دوفصلنامه محاسبات و سامانه‌های توزیع شده
www.ijdcs.ir

# سیستم‌های توصیه گر گردشگری سفر بر اساس الگوریتم خفاش و فیلترینگ ترکیبی

آرش خسروی*[1]، محمدعلی صادقی[2]
گروه مهندسی کامپیوتر، مرکز آموزش عالی محلات[1]
دانشکده مهندسی برق و کامپیوتر، دانشگاه غیردولتی شهاب دانش، قم[2]

## چکیده

حجم بسیار و روبه رشد اطلاعات بر روی اینترنت، فرآیند تصمیم‌گیری و انتخاب اطلاعات، داده یا کالاهای موردنیاز را، برای بسیاری از کاربران وب دشوار کرده است. سامانه‌های پیشنهاددهنده (توصیه گر)[1] ، باهدف رفع این چالش به وجود آمده‌اند و تلاش می‌کنند تا از میان حجم عظیم اطلاعات، اطلاعات خاص و مفید را با توجه به علاقه و سلیقه کاربر و تجربیات کاربران گذشته به وی پیشنهاد دهند. تاکنون سامانه‌های پیشنهاددهنده زیادی در زمینه‌های کاربردی متنوع ازجمله فیلم، موسیقی، کتاب و ... ایجادشده‌اند. انتخاب یک سفر مناسب، پیشنهاد هتل و ... با توجه به بودجه‌ی فرد، معمولاً سختی‌ها و نگرانی‌های زیادی را برای کاربران به همراه دارد و عموماً با صرف زمان و انرژی زیادی انجام می‌گیرد. لذا در این مقاله یک سیستم پیشنهاددهنده سفر و هتل ارائه می‌شود که از ترکیب روش فیلترهای مختلف ساخته‌شده است تا دقت آن دوچندان شود. این سیستم برای ارائه پیشنهادهای نهایی خود، سلایق کاربر جاری، کیفیت مجموعه‌های خدمات دهنده و تجربیات گذشته کاربران مشابه با کاربر جاری را مدنظر قرار داده و بدین ترتیب علاوه بر ارائه پیشنهادهای دقیق‌تر، مشکل شروع سرد[2] را که معمولاً برای کاربران جدید بروز می‌کند که در سیستم ثبت‌نام می‌کنند و سیستم هیچ اطلاعاتی از نظرات یا علایق کاربر ندارد، نیز برطرف می‌نماید. در چنین شرایطی، سامانه‌ها معمولاً از یادگیری فعال[3] یا استفاده از ویژگی‌های شخصیتی کاربر، برای حل مشکل استفاده می‌کنند.

کلمات کلیدی: سیستم پیشنهاددهنده، فیلترینگ ترکیبی، الگوریتم خفاش، فیلترینگ مشارکتی، فیلترینگ مبتنی بر محتوی

# Travel Tourism Recommendation Systems Based on Bat Algorithm and Hybrid Filtering

Arash Khosravi[1], Mohammad Ali Sadeghi[2]
Mahallat Institute of Higher Education, Mahallat, Iran[1]
Shahab Danesh University, Qom, Iran[2]

**Abstract**

The growing amount of information on the internet has made it difficult for many web users to make the decision-making and selection of information, data or goods. Recommended systems are designed to address this challenge and try to offer specific and useful information with respect to user tastes and past user experiences. So far, many offering systems have been developed in a variety of applications including movies, music, books, hotels etc. Choosing the right trip, the hotel proposal and so on, with regard to the individual's budget usually have a lot of difficulties and concerns for users and generally takes a lot of time and energy. In this paper, a travel and hotel recommendation system is developed which is constructed from combination of different filtering methods to maximize accuracy. The system is considering the current user's preferences, the quality of the service packages and past experiences of the same users with the current user in order to providing more accurate suggestions. It also eliminates the cold start problem.

**تاریخچه مقاله:**
تاریخ ارسال: ۹۸/۱۰/۰۱
تاریخ اصلاحات: ۹۸/۱۲/۰۱
تاریخ پذیرش: ۹۸/۱۲/۱۵
تاریخ انتشار: ۹۸/۱۲/۲۰

**Keywords:**
*Recommended System*
*Bat Algorithm*
*Hybrid Filtering*
*Collaboration Filtering*
*Content-Based Filtering*
*Meta-Heuristic Algorithm*

روش ارجاع به مقاله: سیستم‌های توصیه‌گر گردشگری سفر بر اساس الگوریتم خفاش و فیلترینگ ترکیبی آ. خسروی، م. صادقی، دوفصلنامه محاسبات و سامانه های توزیع شده، سال دوم، شماره دوم، شماره پیاپی ۴، سال ۱۳۹۸، ص ۸۸ تا ۱۱۰

Email: Khosravi.280@gmail.com

[1] Recommend Systems [2] Cold Start Problem [3] Active Learning

## ۱ – مقدمه

اینترنت با سرعت بسیار زیادی شروع به رشد کرده و فرصتی برای به اشتراک گذاری دانش و همچنین به وجود آمدن شبکه‌های اجتماعی شده است. هدف اصلی سامانه‌های توصیه گر، تولید توصیه‌های بامعنی به گروهی از کاربران است که آن کاربران به آن گروه از محصولات یا اقلام و... علاقه دارند.

سامانه‌های توصیه گر، اطلاعات کاربران را در مورد مجموعه‌ای از آیتم ها که به آن ها علاقه دارند را به‌صورت صریح[1] (معمولاً از طریق امتیازدهی به آیتم‌ها توسط کاربران) یا ضمنی[2] (معمولاً، با کنترل رفتار کاربران؛ مانند، آخرین آهنگ شنیده‌شده توسط کاربر، آخرین فیلم دیده شده و یا وب سایت‌های بازدید شده توسط کاربر) جمع‌آوری کرده و از آن‌ها برای پیشنهاد دادن آیتم‌ها به کاربران در آینده استفاده می‌نماید. این سامانه‌ها به‌طور مستقیم به کاربران کمک می‌کند تا به محتوا، محصولات و یا سرویس‌های موردنظرشان دست یابند ( مانند، کتاب، محصولات دیجیتالی، فیلم، موسیقی، برنامه‌های تلویزیونی و...) و ازجمله ابزارهایی است که می‌تواند کاربران را در محیط‌های الکترونیکی به سمت یافتن اطلاعات، خدمات و آیتم های موردنظرشان هدایت کند [۵، ۶، ۷].

به علت گسترش اینترنت، ضرورت استفاده از سامانه‌های توصیه گر برای فیلتر کردن اطلاعات تا حد زیادی افزایش‌یافته است. برخی شرکت‌های بزرگ مانند نت فلیکس و آمازون که به‌صورت آنلاین می باشند، از این سیستم برای آسان‌تر پیدا کردن آیتم‌هایشان برای کاربران استفاده می‌کنند.

سامانه‌های توصیه گر باقابلیت کشف علایق کاربران و پیش‌بینی اولویت‌های آن‌ها، آیتم‌هایی که احتمال می‌رود موردتوجه کاربر باشد را از بین حجم بالای داده‌ها پالایش کرده و با پیشنهاد آن‌ها، در زمان او صرفه‌جویی می‌کند. از طرف دیگر این سامانه‌ها با توانایی ذخیره و تحلیل رفتارهای گذشته کاربر، خدمات و اطلاعاتی را که موردتوجه کاربران نبوده ولی احتمالاً به آن‌ها علاقه‌مند هستند[3] را نیز استنتاج کرده و نتایج جالب‌توجهی به کاربران ارائه می‌کند. درواقع سامانه‌های توصیه گر یکی از ابزارهای اصلی برای غلبه بر مشکل افزونگی اطلاعات[4] بوده و با داشتن قدرت تحلیل رفتارهای کاربر، مکمل هوشمندی برای مفاهیم بازیابی و پالایش اطلاعات است.

منظور از افزونگی، به حالتی می گویند که فرد تصمیم‌گیرنده با یک مجموعه اطلاعاتی مواجه می‌شود که آن اطلاعات حجیم، پیچیده و بعضاً هم حاوی تناقضاتی هستند به‌طوری‌که این پیچیدگی و تنوعِ بیش‌ازحد اطلاعات باعث می‌شود فرد تصمیم‌گیرنده در مشخص کردن بهترین تصمیم ممکن دچار مشکل بشود.

[1] Explicit
[2] Implicit
[3] Cross Selling
[4] Redundancy Of Information

در اوا سط دهه ۱۹۸۰، آژانس‌های م سافرتی شروع به شـکایت از بازاریابی مسـتقیم[5] کردند که این شـیوه درواقع گونه‌ای از تبلیغات اسـت که به کسـب‌وکارها و سازمان‌ها اجازه می‌دهد تا از طریق انواع ر سانه‌ها نظیر ت ماس های تلفنی، پ یام های متنی، ایم یل ها، توزیع کا تالوگ، نا مه های تبلیغاتی و کمپین های هدفمند تلویزیونی به صورت م ستقیم با م شتریان ارتباط برقرار کنند؛ زیرا آن‌ها عقیده داشتند که این کار خطر بزرگی برای شبکه آژانس‌ها است.

در دهـه ۱۹۹۰، ایـن فشـار همچنـان وجـود داشـت. برای مثـال، سـفرهای قـایقی در تبلیغـات، مشـتریان را بـه گـرفتن رزرو بـا اسـتفاده از کانـال‌هـای تلویزیـونی تشـویق مـی‌کردنـد. بـه‌عـلاوه، در اواخـر سال ۱۹۹۵، خطـوط هواپیمـایی آمریکـا یـک آگهـی تلویزیـونی را بـه کـاربرد کـه سیسـتم خریـد بلیت‌های الکترونیکی را تبلیغ می‌کرد.

مـدیریت بسـیاری از هتـل‌هـا بـه‌منظـور اسـتفاده از پتانسـیل رو بـه رشـد اینترنـت و بـه کـار بـردن آن بـرای اهـداف مختلـف، وب‌سـایت‌هـای خـود را طراحـی کردنـد. ورای ایـن حقیقـت کـه اینترنـت بیشـتر بـه‌عنـوان ابـزاری بـرای بازاریـابی و فـروش استفاده مـی‌شـود، بـه‌عنـوان یـک کانـال رزرو کـردن هـم مورداسـتفاده قـرار مـی‌گیـرد. علـی‌رغـم ایـن پیشـرفت‌هـا، بسـیاری از هتـل‌هـا همچنـان عقیـده دارنـد کـه روش‌هـای قـدیمی تبلیـغ ماننـد رادیـو، تلویزیـون و رسـانه‌هـای چـاپی مـؤثرترین روش‌هـای معرفی هستند.

در منـابع نیـز تعـداد قابـل‌تـوجهی از مطالعـات، دربـاره رابطـه بـین بخـش گردشـگری و اینترنـت وجـود دارد. پیشـرفت نمونـه‌هـای اولیـه خـدمات اطلاعـاتی را کـه در وب‌سـایت‌هـا ایجادشـده اسـت، مورد بحث قرار می‌دهند.

ابراهیم پور، پور حسین کلایی و همکاران (۱۳۹۶) تحقیقی تحت عنوان «بررسی تأثیر عوامل توصیه گر بر رفتار خرید برنامه‌ریزی نشده (مطالعه موردی: فروشگاه های زنجیره‌ای ایران کتان استان مازندران)» انجام داده‌اند. جامعه آماری آن، کلیه مشتریان فروشگاه‌های زنجیره‌ای ایران کتان در استان مازندران هستند؛ نمونه آماری با استفاده از فرمول جامعه آماری نامحدود و روش نمونه‌برداری تصادفی ساده ۳۸۴ نفر تخمین زده و برای گردآوری داده‌ها از روش میدانی که در طراحی پرسش‌نامه از مقیاس هاستلر استفاده گردید. آزمون فرضیه‌های پژوهش با استفاده از مدل‌سازی معادلات ساختاری و با به‌کارگیری نرم‌افزار لیزرل[6] انجام شد. درنهایت، هر شش فرضیه پژوهش تائید شد. یافته‌های پژوهش نشان می‌دهد استفاده از عامل توصیه گر بر اثربخشی ارتقاء محصول و اثربخشی جستجوی محصولات تأثیر مثبت و معنی‌داری دارد. اثربخشی ارتقاء محصول نیز بر اثربخشی جستجوی محصولات و رضایت مشتری تأثیر مثبت و معنادار دارد. تأثیر مثبت و معنادار اثربخشی جستجوی محصولات بر رفتار خرید برنامه‌ریزی نشده نیز تائید شد. درنهایت نیز نشان داده شد که رضایت مشتری بر رفتار خرید برنامه‌ریزی نشده تأثیر مثبت و معناداری دارد. [۱]

[5] Direct Marketing

[6] Lisrel

رجب‌زاده و رافع (۱۳۹۶) به پژوهشی تحت عنوان «ارائه یک سیستم توصیه گر ترکیبی برای تجارت الکترونیک» پرداختند. رشد روزافزون تجارت الکترونیک در فضای مجازی باعث معرفی محصولات متنوع و زیادی از سوی شرکت‌های فعال در این حوزه شده است. در چنین شرایطی انتخاب مناسب و بهینه محصولات از میان حجم انبوه اطلاعات ارائه‌شده برای کاربران کاری مشکل است. [۲]

برای پایه‌ریزی و ایجاد یک سیستم پیشنهاددهنده کارا مواردی وجود دارند که باید به آن‌ها توجه نمود:

۱- نوع داده‌های موجود در بستر سیستم: بنا به کاربرد سیستم ممکن است انواع مختلفی از منابع اطلاعاتی در سیستم وجود داشته باشد. این اطلاعات می‌توانند امتیازهای کاربران به آیتم‌ها، اطلاعات شخصی کاربران، محتوای مربوط به آیتم‌های سیستم ارتباطات موجود در شبکه‌های اجتماعی و اطلاعات مربوط به موقعیت کاربر، باشند. طبیعی است که در پروسه طراحی یک سیستم پیشنهاددهنده باید به نوع داده‌های در اختیار توجه بسیار نمود.

۲- الگوریتم پالایش[۷] مورداستفاده: هدف سامانه‌های پیشنهاددهنده درواقع رتبه‌بندی آیتم‌های سیستم به لحاظ نزدیک بودن به علایق کاربران هست تا در هنگام ارائه پیشنهاد
آیتم‌هایی با رتبه بالاتر را به کاربر پیشنهاد دهند.
برای این منظور الگوریتم‌های متعددی پیشنهادشده‌اند که مهم‌ترین آن‌ها عبارت‌اند از:

۱. الگوریتم مبتنی بر پالایش مشارکتی[۸]
۲. الگوریتم مبتنی بر محتوا[۹]
۳. الگوریتم‌های مبتنی بر دانش[۱۰]
۴. الگوریتم پالایش جمعیت شناختی[۱۱]
۵. الگوریتم پالایش ترکیبی[۱۲]

۳-مدل انتخابی برای سیستم: در حال حاضر برای پیاده‌سازی سامانه‌های پیشنهاددهنده دو راهکار استفاده می‌شود. یکی استفاده مستقیم از داده‌های موجود سیستم است که به این راهکار، روش مبتنی بر حافظه گفته می‌شود و دیگری که کمی هوشمندتر است استفاده از یک مدل در سیستم است که بدان راهکار مبتنی بر مدل گفته می‌شود.

۴- تکنیک مورداستفاده در پیشنهاد دادن: راهکارها و فن‌های مختلفی برای پیاده‌سازی هسته یک سیستم پیشنهاددهنده وجود دارد. به‌عنوان‌مثال می‌توان از الگوریتم‌های ژنتیک، شبکه‌های عصبی، روش‌های احتمالی و با الگوریتم‌های همسایگی برای پیاده‌سازی هسته چنین سامانه‌هایی استفاده می‌شوند. در جلوتر در مورد هر یک از این الگوریتم‌های پالایش اشاره‌شده، توضیحاتی را خواهیم داد.

۵- مقیاس‌پذیری مورد انتظار سیستم، کارایی مطلوب سیستم (ازنظر میزان حافظ مصرفی و زمان).

### ۱-۱- کاربردهای سامانه‌های توصیه گر

سامانه‌های توصیه گر کاربردهای فراوانی دارند که برخی از زمینه‌های کاربردی آن به شرح زیر است:

[7] Filtering Algorithm
[8] Collaboration Filtering
[9] Content-Based Filtering
[10] Knowledge-Based Filtering
[11] Demographic filtering
[12] Hybrid Filtering

۱. اینترانت های بنگاهی: برای پیدا کردن افراد خبره در یک زمینه خاص یا افرادی که در رویارویی با شرایط مشابه، تجاربی کسب کرده و راه‌حل‌هایی یافته‌اند (بیشتر داخل یک سازمان کاربرد دارد).
۲. کتابخانه دیجیتال: پیدا کردن کتاب، مقاله و ...
۳. کاربردهای پزشکی: انتخاب پزشک متناسب با شرایط (مکان، نوع بیماری، زمان و ...) بیمار، انتخاب دارو و...
۴. مدیریت ارتباط با مشتری CRM: برای ارائه راهکارهایی برای حل مشکلات تولیدکننده و مصرف‌کننده در زنجیره تأمین.
۵. تجارت الکترونیک: برای توصیه محصولات و خدمات مختلف.

در ادامه این مقاله در بخش دوم، به‌مرور کارهای گذشته در این زمینه خواهیم پرداخت. در بخش سوم مدل پیشنهادی بیان‌شده و در بخش چهارم مدل بر روی مجموعه داده‌های موجود پیاده‌سازی و آزمایش می‌شود. در انتها نتایج ارزیابی مدل پیشنهادی بیان و نتیجه‌گیری و پیشنهادهای آینده مطرح می‌شود.

## ۲ – کارهای پیشین

تاکنون کارهای مختلفی درزمینه سامانه‌های توصیه گر درزمینه گردشگری صورت گرفته است که در این بخش با برخی از آن پژوهش‌ها آشنا می‌شویم.

همان‌طور که در پیش‌تر ذکر شد، سامانه‌های توصیه گر از دهه ۱۹۹۰ مورد استقبال دیگران قرارگرفته و توسعه یافتند.[۹] اولین سامانه توصیه گر با عنوان Guide، توسط دیویس و همکاران به‌عنوان یکی از اولین سامانه‌های توصیه گر گردشگری مطرح گردید که برای شهر لنکستر انگلستان پیاده‌سازی و اجرا گردید. این سامانه برای استفاده نیازمند به سیستم تعیین موقعیت (GPS) برای تعیین محل کاربر و ارائه یکسری اطلاعات پویا از دستگاه‌های تلفن همراه با زیرساخت مبتنی بر سلول با پهنای باند بالا که در سرتاسر شهر موجود است، استفاده می‌کند. اطلاعات ارائه‌شده به بازدیدکنندگان بر اساس نمایه کاری، اطلاعات بافتی و محل فیزیکی کاربر هست. همچنین بر اساس بافت فعلی، سامانه یک تور مناسب را که می‌تواند به‌صورت پویا و بر اساس محدودیت‌های سفر تغییر کند، ارائه می دهد[۸]

در سال ۲۰۰۱،Poslad و همکارانش سامانه‌یCRUMPET را به‌منظور ارائه اطلاعات و خدمات شخصی به گردشگران ارائه دادند. این سامانه برای هر جاذبه، سامانه‌ی توضیحات، نقشه‌ها و مسیرها و تصاویر را نشان می‌دهد که باهدف تسهیل تعاملات بین کاربر و پیشنهادها ارائه گردید. هنگام استفاده از این سامانه برای نخستین بار، سامانه برای یادگیری علایق کاربر، اطلاعات جمعیت شناختی گردشگر را دریافت نموده و از آن برای پیشنهاد دادن در آینده استفاده می‌کرد. این سامانه معمولاً توصیه‌های خود را بر اساس مکان فعلی کاربر و علایق شخصی وی ارائه می‌دهد. [۱۰]

در سال ۲۰۱۵، XU و همکارانش با در نظر گرفتن موضوع برای استخراج ترجیحات کاربر روشی را که

برای پیشنهاد مسیر گردشگری ارائه دادند بر اساس تراکم سوابق سفر به سایر شهرها بود. در این روش برای توصیه‌های خود از روش پالایش گروهی بر اساس مشارکت جامعه در عکس‌هایی با برچسب جغرافیایی و محاسبه‌ی بافت کاربر استفاده می‌کند.[۱۱]

سامانه دیگری در سال ۲۰۱۵ به نام Smart Tourism توسط Fermoso و همکارانش بر روی سیستم‌عامل اندروید ارائه گردید. سامانه مذکور از تلفن همراه و فنّاوری معنایی نظیر هستی‌شناسی‌ها و داده-هایی با دسترسی آزاد (LOD) به‌عنوان ابزاری قدرتمند برای به اشتراک‌گذاری اطلاعات استفاده می‌کند. همچنین از همکاری میان آن‌ها برای تولید دانش و یادگیری درباره میراث فرهنگی بهره می‌گیرد.[۱۲]

## ۳ – مدل پیشنهادی

### ۳-۱- اصول سامانه‌های پیشنهاددهنده

تعاریف متفاوتی برای سامانه‌های توصیه گر ارائه‌شده است. از آن جمله، تعریف کلی‌نگر و خلاصه آقای Ting-peng liang در سال 2007 است که سامانه‌های توصیه گر[13]RS را زیرمجموعه‌ای از سیستم پشتیبانی از تصمیم یا[14]DSS ها می‌داند و آن‌ها را سامانه‌های اطلاعاتی تعریف می‌کند که توانایی تحلیل رفتارهای گذشته و ارائه توصیه‌هایی برای مسائل جاری را دارا هستند. به زبان ساده‌تر در سامانه‌های توصیه گر تلاش بر این است تا با حدس زدن شیوه تفکر کاربر (به کمک اطلاعاتی که از نحوه رفتار وی یا کاربران مشابه وی و نظرات آن‌ها داریم) به وی مناسب‌ترین و نزدیک‌ترین کالا، اقلام و... به سلیقه او راشناسایی و پیشنهاد کنیم. این سامانه‌ها در حقیقت همان فرایندی که ما در زندگی روزمره خود به کار می‌بریم و طی آن تلاش می‌کنیم تا افرادی با سلایق نزدیک به خود را پیداکرده و از آن‌ها در مورد انتخاب‌هایمان نظر بخواهیم. توصیه‌هایی که از سوی سامانه‌های توصیه گر ارائه می‌شوند به‌طورکلی می‌توانند دو نتیجه در برداشته باشند

- کاربر را در اخذ تصمیمی یاری می‌کنند (که مثلاً از میان چندین گزینه پیش رو کدام بهتر است و آن را انتخاب کند).
- موجب افزایش آگاهی کاربر، درزمینه موردعلاقه وی می‌شود (مثلاً در حین ارائه توصیه به کاربر موجب می‌شود تا وی با اقلام و اشیاء جدیدی را که قبلاً آن‌ها را نمی‌شناخته، آشنا شود).

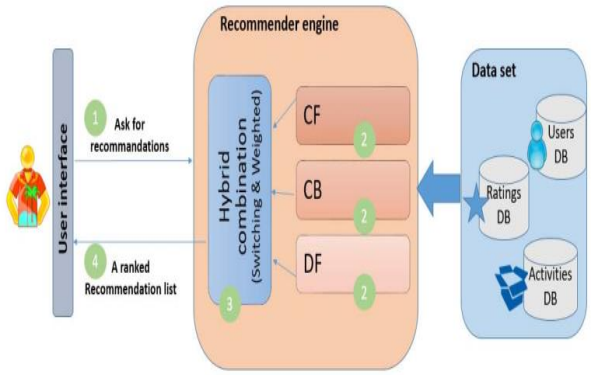
سیستم توصیه گر پیشنهادی دارای سه ماژول مجموعه داده‌ها، موتور پیشنهاددهنده و رابط کاربری است. هنگامی‌که یک کاربر وارد سیستم می‌شود، او باید نوع فعالیت‌های مورد جستجو را انتخاب کند (به‌عنوان‌مثال هتل‌ها، رستوران‌ها و مقصدها). هنگامی‌که نوع جستجو را انتخاب می‌کند، موتور پیشنهاددهنده ترکیبی یک توصیه شخصی‌سازی‌شده را با توجه به علایق کاربر ایجاد می‌کند.

این موتور با استفاده از سه روش پالایش مشارکتی، محتوا و جمعیت شناختی اجرا می‌شود و ترکیب نتایج

[13] Recommend Systems

[14] Decision Support System

به‌دست‌آمده از آن‌ها به‌منظور پیش‌بینی درجه‌ی فعالیت-ها استفاده می‌شود. به‌عنوان یک خروجی، کاربر یک لیست نهایی رتبه‌بندی شده از فعالیت‌های توصیه‌شده را بر اساس نرخ پیش‌بینی‌شده برای فعالیت‌ها دریافت می-کند [۱۳].




(شکل-۱): نحوه کار سیستم پیشنهاد دهنده ارائه‌شده

در ادامه هر یک از سه الگوریتم پالایش استفاده‌شده در این پژوهش و نحوه به‌کارگیری آن‌ها توضیح داده‌شده است.

### ۳-۱-۱- الگوریتم پالایش مشارکتی

شناخته‌شده‌ترین نوع از الگوریتم‌های پالایش مشارکتی الگوریتم نزدیک‌ترین همسایگی است. در این پژوهش از الگوریتم پالایش غیر احتمالی نزدیک‌ترین همسایگی مبتنی بر کاربر استفاده‌شده است. این روش باهدف پیش‌بینی امتیازدهی کاربران فعال در محصولات غیر رتبه‌بندی شده بر اساس علاقه کاربر و علاقه کاربران همسایه (مشابه) است. معادله‌ی ضریب تانیموتو در رابطه (۱) ارائه‌شده است.

$$sim(i,j)=\frac{|f_i \cap f_j|}{|f_i|+|f_j|-|f_i \cap f_j|}$$

(رابطه-۱): فرمول ضریب تانیموتو

در این رابطه، $f_i$ مجموعه آیتم‌های ترجیح داده‌شده توسط کاربر i است. $f_j$ مجموعه آیتم‌های ترجیح داده‌شده توسط کاربر j است. $f_i \cap f_j$ مجموعه آیتم-هایی است که توسط هر دو کاربر به‌صورت مشترک ترجیح داده‌شده است [۱۴].

مهم‌ترین چالش‌های پیش رو در سامانه‌های پیشنهاددهنده مبتنی بر فیلترینگ مشارکتی که در سال‌های اخیر موردتوجه محققان مختلف قرارگرفته است عبارت‌اند از:

- **مشکل شروع سرد:**

آیتم‌ها و کاربران جدید، باعث بروز چالش مهمی در سامانه‌های توصیه گر می‌شوند. شروع سرد اشاره به حالتی دارد که در آن، یک آیتم نمی‌تواند پیشنهاد داده شود، مگر اینکه توسط تعداد قابل‌توجهی از کاربران، امتیازدهی شده باشد. اولین دسته از این مشکلات، در سامانه‌های فیلترینگ مشارکتی رخ می‌دهند؛ جایی که یک آیتم قابلیت مطرح‌شدن به‌عنوان پیشنهاد را ندارد مگر آنکه قبلاً برخی از کاربران به آن رأی داده باشند. این مشکل به آیتم‌های جدید و مبهم اِعمال می‌شود و به‌ویژه برای کاربرانی باسلیقه‌های مختلف خوشایند نیست و از این مشکل آیتم جدید، اغلب به‌عنوان مشکل درجه اول یاد می‌شود. علاوه بر این، یک کاربر جدید بایستی میزان مناسبی از آیتم‌ها را امتیازدهی کرده باشد تا الگوریتم پیشنهاددهنده بتواند توصیه‌های قابل‌اعتماد و دقیقی ارائه دهد.

مطالعات در این زمینه، در درجه‌ی اول بر روی مسئله ی انتخاب مؤثر آیتم‌هایی که توسط یک کاربر امتیازدهی می‌شوند متمرکزشده است تا با داشتن

حداقل بازخورد از کاربر، کارایی پیشنهاد را به‌سرعت بهبود ببخشد. در این حالت، می‌توان از تکنیک‌های کلاسیک پیرامون یادگیری فعال برای عمل انتخاب آیتم بهره گرفت. به‌علاوه، می‌توان از اطلاعات شبکه‌های اجتماعی نیز استفاده کرد تا ضعف موجود در مشکل شروع سرد را حل کرده و شباهت‌های بین کاربران را پیدا کرد.

- **اعتماد:**

اعتبار کاربرانی که تاریخچه طولانی‌تری در استفاده از سیستم دارند نسبت به کاربرانی که تاریخچه حضور آن‌ها در سیستم کمتر است، بیشتر است. لذا امتیازات آن‌ها به هتل‌ها از ارزش بالاتری برخوردار است.

- **پراکندگی داده‌ها:**

معمولاً سایت‌های تجارت الکترونیک شامل محصولات زیادی هستند و هر کاربر تنها بخشی از محصولات سایت را ارزیابی می‌کند. اگر دو کاربر به کالای مشابه ای رتبه ندهند، در صورت داشتن علایق مشابه، سیستم نمی‌تواند مشابهت بین آن‌ها را تشخیص دهد و نمی‌تواند هیچ محصولی را برای پیشنهاد انتخاب کند.

- **مقیاس‌پذیری:**

با گسترش سامانه‌های نرم‌افزاری تعداد کاربران و اقلام در فضای کاربر هدف برای جستجوی همسایه‌ها، افزایش می‌یابد. در فیلترسازی مشارکتی مبتنی بر کاربر یا هتل، برای محاسبه مشابهت باید کل پایگاه داده بررسی شود. بنابراین با افزایش رکوردهای ثبت‌شده، پیچیدگی محاسباتی به‌طور چشم‌گیری افزایش خواهد یافت که منجر به کاهش کارایی سیستم می‌شود.

در این مقاله سیستم پیشنهاددهنده‌ای که طراحی‌شده است با استفاده از یک روش ترکیبی و با در نظر گرفتن امتیاز برای کاربران قدیمی مشکلات شروع سرد و اعتماد را تا حد زیادی حل می‌نماید. روش پیشنهادی شامل سه مرحله خوشه‌بندی اولیه کلیه کاربران و تعیین خوشه، تعیین وزن مناسب برای هرکدام از ویژگی‌ها و تعیین همسایگان نزدیک کاربر جدید، تشکیل ماتریس مجاورتی و محاسبه امتیاز کاربر جدید به هر هتل. روش ارائه‌شده به‌منظور خوشه‌بندی اولیه، از سیستم محتوا محور بر اساس ارائه مدلی برای سامانه‌های توصیه گر هتل مبتنی بر رویکرد مشارکت محور اطلاعات دموگرافیک کاربران استفاده می‌نماید. جهت شناسایی همسایگان نزدیک و ارائه پیشنهادها از فیلترینگ مشارکتی استفاده می‌کند.

### ۳-۱-۲- الگوریتم پالایش مبتنی بر محتوا

برخلاف الگوریتم پالایش مشارکتی که امتیازدهی یک کاربر به یک محصول از طریق امتیازدهی همسایه‌های آن کاربر پیش‌بینی می‌شد، در پالایش مبتنی بر محتوا پیش‌بینی، بر اساس مشابهت بین آیتم‌هایی خواهد بود که یک کاربر قبلاً امتیازدهی کرده است. در این پژوهش از الگوریتم نزدیک‌ترین همسایگی برای پیدا کردن میزان مشابهت بین یک آیتم خاص و آیتم‌های رتبه‌بندی شده توسط کاربر استفاده می‌شود. هر آیتم به‌عنوان یک بردار از ویژگی‌ها نمایش داده می‌شود. برای محاسبه میزان مشابهت بین آیتم‌ها از فاصله اقلیدسی استفاده می‌شود.

الگوریتم پالایش محتوا به مشکل شروع سرد برای آیتم-های جدید که هنوز توسط کاربران رتبه‌بندی نشده‌اند، غلبه می‌کند، اما هنوز از مشکل شروع سرد برای کاربرهای جدید رنج می‌برد. در این حالت نمی‌تواند نتایج را ارائه دهد، زیرا یک کاربر جدید در سیستم هیچ-گونه امتیازدهی برای آیتم‌ها نکرده است [۱۵].

مراحل اصلی این روش عبارت‌اند از:

۱. شناسایی آیتم‌های بدون امتیازدهی شده توسط کاربر فعال
۲. برای هر آیتم بدون امتیاز، از طریق فاصله اقلیدسی میزان مشابهت آن آیتم به آیتم‌های امتیازدهی شده توسط کاربر فعال محاسبه می‌شود
۳. از طریق الگوریتم نزدیک‌ترین همسایگی، آیتم‌هایی که بیش‌ترین مشابهت به آیتم موردنظر را دارد پیدا می‌شود و سپس از طریق امتیاز آن‌ها، امتیاز آیتم مذکور پیش‌بینی می‌شود [۱۶].

### ۳-۱-۳- الگوریتم پالایش اطلاعات جمعیت شناختی

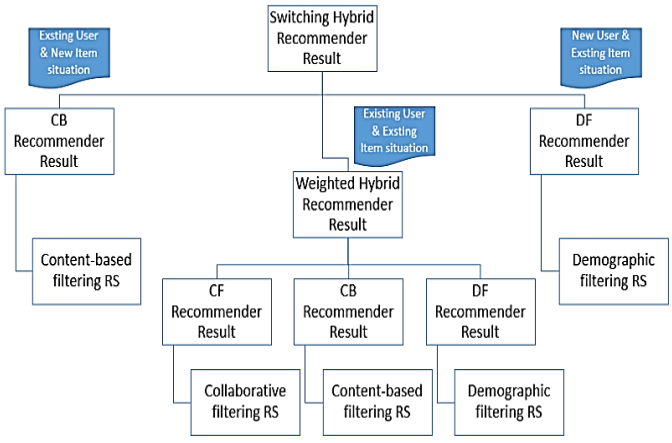
از درخت تصمیم‌گیری برای اجرای پالایش اطلاعات جمعیت شناختی استفاده می‌شود. درخت تصمیم‌گیری کاربر را بر اساس اطلاعات کاربری (سن، جنس، منطقه، سبک سفر) به‌منظور به دست آوردن امتیاز یک آیتم ازنظر کاربر طبقه‌بندی می‌کند. اطلاعات کاربری را به‌صورت گره‌ها و امتیازها به آیتم‌ها را به‌صورت برگ در درخت تصمیم در نظر گرفته می‌شود. برای درخت‌های تصمیم‌گیری الگوریتم‌های زیادی وجود دارد. در این پژوهش از درخت تصمیم‌گیری استفاده می‌شود. به این خاطر از این روش استفاده می‌شود که اغلب از این روش برای ویژگی‌های گسسته استفاده می‌شود و بیش‌ترین سرعت برای ساخت درخت را دارد [۱۳].

این الگوریتم پالایش شامل سه مرحله اصلی زیر است:

۱. شناسایی آیتم‌های بدون امتیازدهی شده توسط کاربر فعال
۲. برای هر آیتم رتبه‌بندی نشده یک درخت تصمیم ایجاد می‌شود که رتبه آن آیتم را پیش‌بینی می-کند.
۳. درخت تصمیم‌گیری ساخته‌شده را بر روی پروفایل کاربر فعال برای به دست آوردن امتیاز پیش‌بینی شده کاربر برای یک آیتم، اعمال می‌شود [۱۷].

### ۳-۱-۴- الگوریتم پالایش ترکیبی

در این پژوهش یک مدل ترکیبی بر اساس سه الگوریتم پالایش مشارکتی، پالایش مبتنی بر محتوا و پالایش جمعیت شناختی برای غلبه بر نقایص هر یک از این روش‌ها به‌خصوص غلبه بر مشکل شروع سرد ارائه‌شده است. علاوه بر این، این روش ترکیبی تلاش می‌کند تا بهترین ترکیب از رویکردهای ذکرشده را پیدا کند تا دقت پیش‌بینی را افزایش دهد.




(شکل-۲): مدل مفهومی الگوریتم پالایش ترکیبی پیشنهادی [۱۴]

در الگوریتم ترکیبی پیشنهادی اگر آیتم جدیدی داشته باشد، شاخه چپ این مدل انتخاب می‌شود. در این شاخه با استفاده از الگوریتم پالایش محتوا، امتیاز آیتم جدید به دست می‌آید. اگر یک کاربر جدید داشته باشد، شاخه راست مدل انتخاب می‌شود. در این شاخه با استفاده از الگوریتم پالایش اطلاعات جمعیت شناختی، امتیاز هر آیتم برای یک کاربر جدید به دست می‌آید. درصورتی‌که با مشکل کاربر جدید و آیتم جدید روبرو نباشد، شاخه میانی مدل انتخاب می‌شود. در این حالت از رویکرد ترکیب‌سازی وزن دهی استفاده می‌شود.

همچنین، رویکرد ترکیب‌سازی وزن دهی به‌صورت یک ترکیب معادله خطی خواهد بود. متغیرهای این معادله خطی هر یک از جواب‌های به‌دست‌آمده برای پیش‌بینی امتیاز یک آیتم برای یک کاربر خاص خواهد بود و وزن‌های این متغیرها از طریق یک معادله برنامه‌ریزی خطی به دست می‌آید. معادله برنامه‌ریزی خطی برای پیدا کردن وزن‌های بهینه به‌صورت رابطه‌ی (۲) تعریف می‌شود.

$$\min imize \frac{\sum_{k=1}^{n} |\alpha.rDF + \beta.rCB + \gamma.rCF - y_i|}{n}$$
$$\alpha + \beta + \gamma = 1$$
$$\alpha \geq 0$$
$$\beta \geq 0$$
$$\gamma \geq 0$$

(رابطه-۲): معادله‌ی برنامه‌ریزی خطی برای پیدا کردن وزن بهینه

در این رابطه $rDF$، $rCB$ و $rCF$ امتیاز به‌دست‌آمده برای یک آیتم خاص برای یک کاربر خاص از سه الگوریتم پالایش است. در این معادله بر اساس امتیازهای پیش‌بینی‌شده برای هر آیتم و مقایسه آن با نتایج واقعی سعی می‌شود، اختلاف بین مقادیر پیش‌بینی‌شده و مقادیر واقعی آن به حداقل مقدار برسد [۱۲].

## ۳-۲- بهبود روش پیشنهادی توسط خوشه‌بندی فازی و الگوریتم خفاش[15]

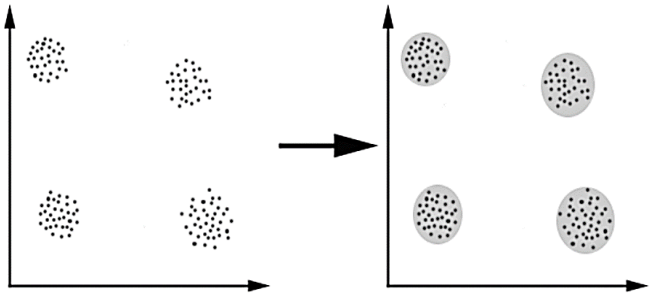
### ۳-۲-۱- تعریف خوشه‌بندی[16]:

خوشه‌بندی یکی از شاخه‌های یادگیری بدون نظارت هست و فرآیند خودکاری است که در طی آن، نمونه‌ها به دسته‌هایی که اعضای آن مشابه یکدیگر می‌باشند تقسیم می‌شوند که به این دسته‌ها خوشه گفته می‌شود. بنابراین خوشه مجموعه‌ای از اشیاء هست که در آن اشیاء با یکدیگر مشابه بوده و با اشیاء موجود در خوشه‌های دیگر غیرمشابه می‌باشند. برای مشابه بودن می‌توان معیارهای مختلفی را در نظر گرفت مثلاً می توان

معیار فاصله را برای خوشه‌بندی مورداستفاده قرارداد و اشیائی را که به یکدیگر نزدیک‌تر هستند را به‌عنوان یک خوشه در نظر گرفت که به این نوع خوشه‌بندی، خوشه‌بندی مبتنی بر فاصله[17] نیز گفته می‌شود. به‌عنوان‌مثال در شکل (۳) نمونه‌های ورودی در سمت چپ به چهار خوشه مشابه شکل سمت راست تقسیم می‌شوند. در این مثال هر یک از نمونه‌های ورودی به

[15] Bat Algorithm
[16] Clustering
[17] Distance-based Clustering

یکی از خوشه‌ها تعلق دارد و نمونه‌ای وجود ندارد که متعلق به بیش از یک خوشه باشد



(شکل-۳): خوشه‌بندی نمونه‌های ورودی

خوشه‌بندی با طبقه‌بندی[18] متفاوت است. در طبقه‌بندی نمونه‌های ورودی برچسب‌گذاری شده‌اند ولی در خوشه‌بندی نمونه‌های ورودی دارای برچسب اولیه نمی‌باشند و درواقع با استفاده از روش‌های خوشه‌بندی است که داده‌های مشابه مشخص و به‌طور ضمنی برچسب‌گذاری می‌شوند. درواقع می‌توان قبل از عملیات طبقه‌بندی داده‌ها یک خوشه‌بندی روی نمونه‌ها انجام داد و سپس مراکز خوشه‌های حاصل را محاسبه کرد و یک‌بر چسب به مراکز خوشه‌ها نسبت داد و سپس عملیات طبقه‌بندی را برای نمونه‌های ورودی جدید انجام داد.

هدف خوشه‌بندی یافتن خوشه‌های مشابه از اشیاء در بین نمونه‌های ورودی هست اما چگونه می‌توان گفت که یک خوشه‌بندی مناسب است و دیگری مناسب نیست؟ می‌توان نشان داد که هیچ معیار مطلقی برای بهترین خوشه‌بندی وجود ندارد بلکه این بستگی به مسئله و نظر کاربر دارد که باید تصمیم بگیرد که آیا نمونه‌ها به‌درستی خوشه‌بندی شده‌اند یا خیر. بااین‌حال معیارهای مختلفی برای خوب بودن یک خوشه‌بندی ارائه‌شده است که می‌تواند کاربر را برای رسیدن به یک خوشه‌بندی مناسب راهنمایی کند. یکی از مسائل مهم در خوشه‌بندی انتخاب تعداد خوشه‌ها هست. در بعضی از الگوریتم‌ها تعداد خوشه‌ها از قبل مشخص شده است و در بعضی دیگر خود الگوریتم تصمیم می‌گیرد که داده‌ها به چند خوشه تقسیم شوند.

### ۳-۲-۲- خوشه‌بندی فازی (FCM)[19]:

برای درک بهتر خوشه‌بندی فازی و الگوریتم های مختلف آن لازم است تا ابتدا با مفهوم مجموعه‌های فازی و تفاوت آن‌ها با مجموعه‌های کلاسیک آشنا شویم. در مجموعه‌های کلاسیک یک عضو از مجموعه مرجع یا عضوی از مجموعه A است یا عضو مجموعه A نیست. مثلاً مجموعه مرجع اعداد حقیقی را در نظر بگیرید. عدد ۲,۵ عضو مجموعه اعداد صحیح نیست حال‌آنکه عدد ۲ عضو این مجموعه است. به زبان دیگر تعلق عدد ۲,۵ به مجموعه اعداد صحیح ۰ است و تعلق عدد ۲ به این مجموعه ۱ است. درواقع می‌توان برای هر مجموعه یک تابع تعلق تعریف کرد که مقدار این تابع تعلق برای اعضای مجموعه ۱ هست و برای بقیه ۰. در مجموعه‌های کلاسیک مقدار این تابع تعلق یا ۰ است یا ۱. حال مجموعه انسان‌های جوان و پیر را در نظر بگیرید. سؤالی که در اینجا مطرح می‌شود این است که آیا فردی با سن ۲۵ جزء این مجموعه است یا خیر؟ سن ۳۰ چطور؟ ۳۵؟ همان‌طور که حدس زدید نمی‌توان به‌طورقطع و یقین مرزی برای انسان های جوان و پیر در نظر گرفت. دلیل آن‌هم این است که اگر

[18] Classification

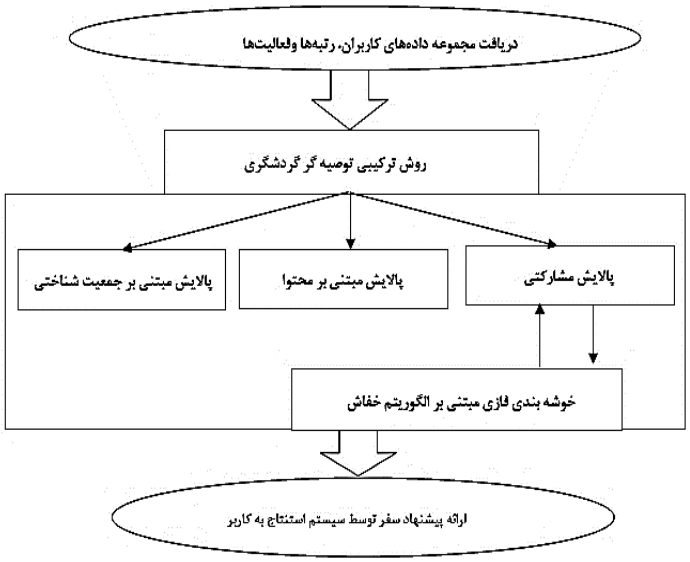
[19] Fuzzy c-means

فرضاً ۳۵ جوان محسوب شود ۳۶ نیز می‌تواند جوان باشد و همین‌طور ۳۷ و ۳۸ و غیره. درواقع در اینجا با مفهوم عدم قطعیت[20] مواجه هستیم. ما خودمان نیز از عدم قطعیت در زندگی روزمره بارها استفاده کرده‌ایم مثلاً هوای سرد، آب داغ و غیره. درواقع تمامی مثال‌های بالا مثال‌هایی از مجموعه‌های فازی می‌باشند. تفاوت اصلی مجموعه‌های فازی و مجموعه‌های کلاسیک در این است که تابع تعلق مجموعه‌های فازی دو مقداری نیست (۰ یا ۱) بلکه می‌تواند هر مقداری بین ۰ تا ۱ را اختیار کند. حال مجموعه انسان‌های جوان و پیر را در نظر بگیرید اگر ۲۵ سال را سن جوانی در نظر بگیریم می‌توانیم به ۲۵ تعلق ۱ بدهیم و مثلاً به ۳۰ تعلق ۰,۸ و به ۳۵ تعلق ۰,۷۵ و به ۹۰ تعلق ۰,۱ را بدهیم. اگر اعضای یک مجموعه فازی تنها دارای تابع تعلق ۰ و ۱ باشند این مجموعه فازی یک مجموعه کلاسیک خواهد بود. نکته جالب‌توجه این است که مثلاً سن ۵۰ می‌تواند با تعلق ۰,۵ عضو مجموعه جوان باشد و با تعلق ۰,۵ عضو مجموعه پیر یعنی یک عضو مجموعه مرجع می‌تواند با درجه‌های تعلق مختلف عضو مجموعه‌های فازی تعریف‌شده روی مجموعه مرجع باشد.

خوشه‌بندی Fuzzy c-means یک موضوع مطالعاتی مهم در خوشه‌بندی داده‌ها در فنّاوری داده‌کاوی است که کاربردهای عملی در زمینه‌های مختلف دارد. بزرگ‌ترین عیب این الگوریتم خوشه‌بندی فازی حساسیت آن به مقداردهی اولیه به تعداد خوشه‌ها و مراکز اولیه خوشه‌ها، انتخاب نامناسب مراکز اولیه خوشه‌ها است که در این الگوریتم منجر به مینیمم محلی می‌شود [۱۸]. همچنین، FCM یک کاربر خاص را به بیش از یک خوشه تخصیص می‌دهد. مورداستفاده درنهایت، این محدودیت توسط الگوریتم خفاش که یک الگوریتم فرا ابتکاری است، حل‌شده است. بهینه‌سازی خفاش برای تعیین بهینه مرکز خوشه بکار می‌رود. بهینه‌سازی خفاش از سایر روش‌هایی همانند الگوریتم ژنتیک (GA[21])، بهینه‌سازی توده ذرات (PSO[22]) و جست‌وجوی هارمونی[23] بهتر عمل می‌کند، زیرا از مزایای آن‌ها به‌صورت ترکیبی استفاده می‌کند و مشکل مینیمم محلی را نیز برطرف می‌کند؛ بنابراین، درروش پیشنهادی جهت بهبود الگوریتم پالایش مشارکتی از خوشه‌بندی فازی مبتنی بر الگوریتم خفاش استفاده‌شده است که کاربران را به گروه‌های مختلفی دسته‌بندی می‌کند که باعث کاهش مقیاس‌پذیری و بهبود الگوریتم پالایش مشارکتی می‌شود.

این پژوهش بر روی ایجاد یک سیستم توصیه گر شخصی‌سازی‌شده درزمینه گردشگری تمرکز دارد. این نرم‌افزار بهترین جاذبه‌های گردشگری (صرفاً هتل) را در یک محل خاص با توجه به ترجیحات فرد، مشخصات فرد و امتیازدهی فرد به مکان‌های قبلی مشاهده‌شده پیشنهاد می‌دهد. در شکل (۴) معماری پیشنهادی نشان داده‌شده است.

[20] Uncertainty
[21] Genetic Algorithm
[22] *Particle swarm optimization*
[23] Harmony search




(شکل-۴): معماری سیستم پیشنهادی

حال در گام بعدی، برای اجرای این الگوریتم‌های توصیه گر گردشگری، از الگوریتم‌های مختلف یادگیری ماشین استفاده می‌شود. از الگوریتم KNN برای پالایش مشارکتی و پالایش مبتنی بر محتوا و از درخت تصمیم-گیری برای الگوریتم پالایش جمعیت شناختی استفاده می‌شود [۱۱].

در گام بعد، برای کاهش مقیاس‌پذیری درروش پیشنهادی از الگوریتم بهینه یافته خوشه‌بندی فازی مبتنی بر الگوریتم خفاش جهت بهبود روش پالایش مشارکتی و دسته‌بندی کاربران به گروه‌های مختلف استفاده‌شده است. این خوشه‌بندی باعث دسته‌بندی کاربران به گروه‌های مختلف برحسب طول و عرض جغرافیایی می‌شود که باعث سهولت جست‌وجو می‌شود.

در گام نهایی، پس از تجزیه‌وتحلیل و استنتاج توسط روش ترکیبی توصیه گر گردشگری، پیشنهاد مناسب جهت سفر کاربر پیشنهاد می‌شود.

ترکیب‌سازی روش‌ها انتخاب خوبی برای بهره‌مندی از مزایا و غلبه بر مشکل شروع سرد در این الگوریتم‌ها است. برای افزایش دقت توصیه‌ها، از دو تکنیک تعویض کردن و وزن دهی ترکیب‌سازی استفاده می‌کنیم [۱۹].

خوشه‌بندی داده‌ها به‌عنوان مهم‌ترین روش تجزیه‌وتحلیل اطلاعات اکتشافی محسوب می‌شود که برای استخراج اطلاعات باارزش ناشناخته از حجم زیادی از داده‌ها برای بسیاری از برنامه‌های کاربردی در زمان واقعی در فنّاوری داده‌کاوی[۲۴] استفاده می‌شود [۱۶]. بیشتر تکنیک‌های خوشه سازی کارایی خود را در بسیاری از زمینه‌ها مانند سامانه‌های تصمیم‌گیری، علوم پزشکی، علوم زمین و غیره نشان دادند [۱۷]. خوشه‌بندی مبتنی بر پارتیشن یکی از رویکردهای اصلی در خوشه‌بندی است.

در الگوریتم خوشه‌بندی فازی[۲۵] باید تعداد و مراکز خوشه‌ها توسط کاربر در مشخص شوند. کیفیت این الگوریتم بشدت به تعداد اولیه خوشه‌ها و مکان اولیه مراکز خوشه‌ها بستگی دارد. هدف از خوشه‌بندی فازی استخراج مدل‌های فازی از داده‌هاست. کاربردهای متعدد خوشه‌بندی فازی در تحلیل داده‌ها و تشخیص الگو و نیز زمینه‌های پژوهشی موجود در این زمینه ازجمله استفاده از آن در حل مسائل مسیریابی، تخصیص و زمان‌بندی نیاز به مطالعه الگوریتم‌های موجود و بهبود و اصلاح آن‌ها را آشکارتر می‌نماید. در این روش، چندین

[24] Data Mining

[25] Fuzzy c-means Technique (FCM)

روش خوشه‌بندی فازی بر پایه روش خوشه‌بندی استاندارد FCM موردبررسی و مطالعه قرارگرفته است [۱۶].

در این الگوریتم نمونه‌ها به c خوشه تقسیم می‌شوند و تعداد c از قبل مشخص‌شده است. در نسخه فازی این الگوریتم نیز تعداد خوشه‌ها از قبل مشخص‌شده است (در فرمول‌های زیر تعداد خوشه‌ها g در نظر گرفته‌شده است). الگوریتم میانگین فازی برای پارامترهای $y_{ji}(i=1,....,n;\ j=1,....,g)$ به‌صورت رابطه‌ی (۳) قابل‌حل است (تابع هدف) [۲۰]

تابع هدف برای خوشه‌بندی فازی از رابطه زیر به دست می‌آید:

$$J_m = \sum_{i=1}^{n}\sum_{j=1}^{g} y_{ij}^{m}\left\|x_i - m_j\right\|^2$$

**(رابطه-۳): تابع هدف**

که تحت شرایط زیر مینیمم می‌شود (از روی $y_{ji}$ می توان یک ماتریس y تعریف کرد که دارای n سطر و g ستون است مؤلفه‌های آن هر مقداری بین صفرتا یک را می‌توانند اختیار کنند. اگر تمامی مؤلفه‌های ماتریس y به‌صورت صفر و یا یک باشند الگوریتم مشابه میانگین فازی کلاسیک خواهد بود. بااینکه مؤلفه‌های ماتریس y می‌توانند هر مقداری بین صفرتا یک را اختیار کنند اما مجموع مؤلفه‌های هر یک از ستون‌ها باید برابر یک باشد)[ ۲۰]

$$D_z \overset{\Delta}{=} \left\{\sum_{j=1}^{g}\left|m_j(k)-m_j(k-1)\right|^2\right\}^{\frac{1}{2}} < \varepsilon$$

**(رابطه-۴)**

معنای این شرط این است که مجموع تعلق هر نمونه به c خوشه باید برابر یک باشد. پارامتر $y_{ji}$ درجه تابع عضویت i امین شیء با j امین خوشه را نشان می‌دهد. پارامتر r یک عدد حقیقی است که فازی بودن را کنترل می‌کند ($r \geq 1$) اگر $r=1$ قرار دهیم تابع هدف خوشه‌بندی میانگین غیر فازی به دست می‌آید. پارامتر $x_i$ نمونه i ام و پارامتر $m_j$ مرکز j امین خوشه است و n تعداد نمونه‌ها (اشیاء) است.

برای به دست آوردن فرمول‌های مربوط به $y_{ji}$ و $m_j$ باید تابع هدف تعریف‌شده را مینیمم کنیم. با استفاده از شرط

$$\alpha_i = \max_{1\le j\le g} y_{ji}^{r-1}\left|x_i - m_j\right|^2 - \min_{1\le j\le g} y_{ji}^{r-1}\left|x_i - m_j\right|^2$$

فوق و برابر صفر قرار دادن مشتق تابع هدف‌داریم.

$$y_{ji} = \frac{1}{\sum_{k=1}^{g}\left(\frac{d_{ij}}{d_{ik}}\right)^{\frac{2}{r-1}}}$$

**(رابطه-۵)**

$$m_j = \frac{\sum_{i=1}^{n} y_{ji}^{r} x_i}{\sum_{i=1}^{n} y_{ji}^{r}}$$

**(رابطه-۶)**

با استفاده از دو فرمول محاسبه‌شده، الگویتم خوشه‌بندی میانگین فازی و مراحل اجرای آن بدین‌صورت است:

۱. انتخاب $r(1<r<\infty)$، مقداردهی اولیه برای مقادیر تابع عضویت $y_{ji}(i=1,....,n;\ j=1,....,g)$

۲. محاسبه‌ی مراکز خوشه‌ها (محاسبه $m_j$ ها).

۳. محاسبه $d_{ij} = |x_i - m_j|$ که $d_{ij}, i = 1,...,n; j = 1,...,g$.

۴. محاسبه تابع عضویت: اگر $d_{il}=0$ باشد برای مقادیری از l آنگاه $y_{li}=1$ است و $y_{ji} = 0$ است برای همه، $j \neq l$ در غیر این صورت با توجه به رابطه (۵) حل می‌شود.

۵. اگر همگرا نبود، به مرحله‌ی ۲ برگرد.

یکی از شرط‌های توقف الگوریتم زمانی است که تغییرات نسبی در مقادیر مرکز کوچک می‌شود.

$$\sum_{j=1}^{g} y_{ji} = 1 \quad 1 \le i \le n$$

$$y_{ji} \ge 0 \quad i = 1,...,n; \quad j = 1,...,g$$

**(رابطه-۷)**

که $m_j(k)$ مقدار j امین مرکز روی k امین تکرار و ε حد مشخص‌شده است. شرط‌های توقف دیگر بر اساس مقادیر تابع عضویت، $y_{ji}$ یا تابع هزینه، $j_r$ است. شرط دیگر بر اساس بهینه‌سازی محلی تابع هزینه که سلیم و اسماعیل (۱۳۹۶) مطرح کرده‌اند.

در اجرای این الگوریتم زمانی توقف داریم که:

$$\max_{1 \le i \le n} \alpha_i < \varepsilon \qquad \textbf{(رابطه-۸)}$$

**که:**

تا این‌که در یک مینیمم محلی در $\alpha_i = 0, i = 1,...,n$. این حالت همگرایی حاصل می‌شود. **(رابطه-۹)**

### ۳-۲-۳-الگوریتم خفاش[۲۶]

روش‌ها و الگوریتم‌های بهینه‌سازی به دودسته الگوریتم های دقیق (exact) و الگوریتم های تقریبی (approximate algorithms) تقسیم‌بندی می شوند. الگوریتم‌های دقیق قادر به یافتن جواب بهینه به‌صورت دقیق هستند اما در مورد مسائل بهینه‌سازی سخت کارایی کافی ندارند و زمان اجرای آن‌ها متناسب با ابعاد مسائل به‌صورت نمایی افزایش می یابد. الگوریتم های تقریبی قادر به یافتن جواب های خوب (نزدیک به بهینه) در زمان حل کوتاه برای مسائل بهینه‌سازی سخت هستند. الگوریتم‌های تقریبی نیز به سه دسته الگوریتم‌های ابتکاری (heuristic) و فرا ابتکاری (meta-heuristic) و فوق ابتکاری ( hyper heuristic) بخش‌بندی می‌شوند. دو مشکل اصلی الگوریتم های ابتکاری، گیر افتادن آن ها در نقاط ینه محلی، همگرایی زودرس به این نقاط است. الگوریتم های فرا ابتکاری برای حل این مشکلات الگوریتم‌های ابتکاری ارائه شده‌اند. درواقع الگوریتم‌های فرا ابتکاری، یکی از انواع الگوریتم های بهینه‌سازی تقریبی ه ستند که دارای راهکارهای برون‌رفت از نقاط بهی نه محلی هست ند و قابل یت کاربرد در طیف گسترده‌ای از مسائل رادارند. [۳ و ۴].

الگوریتم خفاش یا Bat algorithm جزء الگوریتم‌های فرا ابتکاری هست که در سال ۲۰۱۰ توسط Yang معرفی گردید. این الگوریتم بر اساس اصول زندگی خفاش‌ها طراحی‌شده است. این الگوریتم بهینه‌سازی با استفاده از اصل انعکاس صدای خفاش و دریافت صدا توسط این پرنده عمل می‌نماید [۱۵].

هر خفاش با سرعت تصادفی $V_i$ در موقعیت $X_i$ پرواز می‌کند. پس از پیدا کردن طعمه نرخ انتشار پالس

[26] Bat Algorithm

را با r تغییر می‌دهد. جستجوی محلی با استفاده از روش تصادفی تقویت می‌شود. روند تکرار جست‌وجو تا زمانی که بهترین راه‌حل انتخاب‌شده یا مشخص باشد، ادامه می‌یابد.

الگوریتم خفاش به‌صورت زیر خلاصه‌شده است:

- **مقداردهی اولیه:**

جمعیتی از d خفاش به‌صورت $X=(x_1, x_{2,...} x_d)$ به‌صورت تصادفی و برای $x_i$ راه‌حل، $f_i$ فرکانس پالس به‌صورت تصادفی در محدوده $[f_{min}, f_{max}]$ است. نرخ پالس $r_i$ و بلندی صدای $A_i$ مقداردهی می‌شوند. پس‌ازآن جمعیت اولیه ارزیابی می‌شود و سرانجام بهترین راه‌حل $X_{best}$ تعیین می‌شود.

- **تولید راه‌حل جدید:**

راه‌حل‌های جدید، $X_i^{(t+1)}$ و سرعت‌های $V_i^{(t+1)}$ برای هر راه‌حلی که توسط تغییر فرکانسی، بروز رسانی سرعت و موقعیت‌هایی از جمعیت اولیه داده می‌شود.

$$f_i^{(t)} = f_{min} + (f_{max} - f_{min}) \cup (0,1)$$

**(رابطه-۱۰): انتخاب و تعیین فرکانس اولیه**

$$V_i^{(t+1)} = V_i^{(t)} + (X_i^{(t)} - X_{best}) f_i^t$$

**(رابطه-۱۱): انتخاب بهترین راه‌حل**

$$X_i^{(t+1)} = X_i^{(t)} + V_i^{(t)}$$

**(رابطه-۱۲): تولید راه‌حل جدید**

این در حالی است که $(0,1) \cup$ یک بردار تصادفی بین ۰ تا ۱ با توزیع یکنواخت است. $X_{best}$ بهترین راه‌حل سراسری از n خفاش را نشان می‌دهد. $f_i$ نرخ افزایش فرکانسی است. $f_{min}$ و $f_{max}$ مقداری است که توسط اندازه مسئله تعیین می‌شود. برای هر خفاش مقدار اولیه فرکانسی بین $f_{min}$ و $f_{max}$ تعیین می‌شود.

- **جست‌وجوی محلی:**

این جستجو برای انتخاب بهترین راه‌حل محلی در بین راه‌حل‌های جدید بکار گرفته می‌شود و با استفاده از نرخ پالس ri و بهترین راه‌حل بروز شده، انجام می‌شود. در میان جست‌وجوی محلی، راه‌حل بروز شده جدید، برای بهترین راه‌حل با استفاده از رابطه (۱۳) تولید می‌شود:

$$X_{new} = X_{old} + \in A^t$$

**(رابطه-۱۳): انتخاب بهترین راه‌حل توسط جست‌وجوی محلی**

که $\in$ فاکتور مقیاس که همگرایی از الگوریتم را، کنترل می‌کند. $A^t$ متوسط صدای خفاش در نسل است.

- **ارزیابی راه‌حل جدید:**

ارزیابی راه‌حل‌ها با استفاده از تابع برازش که درواقع تابع برازش خوب یا بد بودن یک نتیجه را مشخص می‌کند.

- **پذیرفتن بهترین راه‌حل:**

بهترین راه‌حل عمومی تحت تنظیم راه‌حل انتخاب می‌شود. در هر نسل صدای $A_i$ و نرخ پالس $r_i$ بروز می‌شود:

$$A_i^{t+1} = aA_i^t$$

**(رابطه-۱۴)**

$$r_i^{t+1} = r_i^0 X(1 - \exp(-\gamma))$$

**(رابطه-۱۵)**

- **نزدیک‌ترین انتخاب خوشه:** هنگامی‌که کاربر جدیدی وارد می‌شود، سیستم توصیه گر

خوشه‌ای را که بیشترین شباهت به آن را دارد، با فاصله اقلیدسی[۲۷] به آن تخصیص می‌دهد.

- **محاسبه شباهت:**

تشابه بین کاربر هدف و بقیه کاربران در نزدیک‌ترین خوشه برای محاسبه نزدیک‌ترین همسایه کاربر هدف محاسبه می‌شود. روش همبستگی پیرسون برای محاسبه شباهت مورداستفاده قرار می‌گیرد.

- **توصیه گر:**

در فرآیند توصیه، با استفاده از کاربران مورد هدف صورت می‌گیرد که توسط دیگر همسایگان انتخاب‌شده‌اند. پیش‌بینی رتبه‌بندی موارد رتبه‌بندی نشده برای کاربر فعال

بر اساس میانگین وزنی رتبه‌ای از آیتم‌ها از همان خوشه با استفاده از رابطه (۱۶) محاسبه می‌شود و سپس از میان فهرست N لیست بالا را به کاربر فعال توصیه می‌کند. رتبه‌بندی از مقاصد رتبه‌بندی نشده برای کاربر فعال a با استفاده از رابطه (۱۶) پیش‌بینی می‌شود.

$$P_a(i) = \overline{R_a} + \frac{\sum_{N \in C_x} sim(a,N) \times (R_N(i) - \bar{R}_N}{\sum_{N \in C_x} sim(a,N)}$$

**(رابطه-۱۶): محاسبه‌ی میانگین وزنی**

در رابطه‌ی فوق، a نشان‌دهنده کاربر فعال، $\overline{R_a}$ میانگین a کاربر فعال، $C_x$ مجموعه‌ای از نزدیک‌ترین همسایه‌ی a کاربر فعال متعلق به یک خوشه، $\bar{R}_N$ میانگین رتبه داده‌شده توسط N همسایه‌ی کاربر فعال، $sim(a,N)$ شباهت بین a کاربر فعال و N همسایه است.

## ۴ – نتایج تجربی

### ۴-۱- محیط و شرایط شبیه سازی

الگوریتم پیشنهادی در محیط نرم‌افزاری Matlab 2018 پیاده‌سازی شده است. تمام آزمایش‌های این پژوهش در کامپیوتری با پردازنده Intel® corei5 و رم ۴ گیگابایت انجام‌شده است.

### ۴-۲- مجموعه داده[۲۸]

همان‌طور که در شکل (۴) نشان داده شد، مجموعه داده‌های فراهم‌شده فهرستی از هتل‌ها و نظرات کاربران راجع به آن‌ها است که از سایت اینترنتی[۲۹] فراهم آورده شده است. این مجموعه داده‌ها شامل مکان، کشور، نام، رتبه، نام کاربری، عنوان، طول و عرض جغرافیایی و غیره است.

مجموعه داده‌های دریافت شده شامل:

- ✓ کاربران (شناسه، نام کاربری، رمز عبور، مبدأ، ناحیه، سبک مسافرت)
- ✓ فعالیت‌ها (شناسه، نام فعالیت، موضوع فعالیت، قیمت فعالیت، عرض و طول جغرافیایی)
- ✓ رتبه‌ها (شناسه فعالیت، شناسه کاربر، رتبه)

### ۴-۳- تعیین پارامترهای ارزیابی

برای تعیین کیفیت سامانه‌های توصیه گر، تعدادی از معیارهای ارزیابی شامل میانگین خطای مطلق (MAE)[۳۰]، دقت[۳۱]، صحت[۳۲]، recall و F-measure به‌طور گسترده‌ای برای مقایسه و اندازه‌گیری عملکرد سامانه‌های توصیه گر بکار برده

[27] Euclidean distance
[28] DataSet
[29] https://www.kaggle.com/datafiniti/hotel-reviews
[30] Mean Absolute Error
[31] Accuracy
[32] Precision

می‌شوند که در ادامه هر یک از معیارها توصیف خواهند شد.

MAE یک متریک دقت آماری است که میانگین تفاضل مطلق بین نرخ‌های پیش‌بینی‌شده بر اساس تکنیک بکار برده شده و نرخ‌های واقعی کاربران را اندازه‌گیری می‌نماید که به‌صورت رابطه (۱۷) تعریف می‌شود:

$$MAE = \frac{\sum_{i,m}|R_{i,m} - P_{i,m}|}{N}$$

**(رابطه-۱۷)**

که $R_{i,m}$ و $P_{i,m}$ به ترتیب نرخ‌های واقعی و پیش‌بینی‌شده کاربر i برای آیتم m را نشان می‌دهند و N تعداد کل هتل‌های پیش‌بینی‌شده را نشان می دهد. مقدار MAE کمتر متناظر با پیش‌بینی دقیق‌تر است.

برای اینکه بفهمیم آیا کاربران به توصیه گردشگری و پیشنهاد هتل علاقه‌مند هستند معیارهای precision، recall و دقت را بکار می‌بریم که به‌طور گسترده‌ای در سامانه‌های توصیه گر بکار برده می‌شوند تا سطح هوشمندی توصیه‌ها را تعیین نمایند.

برای تعریف هر یک از معیارهای precision، recall و دقت، ابتدا باید تعریف نماییم که کدام مقدار نرخ نشان می‌دهد که یک آیتم به یک کاربر توصیه‌شده است یا نه. در این تحقیق، اگر مقدار نرخ بین ۳ تا ۵ باشد به معنای آن است که آیتم به کاربر توصیه می شود؛ در غیر این صورت آیتم توصیه نمی‌شود.

علاوه براین، چهار اصطلاح مثبت درست (TP)، منفی نادرست (FN)، مثبت نادرست (FP) و منفی درست (TN) وجود دارد که به‌صورت زیر تعریف می‌شوند:

TP: نرخ واقعی آیتم را توصیه می‌نماید و نرخ پیش‌بینی‌شده همچنین این توصیه را پشتیبانی می نماید.

FN: نرخ واقعی آیتم را توصیه می‌نماید و نرخ پیش‌بینی‌شده این توصیه را پشتیبانی نمی‌کند.

FP: نرخ واقعی آیتم را توصیه نمی‌کند اما نرخ پیش‌بینی‌شده آیتم را توصیه می‌نماید.

TN: نرخ واقعی و نرخ پیش‌بینی‌شده آیتم را توصیه نمی‌کنند.

بر اساس این چهار تعریف انجام‌شده ماتریس confusion به‌صورت جدول (۴-۱) تعریف می‌شود.

| | | Reality | |
|---|---|---|---|
| | | Positive | Negative |
| Test out-come | Positive | True Positive | False positive(error) |
| | Negative | False Negative (error) | True Negative |

(جدول ۴-۱) ماتریس confusion.

بر اساس این چهار تعریف انجام‌شده معیارهای precision، recall و دقت را در ادامه تعریف می‌نماییم.

نرخ دقت بر اساس انتخاب نمونه‌های باکیفیت بالا از مجموعه تمام نمونه‌ها است و به‌صورت رابطه (۱۸) تعریف می‌شود:

علاوه بر شاخص‌های ذکرشده، شاخص کارایی F-measure به‌طور گسترده برای ارزیابی سیستم بکار

برده می‌شود تا دقت precision-recall سیستم را ارزیابی نماید و طبق رابطه (۲۱) تعریف می‌شود.

$$F-Measure = 2\frac{precision \times recall}{precision + recall} = \frac{2TP}{2TP + FP + FN}$$

**(رابطه-۲۱)**

مقدار F-measure بالاتر دقت سیستم بهتر را نشان می‌دهد.

$$\text{precision} = \frac{TP}{TP + FP}$$

**(رابطه-۱۸)**

بالا بودن مقدار معیارهای precision، recall و دقت به معنای سیستم توصیه‌کننده بهتر است. معیار precision کسری از نمونه‌های بازیابی است که مرتبط هستند و طبق رابطه (۱۹) تعریف می‌شود:

رابطه (۱۹)

$$recall = \frac{TP}{TP + FN}$$

معیار recall کسری از نمونه‌های مرتبط است که بازیابی شده‌اند و طبق رابطه (۲۰) تعریف می‌شود:

در این قسمت ابتدا، به بررسی تأثیر تعداد خوشه‌ها بر روی روش پیشنهادی و الگوریتم‌های خوشه‌بندی Kmeans و FCM می‌پردازیم و مقدار بهینه تعداد خوشه‌ها را در این روش‌ها بر اساس معیار MAE تعیین می‌نماییم. سپس عملکرد روش پیشنهادی، الگوریتم‌های خوشه‌بندی Kmeans و FCM را باهم مقایسه می‌کنیم.

در این آزمایش در ابتدا تعداد خوشه‌ها برابر با اعداد صحیح بازه [۳،۱۵] قرار داده می‌شود و مقادیر معیارهای مختلف به ازای تعداد خوشه‌های مختلف برای هر روش محاسبه می‌شوند.

جدول (۴-۲) مقادیر MAE روش‌های مختلف را نشان می‌دهد. همان‌گونه که در جدول مشاهده می‌شود نتایج روش پیشنهادی بهتر از سایر روش‌ها است. بهترین مقدار MAE برای روش پیشنهادی، الگوریتم Kmeans و الگوریتم FCM به ترتیب برابر ۷۳،۸۵٪، ۷۴،۰۷٪ و ۷۴،۰۱٪ است که درنتیجه روش پیشنهادی به کمترین مقدار MAE و بهترین عملکرد ازنظر این معیار دست‌یافت. روش پیشنهادی به بهترین مقدار MAE با تعداد ۴ خوشه و روش FCM با تعداد ۵ خوشه و روش Kmeans با تعداد ۴ خوشه دست یافتند.

| نام روش ها | تعداد خوشه‌ها | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 3 | 4 | 5 | 7 | 9 | 11 | 13 | 15 |
| پیشنهادی | 88.22 | 88.28 | 88.14 | 88.11 | 88.07 | 88.10 | 88.15 | 88.11 |
| FCM | 88.01 | 88.10 | 88.05 | 88.10 | 88.12 | 88.01 | 88.04 | 88.04 |
| Kmeans | 87.98 | 88.04 | 88.03 | 88.07 | 88.06 | 88.01 | 88.04 | 88.03 |

**(جدول ۴-۲)مقادیر معیار MAE (٪) برای روش پیشنهادی و الگوریتم‌های FCM و Kmeans به ازای تعداد خوشه‌های مختلف.**

نتایج مربوط به مقادیر معیار precision برای روش‌های مختلف در جدول (۴-۳) ارائه‌شده است. همان‌طور که در جدول مشاهده می‌شود روش پیشنهادی به بالاترین مقدار این معیار یعنی ۸۹،۸۰٪ با تعداد ۹ خوشه دست‌یافت. بالاترین مقدار معیار precision روش پیشنهادی برابر ۸۹،۸۰٪، روش FCM برابر ۸۹،۷۹٪ و برای روش Kmeans برابر ۸۹،۷۹٪ است.

| نام روش ها | تعداد خوشه‌ها | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 3 | 4 | 5 | 7 | 9 | 11 | 13 | 15 |
| پیشنهادی | 80.86 | 80.95 | 80.76 | 80.73 | 80.67 | 80.71 | 80.78 | 80.73 |
| FCM | 80.72 | 80.73 | 80.71 | 80.68 | 80.62 | 80.70 | 80.72 | 80.71 |
| Kmeans | 80.56 | 80.63 | 80.61 | 80.65 | 80.65 | 80.58 | 80.62 | 80.61 |

**(جدول ۴-۳) مقادیر معیار Precision (٪) برای روش پیشنهادی و الگوریتم‌های FCM و Kmeans به ازای تعداد خوشه‌های مختلف.**

دقت به‌دست‌آمده از روش‌های مختلف در جدول (۴-۴) ارائه‌شده است. بالاترین دقت به‌دست‌آمده از روش‌های پیشنهادی، FCM و Kmeans به ترتیب برابر ۸۰,۹۵٪، ۸۰,۷۳٪ و ۸۰,۶۵٪ است که همان‌طور که مشاهده می‌نماییم روش پیشنهادی به بالاترین مقدار دقت یعنی ۸۰,۹۵٪ با ۴ خوشه دست‌یافت.

| نام روش ها | تعداد خوشه‌ها | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 3 | 4 | 5 | 7 | 9 | 11 | 13 | 15 |
| پیشنهادی | 89.68 | 89.69 | 89.75 | 89.79 | 89.80 | 89.76 | 89.79 | 89.79 |
| FCM | 89.71 | 89.79 | 89.69 | 89.79 | 89.73 | 89.75 | 89.79 | 89.76 |
| Kmeans | 89.79 | 89.78 | 89.75 | 89.70 | 89.74 | 89.77 | 89.75 | 89.77 |

**(جدول ۴-۴) مقادیر معیار دقت (٪) برای روش پیشنهادی و الگوریتم‌های FCM و Kmeans به ازای تعداد خوشه‌های مختلف.**

در جدول (۴-۵) مقادیر معیار recall برای روش‌های مختلف ارائه شد. روش پیشنهادی به بالاترین مقدار rcall یعنی ۸۶,۹۰٪ با تعداد ۴ خوشه دست‌یافت. بالاترین مقدار معیار recall برای روش FCM برابر مقدار ۸۶,۵۷٪ و روش Kmeans برابر مقدار ۸۶,۴۹٪ هستند.

| نام روش ها | تعداد خوشه‌ها | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 3 | 4 | 5 | 7 | 9 | 11 | 13 | 15 |
| پیشنهادی | 74.00 | 73.85 | 73.95 | 74.01 | 74.08 | 74.06 | 73.98 | 74.02 |
| FCM | 74.04 | 74.02 | 74.01 | 74.03 | 74.12 | 74.08 | 74.21 | 74.32 |
| Kmeans | 74.09 | 74.07 | 74.23 | 74.13 | 74.14 | 74.23 | 74.27 | 74.44 |

**(جدول ۴-۵) مقادیر معیار recall (٪) برای روش پیشنهادی و الگوریتم‌های FCM و Kmeans به ازای تعداد خوشه‌های مختلف.**

مقادیر معیار F-Measure روش‌های مختلف، در جدول (۴-۶) گزارش‌شده است. همان‌طور که در جدول مشاهده می‌نمایید بالاترین مقدار این معیار برابر ۸۸,۲۸٪ هست که متعلق به روش پیشنهادی است که با تعداد ۴ خوشه به آن دست‌یافت. بالاترین مقادیر معیار F-Measure برای روش‌های FCM و Kmeans به ترتیب برابر ۸۸,۱۲٪ و ۸۸,۰۷٪ هست.

| نام روش ها | تعداد خوشه‌ها | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 3 | 4 | 5 | 7 | 9 | 11 | 13 | 15 |
| پیشنهادی | 86.81 | 86.90 | 86.58 | 86.50 | 86.40 | 86.50 | 86.57 | 86.50 |
| FCM | 86.38 | 86.48 | 86.47 | 86.48 | 86.57 | 86.32 | 86.36 | 86.39 |
| Kmeans | 86.19 | 86.37 | 86.38 | 86.49 | 86.45 | 86.31 | 86.39 | 86.36 |

**(جدول ۴-۶) مقادیر معیار F-Measure (٪) برای روش پیشنهادی و الگوریتم‌های FCM و Kmeans به ازای تعداد خوشه‌های مختلف.**

در ادامه این بخش، به تحلیل و بررسی روش پیشنهادی در مقایسه با الگوریتم‌های خوشه‌بندی Kmeans و FCM خواهیم پرداخت.

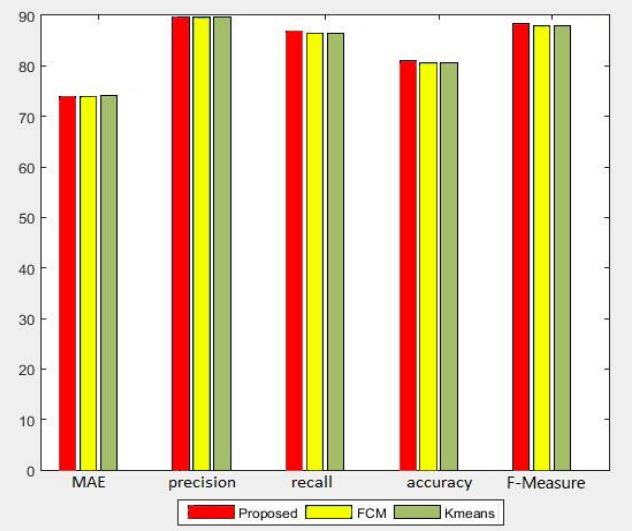
در جدول (۴-۷) مقادیر معیارهای مختلف برای روش‌های مختلف گزارش شد.

| نام روش ها | نام معیارها | | | | |
|---|---|---|---|---|---|
| | MAE | precision | recall | accuracy | F-Measure |
| پیشنهادی | 73.85 | 89.69 | 86.90 | 80.95 | 88.28 |
| FCM | 74.01 | 89.69 | 86.47 | 80.71 | 88.05 |
| Kmeans | 74.07 | 89.78 | 86.37 | 80.63 | 88.04 |

**(جدول ۴-۷) مقایسه روش پیشنهادی با سایر روش‌ها ازنظر معیارهای مختلف.**

### ۴-۴- مقایسه‌ی روش پیشنهادی با سایر روش‌ها

با توجه به جدول (۴-۷) مشاهده می‌شود که روش پیشنهادی به بهترین مقدار MAE در مقایسه با سایر روش‌ها دست‌یافت. همچنین روش پیشنهادی ازنظر معیار recall به بالاترین مقدار یعنی ۸۶,۹۰٪ و ازنظر معیار دقت به بهترین مقدار یعنی ۸۰,۹۵٪ و ازنظر معیار F-Measure به بالاترین مقدار یعنی ۸۸,۲۸٪ دست‌یافت اما ازنظر معیار precision همانند روش FCM عمل نمود و بعد روش Kmeans خوب عمل نمود. شکل (۴-۲) روش‌های مختلف را ازنظر معیارهای مختلف بر اساس نمودار میله‌ای باهم مقایسه می‌نماید.




(شکل-۵): نمودار میله‌ای روش‌های مختلف ازنظر معیارهای مختلف.

## ۵- بحث و نتیجه‌گیری

پالایش مشارکتی موفقیت خود را در سیستم‌های توصیه و پیشنهاد سنتی مانند پیشنهاد کتاب یا فیلم نشان داده‌اند. بااین‌وجود، نشان داده شد که این سیستم برای توصیه گردشگری مناسب نیست. علت این موضوع از سه بخش تشکیل‌شده است. اول، برخلاف سیستم-های توصیه سنتی که افراد می‌توانند به شکلی واضح به اشیاء امتیاز دهند، سیستم‌های توصیه گردشگری تنها دارای روش‌های امتیازدهی ضمنی هستند ازجمله بازدید از محصول‌هایی که نمی‌توانند اولویت‌های مشتری را به‌خوبی نشان دهند. دوم، اگرچه انواع بسیاری از منابع تولیدشده توسط کاربران برای ارائه توصیه‌های گردشگری موردمطالعه قرارگرفته‌اند، بااین‌وجود همه‌ی آن‌ها با مشکل پراکندگی داده روبه‌رو هستند و در آخر، محبوبیت شیء، نقش مهم‌تری در سیستم‌های توصیه گردشگری بازی می‌کند، ازآنجایی‌که مشتریان معمولاً از محبوب‌ترین پیشنهادهای گردشگری ناشناخته بازدید می‌کنند، بااین‌وجود، نمی‌توان از محبوبیت مکان به‌عنوان تنها معیار اندازه‌گیری استفاده شود زیرا برای سیستم توصیه به افراد مناسب نیست، چه برسد به توصیه‌های ترکیبی که پیچیده‌تر نیز می‌باشند. مزیت اصلی روش مایکی کردن محبوبیت محصولی، شباهت کاربر و شباهت ویژگی به‌صورت یک مدل تنها است که باعث بهبود عملکرد می‌شود. همچنین، با کمک خوشه‌بندی فازی بهینه‌شده، سه فاکتور به شکلی مؤثر برای کمک به توصیه‌های شخصی ترکیب‌شده‌اند. علاوه بر این، فازی زمانی که در مورد توصیه‌های ترکیبی به کار می‌رود (در حالتی که تعداد خوشه‌ها و رتبه‌ها به یک اندازه باشند)

در مقایسه با سایر الگوریتم‌های مرتبه بندی مؤثرتر است.

## ۶ – مراجع

[1] پورحسن کلایی، ابراهیم، ثریایی، علی و باقری، محمد (۱۳۹۶). بررسی تأثیر عوامل توصیه گر بر رفتار خرید برنامه‌ریزی نشده مطالعه موردی: فروشگاه‌های زنجیره‌ای ایران کتان استان مازندران. سیویلکا، ناشر تخصصی مقالات کنفرانس‌ها و ژورنال‌ها

[2] رجب‌زاده، مصطفی و رافع، رضا (۱۳۹۶). ارائه گر ترکیبی برای تجارت یک سیستم توصیه الکترونیک. مجله مهندسی برق دانشگاه تبریز، جلد ۵۴، شماره ۵، زمستان ۴

[3] الگوریتم‌های بهینه‌سازی فرا ابتکاری (همراه با کاربردهایی در مهندسی برق)/تألیف دکتر فرشاد مریخ بیات؛ انتشارات جهاد دانشگاهی؛ ۱۳۹۳

[4] الگوریتم‌های بهینه‌سازی فرا ابتکاری/تألیف مسعود یقینی، محمدرحیم اخوان کاظم‌زاده. جهاد دانشگاهی واحد صنعتی امیرکبیر شابک ۹۷۸-۹۶۴-۲۱۰-۰۷۸

[5] *Konstan, J.A., 2008, June. Introduction to recommender systems. In Proceedings of the 2008 ACM SIGMOD international conference on Management of data (pp. 1373-1374).*

[6] *Resnick, P. and Varian, H. R. 1997. Recommender systems, Commun. ACM 40, 3 (Mar, 1997).*

[7] *Schafer, J. B. Konstan, J. and Riedi, J. Recommender systems in e-commerce . In Proceedings of the 1st ACM Conference on Electronic Commerce, Denver, Colorado, United States, (Nov, 1999).*

[8] *Zheng Y, Xie X. (2011). Learning travel recommendations from user-generated GPS traces. ACM Transactions on Intelligent Systems and Technology. 2(1) [article no. 2].*

[9] *D. Gavalas, C. Konstantopoulos, K. Mastakas, and G. Pantziou. (2014). Mobile recommender systems in tourism, Journal of Network and Computer Applications, vol. 39, pp. 319-333*

[10] *Di Bitonto P, Di Tria F, Laterza M, Roselli T, Rossano V, Tangorra F. (2010). Automated generation of itineraries in recommender systems for tourism. In: Proceedings of the 10th international conference on web engineering (ICWE'10). p. 498–508*

[11] *D. Gavalas, V. Kasapakis, C. Konstantopoulos, G. Pantziou, N. Vathis, and C. Zaroliagis. (2015). the eCOMPASS multimodal tourist tour planner, Expert Systems with Applications, vol. 42, pp. 7303-7316*

[12] *Davies, N. Cheverst, K. Mitchell, K. & Friday, A. (1999, February). 'Caches in the air': disseminating tourist information in the GUIDE system. In Mobile Computing Systems and Applications, 1999. Proceedings. WMCSA'99. Second IEEE Workshop on (pp. 11-19). IEEE*

[13] *C. Emmanouilidis, A. Koutsiamanis and A.Tasidou, Mobile guides: Taxonomy of architectures, context awareness, technologies and applications, Network and Computer Applications, pp. 2013*

[14] *M.Elyes and H.Masri, A personalized hybrid tourism recommender system,*

*Computer Systems and Applications, 2017.*

[15] *X.Yang, A New Metaheuristic Bat-Inspired Algorithm, in: Nature Inspired Cooperative Strategies for Optimization, Studies in Computational Intelligence 2010*

*[16] B. Venkataramana, A Study of Fuzzy and Non-fuzzy clustering algorithms on Wine Data, Adikavi Nannaya Universit, January 2017*

*[17] S.Gokten, "Using fuzzy c-means clustering algorithm in financial health scoring, Baskent University, 2017*

[18]*V. Vellaichamy and V.Kalimuthu, "Hybrid Collaborative Movie Recommender System Using Clustering and Bat Optimization, vol3, pp2017*

[19]*S.Bagchi, Performance and quality assessment of similarity measures in collaborative filtering using mahout, Procedia Computer Science, vol 50.2015*

*[20] Andrew R. Webb. QinetiQ Ltd. Malvern, UK. "Statistical Pattern Recognition", second edition, John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, (2002)*

آرش خسروی مدرک کارشناسی خود را در رشته مهندسی نرم‌افزار در سال ۱۳۸۲ از دانشگاه صنعتی اصفهان، مدرک کارشناسی ارشد خود را در رشته مهندسی فناوری اطلاعات در سال ۹۲ و مدرک دکتری خود را در رشته مهندسی فناوری اطلاعات، گرایش سیستم‌های اطلاعاتی در سال ۹۶ از دانشگاه صنعتی مالزی اخذ کرده است. ایشان در حال حاضر به‌عنوان هیئت‌علمی مرکز آموزش عالی محلات مشغول به کار هستند. زمینه‌های پژوهشی موردعلاقه ایشان عبارت‌اند از: هوش تجاری، سیستم‌های پیشنهاددهنده، مدیریت دانش مشتری، داده‌کاوی، متن‌کاوی (تحلیل احساسات، شخصیت شناسی، نظر کاوی)، فناوری اطلاعات در پزشکی

محمدعلی صادقی مدرک کارشناسی خود را در رشته مهندسی نرم‌افزار کامپیوتر در سال ۱۳۹۳ با معدل الف اخذ و در حال حاضر در مقطع کارشناسی ارشد رشته مهندسی کامپیوتر گرایش نرم‌افزار دانشگاه شهاب دانش قم مشغول به تحصیل می باشد.