی افزایشی بر اساس تجزیه ماتریس	افزایش مقیاس‌پذیری و صحت پالایش گروهی با استفاده از تجزیه ماتریس	افزایش مقیاس‌پذیری و صحت
[21]	یک روش مبتنی بر زمان برای توصیه‌های مؤثرتر با استفاده از بازخوردهای ضمنی	استفاده از اطلاعات ضمنی برای افزایش صحت توصیه‌ها	افزایش صحت
[22]	یک مکانیزم توصیه اجتماعی برای تجارت الکترونیک با ترکیب اعتماد، شباهت و رابطه	استفاده از اعتماد، رابطه و شباهت برای افزایش کارایی سامانه‌های توصیه‌گر	حل مشکل شروع سرد و افزایش صحت
[25]	یک سامانه توصیه ترکیبی برای رستوران	استفاده از اطلاعات جغرافیایی برای توصیه	افزایش دقت و صحت
[24]	توصیه مقالات علمی با استفاده از پالایش گروهی چند معیاره	استفاده از معیارهای چندگانه برای پالایش گروهی برای افزایش صحت	افزایش صحت
[27]	یک سامانه توصیه‌گر پالایش گروهی چند معیاره برای حوزه گردشگری با استفاده از حداکثر کردن انتظار (EM) و PCA-ANFIS	ارائه یک روش ترکیبی جدید برای توصیه هتل با استفاده از تکنیک‌های کاهش ابعاد و پیش‌بینی. توسعه یک سیستم توصیه‌گر پالایش گروهی چند معیاره برای پیشنهاد هتل با استفاده از مدل مخلوط گوسی با الگوریتم حداکثر کردن انتظار (EM) و سامانه استنتاج تطبیقی عصبی فازی (ANFIS). استفاده از تجزیه و تحلیل مولفه اصلی (PCA) برای کاهش ابعاد	افزایش صحت
[13]	یک مدل پالایش گروهی ترکیبی با الگوریتم ژنتیک برای توصیه‌های برخط	افزایش صحت سامانه‌های توصیه‌گر پالایش گروهی با استفاده از الگوریتم ژنتیک	افزایش صحت و مقیاس‌پذیری
[39]	یک روش توصیه پالایش گروهی به وسیله بررسی اثر متقابل علاقه و اعتماد	افزایش کیفیت توصیه‌ها با استفاده از اعتماد و علاقه	افزایش صحت و حل شروع سرد
[6]	بهبود عملکرد سامانه توصیه‌گر پالایش گروهی توسط ساخت‌وساز شبکه اعتماد	ارائه رویکردی که شامل اطلاعات اعتماد برای پیش‌بینی رتبه در سامانه توصیه‌گر است	افزایش صحت و حل شروع سرد
[41]	بهبود سامانه‌های توصیه‌گر آگاه به اعتماد با استفاده از شبکه اعتماد small-worldness	مزیت استفاده از اعتماد در سامانه‌های توصیه‌گر	افزایش صحت و حل شروع سرد

سامانه جامعی ارائه بدهیم.

## ۲-۲- فرموله‌سازی مسئله توصیه

برای فرموله‌سازی مسئله توصیه، از ماتریس بهره‌وری<sup>۱</sup>، استفاده می‌شود. در یک سامانه توصیه‌گر، دو دسته موجودیت وجود دارد که عبارتند از موارد و کاربران. ترجیحات کاربران برای موارد مختلف با استفاده از این ماتریس نمایش داده می‌شود. در واقع هر زوج کاربر مورد<sup>۲</sup> بیان‌گر درجه علاقه کاربر به آن مورد است که با یک مقدار عددی نشان داده می‌شود که همان رتبه داده‌شده از سوی کاربر برای مورد است. با توجه به تعداد بالای موارد و کاربران، این ماتریس یک ماتریس خلوت<sup>۳</sup> خواهد بود. مسئله سامانه توصیه‌گر، پیش‌بینی هرچه

با توجه به آنچه بیان شد، به دلیل مشکلات و نقص‌های موجود، به‌طورعمومی سامانه‌های توصیه‌گر را به‌صورت ترکیبی با سایر راه‌کارها به‌کار می‌برند تا از مزایای آن‌ها بهره‌مند شده و در عین حال معایب آن را نیز بر طرف کنند [4,21].

از آنجایی که سامانه پالایش گروهی پرکاربردترین سامانه توصیه‌گر محسوب می‌شود، تمرکز این مقاله بر روی بهبود محدودیت‌ها و مشکلات آن قرار دارد. هیچ یک از سامانه‌های موجود، بر کلیه مشکلات تمرکز نکرده و تنها بر رفع بخشی از مشکلات تمرکز داشته‌اند. هدف این پژوهش ارائه یک سامانه توصیه‌گر ترکیبی جدید است که با بهره‌گیری از اعتماد میان کاربران و خوشه‌بندی، تا جای ممکن به رفع این محدودیت‌ها کمک کند و مشکلات موجود را بهبود بخشد. ما در این سامانه سعی داریم بر همه مشکلات تمرکز کرده و

<sup>1</sup> Utility Matrix

<sup>2</sup> User-Item

<sup>3</sup> Sparse Matrix

دقیق‌تر این مقادیر ناشناخته است تا بر اساس آن بتواند مواردی را که به احتمال زیاد مورد علاقه کاربر هستند به او توصیه کند.

مسئله توصیه به این شکل فرموله می‌شود: فرض کنید  $C$  و  $S$  به ترتیب مجموعه تمام کاربران و موارد باشند (هر یک از این مجموعه‌ها تا میلیون‌ها عضو می‌تواند داشته باشد)، در آن صورت تابع سودمندی  $u$  که میزان مفید بودن مورد  $s$  برای کاربر  $c$  را اندازه‌گیری می‌کند، به صورت رابطه (۱) خواهد بود [2]:

$$u: C \times S \rightarrow R \quad (1)$$

در اینجا  $R$  یک مجموعه مرتب از اعداد صحیح یا اعداد حقیقی با یک دامنه مشخص است. بنابراین سامانه توصیه‌گر باید برای هر کاربر  $c \in C$  موردی مثل  $s^c \in S$  را انتخاب کند که سودمندی کاربر را به بیشینه رساند. این مسئله به زبان ریاضی در رابطه (۲) بیان شده است [2]:

$$\forall c \in C, s^c = \operatorname{argmax} u(c, s) \quad (2)$$

### ۳-۲- رویکردهای پالایش گروهی

اشاره شد که پالایش گروهی مبتنی بر این حقیقت است که عقاید دیگران به طور قابل توجهی بر تصمیم‌گیری خریداران مؤثر است [39]. الگوریتم‌های پالایش گروهی بر شباهت میان کاربران یا شباهت میان موردها با استفاده از رتبه‌های کاربران تمرکز می‌کند. برای سامانه‌های توصیه‌گر مبتنی بر پالایش گروهی، کیفیت توصیه‌ها وابسته به کیفیت بازخوردها<sup>۱</sup> (رتبه‌های کاربران) است [21]. رویکردهای پالایش گروهی بر دو دسته‌اند: [3,4,21]

الف) پالایش گروهی مبتنی بر مدل که به استخراج یک مدل از داده‌های رتبه می‌پردازد و از آن برای انجام پیش‌بینی بهره می‌برد. تاکنون در پژوهش‌های مختلف، روش‌های متنوعی برای یادگیری مدل استفاده شده است؛ مانند شبکه‌های عصبی [14] سامانه‌های فازی [38]، الگوریتم‌های ژنتیک [8,13]، تجزیه ماتریس [23] شبکه‌های بیزی [28].  
ب) پالایش گروهی مبتنی بر حافظه که از مجموعه تمام مواردی که در قبل توسط کاربران رتبه داده شده‌اند، برای انجام پیش‌بینی استفاده می‌کند. به این روش، پالایش گروهی مبتنی بر اکتشاف یا مبتنی بر همسایگی نیز اطلاق می‌شود [3]. پالایش گروهی مبتنی بر حافظه خود به دو گروه:

پالایش گروهی مبتنی بر کاربر و پالایش گروهی مبتنی بر مورد تقسیم می‌شود [30]. مراحل هر یک از این رویکردها به صورت زیر است:

۱. محاسبه شباهت<sup>۲</sup>: در این مرحله با توجه به این که رویکرد مورد نظر مبتنی بر کاربر<sup>۳</sup> یا مبتنی بر مورد<sup>۴</sup> است، به ترتیب شباهت بین کاربران یا شباهت بین موارد محاسبه و از این شباهت‌ها برای تشکیل همسایگی و نیز وزندهی به هنگام انجام پیش‌بینی استفاده می‌شود. یک رویکرد بسیار معروف برای محاسبه شباهت، همبستگی پیرسون<sup>۵</sup> است. در پالایش گروهی مبتنی بر کاربر، همبستگی پیرسون میان کاربر فعال  $a$  با کاربر  $u$  به شکل رابطه ۳ محاسبه می‌شود [5]:

$$\operatorname{sim}(a, u) = \frac{\sum_{i \in I} (r_{a,i} - \bar{r}_a)(r_{u,i} - \bar{r}_u)}{\sqrt{\sum_{i \in I} (r_{a,i} - \bar{r}_a)^2} \sqrt{\sum_{i \in I} (r_{u,i} - \bar{r}_u)^2}} \quad (3)$$

در رابطه ۳،  $I$  مجموعه تمام مواردی است که توسط هر دو کاربر رتبه داده شده‌اند؛  $r_{a,i}$  رتبه داده شده از سوی کاربر  $a$  برای مورد  $i$  و  $\bar{r}_a$  میانگین رتبه‌های داده شده از طرف کاربر  $a$  است.

در رویکرد مبتنی بر مورد، به جای محاسبه شباهت میان کاربران از شباهت میان موارد استفاده می‌شود. شباهت میان موارد  $i$  و  $j$  با استفاده از همبستگی پیرسون به شکل رابطه (۴) محاسبه می‌شود [30]:

$$\operatorname{sim}(a, u) = \frac{\sum_{i \in I} (r_{a,i} - \bar{r}_a)(r_{u,i} - \bar{r}_u)}{\sqrt{\sum_{i \in I} (r_{a,i} - \bar{r}_a)^2} \sqrt{\sum_{i \in I} (r_{u,i} - \bar{r}_u)^2}} \quad (4)$$

در اینجا،  $U$  مجموعه تمامی کاربرانی است که به هر دو مورد رتبه داده‌اند؛  $r_{u,i}$  رتبه داده شده از سوی کاربر  $u$  برای مورد  $i$  و  $\bar{r}_i$  میانگین رتبه‌های داده شده برای مورد  $i$  در بین کاربران است. ضریب همبستگی پیرسون مقدار عددی بین ۱ (همبستگی مثبت شدید) تا -۱ (همبستگی منفی شدید) است. اگرچه از سنج‌های دیگر نیز می‌توان استفاده کرد، اما تحلیل‌های تجربی نشان می‌دهد که ضریب پیرسون نتیجه بهتری از دیگر سنج‌ها برای سامانه‌های توصیه‌گر کاربرمحور تولید می‌کند [3]. برای روش توصیه مبتنی بر مورد گزارش شده است که سنج شباهت کسینوسی بهتر از سنج همبستگی پیرسون است [15]. شباهت کسینوسی بین دو محصول  $a$  و  $b$  به صورت رابطه (۵) محاسبه می‌شود [15]:

<sup>2</sup> Similarity

<sup>3</sup> User-Base

<sup>4</sup> Item-Base

<sup>5</sup> Pearson Correlation

<sup>1</sup> Feedback



هدف از خوشه بندی تقسیم یک مجموعه داده به چند گروه بر اساس شباهت ها می باشد [10].

رشد مشتریان و محصولات در سال های اخیر، چالش های کلیدی را برای پالایش گروهی به وجود آورده است. در چنین وضعیتی تولید توصیه ها با توجه به تعداد زیاد مشتریان و محصولات، زمان بر خواهد بود و منجر به توصیه های ضعیف خواهد شد. با استفاده از خوشه بندی دوطرفه هم مشتریان و هم محصولات، خوشه بندی می شوند تا مقیاس پذیری پالایش همکاری بالا رود و توصیه های بهتر تولید کند [43].

درواقع خوشه بندی دوجته یک رویکرد خوشه بندی است که در آن ردیف ها و ستون های یک ماتریس داده به طور هم زمان گروه بندی می شوند. این نوع خوشه بندی، مقیاس پذیرتر از خوشه بندی سنتی است و حاوی اطلاعات بیش تری است. از الگوریتم های خوشه بندی دوجته به الگوریتم های بایمکس<sup>1</sup> [29] که مبتنی بر یک استراتژی تقسیم و حل، الگوریتم ایکس موتیف<sup>2</sup> [26] که از طریق استخراج الگوهای ژنتیک اصلی از میان داده ها به دنبال خوشه های دوطرفه، الگوریتم امضای تکراری<sup>3</sup> (ISA) که به دنبال زیرماتریس هایی می شود که دارای ردیف ها و ستون های همبسته هستند؛ و الگوریتم بزرگترین میانگین زیرماتریس<sup>4</sup> (LAS) که ماتریس را با استفاده از میانگین زیرماتریس ها تقسیم می کند، می توان اشاره کرد [32].

تاکنون پژوهش های اندکی درخصوص به کارگیری خوشه بندی دوجته در فضای CF صورت گرفته است. به طور مثال نویسندگان [9] با توسعه نسخه های افزایشی و موازی خوشه بندی دو طرفه، یک CF مقیاس پذیر و بلادرنگ طراحی کردند. بزرگترین محدودیت این روش ها این است که هر کاربر یا مورد، فقط می تواند در یک خوشه قرار بگیرد. روش<sup>5</sup> و دیگران این مشکل را بر طرف کرده اند [37]. آنها با گروه بندی کاربران و موارد در زیرگروه های متعدد، روش های قدیمی خوشه بندی دوجته را توسعه داده اند. در پژوهشی دیگر، نویسندگان [35] یک روش جدید موسوم به "پالایش گروهی مبتنی بر نزدیکترین خوشه های دوجته" (NBCF<sup>6</sup>) ارائه کردند. در این روش، همسایگی کاربر فعال حاوی k تا از نزدیکترین خوشه های دوجته به وی است.

$$sim(\vec{a}, \vec{b}) = \frac{\vec{a} \cdot \vec{b}}{(|\vec{a}| * |\vec{b}|)} \quad (5)$$

مقدارهای ممکن برای شباهت کسینوسی عددی بین صفر و یک است که مقدارهای نزدیک یک، نشان دهنده شباهت بالا و مقدارهای نزدیک صفر، نشان دهنده شباهت پایین است.

۲. تشکیل همسایگی: در این مرحله اگر از روش مبتنی بر کاربر استفاده کرده باشیم، K تا از کاربرانی را که بیشترین شباهت را با کاربر فعال دارند، به عنوان همسایه های وی انتخاب می کنیم. اگر از روش مبتنی بر مورد استفاده کرده باشیم، K تا از مواردی که بیشترین شباهت را با مورد مورد نظر دارند، به عنوان همسایه های آن انتخاب می کنیم. یک تحلیل بر روی مجموعه داده movilens نشان می دهد که در بیشتر موقعیت های جهان واقعی انتخاب بیست تا پنجاه همسایه، منطقی است [12].

۳. پیش بینی: در این مرحله نوبت به پیش بینی رتبه کاربر فعال برای یک مورد می رسد. در روش مبتنی بر کاربر، پیش بینی بر اساس رتبه های داده شده از سوی همسایگان کاربر فعال به آن مورد انجام می شود؛ اما در روش مبتنی بر مورد، این کار از طریق رتبه های داده شده از سوی کاربر فعال به همسایگان آن مورد صورت می پذیرد. برای انجام پیش بینی از روش مجموع وزن دهی شده می توان استفاده کرد که برای حالت مبتنی بر کاربر به صورت رابطه (۶) خواهد بود [2]:

$$p_{a,i} = \frac{\sum_{u \in K} r_{u,i} \times sim(a,u)}{\sqrt{\sum_{u \in K} sim(a,u)}} \quad (6)$$

که در آن،  $p_{a,i}$  پیش بینی رتبه کاربر فعال a برای مورد i بوده و K تعداد همسایگان کاربر فعال است. درواقع شباهت محاسبه شده در مرحله نخست به عنوان اوزان پیش بینی استفاده می شود و در نتیجه کاربرهای مشابه تر، وزن بیشتری در پیش بینی خواهند داشت. معادل رابطه (۶) برای حالت مبتنی بر مورد به صورت رابطه (۷) خواهد بود [2]:

$$p_{a,i} = \frac{\sum_{u \in K} r_{u,i} \times sim(a,u)}{\sqrt{\sum_{u \in K} sim(a,u)}} \quad (7)$$

## ۴-۲- خوشه بندی دوجته

ایده اصلی خوشه بندی اطلاعات، جدا کردن نمونه ها از یکدیگر و قرار دادن آن ها در گروه های شبیه به هم است؛ به این معنا که نمونه های شبیه به هم در یک گروه قرار گرفته و با نمونه های دیگر بیشینه تفاوت را داشته باشند [1]. درواقع

<sup>1</sup> Bimax

<sup>2</sup> XMotif

<sup>3</sup> Iterative Signature Algorithm

<sup>4</sup> Large Average Submatrices

<sup>5</sup> Xu

<sup>6</sup> Nearest-Biclusters CF

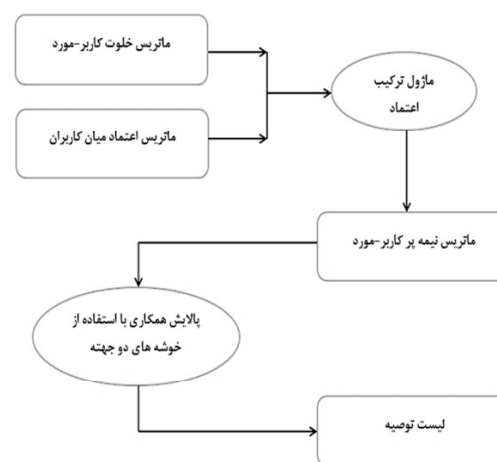
## ۵-۲- اعتماد

سامانه‌هایی که به جای پیداکردن شباهت بین کاربران (مانند سامانه‌های پالایش‌گروهی)، کاربران مورد اعتماد را به‌وسیله تخمین میزان اعتماد در شبکه پیدا می‌کنند و محصولاتی را که به‌وسیله کاربران مورد اعتماد پسندیده شده‌اند، به کاربر فعال توصیه می‌کنند، سامانه‌های توصیه آگاه به اعتماد (TARS) نامیده می‌شوند.<sup>۱</sup> در این سامانه‌ها، توصیه‌ها بر اساس اطلاعات به‌دست آمده از شبکه‌های اعتماد تولید می‌شوند. قدرت اصلی بیشتر این سامانه‌ها استفاده از عملکرد انتشار<sup>۲</sup> اعتماد است؛ سازوکاری که میزان اعتماد بین دو کاربر ناشناس در شبکه‌های اجتماعی را تخمین می‌زند [36].

سنجه انتشار اعتماد مبتنی بر این اصل است که رابطه اعتماد، یک رابطه تعدی است. بنابر رابطه تعدی، اگر کاربر a به کاربر b (که آن را سوم شخص مورد اعتماد می‌نامند) اعتماد داشته باشد و سوم شخص مورد اعتماد b به کاربر c اعتماد کند، پس منطقی است، فرض کنیم که a باید به c با یک درجه مشخص اعتماد داشته باشد [36].

## ۳- راه‌کار پیشنهادی

به‌منظور افزایش کارایی پالایش‌گروهی، در این مقاله یک رویکرد ترکیبی که متشکل از دو ماژول ترکیب اعتماد و ماژول پالایش‌گروهی با استفاده از خوشه‌بندی دوجته (به‌منظور افزایش مقیاس‌پذیری و صحت) است، پیشنهاد می‌شود. مدل پیشنهادی در شکل (۱) نشان داده شده است.



(شکل ۱-): مدل پیشنهادی  
(Figure-1): The proposed model

<sup>1</sup> Trust-Aware RS

<sup>2</sup> Propagation

## ۱-۳- ماژول نخست: ترکیب اعتماد با رتبه‌های

### کاربران

الگوریتم پیشنهادی ماژول ترکیب اعتماد با رتبه‌های کاربران در گام‌های زیر می‌تواند خلاصه شود:

گام نخست: با توجه به ماتریس اعتماد بین کاربران، فهرست افرادی را که مورد اعتماد کاربر هستند، به‌دست می‌آوریم. برای یافتن مجموعه افراد مورد اعتماد کاربر، از ماتریس اعتماد مابین کاربران استفاده می‌کنیم و فهرست افرادی را که به‌طور صریح توسط کاربر به عنوان معتمد شناخته شده‌اند، استخراج می‌کنیم. چون این مجموعه دارای اعضای کمی است، منجر به آن خواهد شد که دوباره، ماتریس کاربر مورد تاحدودی خلوت باشد. با استفاده از اصل انتشار اعتماد، اعتماد بین کاربر و دیگر کاربران را نیز به‌دست می‌آوریم. با توجه به این اصل که در دنیای مجازی با شش گام به هر کاربری می‌توان وصل شد،  $3^6$  و همکاران سه گام را برای انتشار اعتماد در شبکه‌های مجازی منطقی دانسته‌اند [11]. بنابراین ما نیز سه گام برای انتشار اعتماد در نظر می‌گیریم. گام دوم: در این مرحله موردهایی را که مجموعه کاربران مورد اعتماد (که در گام قبل استخراج شده‌اند)، رتبه‌بندی کرده‌اند از ماتریس کاربر مورد استخراج می‌کنیم. گام سوم: مجموعه رتبه‌بندی‌شده توسط کاربر فعال را از مجموعه استخراج‌شده در گام دوم کم می‌کنیم؛ زیرا این محصولات توسط خود کاربر رتبه‌بندی شده‌اند و نیازی به رتبه‌بندی مجدد آنها نیست.

گام چهارم: در این مرحله میانگین رتبه موردهایی را که تکرار شده‌اند، به‌دست می‌آوریم. گام پنجم: در این مرحله رتبه موردها که از مراحل قبل استخراج شده است، برای کاربر فعال در ماتریس کاربر مورد وارد می‌شود. بدین ترتیب ماتریس خلوت کاربر مورد به ماتریس نیمه‌پر تبدیل می‌شود.

## ۲-۳- ماژول دوم: پالایش‌گروهی

### بهبود یافته و مقیاس‌پذیر با استفاده

### از خوشه‌بندی دوجته

این ماژول، ماتریس کاربر مورد را به‌عنوان ورودی دریافت و هم‌زمان کاربران و موارد را با استفاده از الگوریتم بزرگترین میانگین زیرماتریس (LAS) خوشه‌بندی می‌کند؛ سپس

<sup>3</sup> Guo



دارند و  $P_{ik}$  زیرگروه  $k$  است که محصول در آن پیش‌بینی شده است. پس از محاسبه تمامی رتبه‌ها می‌توان  $N$  تا از محصولات را با بیشترین رتبه برای کاربر مورد نظر پیشنهاد داد [37].

#### ۴- پیاده‌سازی بر روی داده‌های واقعی

##### ۴-۱- مجموعه داده مورد استفاده

یک ویژگی مهم سایت‌های بررسی محصول مانند Epinions، وجود شبکه اعتماد در میان کاربران آن است. این سایت‌ها یک پلتفرم معقول برای مطالعه اعتماد در جهان برخط فراهم می‌کنند؛ بنابراین برای ارزیابی راه‌کار پیشنهادی از داده‌های موجود در سایت Epinions استفاده می‌کنیم. داده‌های این سایت به‌صورت برخط از طریق وب جهان گستر<sup>۱</sup> برای مطالعات دانشگاهی در دسترس است. داده‌های موجود، شامل دو بخش است که یک بخش مربوط به داده‌های رتبه‌دهی کاربران و دیگری مربوط به داده‌های اعتماد است. برای هر کاربر رتبه‌های وی و همچنین افراد مورد اعتماد وی مشخص شده‌اند. برای هر رتبه‌دهی، نام محصول، گروه محصول و رتبه‌ای که کاربر به محصول داده مشخص شده است. در این داده‌ها رتبه‌ها از یک تا پنج می‌باشند که یک نشان‌گر نپسندیدن و پنج نشان‌گر پسندیدن است. یک ردیف از داده‌های مربوط به اعتماد کاربران به‌صورت (پنج و چهار) است که نشان دهنده این است که کاربر چهار به کاربر پنج اعتماد دارد، اما برعکس آن ممکن است، صادق نباشد.

در این مجموعه داده، محصولات به ۲۷ گروه تقسیم‌بندی شده‌اند که ما برای ارزیابی راه‌کار پیشنهادی خود از داده‌های گروه سه که مربوط به گروه فیلم است، استفاده می‌کنیم. این مجموعه‌داده دارای برخی ناسازگاری‌ها نیز بود؛ به‌عنوان مثال برای بعضی محصولات در برخی موارد رتبه‌های تکراری وجود داشت که برای برطرف کردن این ناسازگاری‌ها، پیش‌پردازش‌هایی بر روی داده‌ها انجام و داده‌های تکراری حذف شدند. پس از پیش‌پردازش داده‌ها و حذف کاربرانی که کمتر از ده فیلم رتبه داده‌اند و محصولاتی که کمتر از ده بار دیده شده‌اند، مجموعه‌داده به ۵۱۹۶۲ رتبه از سوی ۱۲۴۰ کاربر برای ۱۵۲۸ محصول کاهش یافت. میزان خلوتی این داده‌ها برابر با ۹۷،۱۵ است که از تقسیم تعداد عناصر صفر ماتریس بر تعداد کل عناصر ماتریس کاربر مورد به‌دست می‌آید. تعداد کل عناصر ماتریس برابر ۱۸۲۴۷۲۰ است. لازم به ذکر است، خلوتی داده‌ها اغلب بیشتر از ۹۹٪ است [40].

پالایش گروهی به‌صورت جداگانه در زیرماتریس‌ها انجام و پیش‌بینی‌های حاصل از پالایش گروهی در زیرماتریس‌های جداگانه با هم ترکیب می‌شود و توصیه‌های نهایی را تولید می‌کند. از آنجایی که خوشه‌بندی دوجته بر روی ماتریس نیمه‌پر به‌جای ماتریس خلوت اعمال می‌شود، نتایج خوشه‌بندی دقیق‌تر خواهد بود [32].

پس از آن که خوشه‌ها مشخص شدند، الگوریتم پالایش گروهی را برای هر زیرماتریس که از خوشه‌بندی دوجته به‌دست آمده است، اجرا می‌کنیم. روش پالایش گروهی خالص یک ماتریس کاربر مورد را به‌عنوان ورودی دریافت کرده و پیش‌بینی‌هایی برای رتبه محصولاتی که توسط کاربر رتبه داده نشده‌اند تولید می‌کند. در اینجا زیرماتریس‌ها به‌عنوان ماتریس کاربر مورد به الگوریتم پالایش گروهی داده می‌شود و در خروجی رتبه‌هایی برای محصولاتی که توسط کاربر دیده نشده‌اند، تولید می‌شود. باید به این موضوع توجه کرد که ممکن است، کاربران یا محصولات، متعلق به یک یا دو یا چند زیرماتریس باشند و برای یک محصول رتبه‌های متفاوتی در زیرماتریس‌های مختلف تولید شود. نویسندگان مقاله [37] یک چارچوب یکپارچه‌سازی برای مدیریت این مسئله ارائه داده‌اند که در رابطه (۸) نشان داده شده است.

اگر  $pre = (u_i, y_j, k)$  را به‌عنوان پیش‌بینی کاربر  $i$  برای محصول  $z$  در زیر گروه  $k$  در نظر بگیریم که توسط پالایش گروهی محاسبه شده است و  $Y_{i,j}$  را پیش‌بینی نهایی کاربر  $i$  برای آیتم  $z$  در نظر بگیریم، برای محاسبه  $Y_{i,j}$  از رابطه (۸) می‌توانیم استفاده کنیم:

$$Y_{i,j} = \begin{cases} \sum_k pre(u_i, y_j, k) \cdot \delta_{i,k} & \text{اگر } u_i \text{ و } y_j \text{ متعلق به یک یا بیش از یک زیر ماتریس باشد} \\ 0 & \text{در غیر اینصورت} \end{cases} \quad (8)$$

در رابطه (۸)،  $\delta_{i,k}$  یک مقدار شاخص است که میزان علاقه‌مندی کاربر  $i$  به زیرگروه  $k$  را نشان می‌دهد که از رابطه (۹) محاسبه می‌شود:

$$\delta_{i,k} = \begin{cases} 1 & \text{if } P_{ik} \text{ is } \max(\frac{|I_i \cap I_k|}{|I_k|}) \\ 0 & \end{cases} \quad (9)$$

در رابطه (۹)  $I_i$  محصولاتی است که توسط کاربر رتبه‌دهی شده‌اند و  $I_k$  محصولاتی است که در گروه  $k$  وجود

<sup>۱</sup> [http://www.trustlet.org/wiki/Trust\\_network\\_datasets](http://www.trustlet.org/wiki/Trust_network_datasets)

## ۴-۲- معیار ارزیابی

معیار ارزیابی مورد استفاده، میانگین خطای مطلق (MAE) است. MAE، میزان انحراف رأی‌های پیش‌بینی شده از رأی‌های درست را مشخص می‌کند و با رابطه (۱۰) محاسبه می‌شود [12].

$$MAE = \frac{\sum_{i,j \in P} |pred_{i,j} - r_{i,j}|}{|p|} \quad (10)$$

P مجموعه آزمایش و حاوی مجموعه‌ای از پیش‌بینی‌ها برای زوج های کاربر مورد ناشناخته است.  $r_{i,j}$  و  $pred_{i,j}$  به ترتیب عبارتند از رتبه پیش‌بینی شده و رتبه واقعی برای کاربر i و مورد j. هرچقدر مقدار خطای مطلق میانگین کم باشد، سامانه از صحت بیشتری برخوردار است.

## ۴-۳- نتایج آزمایش‌ها

### ۴-۳-۱- نتایج آزمون MAE

به منظور ارزیابی سامانه پیشنهادی، به مقایسه کارایی آن با الگوریتم‌های پالایش گروهی مبتنی بر کاربر (CF-UB) و پالایش گروهی ترکیبی با اعتماد (CF-Trust) می‌پردازیم. ابتدا مجموعه کاربران به دو مجموعه آموزش و آزمون تقسیم می‌شوند؛ سپس داده‌های مجموعه آموزش برای ساختن مدل احتمالی یا پایگاه داده پالایش گروهی استفاده می‌شود. پس از آن کاربران در مجموعه آزمون را به عنوان کاربر فعال (کاربری که قصد داریم توصیه‌ها را برای وی ایجاد کنیم) در نظر می‌گیریم. بعد، رأی‌های داده شده توسط کاربران مجموعه آزمون را به دو مجموعه مشاهده شده Ia و مجموعه‌ای که باید پیش‌بینی شوند Pa تقسیم می‌کنیم. در آخر، با استفاده از رأی‌های داده شده در Ia رتبه‌های موجود در Pa را پیش‌بینی می‌کنیم. مواردی که در مجموعه آزمون مخفی می‌شوند، به طور تصادفی انتخاب می‌شوند. برای این کار بریز<sup>۱</sup> و همکاران پروتکلی به نام givenx را معرفی کردند که در آن برای هر کاربر آزمون، به طور تصادفی x مورد از مجموعه مشاهده نشده انتخاب و به عنوان داده‌های آموزش به الگوریتم داده می‌شود و بقیه موارد جهت ارزیابی مخفی می‌شود [5]. جدول (۲) نتایج آزمون MAE را برای ۲، ۵، ۱۰، ۱۵ برای هر ۳ روش نشان می‌دهد. K در هر سه روش، ۲۰ در نظر گرفته شده است.

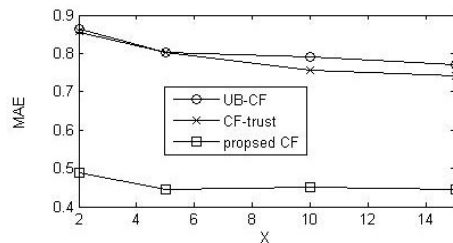
(جدول ۲): نتایج آزمون برای MAE

(Table-2): Test results for MAE

	MAE		
	CF-UB	CF-Trust	Proposed CF
Given2	0.8643	0.8564	0.4901
Given5	0.8023	0.8028	0.4464
Given10	0.7903	0.7547	0.4508
Given15	0.7715	0.7416	0.4445

<sup>۱</sup> Breese

در نمودار شکل (۲)، نتایج آزمون MAE برای ۲، ۵، ۱۰، ۱۵ x برای هر ۳ روش آورده شده است.



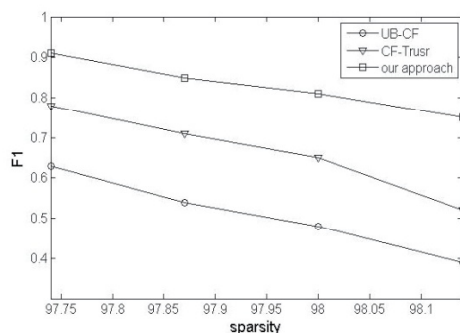
(شکل ۲): نتایج آزمون MAE  
(Figure-2): MAE test results

همان‌طور که مشاهده می‌شود، میانگین خطا در روش پیشنهادی به طور چشم‌گیری کاهش یافته است و سامانه پیشنهادی کارایی بیشتری نسبت به دو الگوریتم دیگر دارد.

### ۴-۳-۲- تأثیر بر روی مسئله خلوتی

در اینجا عملکرد سامانه پیشنهادی در برابر الگوریتم‌های محک‌زنی به ازای سطوح مختلف خلوتی مورد مقایسه قرار داده می‌شود. برای هر سطح خلوتی، به طور متوالی ۲۵۰۰ رتبه از مجموعه آموزش حذف و معیار f1 برای آن محاسبه می‌شود. آن‌چنان که در شکل (۳) نشان داده شده است، راه‌کار پیشنهادی، بالاترین عملکرد را برای سطوح مختلف خلوتی دارد.

همان‌طور که انتظار می‌رود با افزایش سطح خلوتی، عملکرد تمام الگوریتم‌ها کاهش می‌یابد؛ این به دلیل کاهش اندازه مجموعه آموزش است. مشاهده می‌شود که کارایی الگوریتم پیشنهادی تغییر کمتری نسبت به الگوریتم‌های دیگر دارد. بنابراین الگوریتم پیشنهادی در مقابل افزایش سطح خلوتی بهتر عمل کرده و بر این مشکل غلبه می‌کند.

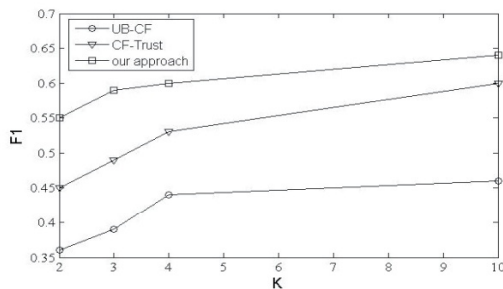


(شکل ۳): نتایج آزمون خلوتی  
(Figure-3): Sparsity test results

### ۴-۳-۳- تأثیر بر روی مسئله مقیاس پذیری

برای مقایسه مقیاس‌پذیری رویکردهای مختلف، زمان اجرای

با افزایش  $k$  تعداد کاربران جدید کاهش و بنابراین عملکرد تمامی الگوریتم‌ها افزایش و همان‌طور که مشاهده می‌شود رویکرد پیشنهادی، کارایی بهتری نسبت به الگوریتم‌های دیگر مورد آزمون، برای کاربران جدید دارد؛ زیرا به دلیل استفاده از اعتماد تعداد محصولات دیده شده توسط کاربر را می‌تواند افزایش دهد؛ البته به شرط آن که کاربر جدید دارای رابطه اعتماد با دیگر کاربران باشد.



(شکل-۵): نتایج آزمون برای کاربر جدید  
(Figure-5): Test results for the new user

جدول (۳) کارایی الگوریتم‌ها را برای مشکلات مقیاس‌پذیری، خلوتی و کاربر جدید نشان می‌دهد.

(جدول-۳): مقایسه الگوریتم‌ها  
(Table-3): Comparison of algorithms

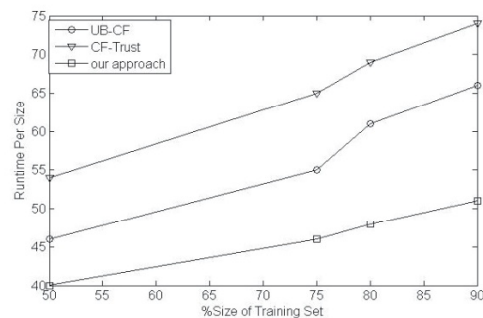
الگوریتم	مقیاس‌پذیری	خلوتی	کاربر جدید
پالایش گروهی مبتنی بر کاربر	۲*	۳	۳
پالایش گروهی ترکیبی با اعتماد	۳	۲	۲
الگوریتم پیشنهادی	۱	۱	۱

\*اعداد نشان‌دهنده رتبه کارایی الگوریتم‌ها هستند. ۱ نشانگر کارایی بالا و ۳ نشانگر کارایی پایین است.

## ۵- نتیجه‌گیری و پیشنهادهای آینده

تا به حال انواع مختلفی از سامانه‌های توصیه‌گر به وجود آمده‌اند که هر کدام از این سامانه‌ها مزایا و معایبی دارند. بیشتر سامانه‌های توصیه‌گر، مبتنی بر پالایش گروهی هستند. در پالایش گروهی از شباهت میان کاربران یا موارد برای توصیه محصولات استفاده می‌شود. در این مقاله، یک سامانه توصیه‌گر ترکیبی با استفاده از اعتماد و خوشه‌بندی دوچهار پیشنهاد شد که کارایی پالایش گروهی را بهبود بخشد. برای ارزیابی سامانه پیشنهادی با پالایش گروهی کاربرمحور و پالایش گروهی ترکیبی با اعتماد مقایسه شد که نتایج ارزیابی سامانه پیشنهادی و مقایسه آن با دو راه‌کار دیگر در مواجهه با مشکلات خلوتی، مقیاس‌پذیری و شروع سرد، نشان از کارایی بالای سامانه پیشنهادی دارد.

بخش برخی از آن‌ها اندازه‌گیری می‌شود. از آنجا که محاسبات مربوط به مازول ترکیب اعتماد و خوشه‌بندی دوچهار به صورت غیربرخط محاسبه می‌شود، لذا در محاسبه زمان اجرای الگوریتم پیشنهادی تأثیر نداشته و تنها زمان اجرای پالایش گروهی بر زیرماتریس‌ها به عنوان زمان اجرای الگوریتم در نظر گرفته می‌شود. برای آزمون مسئله مقیاس‌پذیری، متوسط زمان لازم (به میلی ثانیه) برای ارائه فهرست توصیه‌ها به یک کاربر آزمون، برای اندازه‌های مجموعه آموزش متفاوت اندازه‌گیری شده است. در اینجا برای ایجاد مجموعه آموزش با اندازه‌های مختلف از اعتبارسنجی متقابل  $k$  تایی (۲ و ۳ و ۴ و ۱۰) استفاده شده است. در شکل (۴) نتایج آزمون آورده شده است. همان‌طور که انتظار می‌رود با افزایش اندازه مجموعه آموزش، زمان اجرا برای تمام رویکردها افزایش می‌یابد. در الگوریتم پیشنهادی چون خوشه‌ها به صورت غیربرخط ساخته می‌شوند و اندازه زیرماتریس‌ها نسبت به ماتریس اصلی کوچک‌تر است، زمان کمتری برای ایجاد فهرست توصیه‌ها نیاز است. بنابراین الگوریتم پیشنهادی مقیاس‌پذیرتر از الگوریتم‌های دیگر مورد آزمون است.



(شکل-۴): نتایج آزمون برای مقیاس‌پذیری  
(Figure-4): Scalability test results

## ۴-۳-۴- تأثیر بر روی مسئله کاربر جدید

تعداد محصولات دیده‌شده توسط کاربر جدید بسیار کم است که این مسئله پیدا کردن کاربران مشابه را با مشکل روبه‌رو می‌کند. برای آزمون مسئله کاربر جدید، در ابتدا با استفاده از اعتبارسنجی متقابل  $k$  تایی، کاربران به مجموعه آموزش و آزمون تفکیک شده‌اند. کاربرانی که کمتر از بیست رتبه دارند، به عنوان کاربر جدید (آزمون) در نظر گرفته شده‌اند و از پروتکل given2 استفاده شده است و به طور تصادفی دو رتبه به عنوان مشاهده‌شده در مجموعه آموزش قرار داده شده است. پس از آن فهرست توصیه‌ها برای کاربران جدید تولید شده است. شکل (۵) سنجه  $f1$  را به ازای  $k$ های مختلف نشان می‌دهد.

- [6] D. U. Yongping, and L. HUANG, "Improve the Collaborative Filtering Recommender System Performance by Trust Network Construction", *Chinese Journal of Electronics*, 25(3):418-23, 2016 May.
- [7] A. Felfering, G. Friedrich, and L. S. Thieme, "Recommender Systems", *Intelligent Systems IEEE*, 22, pp. 18-22, 2007.
- [8] L. Gao, and C. Li, "Hybrid personalized recommended model based on genetic algorithm", *In 2008 4th International Conference on Wireless Communications, Networking and Mobile Computing*, IEEE, 12, pp. 1-4, 2008 Oct.
- [9] T. George, and S. Merugu, "A scalable collaborative filtering framework based on co-clustering", *In Fifth IEEE International Conference on Data Mining (ICDM'05)*, IEEE, pp. 4-pp, 2005 Nov 27.
- [10] F. Gorunescu, "Data Mining: Concepts, models and techniques", *Springer Science & Business Media*, Vol. 12, 2011.
- [11] G. Guo, J. Zhang, and D. Thalmann, "Merging trust in collaborative filtering to alleviate data sparsity and cold start", *Knowledge-Based Systems*, 57:57-68, 2014.
- [12] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating collaborative filtering recommender systems", *ACM Transactions on Information Systems (TOIS)*, 22(1):5-3, 2004 Jan 1.
- [13] Y. Ho, D. Fong, and Z. Yan, "A Hybrid GA-based Collaborative Filtering Model for Online Recommenders", *In Ice-b*, pp. 200-203, 2007.
- [14] H. Ingoo, J. O. Kyong, and H. R. Tae, "The collaborative filtering recommendation based on SOM cluster-indexing CBR", *Expert Systems with Applications*, 25(3):413-23, 2003.
- [15] D. Jannach, and M. Zanker, A. Felfering, and G. Friedrich, "Recommender Systems an introduction", *Cambridge University Press*, New York, 2013.
- [16] R. Katarya, and O. P. Verma, "A collaborative recommender system enhanced with particle swarm optimization technique", *Multimedia Tools and Applications*, 1-5, 2016.
- [17] K. J. Kim, and H. Ahn, "A recommender system using GA K-means clustering in an online shopping market", *Expert systems with applications*, 34(2):1200-9, 2008 Feb 29.
- [18] S. C. Kim, C. S. Park, and S. K. Kim, "A Hybrid Recommendation System Using Trust Scores in a Social Network", *Embedded and Multimedia Computing Technology and Service*, 107-112, 2012.
- [19] U. Kuzelewska, and K. Wichowski, "A Modified Clustering Algorithm DBSCAN Used in a

چندین مورد در پژوهش‌های آینده می‌تواند لحاظ شود: در این مقاله، الگوریتم پیشنهادی تنها با الگوریتم پالایش‌گروهی مبتنی بر کاربر و پالایش‌گروهی ترکیبی با اعتماد مقایسه شد. پژوهش‌های آینده در راستای مقایسه الگوریتم پیشنهادی با دیگر الگوریتم‌های موجود در این زمینه مانند سامانه‌های توصیه‌گر مبتنی بر جمعیت یا سامانه‌های توصیه‌گر ترکیبی می‌تواند انجام شود. از سویی برای ارزیابی الگوریتم پیشنهادی، از داده‌های موجود در گروه فیلم تارنمای epinions استفاده شده است. این پژوهش را بر روی مجموعه داده‌های مختلف و با حجم‌های بیشتر می‌توان انجام داد تا کارایی و اثربخشی آن برای مجموعه داده‌های مختلف مورد بررسی قرار گیرد. از سویی الگوریتم خوشه‌بندی که برای این پژوهش استفاده شد، الگوریتم LAS است. پژوهش را با الگوریتم‌های خوشه‌بندی دوجهته دیگر می‌توان انجام داد تا مشخص شود کدام یک از الگوریتم‌های خوشه‌بندی دارای کارایی و اثربخشی بالا هستند. همچنین بر روی الگوریتم خوشه‌بندی که مختص سامانه‌های توصیه‌گر باشد می‌توان پژوهش کرد.

## 6- References

## ۶- مراجع

- [۱] علیزاده، حسین، مشکی، محسن، پروین، حمید، مینایی بیدگلی، بهروز، "خوشه‌بندی ترکیبی مبتنی بر زیرمجموعه‌ای از خوشه‌های اولیه"، پردازش علائم و داده‌ها، ۷ (۱): ۳۲-۱۹، ۱۳۸۹.
- [1] H. Alizadeh, M. Moshki, H. Parvin, B. Minaei Bidgoli, "Clustering Ensemble based on combination of subset of primary clusters", *JSDP*, 7 (1):19-32, 2010.
- [2] G. Adomavicius, and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions", *IEEE transactions on knowledge and data engineering*, 17(6):734-49, 2005 Jun,
- [3] C. Basu, H. Hirsh, and W. Cohen, "Recommendation as classification: Using social and content-based information in recommendation", *In Aaai/iaai*, pp. 714-720, 1998 Jul 1.
- [4] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey", *Knowledge-Based Systems*, 46:109-32, 2013.
- [5] J. S. Breese, D. Heckerman, and C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering". *In Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, Morgan Kaufmann Publishers Inc, pp. 43-52, 1998 Jul 24.

- "Item-based collaborative filtering recommendation algorithms", *InProceedings of the 10th international conference on World Wide Web*, ACM, pp. 285-295, 2001 Apr 1.
- [31] J. B. Schafer, J. Konstan, and J. Riedl, "Recommender systems in e-commerce", *InProceedings of the 1st ACM conference on Electronic commerce*, ACM, pp. 158-166, 1999 Nov 1.
- [32] A. A. Shabalin, V. J. Weigman, C. M. Perou, and A. B. Nobel, "Finding large average submatrices in high dimensional data", *The Annals of Applied Statistics*, 985-1012, 2009 Sep 1.
- [33] X. Shen, H. Long, and C. Ma, "Incorporating trust relationships in collaborative filtering recommender system", *InSoftware Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, 2015 16th IEEE/ACIS International Conference, pp. 1-8, 2015 Jun 1.
- [34] T. H. Soliman, S. A. Mohamed, and A. A. Sewisy, "Developing a mobile location-based collaborative Recommender System for GIS applications", *InComputer Engineering & Systems (ICCES), 2015 Tenth International Conference*, IEEE, pp. 267-273, 2015 Dec 23.
- [35] P. Symeonidis, A. Nanopoulos, A. N. Papadopoulos, and Y. Manolopoulos, "Nearest-biclusters collaborative filtering based on constant and coherent values", *Information retrieval*, 11(1):51-75, 2008 Feb 1;
- [36] P. Victor, M. De Cock, and C. Cornelis, "Trust and recommendations", *InRecommender systems handbook 2011*, Springer US, pp. 645-675, 2011.
- [37] B. Xu, J. Bu, C. Chen, and D. Cai, "An exploration of improving collaborative recommender systems via user-item subgroups", *InProceedings of the 21st international conference on World Wide Web*, ACM, pp. 21-30, 2012 Apr 16.
- [38] R. R. Yager, "Fuzzy logic methods in recommender systems", *Fuzzy Sets and Systems*, 136(2):133-49, 2003 Jun 1.
- [39] S. Yan, "A Collaborative Filtering Recommender Approach by Investigating Interactions of Interest and Trust", *InKnowledge Engineering and Management*, Springer Berlin Heidelberg, pp. 173-188, 2014.
- [40] X. Yang, Y. Guo, Y. Liu, and H. Steck, "A survey of collaborative filtering based social recommender systems", *Computer Communications*, 41:1-0, 2014 Mar 15.
- [41] W. Yuan, D. Guan, Y. K. Lee, S. Lee, and S. J. Hur, "Improved trust-aware recommender system using small-worldness of trust networks", *Knowledge-Based Systems*, 23(3):232-8, 2010 Apr 30.
- Collaborative Filtering Recommender System for Music Recommendation", *InTheory and Engineering of Complex Systems and Dependability*, Springer International Publishing, pp. 245-254, 2015.
- [20] N. Lathia, S. Hailes, and L. Capra, "Trust-based collaborative filtering", *InIFIP International Conference on Trust Management 2008 Jun 18*, Springer US, pp. 119-134, 2008.
- [21] T. Q. Lee, Y. Park, and Y. T. Park, "A time-based approach to effective recommender systems using implicit feedback", *Expert systems with applications*; 34(4):3055-62, 2008 May 31
- [22] Y. M. Li, C. T. Wu, and C. Y. Lai, "A social recommender mechanism for e-commerce: Combining similarity, trust, and relationship", *Decision Support Systems*, 55(3):740-52, 2013 Jun 30.
- [23] X. Luo, Y. Xia, and Q. Zhu, "Incremental collaborative filtering recommender based on regularized matrix factorization", *Knowledge-Based Systems*, 27:271-80, 2012 Mar 31.
- [24] N. Manouselis, and K. Verbert, "Layered evaluation of multi-criteria collaborative filtering for scientific paper recommendation", *Procedia Computer Science*, 18:1189-97, 2013 Dec 31.
- [25] L. Martinez, R. M. Rodriguez, and M. Espinilla, "Reja: a georeferenced hybrid recommender system for restaurants", *InProceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*, IEEE Computer Society, Volume 03, pp. 187-190, 2009 Sep 15.
- [26] T. M. Murali, and S. Kasif, "Extracting conserved gene expression motifs from gene expression data", *InPacific symposium on Biocomputing*, Vol. 8, pp. 77-88, 2003.
- [27] M. Nilashi, O. Ibrahim, N. Ithnin, and N. H. Sarmin, "A multi-criteria collaborative filtering recommender system for the tourism domain using Expectation Maximization (EM) and PCA-ANFIS", *Electronic Commerce Research and Applications*, 14(6):542-62, 2015 Nov 30.
- [28] M. H. Park, J. H. Hong, and S. B. Cho, "Location-based recommendation system using bayesian user's preference model in mobile devices", *InInternational Conference on Ubiquitous Intelligence and Computing*, Springer Berlin Heidelberg, pp. 1130-1139, 2007 Jul 1.
- [29] A. Prelić, S. Bleuler, P. Zimmermann, A. Wille, P. Bühlmann, W. Gruissem, L. Hennig, L. Thiele, and E. Zitzler, "A systematic comparison and evaluation of biclustering methods for gene expression data", *Bioinformatics*, 22(9):1122-9, 2006 May 1.
- [30] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl,



الگوریتم‌های فراابتکاری و شبکه‌های اجتماعی است.  
نشانی رایانامه ایشان عبارت است از:

[a.baghaci@mail.kntu.ac.ir](mailto:a.baghaci@mail.kntu.ac.ir)

[42] N. Zheng, and Q. Li, "A recommender system based on tag and time information for social tagging systems", *Expert Systems with Applications*, 38(4):4575-87, 2011 Apr 30.

[43] R. Zhu, and S. Gong, "Analyzing of collaborative filtering using clustering technology", *In Computing, Communication, Control, and Management. CCCM 2009, ISECS International Colloquium on Vol. 4*, pp. 57-59, IEEE, 2009.



**منیره حسینی** مدرک کارشناسی خود

را در مهندسی برق گرایش مخابرات از  
دانشگاه علم و صنعت در سال ۱۳۷۵ و

مدارک کارشناسی ارشد و دکترای خود را  
در رشته‌های مهندسی صنایع گرایش

مدیریت سیستم و مهندسی صنایع گرایش فناوری اطلاعات  
به ترتیب از دانشگاه صنعتی شریف و دانشگاه تربیت مدرس در  
سال‌های ۱۳۸۲ و ۱۳۸۹ دریافت کرد. وی استادیار و عضو  
هیئت علمی گروه مهندسی فناوری اطلاعات دانشگاه صنعتی  
خواجه نصیرالدین طوسی است و مقالات متعددی در  
زمینه‌های فناوری اطلاعات در نشریات علمی پژوهشی و  
آی-اس-آی دارد. زمینه‌های پژوهشی مورد علاقه ایشان  
شبکه‌های اجتماعی، مدیریت ارتباط با مشتری و بازاریابی  
اینترنتی است.

نشانی رایانامه ایشان عبارت است از:

[hosseini@kntu.ac.ir](mailto:hosseini@kntu.ac.ir)



**مقصود نصرالهی** در سال ۱۳۹۳ مدرک

کارشناسی ارشد خود را در رشته مهندسی

فناوری اطلاعات گرایش مدیریت

سامانه‌های اطلاعاتی از دانشگاه صنعتی

خواجه نصیرالدین طوسی دریافت کرد.

زمینه‌های پژوهشی مورد علاقه ایشان سامانه‌های توصیه‌گر و  
بازاریابی اینترنتی است.

نشانی رایانامه ایشان عبارت است از:

[maghsod68@gmail.com](mailto:maghsod68@gmail.com)



**علی بقائی** در سال ۱۳۹۲ مدرک

کارشناسی خود را در رشته مهندسی

کامپیوتر گرایش نرم‌افزار از دانشگاه

گیلان و در سال ۱۳۹۵ کارشناسی ارشد

خود را در رشته مهندسی فناوری اطلاعات

گرایش تجارت الکترونیک از دانشگاه صنعتی خواجه  
نصیرالدین طوسی دریافت کرد. زمینه‌های پژوهشی مورد  
علاقه ایشان، مدیریت ارتباط با مشتری، داده‌کاوی،

فصل ۳

