

## سیستم توصیه‌گر زمینه آگاه سفر با بهره‌گیری از اطلاعات عکس‌های برچسب‌دار جغرافیایی

\*رضوان محمدرضایی لرکی      \*\*رضا روانمهر      \*\*\*میلاد امرالهی

\*دانشجوی دکتری، گروه مهندسی کامپیوتر، واحد تهران مرکزی، دانشگاه آزاد اسلامی، تهران  
\*\*استادیار، گروه مهندسی کامپیوتر، واحد تهران مرکزی، دانشگاه آزاد اسلامی، تهران  
\*\*\*کارشناسی ارشد، گروه مهندسی کامپیوتر، واحد تهران مرکزی، دانشگاه آزاد اسلامی، تهران  
تاریخ دریافت: ۱۳۹۸/۱۲/۲۲      تاریخ پذیرش: ۱۳۹۹/۰۶/۲۹

### چکیده

سیستم‌های توصیه‌گر بر اساس اطلاعات کاربران و آیتم‌ها، سلايق و علاقه‌مندی‌های مورد نظر کاربران را پیش‌بینی کرده و بر اساس آن‌ها آیتم‌های مناسب را به کاربران پیشنهاد می‌دهند. اکثر رویکردهای موجود برای سیستم‌های توصیه‌گر بر روی پیشنهاد دادن مرتبط‌ترین آیتم‌ها تمرکز می‌کنند و اطلاعات زمینه‌ای مانند زمان یا مکان را در نظر نمی‌گیرند. در این مقاله یک روش برای سیستم توصیه‌گر زمینه آگاه سفر با بهره‌گیری از اطلاعات عکس‌های برچسب‌دار جغرافیایی ارائه شده است که از دقت بالایی برخوردار است. این روش نسبت به مقالات مشابه، تعداد زمینه‌های بیشتری مانند (وضعیت آب و هوا، وضعیت روحی کاربر، میزان ترافیک و ...) را در نظر می‌گیرد. با این کار باعث نزدیک شدن کاربران یک خوشه و در نتیجه افزایش دقت روش پیشنهادی به جای استفاده از الگوریتم خوشه‌بندی کلاسیک از ادغام دو الگوریتم رقابت استعماری و خوشه‌بندی C-Means فازی استفاده می‌کند. ارزیابی روش ارائه شده بر روی مجموعه داده فلیکر انجام شده و نتایج ارزیابی حاکی از آن است که روش پیشنهادی قادر به فراهم کردن توصیه‌های موثر و دقیق مطابق با علاقه‌مندی‌های کاربر و موقعیت فعلی بازدید او می‌باشد. رویکرد پیشنهادی در معیارهای دقت، متوسط مطلق دقت و میانگین قدر مطلق خطا به ترتیب به ۰.۶۹ و ۰.۵۳ و ۰.۳۱ دست یافته است.

**کلمات کلیدی:** سیستم توصیه‌گر زمینه آگاه، سیستم توصیه‌گر سفر، زمینه، الگوریتم رقابت استعماری، خوشه‌بندی C-Means فازی.

## ۱- مقدمه

سازی به فرایندی گفته می‌شود که در طی آن رفتار سامانه بر اساس دانش استخراج شده از رفتارهای کاربران، تغییر می‌کند [۵]. این مقاله سعی دارد با توجه به فرآیند شخصی سازی، رویکردهای مرتبط با سیستم‌های توصیه‌گر و چگونگی به‌کارگیری و کارایی این الگوریتم‌ها را در محیط‌های گردشگری الکترونیک بررسی نماید.

با توجه به این که کیفیت توصیه‌ها در سیستم‌های توصیه‌گر گردشگری بسیار حائز اهمیت است، اگر توصیه‌ای ارائه شود که کاربر قبلاً آن را دریافت کرده باشد و یا به آن نیازی نداشته باشد به کاهش نگرش مثبت وی به توصیه‌ها و بی‌توجهی به آن‌ها می‌انجامد. به همین دلیل لازم است سیستم‌های توصیه‌گر در ارائه توصیه به گردشگران از دقت بالایی برخوردار باشند که بکارگیری منطق فازی در کاهش عدم قطعیت و ارائه توصیه‌های موثر، می‌تواند مفید واقع شود [۶]. در این پژوهش ترکیب دو الگوریتم رقابت استعماری و خوشه‌بندی فازی به منظور پردازش بیشتر و قوی‌تر در مقابل استفاده از خوشه‌بندی کلاسیک، موجب افزایش دقت پیشنهادهای ارائه شده توسط سیستم است.

آنچه موجب متمایز شدن روش پیشنهادی این مقاله شده است افزودن تعداد زمینه‌های<sup>۱</sup> بیشتر (وضعیت آب و هوا، وضعیت روحی کاربر، میزان ترافیک و ...) به نسبت مقالات مشابه است که فقط از زمینه‌های زمانی و مکانی استفاده کرده‌اند. این نکته موجب نزدیک تر شدن کاربران درون یک خوشه و به موجب آن افزایش دقت گردید.

در روش ارائه شده به منظور پارامترهای کیفیت سرویس دقت، متوسط مطلق دقت و میانگین قدر مطلق خطا اندازه‌گیری و ارزیابی می‌شوند. در این تحقیق سعی می‌شود پارامترهای ذکر شده بهبود یابند. به منظور ارزیابی روش ارائه شده از دیتاست فلیکر، یکی از بزرگ‌ترین سلیت‌های اشتراک‌گذاری تصویر و ویدیو آنلاین استفاده شده است. به طور خلاصه نوآوری‌های این پژوهش به ترتیب زیر عبارتند از:

- ارائه یک روش مناسب برای سیستم زمینه آگاه سفر بر اساس ترکیب دو الگوریتم رقابت استعماری و خوشه‌بندی C-Means فازی

شاید تا به حال برای شما این سؤال پیش آمده باشد که شرکت گوگل با چه الگوریتمی تبلیغات هوشمند را به کاربران نمایش می‌دهد؟ یا فیس‌بوک چگونه هنگام آپلود عکس‌ها و برچسب‌زنی آن‌ها، نام دوستانمان را به ما پیشنهاد می‌دهد؟ پاسخ این سوالات در استفاده این سایت‌ها از سامانه‌های توصیه‌گر است. سیستم‌های توصیه‌گر با توجه به رشد روزافزون اطلاعات در دنیای حاضر، افزایش چشمگیر کاربران آنلاین و همچنین ضرورت تحویل سریع اطلاعات درست به کاربران، اهمیتی حیاتی دارند و با هدف حذف سربارهای اطلاعاتی، یکی از ابزارهای مهم تجارت الکترونیک شده‌اند [۱]. مهم‌ترین هدف در این سیستم‌ها، جذب مشتریان و جلب اعتماد آن‌ها از طریق ارائه بهترین و مناسب‌ترین پیشنهاد خرید محصولات، با توجه به علایق و سلاقی آن‌ها می‌باشد [۲]. تحقیقات در حوزه الگوریتم‌های توصیه‌گر به صورت روز افزون در حال افزایش هستند.

شرکت‌ها با ایجاد سامانه‌های هوشمندی که رفتار گذشته کاربر و نیز رفتار کاربران مشابه را بررسی می‌کنند، پیشنهادات و توصیه‌های مناسب و مورد توجه کاربر را به او ارائه می‌دهند [۱]. مواردی از قبیل فرصت‌های شغلی مناسب کاربر، فیلم‌های مورد علاقه کاربر، ویدئوهای پیشنهادی، دوستان فیس بوکی که احتمالاً آن‌ها را می‌شناسید و ... از جمله موارد پیشنهادی به کاربران می‌تواند باشد.

با توجه به جایگاه فناوری اطلاعات در صنعت گردشگری و پیدایش گردشگری الکترونیک، نقطه‌ی عطفی میان این صنعت و فناوری اطلاعات به وجود آمده است. توجه به گردشگری و استفاده از فناوری‌های نوین در حال گسترش آن، زمانی روشن می‌شود که بدانیم درآمد حاصل از آن، هم‌ردیف درآمدهای نفتی می‌باشد [۳]. بنابراین یکی از مهم‌ترین کاربردهای فناوری اطلاعات و ارتباطات، صنعت گردشگری است که نقش چشم‌گیری در جهت توسعه‌ی منافع ملی ایفا می‌کند.

از نکات قابل توجه در این زمینه تأمین نیازهای کاربران و ارائه‌ی خدمات مطلوب‌تر به آن‌ها است. در سال‌های اخیر الگوریتم‌های مختلفی در جهت بهبود این سیستم‌ها ابداع شده‌اند که در اکثر کاربردها و زمینه‌های مبتنی بر وب، شخصی سازی نقش مهمی را ایفا می‌کند [۴]. شخصی

<sup>1</sup> Context

را با بررسی و دنبال کردن رفتارها و فعالیت‌های او بیابد. (به عنوان مثال باید ببینید کاربر به چه موزیک‌هایی بیشتر گوش می‌دهد، چه صفحاتی را بازدید می‌کند، با چه کسانی در ارتباط است و ...).

علاوه بر اطلاعات ضمنی و صریح برخی از سیستم‌ها نیز هستند که از اطلاعات شخصی کاربران استفاده می‌کنند. به عنوان مثال سن، جنسیت و ملیت کاربران می‌توانند منابع خوبی برای شناخت کاربر و ارائه پیشنهاد به وی باشند. به این‌گونه از اطلاعات، اطلاعات آماری<sup>۲</sup> گفته می‌شود که گروهی از سیستم‌های پیشنهاددهنده مبتنی بر همین اطلاعات بنا شده‌اند. با ظهور وب ۲ و گسترش شبکه‌های اجتماعی در سال‌های اخیر محققین به منبع اطلاعاتی دیگری برای بهبود کیفیت پیشنهادها پی بردند که همان اطلاعات موجود در شبکه‌های اجتماعی بود و بر همین اساس کارهای تحقیقاتی زیادی در این حوزه انجام شد [۱].

## ۲-۲- داده‌های زمینه‌ای و انواع آن‌ها

آدوماویکویس و توژیلین [۸] بر این واقعیت تاکید دارند که با توجه به این که مفهوم زمینه بسیار گسترده است، تمرکز زمینه باید به طور مستقیم بر حوزه‌های مرتبط با سیستم‌های توصیه‌گر نظیر داده کاوی، شخصی‌سازی تجارت الکترونیک، پایگاه‌های داده، بازیابی اطلاعات، سیستم‌های فراگیر آگاه از زمینه، بازاریابی و مدیریت باشد. نیاز به تعریف و مدل‌سازی یک زمینه در سیستم‌های توصیه‌گر از طریق رویه‌ای سازگار، توسط پژوهشگران مختلف شناسایی شده است. بنا به نظر وربرت و همکارانش [۹] تعریف و مدل‌سازی دقیق زمینه، شناسایی آن چه که زمینه را می‌سازد آسان نموده و استفاده مجدد و تبادل داده‌های زمینه‌ای در برنامه‌های کاربردی را امکان‌پذیر می‌کند. توصیف رده‌های مختلف زمینه و انواع قابل به‌کارگیری آن در سیستم‌های توصیه‌گر در زیر ارائه شده است:

✓ فیزیکی: زمینه فیزیکی، موقعیت‌های محیطی کاربر یا سیستم در یک مکان خاص را نشان می‌دهد. زمینه فیزیکی شامل ویژگی‌هایی مثل نور، گرما و صدا است. بر اساس نظر وربرت و همکارانش [۹]، زمینه فیزیکی به طور گسترده در تحقیقات مربوط به خودکارسازی منازل مورد بررسی قرار گرفته است.

- بهره‌گیری از اطلاعات عکس‌های برچسب دار جغرافیایی جهت خوشه بندی و مدل‌سازی زمینه‌ای داده‌ها
- استفاده از اطلاعات زمینه‌ای مکان، تعداد بازدید (میزان محبوبیت)، وضعیت آب و هوا، موقعیت زمانی در روز، حالت روحی کاربر، میزان ترافیک، فصل و دما

در ادامه مقاله، در بخش ۲ به مبانی تحقیق و معرفی سیستم‌های پیشنهاددهنده، دسته‌ها و انواع زمینه و مسائل باز در این حوزه خواهیم پرداخت و در بخش ۳ کارهای مرتبط در زمینه سیستم‌های توصیه‌گر زمینه آگاه بررسی می‌شود. در بخش ۴ روش پیشنهادی مطرح شده، نتایج ارزیابی روش پیشنهادی به تفصیل در بخش ۵ بیان می‌شود و در پایان نتیجه‌گیری و کارهای آینده بیان شده است.

## ۲-مبانی تحقیق

سیستم توصیه‌گر سفر زمینه آگاهی که در این مقاله معرفی شده است، مبتنی بر الگوریتم خوشه‌بندی فازی و الگوریتم بهینه‌سازی رقابت استعماری است. در این بخش ابتدا مفاهیم سیستم‌های توصیه‌گر و داده‌های زمینه‌ای و سپس الگوریتم‌های خوشه‌بندی C-Means فازی و الگوریتم بهینه‌سازی رقابت استعماری به عنوان مبانی و بلاک‌های سازنده روش پیشنهادی مورد بررسی قرار می‌گیرد.

## ۲-۱- سیستم‌های توصیه‌گر

سیستم‌های توصیه‌گر سیستم‌هایی هستند که در پیدا کردن و انتخاب نمودن آیتم‌های موردنظر کاربران به آن‌ها کمک می‌کنند. طبیعی است که این سیستم‌ها بدون در اختیار داشتن اطلاعات کافی و صحیح در مورد کاربران و آیتم‌های موردنظر آن‌ها (به عنوان مثال فیلم، موزیک، کتاب و...) قادر به تولید پیشنهاد مناسب نمی‌باشند. بنابراین یکی از اساسی‌ترین اهداف آن‌ها جمع‌آوری اطلاعات گوناگون در رابطه با سلاقی کاربران و آیتم‌های موجود در سیستم است [۱]. منابع و روش‌های گوناگونی برای جمع‌آوری چنین اطلاعاتی وجود دارد. یکی از روش‌های جمع‌آوری اطلاعات به صورت صریح است که در آن کاربر با صراحت اعلام می‌کند که به چه آیتم‌هایی علاقه دارد (به عنوان نمونه با امتیاز دادن به یک موزیک) و روش دیگر روش ضمنی است که کمی دشوارتر است [۷] و در آن سیستم باید سلاقی کاربر

<sup>2</sup> Demographic

باشد. بعضی از محققین روابط اجتماعی را به عنوان بعد زمینه‌ای مهم شناسایی نموده‌اند [۹].

### ۲-۳- خوشه‌بندی C-Means فازی

خوشه‌بندی با فراهم کردن قابلیت ورود به فضای داده و تشخیص ساختار آن‌ها، یکی از ایده‌آل‌ترین روش‌ها برای کار با دنیای عظیم داده‌ها است [۱۶]. خوشه‌بندی یکی از شاخه‌های یادگیری بدون نظارت است و فرایند خودکاری است که 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 است که در آن اشیاء با یکدیگر مشابه بوده و با اشیاء موجود در خوشه‌های دیگر نامشابه می‌باشند [۱۷]. برای تشابه می‌توان معیارهای مختلفی را در نظر گرفت مثلاً می‌توان معیار فاصله را برای خوشه‌بندی مورد استفاده قرارداد و اشیائی که به یکدیگر نزدیک‌تر هستند را به‌عنوان یک خوشه در نظر گرفت که به این نوع خوشه‌بندی، خوشه‌بندی مبتنی بر فاصله نیز گفته می‌شود. روش‌های موجود خوشه‌بندی را می‌توان در پنج دسته، روش‌های سلسله مراتبی، جزءبندی، مبتنی بر تراکم، مبتنی بر شبکه و مبتنی بر مدل دسته‌بندی کرد [۱۸]. در میان روش‌های مختلف در خوشه‌بندی داده‌ها، روش خوشه‌بندی K-Means و C-Means فازی به‌طور گسترده در زمینه‌های مختلف مورد استفاده قرار می‌گیرد. الگوریتم K-Means یک از معروف‌ترین روش‌ها برای خوشه‌بندی داده‌ها به شمار می‌رود. در این الگوریتم، داده‌ها پس از تعدادی تکرار به  $k$  خوشه مختلف دسته‌بندی می‌شوند. یکی از مشکلات الگوریتم K-Means و اکثر الگوریتم‌های سنتی خوشه‌بندی این است که تعلق داده به هر خوشه با عدد صفر یا یک مشخص می‌شود. به عبارت دیگر در خوشه‌بندی کلاسیک، هر نمونه ورودی متعلق به یک و فقط یک خوشه است و نمی‌تواند عضو دو خوشه و یا بیشتر باشد. در واقع، خوشه‌ها همپوشانی ندارند.

برای حل چالش‌های روش خوشه‌بندی k-means، الگوریتم خوشه‌بندی فازی ارائه شده است [۱۹]. در خوشه‌بندی فازی یک نمونه می‌تواند متعلق به بیش از یک خوشه باشد. در الگوریتم خوشه‌بندی فازی تعلق هر داده به یک خوشه خاص با یک عدد حقیقی بین صفر و یک مشخص می‌شود. ایده بنیادین در خوشه‌بندی فازی به این ترتیب است که فرض

✓ محاسبه: زمینه محاسبه به طور گسترده توسط ریکی [۱۰] و انجمن تحقیقاتی یادگیری فراگیر [۱۱] مورد بررسی قرار گرفته است. کسب زمینه محاسباتی لازم برای حمایت از واسطه‌های هوشمند که بتوانند منابع مناسب سیستم توصیه‌گر را انتخاب و توصیه کنند مهم است.

✓ مکان: زمینه مکان در تحقیقات زمینه آگاه بسیار فراگیر شده است [۱۲, ۱۳]. به دلیل افزایش استفاده از ابزارهای متحرک، سیستم‌های توصیه‌گر مبتنی بر مکان به طور گسترده مورد استفاده قرار گرفته‌اند [۹].

✓ زمان: زمینه زمان از اطلاعات تاریخ و زمان که معمولاً شامل دقیقه، ساعت، هفته، ماه، ترم، فصل و غیره است تشکیل می‌شود. زمان معمولاً مرکب بوده و در رابطه با دسته‌های دیگری مانند مکان به صورت مدت زمان یا مهر زمانی<sup>۳</sup> مورد استفاده قرار می‌گیرد. زمینه زمان، یک نمونه یا دوره را در مدتی که اطلاعات زمینه‌ای مورد نیاز است یا مربوط به کاربر سیستم توصیه‌گر است نشان می‌دهد [۹, ۱۴].

✓ کاربر: شخصی است که از سیستم توصیه‌گر استفاده می‌کند یا آیتم‌های توصیه برای وی تولید می‌شود. برای مثال کاربران در آموزش الکترونیک، یادگیرنده‌های الکترونیکی هستند و یک مدل آموزشی را برای تولید توصیه‌های دقیق و ارزشمند یاد می‌گیرند. مدل‌های یادگیر به طور گسترده به مدل‌سازی آموزشی کاربر و ابررسانه تطبیقی آموزشی<sup>۴</sup> دسترسی دارند [۹]. زمینه کاربر در چارچوب سیستم توصیه‌گر می‌تواند شامل اطلاعات شخصی، دانش/عملکرد، علائق، روش‌های یادگیری و شناختی و پس زمینه کاربر باشد [۱۵].

✓ رولپط اجتماعی<sup>۵</sup>: رولپط اجتماعی، حلقه‌های اجتماعی، انجمن‌ها، ارتباطات و وابستگی بین دو یا چند شخص را توصیف می‌کنند. برای مثال، روابط اجتماعی می‌تواند اطلاعاتی در مورد دوستان، اشخاص، همسایگان، همکاران، دشمن‌ها و فامیل

<sup>5</sup> Social relations

<sup>3</sup> Time Stamp

<sup>4</sup> Educational Adaptive Hypermedia

استعمارگر  $n$  ام می باشد. با داشتن هزینه نرمال شده، قدرت نسبی نرمال شده هر استعمارگر، به صورت زیر محاسبه شده و بر مبنای آن، کشورهای مستعمره، بین استعمارگران تقسیم می شوند:

$$P_n = \frac{C_n}{\sum_{i=1}^{N_{imp}} C_i} \quad (2)$$

بنابراین تعداد اولیه مستعمرات یک استعمارگر برابر خواهد بود با:

$$N.C_n = \text{round}\{P_n \cdot (N_{col})\} \quad (3)$$

که در آن  $N.C_n$ ، تعداد اولیه مستعمرات یک امپراطوری و  $N_{col}$  نیز تعداد کل کشورهای مستعمره موجود در جمعیت کشورهای اولیه است. با داشتن حالت اولیه تمام امپراطوری‌ها، الگوریتم رقابت استعماری شروع می شود. سیاست کلی در این الگوریتم به این صورت است که کشورهای استعمارگر، کشورهای مستعمره را به طرف خود می کشند. همان طور که در رابطه ۴ نشان داده شده است، هزینه کل هر امپراطوری، علاوه بر هزینه کشور استعمارگر از بخشی از میانگین هزینه مستعمره‌ها نیز تشکیل می شود.

$$C_n = \text{cost}(\text{imperialist}) + \xi \text{mean}(\text{cost}(\text{colonies of impire}_n)) \quad (4)$$

کشور مستعمره، به اندازه  $X$  واحد در جهت استعمارگر حرکت کرده و به موقعیت جدیدی کشانده می شود. اگر  $d$  را فاصله میان استعمارگر و مستعمره و  $X$  نیز عددی تصادفی با توزیع یکنواخت باشد؛ برای  $X$  داریم:

$$x \sim U(0, \beta \times d) \quad (5)$$

که در آن  $\beta$  عددی بزرگ‌تر از یک و نزدیک به ۲ است. مقدار  $\beta=2$  می تواند یک انتخاب مناسب باشد. وجود ضریب  $\beta>1$  باعث می شود تا کشور مستعمره در حین حرکت به سمت کشور استعمارگر، از جهت‌های مختلف به آن نزدیک شود. برای افزایش ناحیه جستجوی اطراف استعمارگر، یک انحراف زاویه ای برابر  $\theta$  که از یک توزیع یکنواخت تصادفی تبعیت می نماید، به بردار اصلی اضافه می شود:

$$\theta \sim U(-\gamma, \gamma) \quad (6)$$

$\gamma$  پارامتری است که محدوده انحراف زاویه‌ای را کنترل می کند. گام بعدی، گام رقابت استعماری می باشد. در این مرحله ضعیف‌ترین مستعمره از ضعی ترین امپراطوری انتخاب شده و به یک امپراطوری قوی (نه لزوماً قوی‌ترین

کنیم هر خوشه مجموعه‌ای از عناصر است، سپس با تغییر در تعریف عضویت عناصر در این مجموعه از حالتی که یک عنصر فقط بتواند عضو یک خوشه باشد (حالت افرازی)، به حالتی که هر عنصر می تواند با درجه عضویت‌های مختلف به چندین خوشه مختلف تعلق داشته باشد، خوشه‌بندی‌هایی که انطباق بیشتری با واقعیت دارند ارائه کنیم [۱۹]. در سال‌های اخیر نسخه‌های بهبودیافته‌ای از این الگوریتم نیز ارائه شده است. برای حل مشکلات ناشی از مقداردهی اولیه خوشه‌ها، استتکو و همکارانش یک طرح جدید در مقداردهی اولیه خوشه‌ها در خوشه‌بندی C-Means فازی ارائه دادند [۲۰]. روش فازی C-Means به طور گسترده در زمینه‌های مختلف مانند سنجش از راه دور، خوشه‌بندی سری‌های زمانی و قطعه‌بندی تصاویر رنگی و ... مورد استفاده قرار می گیرد.

یکی دیگر از مشکلات خوشه‌بندی، تعیین تعداد بهینه خوشه‌ها است. در این مقاله برای حل این مشکل و اجرای حدس اولیه تعیین تعداد خوشه‌ها پیش از اجرای الگوریتم، از الگوریتم بهینه‌سازی رقابت استعماری استفاده شده است که در بخش ۴-۲ توضیح داده شده است.

#### ۲-۴- الگوریتم بهینه سازی رقابت استعماری

الگوریتم بهینه‌سازی رقابت استعماری، یک الگوریتم تکاملی است که با هدف بهینه‌سازی مسائل مختلف از رقابت استعمارگران الهام گرفته شده است. همانند دیگر الگوریتم‌های تکاملی، این الگوریتم، نیز با تعدادی جمعیت اولیه تصادفی که هر کدام از آن‌ها یک کشور نامیده می شوند، شروع می شود. تعدادی از بهترین عناصر به منزله‌ی استعمارگر انتخاب و باقیمانده جمعیت نیز به منزله‌ی مستعمره در نظر گرفته می شوند.

در این الگوریتم، برای حل مساله بهینه‌سازی مفروض،  $N$  کشور در نظر گرفته می شود که هر کدام با یک بردار نمایش داده شده و نشان دهنده یک نقطه در فضای  $n$  بعدی می باشد. از بین این نقاط، نقاطی که کمترین هزینه را طبق تابع بهینه‌سازی داشته باشند، به عنوان استعمارگر و بقیه نقاط به عنوان مستعمره در نظر گرفته می شوند [۲۱]. در ابتدا برای هر استعمارگر هزینه نرمال شده به صورت زیر محاسبه می شود:

$$C_n = c_n - \max_i \{c_i\} \quad (1)$$

در رابطه (۱)  $C_n$ ، هزینه نرمال شده استعمارگر  $n$  ام،  $\max\{c_i\}$  بیشترین هزینه میان استعمارگران و  $c_n$  هزینه

شماره ۱ ارایه داده شده‌اند. در مقایسه با اکثر روش‌های ارایه شده در این زمینه، داده‌های زمینه‌ای روش پیشنهادی در این مقاله علاوه بر داده‌های مکانی که شامل مختصات جغرافیایی مکان بازدید می‌شود، از داده‌های زمانی مانند موقعیت زمانی بازدید در روز، فصل بازدید و همچنین داده‌های مربوط به تعداد بازدید (میزان محبوبیت)، وضعیت آب و هوا، حلت روحی کاربر، میزان ترافیک، و دما نیز جهت افزایش دقت پیشنهادات استفاده می‌کند.

روش پیشنهادی با هدف مدل‌سازی داده‌ها بر اساس داده‌های زمینه‌ای و با استفاده از ادغام دو الگوریتم رقابت استعماری و روش C-Means فازی، خوشه‌بندی انجام می‌دهد و مبتنی بر نتایج به دست آمده، نقاط بازدید مناسب را به کاربران پیشنهاد می‌دهد. روش پیشنهاد داده شده در این مقاله مبتنی بر الگوهای مدل‌سازی زمینه‌ای و پس‌پالایش زمینه‌ای است.

امپراطوری) داده می‌شود. البته نحوه انتخاب امپراطوری موردنظر به این صورت است که هرچه یک امپراطوری قوی‌تر باشد، احتمال انتخاب آن بالاتر می‌رود. در نهایت وقتی یک امپراطوری همه مستعمره‌های خود را از دست داد، آن امپراطوری از لیست امپراطوری‌ها خارج شده و در یک رقابت استعماری به عنوان مستعمره به امپراطوری‌های دیگر داده می‌شود. روند تکامل در یک حلقه قرار دارد که تا برآورده شدن شرط توقف ادامه می‌یابد [۲۱].

الگوریتم رقابت استعماری در روش پیشنهادی این مقاله، برای به دست آوردن تعداد خوشه‌ها و مراکز آن‌ها به کار برده شده است. در واقع مقادیر به دست آمده از این الگوریتم، تعداد خوشه‌ها و مراکز آن‌ها را برای اجرای خوشه‌بندی فازی فراهم می‌کنند.

### ۳-مروری بر تحقیقات پیشین

در گذشته، سیستم‌های توصیه‌گر، پیشنهادات را فقط مبتنی بر اطلاعات کاربران و آیتم‌ها تولید می‌کردند اما سیستم‌های توصیه‌گر مبتنی بر زمینه، برای ارایه پیشنهادات، علاوه بر اطلاعات کاربر و آیتم‌ها از اطلاعات مختلف زمینه‌ای نیز استفاده می‌کنند. بر اساس شیوه و زمان استفاده از اطلاعات زمینه‌ای در سیستم‌های توصیه‌گر، سه نوع الگوی مختلف وجود دارد که عبارتند از [۲۲]:

✓ پیش‌پالایش زمینه‌ای؟ اطلاعات زمینه‌ای پیش از اجرای فرایند پیشنهاد دهنده، به عنوان یک روش فیلترکننده بر روی داده‌ها اعمال می‌شوند.

✓ پس‌پالایش زمینه‌ای<sup>۶</sup>: اطلاعات زمینه‌ای در ابتدای فرایند نادیده گرفته می‌شوند و پس از آن که پیشنهادات بر اساس مدل توصیه‌گر ایجاد شد، از اطلاعات زمینه‌ای برای فیلتر کردن نتایج و ارایه پیشنهادات مطابق با علاقه مندی‌های کاربر، استفاده می‌شود.

✓ مدل‌سازی زمینه‌ای<sup>۸</sup>: از اطلاعات زمینه‌ای به طور مستقیم در مدل تولید پیشنهادات استفاده می‌شود.

تحقیقات بسیاری در ارتباط با سیستم‌های توصیه‌گر زمینه آگاه و کاربردهای آن‌ها در زمینه سفر و گردشگری انجام شده است که در این بخش تعدادی از آن‌ها بر اساس نوع الگو، نوع اطلاعات زمینه‌ای مورد استفاده، زمینه کاربردی و روش به‌کار برده شده برای تولید پیشنهادات، در جدول

<sup>8</sup> Contextual modeling

<sup>6</sup> Contextual Pre-filtering

<sup>7</sup> Contextual post-filtering

## جدول ۱: مقایسه تحقیقات پیشین بر اساس تکنیک پیشنهادی، نوع الگو، نوع اطلاعات زمینه‌ای، کاربرد و رویکرد

مقاله	تکنیک سیستم توصیه‌گر	نوع الگو	نوع اطلاعات زمینه‌ای	کاربرد	رویکرد مقاله نسبت به اطلاعات زمینه‌ای
[۲۳]	مبتنی بر مدل، رگرسیون خطی <sup>۹</sup>	مدل‌سازی زمینه‌ای	شبکه‌های اجتماعی، زمان	هتل، گردشگری	بر اساس مدل رگرسیون خطی و داده‌هایی که ارائه دهنده زمینه مورد علاقه کاربر هستند، علاقه مندی‌های کاربر پیش بینی می‌شود. این داده می‌تواند بر اساس سه روش احتمالی محاسبه شوند: روش مبتنی بر اطلاعات انحصاری، روش مبتنی بر تقویت اطلاعات، روش مبتنی بر احتمال خی-۲
[۲۴]	روش‌های مکاشفه‌ای	پس پالایش زمینه‌ای	مکان، زمان، آب و هوا	هتل و گردشگری	ویژگی‌های زمینه‌ای مسافرت‌های گذشته کاربر بر اساس هر مکان بررسی می‌شود. پیشنهادات مبتنی بر زمینه از طریق یافتن کاربرانی که شباهت بیشتری به یکدیگر دارند استخراج شده و برای هر مکان امتیازی به آن تعلق می‌گیرد و پیشنهاداتی که مکان آن‌ها شرایط زمینه‌ای را ندارند فیلتر و حذف می‌شوند.
[۲۵]	مدل‌سازی موضوعی، شبکه عصبی مصنوعی، ماشین بردار پشتیبان	پیش پالایش و مدل‌سازی زمینه‌ای	مکان، شرایط آب و هوا، داده‌های بلاگ‌های گردشگری	مکان‌های گردشگری	روش تخصیص دریکله پنهان، موضوعات نظرات گردشگران را استخراج کرده و روش ماشین بردار پشتیبان برای آن‌ها تحلیل احساسات انجام می‌دهد. سپس این داده‌ها به همراه امتیازها، رتبه‌های مکان‌ها و اطلاعات تاریخچه‌ای کاربران و مکان‌ها با استفاده از یک شبکه عصبی مصنوعی پردازش شده و مکان‌های موردعلاقه کاربر را که کمتر موردتوجه بوده‌اند، به کاربر پیشنهاد داده می‌شود.
[۲۶]	کاوش داده‌های استخراج شده از وب	مدل‌سازی زمینه‌ای	زمان، مکان	شهرهای گردشگری	در این سیستم پیشنهاد دهنده گردشگری چند سطحی، علاوه بر علاقه مندی‌های کاربر، محدودیت‌های او نیز در نظر گرفته می‌شود. ابتدا بر اساس تاریخ‌های سفر و بودجه گردشگران، تعدادی از شهرها انتخاب می‌شوند. سپس برای هر شهر یک امتیاز در نظر گرفته می‌شود که بر اساس محدودیت‌ها و علاقه مندی‌های کاربران محاسبه می‌شود و بر اساس آن‌ها شهرهای مناسبی به کاربر پیشنهاد داده می‌شود.
[۲۷]	تکنولوژی‌های اینترنت اشیا، آنتولوژی، وب معنایی	پیش پالایش و مدل‌سازی زمینه‌ای	زمان، مکان، شبکه-های اجتماعی، داده‌های سلامت کاربران، سنسور ابزارها	گردشگری	یک سیستم گردشگری هوشمند مبتنی بر آنتولوژی و زمینه آگاه که با استفاده از تکنولوژی اینترنت اشیا و سرویس‌های وب معنایی به گردشگران سرویس‌های متناسبی را پیشنهاد می‌دهد
[۲۸]	ماشین بردار پشتیبان	پیش پالایش و مدل‌سازی زمینه‌ای	داده‌های جمع آوری شده از سایت‌های گردشگری	گردشگری	سیستم توصیه‌گر پروفایل‌های کاربران را از طریق TripAdvisor استخراج کرده و مجموعه داده به دست آمده را با روش ماشین بردار پشتیبان پردازش می‌کند. هدف این سیستم، ایجاد سهولت و بالا بردن دقت در پیشنهادات ارائه شده به گردشگران بر اساس داده‌های باز <sup>۱۰</sup> است.
[۲۹]	پیش بینی لینک نظارت شده و مبتنی بر همسایه	مدل‌سازی زمینه‌ای	مکان	هتل	در این سیستم یک شبکه دو بخشی تعاملات مسافر و هتل ساخته می‌شود که از روش پیش‌بینی لینک مبتنی بر همسایگی برای پیش بینی ارتباطاتی که بر روی شبکه وجود ندارند استفاده می‌شود.
[۳۰]	تنوری گراف، متدهای تحلیل شبکه، فیلترکننده ترکیبی	مدل‌سازی زمینه‌ای	جنسیت، زبان و کشور گردشگر، مکان، امتیاز و نوع هر مکان توریستی	جذابیت‌های گردشگری	این سیستم توصیه‌گر ترکیبی شخصی‌سازی شده مبتنی بر داده‌های اجتماعی، با استفاده از میزان شباهت، اعتماد و روابط اجتماعی میان گردشگران، جذابیت‌های گردشگری موردعلاقه گردشگران را به آن‌ها پیشنهاد می‌دهد.
[۳۱]	فیلترکننده مشارکتی، تجزیه ماتریس، کاهش گرایان تصادفی	پس پالایش زمینه‌ای	صرفه اقتصادی، هتل، مکان هتل، نظرات و احساسات مهمان-های پیشین	هتل	اطلاعات انبوه سپاری هتل‌ها شامل اطلاعات رسمی هتل، امتیازات چند معیاری و نظرات متنی مسافران، استخراج می‌شود تا با استفاده از فیلترکننده مشارکتی برای هتل‌ها و مسافران پروفایل ایجاد شود. سپس براساس روش تجزیه ماتریس کاهش گرایان تصادفی پیشنهادات اولیه پیش بینی شده و بعد از اعمال اطلاعات زمینه‌ای بر روی آن‌ها، پیشنهادات نهایی ارائه می‌شود.

<sup>9</sup> Linear regression<sup>10</sup> Open data

سیستم توصیه‌گر زمینه آگاه سفر با بهره‌گیری از اطلاعات عکس‌های برچسب‌دار جغرافیایی

[۳۲]	خوشه‌بندی P-DBSCAN , تئوری بیز	پس پالایش و مدل-ساز زمینه‌ای	مکان، جنسیت کاربر	گردشگری	از معیار تحرک مبتنی بر آن‌رویی برای دسته‌بندی عکس‌هایی که برچسب جغرافیایی دارند استفاده شده و با تشخیص چهره در عکس‌هایی که مربوط به تورهای گردشگری هستند، تعیین جنسیت انجام می‌شود. سپس پروفایلی از کاربران و مکان-های گردشگری مبتنی بر جنسیت ایجاد شده و با بررسی جنسیت کاربران و شباهت آن‌ها، مکان‌های گردشگری پیشنهاد داده می‌شود.
[۳۳]	فیلترکننده مشارکتی	پس پالایش و مدل-ساز زمینه‌ای	شرایط آب و هوایی، زمان، مکان	مکان گردشگری	با بهره‌گیری از روش فیلترکننده مشارکتی، از تاریخچه سفرهای گردشگران و اطلاعات زمینه‌ای مانند شرایط آب و هوایی، زمان و مکان آن‌ها استفاده می‌شود تا مکان‌هایی در شهرهای جدید به آن‌ها پیشنهاد داده شود که مشابه با مکان‌های قبلی مورد علاقه گردشگران باشد.
[۳۴]	رتبه‌بندی بر اساس فیلترکننده مشارکتی	پیش پالایش و مدل‌سازی زمینه‌ای	شرایط آب و هوایی، زمان، مکان	مکان گردشگری	ارایه مکان‌های گردشگری پیشنهادی مبتنی بر معنا، با بهره‌گیری از اطلاعات زمینه‌ای صورت می‌گیرد. اساس این روش بر الگوریتم فیلترکننده مشارکتی و رتبه‌بندی اطلاعات زمینه‌ای است.
[۳۵]	مدل کردن موضوعی پویا، تجزیه ماتریس	مدل‌سازی و پس پالایش زمینه‌ای	داده شخصی گردشگران، زمان، مکان، اطلاعات ورود و خروج، اطلاعات هواشناسی	مکان گردشگری	مدل‌سازی موضوعی پویا برای بدست آوردن توزیع مبتنی بر زمان و اطلاعات ضمنی کاربر به کار گرفته شده و علاوه بر آن حجم زیادی از داده‌های صریح بر اساس اطلاعات ورود و خروج، محتویات بصری و دسته‌بندی‌های مکان‌های مورد علاقه گردشگران استخراج می‌شوند. سپس این اطلاعات برای تشخیص مشابهت‌های کاربر-کاربر و مکان-مکان به کار گرفته می‌شود و تجزیه ماتریس بر اساس آن‌ها صورت می‌گیرد.
[۳۶]	ترکیب پالایش مشارکتی و مبتنی بر محتوی	پس پالایش زمینه‌ای	اطلاعات اینترنت اشیا و شبکه‌های اجتماعی، اطلاعات زمینه‌ای کاربر	مکان رویداد	این سیستم توصیه‌گر مبتنی بر مکان و گرایش آگاه که به صورت ترکیبی طراحی شده است، برای پیشنهاد رویدادها علاوه بر اطلاعات زمینه‌ای کاربران، از اطلاعات زمینه‌ای دستگاه‌های هوشمند مربوط به کاربران با استفاده از اینترنت اشیا نیز بهره می‌برد. همچنین از اطلاعات مربوط به اجتماعات مجازی در شبکه‌های اجتماعی برای زمانی که علاقه مندی‌های کاربر بر اساس داده‌های اینترنت اشیا قابل دستیابی نیست، استفاده می‌کند.
[۳۷]	تجزیه ماتریس احتمالاتی، تحلیل احساسات	پیش پالایش و مدل‌سازی زمینه‌ای	داده زمینه‌ای احساسات، اطلاعات جغرافیایی مکان	نقطه مورد علاقه	خصوصیات احساسی کاربران در مکان‌های مختلف مشخص می‌شود و برای کاوش مکان مورد علاقه باید اطلاعات زمینه‌ای مکانی و احساسی را با هم ترکیب کند. سپس با مشارکت فاکتورهای فاصله‌های جغرافیایی و شباهت احساسی، مکان مورد علاقه کاربر را پیشنهاد می‌دهد.
[۳۸]	TF-IDF، دسته‌بندی متون، FastText	مدل‌سازی زمینه‌ای	برچسب جغرافیایی عکس‌ها، اطلاعات جغرافیایی مکان	پیشنهاد مکان	این سیستم توصیه‌گر مبتنی بر مدل، برای هر مکان یک پروفایل ایجاد می‌کند تا بر اساس آن‌ها و علاقه‌مندی‌های کاربران، مکان‌هایی را به آن‌ها پیشنهاد دهد. در این روش ابتدا با استفاده از TF-IDF، برچسب‌های مکان‌های استخراج شده از توضیحات عکس‌ها در اینستاگرام، رتبه‌بندی می‌شوند و سپس با روش FastText برچسب‌های توضیحات عکس‌ها دسته‌بندی می‌شوند.
[۳۹]	تجزیه ماتریس، K-means	مدل‌سازی زمینه‌ای، پیش پالایش	داده‌های زمینه‌ای مکانی مربوط به مسیر و مقصد سفر	مکان سفر، مسیر سفر	ابتدا یک روش داده محور برای کاوش سفرها در شبکه‌های اجتماعی مبتنی بر مکان ارایه می‌شود که نحوه سفر کردن کاربران را بررسی می‌کند و الگوهای حرکتی آن‌ها را به دست می‌آورد. از اطلاعات به دست آمده برای خوشه بندی مسافران و همچنین خوشه بندی فاصله‌های مقصدهای سفرها استفاده می‌کند.
[۴۰]	تجزیه ماتریس، تخصیص دریکله پنهان (LDA)	مدل‌سازی زمینه‌ای، پیش پالایش	برچسب جغرافیایی عکس‌ها، اطلاعات متنی، ترتیبی و تصویری	مکان سفر	با استفاده از عکس‌هایی دارای برچسب جغرافیایی، یک سیستم توصیه‌گر سفر شخصی‌سازی شده ارایه شده که مبتنی بر تجزیه ماتریس محدود شده برای اطلاعات چندگانه وزن دار است. با استفاده از اطلاعات چندگانه ابتدا برای کاربران و مکان‌های سفر پروفایل ایجاد شده و برای پیشنهادات مکان سفر، به داده‌های متنی، تصویری و فاصله‌ای، وزن‌های مختلفی بر پایه شباهت آن‌ها اختصاص داده می‌شود و از این وزن‌ها برای تجزیه ماتریس کاربر-مکان استفاده می‌شود.



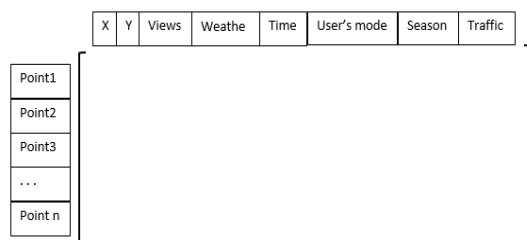


شکل ۱: فرآیند اجرای روش پیشنهادی

شامل طول و عرض جغرافیایی و داده‌های زمینه‌ای زمانی شامل فصل و زمان بازدید هستند. ویژگی‌های تعداد بازدید از مکان موردنظر، وضعیت آب و هوا، میزان ترافیک و وضعیت روحی کاربر در جهت افزایش دقت روش پیشنهادی و ارائه پیشنهادات دقیق‌تر مورد استفاده قرار گرفته است. شکل ۲ ساختار (سطرها و ستون‌های) ماتریس مربوط به داده‌های زمینه‌ای را نشان می‌دهد.

جدول ۲: ویژگی‌های نقطه بازدید

نام ویژگی	شرح
X (Lat)	مختصات جغرافیایی X نقطه بازدید
Y (Lon)	مختصات جغرافیایی Y نقطه بازدید
Views	تعداد بازدید از مکان موردنظر (محبوبیت مکان)
Current Weather	وضعیت هوای در زمان گرفتن عکس
Time Of The Day	زمان بازدید در طول روز
User's Mode	وضعیت روحی کاربر در زمان بازدید
Season	فصل زمان بازدید
Traffic	میزان ترافیک در زمان ترافیک



شکل ۲: ساختار ماتریس ویژگی مبتنی بر اطلاعات زمینه‌ای

#### ۴-۲- تعیین تعداد خوشه‌ها با استفاده از الگوریتم رقابت استعماری

برای خوشه‌بندی مکان‌های مختلف، ابتدا با استفاده از الگوریتم رقابت استعماری، بر اساس ویژگی‌های زمینه‌ای مختلف مربوط

#### ۴-روش پیشنهادی

سیستم توصیه‌گر سفر زمینه آگاهی که در این مقاله ارائه شده است، مبتنی بر الگوریتم خوشه‌بندی فازی و الگوریتم بهینه‌سازی رقابت استعماری است. شکل ۱ نشان دهنده فرآیند اجرای روش پیشنهادی است که هر یک از فازهای فرآیند روش پیشنهادی به تفکیک در این بخش توضیح داده شده است.

#### ۴-۱- استخراج اطلاعات زمینه‌ای و ایجاد ماتریس ویژگی

روش پیشنهادی بر اساس مجموعه داده فلیکر<sup>۱۱</sup>، یکی از بزرگ‌ترین سایت‌های اشتراک‌گذاری تصویر و ویدیو برخط<sup>۱۲</sup> و قابل‌دسترسی برای عموم، ارزیابی شده است. مقاله [۴۰] بر تاثیر اطلاعات و داده‌های جغرافیایی بر روی بهبود کارایی سیستم‌های توصیه‌گر تاکید کرده است. دلیل این نکته نزدیک بودن اغلب مکان‌های توریستی به یکدیگر است، به طور مثال ۷۶ درصد مکان‌های توریستی شهر پاریس کمتر از ۵ کیلومتر از هم فاصله دارند. اما باید این نکته را در نظر بگیریم که فقط ویژگی‌های جغرافیایی و فاصله‌های مکانی حائز اهمیت نیستند. به عنوان مثال اگر به کاربری که در مرکز شهر است و نیاز به استراحت دارد، فقط بر اساس طول و عرض جغرافیایی، مکانی نزدیک پیشنهاد داده شود، ممکن است باعث عدم رضایت کاربر شده و مکان پیشنهادی را انتخاب نکند. بنابراین علاوه بر برجسب‌های جغرافیایی باید ویژگی‌های دیگری نیز به عنوان تکمیل کننده اطلاعات و ارائه پیشنهادات دقیق‌تر در نظر گرفته شوند [۳۷].

برای استخراج اطلاعات زمینه‌ای موردنیاز برای روش پیشنهادی، برجسب‌های نشان داده شده در جدول ۲ به ازای هر مکان بازدید شده، از سایت استخراج شده‌اند. ویژگی‌های استخراج شده به ازای هر نقطه از بازدید عبارتند از: طول و عرض جغرافیایی مکان، تعداد بازدید (میزان محبوبیت)، وضعیت آب و هوا، موقعیت زمانی در روز، حالت روحی کاربر، میزان ترافیک، فصل و دما. بنابراین داده‌های زمینه‌ای مکانی

<sup>11</sup> www.flickr.com/services/api

<sup>12</sup> Online

به مکان‌ها، مکان‌های مشابه در یک خوشه قرار می‌گیرند. بنابراین بر اساس برچسب‌های مکان‌های بازدید و بدون نیاز به مشخص کردن تعداد خوشه‌ها از پیش، به‌طور خودکار مکان‌های بازدید خوشه‌بندی می‌شوند. همان‌طور که در بخش ۲-۳ شرح داده شده است، ابتدا به کشورها مقاداردهی اولیه می‌شود و استعمارگران و مستعمره‌های آن‌ها تعیین می‌شود. سپس حرکت تصادفی کشورهای مستعمره به سمت استعمارگران آن‌ها انجام شده و در صورتی که قدرت مستعمره‌ها بیشتر از استعمارگران نظیر آن‌ها باشد، مستعمره-ای که قدرت بیشتری دارد، به عنوان استعمارگر جدید شناخته شده و استعمارگر قدیم نیز مستعمره آن خواهد شد. سپس ضعیف‌ترین مستعمره از ضعیف‌ترین استعمارگر به طور تصادفی به یکی دیگر از استعمارگرها ملحق می‌شود و بررسی می‌شود که اگر استعمارگری بدون مستعمره وجود دارد حذف شود. این مراحل تا زمانی تکرار می‌شوند که شرط پایان اجرا برقرار شود. شبه کد الگوریتم رقابت استعماری مطابق الگوریتم ۱ است.

۱. یک معیار ارزیابی خوشه‌بندی دو هدف را دنبال می‌کند: ۱. مشخص کردن تعداد خوشه‌ها ۲. به دست آوردن بهترین حالت خوشه‌بندی با توجه به تعداد خوشه‌ها. هر معیار ارزیابی خوشه‌بندی باید دو وجه خوشه‌بندی زیر را مدنظر قرار دهد: ۱. پیوستگی یا فشردگی: الگوهای موجود در یک خوشه باید تا حد امکان به یکدیگر شبیه باشند. واریانس یا پراکندگی الگوهای موجود در یک خوشه نمایانگر پیوستگی یا فشردگی الگوهای درون یک خوشه هستند. ۲. تفکیک: خوشه‌ها باید تا حد امکان از هم فاصله داشته باشند. فاصله بین مراکز خوشه‌ها (به عنوان مثال فاصله اقلیدسی) می‌تواند نمایانگر تفکیک خوشه‌ها باشد [۴۲].

معیارهای ارزیابی مختلفی برای ارزیابی خوشه‌بندی غیر فازی وجود دارند که از جمله آن‌ها می‌توان به معیار PBM<sup>۱۳</sup>، DB<sup>۱۵</sup> [۴۳]، CS<sup>۱۶</sup> [۴۴] اشاره کرد. برای تمام این معیارها، مقدار بیشینه و یا کمینه آن‌ها نشان‌دهنده خوشه‌بندی بهینه مجموعه الگوها و یا داده‌هاست. از این‌رو می‌توان آن‌ها را در الگوریتم‌های بهینه‌سازی فرا ابتکاری استفاده نمود. پیچیدگی محاسباتی معیار DI با افزایش تعداد خوشه‌ها و اندازه داده‌ها به شدت افزایش می‌یابد و همچنین معیار PBM بیشتر در مواردی که تعداد خوشه‌ها کم و داده‌ها دارای پراکندگی پایین و تراکم بالا هستند، مورد استفاده قرار می‌گیرد. معیار DB علاوه بر اینکه در مجموعه داده‌های بزرگ و مجموعه داده‌های سلسله‌مراتبی، پیچیدگی محاسباتی قابل قبولی دارد، برای داده‌هایی با ابعاد متفاوت نیز می‌تواند نتایج خوبی ارائه دهد. معیار CS نیز برای خوشه‌هایی با چگالی‌های متفاوت به صورت موثر عمل می‌کند. بنابراین در این مقاله برای ارزیابی خوشه‌بندی از معیارهای DB و CS استفاده شده است. نتایج ارزیابی این معیارها در بخش ۵-۴ آمده است.

✓ **معیار ارزیابی DB:** این معیار تابعی از نسبت مجموع پراکندگی درون خوشه‌ای به پراکندگی بین خوشه‌هاست. ابتدا پراکندگی خوشه A و سپس

الگوریتم ۱. شبه کد الگوریتم رقابت استعماری

#### The pseudo-code of the Imperialist Competitive Algorithm

**Input:** Numpopulation, NumInitialImperialists

**Output:** Best Imperialists

1. initialize population randomly
2. **For** i=1 **to** Numpopulation
3.     Compute the evaluation cost  $c_i$
4.     Sort the computed cost  $c_i$  in descending order
5.     Select NumInitialImperialists out of Numpopulation
6.     Normalized the cost of each Imperialist  $c_n$
7.     Compute the normalized power of each Imperialist  $p_n$
8.     Assign Numcolonies remined countries to the Imperialists
9. **EndFor**
10. **For** j=1 **to** NumInitialImperialists
11.     Move the colony toward the relevant Imperialist (assimilation)
12.     Compute the cost of assimilated countries
13.     Perform revolution on new colony
14.     If the cost of new colony is less than cost of Imperialist
15.         Then exchange the position of colony and Imperialist
16.     Pick the weakest colony from the weakest empire and assign it to the empire that has most likelihood to possess it
17. **EndFor**
18. **If** there is Imperialist with no colonies
19.     **Then** eliminate the Imperialists
20. **Until** stopping condition is reached

در روش پیشنهادی این مقاله، الگوریتم رقابت استعماری با هدف به دست آوردن تعداد خوشه‌ها و مراکز آن‌ها، برای خوشه‌بندی مکان‌ها بر اساس ویژگی‌های زمینه‌ای آن‌ها به کار برده شده است. در [۴۱] یک الگوریتم رقابت استعماری بهبود یافته ارائه شده است که نسبت به الگوریتم رقابت استعماری

<sup>13</sup> Davies & Boulidin (DB)

<sup>16</sup> Chou, Su, Lai (CS)

<sup>13</sup> Pakhira, Bandyopadhyay & Maulik (PBM)

<sup>14</sup> Dunn index (DI)

اعمال می‌شود و تا جایی این فرایند تکرار می‌شود که تعداد و مراکز بهینه خوشه‌ها مشخص شود.

از آنجا که در ابتدا تعداد خوشه‌ها توسط کاربر مشخص نشده است، نحوه نمایش جواب‌ها باید طوری باشد که بتوان تعداد خوشه‌ها را همزمان با انجام خوشه‌بندی داده‌ها در زمان اجرا به دست آورد. بنابراین با فرض اینکه تعداد بیشینه خوشه‌ها برابر با  $K$  و تعداد ابعاد یا ویژگی‌های مجموعه داده برابر با  $d$  باشد، آنگاه هر یک از جواب‌های ممکن یک ماتریس با ابعاد  $(d+1) \times K$  خواهد بود که برای بدست آوردن تعداد بهینه خوشه‌ها، یک متغیر تصمیم به متغیرهای تصمیم موجود برای حل مسئله خوشه‌بندی اضافه می‌شود. به عبارتی هر یک از مراکز خوشه‌ها یک متغیر تصمیم به نام متغیر تصمیم فعال‌سازی دارد که مشخص می‌کند خوشه مربوطه فعال باشد و یا غیرفعال باشد. مقادیر مجاز برای متغیر تصمیم فعال‌سازی اعداد حقیقی در بازه  $(0, 1)$  است. مرکز یک خوشه در صورتی فعال می‌شود که مقدار متغیر تصمیم مربوط به آن بزرگ‌تر از ۰.۵ باشد. در غیر این صورت مرکز خوشه غیرفعال است.

برای ارزیابی جواب‌ها در الگوریتم رقابت استعماری، هر یک از معیارهای ارزیابی خوشه‌بندی می‌توانند مورد استفاده قرار گیرند. در این مقاله دو معیار ارزیابی خوشه‌بندی  $CS$  و  $DB$  به عنوان تابع هدف طبق رابطه (۱۳) در نظر گرفته شدند که هدف آن‌ها کمینه‌سازی مقدار معیار  $CS$  و  $DB$  تعریف شده است.

$$f_1 = CS(K), f_2 = DB(K) \quad (13)$$

#### ۳-۴- خوشه‌بندی فازی داده‌ها

در خوشه‌بندی  $C$ -Means فازی اگر مجموعه‌ی مکان‌های بازدید را  $L = [l_1, l_2, \dots, l_n]$  و مجموعه خوشه‌ها را برابر با  $C = [c_1, c_2, \dots, c_c]$  در نظر گرفته شود هر عضو مجموعه  $L$  می‌تواند با یک مقدار بین ۰ تا ۱ به عنوان درجه عضویت در بیش از یک خوشه قرار بگیرد. بنابراین، با توجه به این‌که هر مکان بازدید در بیش از یک خوشه قرار می‌گیرد، یکی از چالش‌های سیستم‌های توصیه‌گر که خلوت بودن داده-ها است تا حدود زیادی حل می‌شود و همسایگان بیشتری برای مکان‌های بازدید هر خوشه وجود خواهد داشت. شبه کد الگوریتم خوشه‌بندی فازی مطابق الگوریتم ۲ است:

فاصله بین خوشه  $i$  و خوشه  $j$  طبق روابط (۷) و (۸) محاسبه می‌شوند [۴۳]:

$$S_{i,q} = \left[ \frac{1}{x} \sum \| \vec{X} - \vec{m} \|_2^q \right]^{\frac{1}{q}} \quad (7)$$

$$D_{ij,t} = \left\{ \sum_{p=1}^d |m_{i,p} - m_{j,p}|^t \right\}^{\frac{1}{t}} \| \vec{m}_i - \vec{m}_j \|_t \quad (8)$$

که  $m_i$  مرکز خوشه  $i$ ام،  $t \geq 1$  و  $q, t$  می‌توانند به طور مستقل مقداردهی شوند.  $N_i$  تعداد الگوهای متعلق به خوشه  $i$  است.  $D_{ij,t}$  نرم  $t$ ام برای مراکز خوشه  $C_i (m_i)$  و خوشه  $C_j (m_j)$  را محاسبه می‌کند. مقدار معیار ارزیابی  $DB$  بر اساس رابطه (۹) و (۱۰) محاسبه می‌شود.

$$R_{i,qt} = \max_{j \in K, j \neq i} \left\{ \frac{S_{i,q} + S_{j,q}}{D_{ij,t}} \right\} \quad (9)$$

$$DB(K) = \frac{1}{K} \sum_{i=1}^K R_{i,qt} \quad (10)$$

در رابطه (۱۰)،  $K$  نشان‌دهنده خوشه  $K$ ام است. کمترین مقدار به دست آمده برای معیار  $DB$  نشان‌دهنده خوشه‌بندی بهینه می‌باشد.

✓ معیار ارزیابی  $CS$ : قبل از محاسبه معیار ارزیابی

$CS$ ، مرکز هر خوشه به وسیله میانگین الگوهای آن خوشه طبق رابطه (۱۱) مشخص می‌شود [۴۴]:

$$\vec{m}_i = \frac{1}{N_i} \sum_{x_j \in C_i} \vec{x}_j \quad (11)$$

در رابطه (۱۱)،  $N_i$  تعداد الگوهای متعلق به خوشه‌ی  $C_i$  است. معیار فاصله بین دو نمونه  $\vec{X}_i$  و  $\vec{X}_j$  به صورت  $d(\vec{X}_i, \vec{X}_j)$  نشان داده می‌شود. معیار  $CS$  به صورت رابطه (۱۲) تعریف می‌شود:

$$CS(K) = \frac{\sum_{i=1}^K \left[ \frac{1}{N_i} \sum_{\vec{X}_i \in C_i} \max_{\vec{X}_q \in C_i} \{d(\vec{X}_i, \vec{X}_q)\} \right]}{\sum_{i=1}^K \left[ \min_{j \in K, j \neq i} d(\vec{m}_i, \vec{m}_j) \right]} \quad (12)$$

همانند معیار ارزیابی  $DB$ ، معیار  $CS$  برابر با نسبت فاصله درون خوشه‌ای به فاصله بین خوشه‌ای است و بنابراین باید صورت کسر کمینه و مخرج بیشینه و در کل مقدار  $CS$  کمینه شود.

در واقع در هر تکرار الگوریتم فاصله هر نمونه از مرکز خوشه مشخص می‌شود و نمونه‌ها به خوشه‌ای اختصاص داده می‌شوند که کمترین فاصله را تا مرکز خوشه داشته باشد. سپس الگوریتم رقابت استعماری بهبود یافته بر روی مراکز خوشه‌ها

این شرط به این معنا است که مجموع تعلق هر نمونه به  $C$  خوشه باید برابر ۱ باشد. با استفاده از این شرط و کمینه کردن تابع هدف خواهیم داشت:

$$v_i = \frac{\sum_{k=1}^n u_{ik}^m x_k}{\sum_{k=1}^n u_{ik}^m} \quad (16)$$

$$u_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{d_{ik}}{d_{jk}}\right)^{2/(m-1)}} \quad (17)$$

#### ۴-۴- پیشنهاد نقاط بازدید به کاربر

پس از مشخص شدن مراکز و تعداد خوشه ها توسط الگوریتم رقابت استعماری، مراکز به عنوان مراکز اولیه الگوریتم C-Means فازی در نظر گرفته می شود. سپس الگوریتم C-Means فازی اجرا می شود و در نهایت مکان های خوشه بندی شده به دست می آید. سپس برای هر کاربر خوشه هایی که متعلق به مکان های آن کاربر هستند مشخص می شود. مثلاً ممکن است مکان هایی که یک کاربر از آن ها دیدن کرده است متعلق به خوشه های ۱ و ۲ باشد. خوشه ای که بیشترین مکان بازدید شده توسط کاربر را داراست به عنوان خوشه مورد علاقه کاربر در نظر گرفته می شود. سپس برای پیش بینی مکان های بعدی به کاربر مورد نظر، از مکان های متعلق به آن خوشه استفاده می شود. برای این کار مکان های متعلق به خوشه بر اساس نزدیکترین مکان به مرکز خوشه مرتب می شوند. سپس مکان هایی که توسط کاربر بازدید شده اند و به آن خوشه تعلق دارند حذف می شوند و در نهایت از بقیه مکان ها بر اساس تعداد مکان مورد نیاز، مکان هایی به کاربر پیشنهاد می شود.

#### ۵-۱- ارزیابی

در این بخش، مجموعه داده استفاده شده برای ارزیابی روش پیشنهادی معرفی می شود، سپس نحوه پیش پردازش داده ها و مقاردهی به پارامترها، نتایج حاصل از ارزیابی ها و در نهایت مقایسه نتایج حاصل با دیگر روش های مشابه ارائه می شوند. در نمودارها و جداول ارزیابی از برچسب TRS-GP<sup>۱۸</sup> برای روش پیشنهادی استفاده می شود.

#### ۵-۱- مجموعه داده

فلیکر یکی از بزرگترین سایت های اشتراک گذاری تصویر و ویدئو، خدمات وب و جوامع آنلاین است که توسط شرکت Ludicorp در سال ۲۰۰۴ تاسیس شده و در سال ۲۰۰۵

#### The pseudo-code of the Fuzzy C-Means (FCM)

**Input:** L and C.

**Output:** final Fuzzy C-Means clusters.

- 1: Select an initial fuzzy pseudo-partition, i.e., assign values to all  $u_{i,j}$
- 2: Repeat
- 3: Compute the centroid of each cluster using the fuzzy partition.
- 4: Update the fuzzy partition, i.e., the  $u_{i,j}$ .
- 5: Until the centroids don't change.

همان طور که در شبه کد الگوریتم ۲ مشاهده می شود، اگر  $L$  مجموعه مکان های بازدید و  $C$  مجموعه خوشه و  $i$  و  $j$  به ترتیب شاخص های هر یک از خوشه ها و مکان ها در نظر گرفته شود، ابتدا میزان تعلق هر مکان به خوشه ها که بر اساس  $u_{i,j}$  به دست می آید، محاسبه شده و سپس مرکز هر خوشه مشخص می شود. سپس مجدداً خوشه بندی مکان ها با توجه به مراکز جدید انجام می شود. این کار تا زمانی که تغییرات مراکز خوشه ها از یک حد آستانه مشخص کوچکتر شود، تکرار شده و در نهایت خوشه های نهایی به دست می آید.

در الگوریتم خوشه بندی فازی تعداد خوشه ها ( $C$ ) باید از قبل مشخص شده باشند تابع هدفی که برای این الگوریتم تعریف شده است به صورت زیر می باشد [۱۹]:

$$J = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m \|x_k - v_i\|^2 \quad (14)$$

در رابطه (۱۴)  $m$  یک عدد حقیقی بزرگتر از ۱ است که در اکثر موارد برای  $m$  عدد ۲ انتخاب می شود.  $X_k$  نمونه  $k$  ام است و  $V_i$  مرکز خوشه  $i$  ام و  $n$  تعداد مکان های بازدید است.  $U_{ik}$  میزان تعلق نمونه  $i$  ام به خوشه  $k$  ام را نشان می دهد. علامت  $\|*\|$  میزان تشابه (فاصله) نمونه با (از) مرکز خوشه می باشد که می توان از هر تابعی که بیانگر تشابه نمونه و مرکز خوشه باشد را استفاده کرد. از روی  $U_{ik}$  می توان یک ماتریس  $U$  تعریف کرد که دارای  $C$  سطر و  $n$  ستون می باشد و مؤلفه های آن هر مقداری بین ۰ تا ۱ را می توانند اختیار کنند. اگر تمامی مؤلفه های ماتریس  $U$  به صورت ۰ و یا ۱ باشند الگوریتم مشابه C-Means کلاسیک خواهد بود. با اینکه مؤلفه های ماتریس  $U$  می توانند هر مقداری بین ۰ تا ۱ را اختیار کنند اما مجموع مؤلفه های هر یک از ستون ها باید برابر ۱ باشد و داریم:

$$\sum_{i=1}^c u_{ik} = 1, \forall k = 1, \dots, n \quad (15)$$

ویژگی موردنظر است. در واقع مقادیر تمامی ویژگی‌ها به منظور نرمال سازی به مقادیر صحیح تبدیل شده است. به عنوان مثال مقادیر وضعیت هوا (آفتابی، ابری و بارانی، برفی و سرد، و مرطوب و معتدل) به مقادیر ۱ تا ۴، زمان بازدید در طول روز (صبح، ظهر، عصر و شب) به مقادیر ۱ تا ۴، وضعیت روحی کاربر (نرمال و طبیعی، شاد، عصبی، ناراحت، هیجان زده، خسته و نگران) به مقادیر ۱ تا ۶، و به همین ترتیب فصول سال و میزان ترافیک به اعداد صحیح نگاشت شده است. پس از انجام پیش پردازش بر روی مجموعه داده اولیه، مجموعه داده جدید با ساختار موجود در شکل ۳ بدست می‌آید.

همان گونه که در بخش ۴-۲ به آن اشاره شد اجرای الگوریتم رقابت استعماری نیاز به تعیین یک سری پارامترهای اولیه دارد که باید مشخص شود نظیر پارامترهای  $\beta$ ،  $\gamma$ ، تعداد جمعیت (کشورها)، تعداد استعمارگران اولیه و تعداد تکرارها. این مقادیر در جدول ۳ نشان داده شده است.

جدول ۳: مقادیر پارامترهای الگوریتم رقابت استعماری

پارامتر	مقدار
تعداد کشورها	۱۰۰
تعداد استعمارگران اولیه	۸
$\beta$	۲
$\gamma$	۰.۷
$\xi$	۰.۰۵
تعداد تکرار	۲۰۰

توسط یاهو خریداری شد. این سایت به مرور به یک شبکه اجتماعی مبتنی بر عکس تبدیل شده است [۴۵]. در این مقاله برای ارزیابی روش پیشنهادی، از مجموعه داده‌ای استفاده شده است که یک خزنده<sup>۱۹</sup>، در بازه زمانی سال ۲۰۱۵ تا ۲۰۱۹ مجموعه‌ای از مکان‌های بازدید شده در کشور امریکا و عمدتاً در قاره اروپا را بازیابی کرده و برچسب‌های مختصات جغرافیایی مکان، تعداد بازدید (میزان محبوبیت)، وضعیت آب و هوا، موقعیت زمانی در روز، حالت روحی کاربر، میزان ترافیک، فصل و دما را برای بازدیدها جمع آوری کرده است. جدول ۲ برچسب‌های استخراج شده به ازای هر مکان بازدید شده، را نشان می‌دهد.

### ۵-۲- پیش پردازش داده‌ها و مقداردهی پارامترها

به منظور یکسان سازی ویژگی‌ها و عملیات پاک سازی داده‌ها از میان ۱۰۰۰۰ رکورد، ۲۵۰۰ رکورد که شامل تمام ویژگی‌ها بود، برای شبیه سازی بکار رفته است. در مجموعه داده مورد استفاده ۱۰۰ کاربر وجود دارد که برای هر کاربر ۲۰ مکان ملاقات شده مشخص شده است. از بین این ۲۰ مکان، ۱۵ مکان برای آموزش و ۵ مکان برای آزمایش مدل پیشنهادی در نظر گرفته شده است. به عبارت بهتر، با استفاده از ۱۵ مکان ملاقات شده، ۵ مکان برای هر کاربر پیش بینی می‌شود. ویژگی‌های مجموعه داده مورد استفاده از مقادیر عددی و غیر عددی تشکیل شده است. به منظور استفاده از مقادیر غیر عددی برخی ویژگی‌ها، نگاشتی به کار گرفته می‌شود تا ویژگی‌های غیر عددی با مقادیر عددی مقداردهی شوند که در رابطه (۱۸) مشخص شده است:

$$NormalizedValue = \frac{\max(field) - \min(field)}{number} \quad (18)$$

$field$  مقدار ویژگی مورد نظر،  $number$  تعداد بازه‌هایی است که برای نگاشت مقادیر غیر عددی در نظر گرفته شده و  $NormalizedValue$  مقدار صحیح و نرمال سازی شده

<sup>19</sup> Crawler

1	Lat	Lon	views	date	current weather	time of the day	user's mode	season code	traffic	temperature	time range season
2	51.02815	4.48078	899661	41622	1	2	1	1	14	-15.00	1 winter
3	38.20551	-85.7028	534764	41217	4	2	2	3	4	-9.00	9 autumn
4	39.95653	-75.1722	1117325	42043	1	1	2	1	11	-37.00	4 winter
5	45.53467	10.23746	1534894	42048	3	2	3	1	14	-23.00	5 winter
6	43.17257	2.345066	1784617	40678	2	3	1	4	5	-5.00	6 Spring
7	-1.75856	36.05921	913359	40566	2	1	3	1	14	-60.00	2 winter
8	38.47316	-111.354	94683	42823	1	4	6	4	4	28.00	4 Spring
9	56.10125	-4.56877	92240	42775	4	1	6	1	15	-16.00	5 winter
10	35.57502	-100.963	88318	42858	4	3	5	4	1	17.00	8 Spring
11	51.76378	6.826575	49690	42859	2	2	6	4	4	-1.00	2 Spring
12	58.67722	6.368225	47634	42812	2	1	4	4	4	17.00	9 Spring
13	51.21709	4.421192	47032	41718	3	3	2	4	3	-4.00	5 Spring
14	37.07028	-91.7361	46734	42750	4	2	1	1	13	-25.00	3 winter
15	59.9075	10.75216	59666	42859	2	4	3	4	3	0.00	0 Spring
16	42.41321	-8.76705	86236	42859	3	3	3	4	4	25.00	2 Spring
17	40.29688	-4.66627	99668	40096	4	3	2	3	1	6.00	7 autumn
18	43.08798	-2.54714	100025	42823	4	4	6	4	1	1.00	2 Spring
19	44.36205	4.782507	80172	42860	1	4	3	4	5	23.00	2 Spring
20	47.836	12.89598	96978	42834	2	4	3	4	5	27.00	7 Spring

شکل ۳: بخشی از داده‌های نگاشت شده مجموعه داده فلیکر

مورد علاقه کاربر نسبت به کل پیشنهادات به دست آمده است

$$[46]:$$

$$MAE = 1 - Accuracy$$

$$(21)$$

#### ۴-۵- پیکربندی سخت افزار/نرم افزار

مشخصات پیکربندی نرم افزاری و سخت افزاری برای پیاده‌سازی روش پیشنهادی در جدول ۴ ارائه شده است.

جدول ۴: مشخصات پیکربندی سخت افزاری و نرم افزاری

مشخصات	اجزا	
MATLAB 9.4	زبان برنامه نویسی	نرم افزار
Windows 7	ویندوز	
Intel core(TM) i7-7700HQ CPU- 2.80GHz, 4 Core(s), 8 Logical Processor(s)	CPU	سخت افزار
GetForce GTX 1050 Ti, NVIDIA compatible – 4GB RAM	GPU	
16 GB	RAM	

#### ۵-۵- نتایج ارزیابی TRS-GP

برای ارزیابی TRS-GP در ابتدا دقت فرایند خوشه‌بندی برای روش‌های C-Means فازی و K-Means بر اساس تعداد مختلف خوشه‌ها بررسی می‌شود. سپس معیارهای الگوریتم رقابت استعماری بر اساس فاکتورهای DB و CS بررسی می‌شود. نتایج به دست آمده برای ارزیابی دقت با در نظر گرفتن تعداد خوشه‌های مختلف برای الگوریتم‌های خوشه‌بندی K-Means در جدول ۵ آمده است.

#### ۳-۵- معیارهای ارزیابی

برای ارزیابی روش پیشنهادی از معیارهای دقت ۲۰ و متوسط مطلق دقت ۲۱ و میانگین قدر مطلق خطا استفاده می‌شود. معیار دقت بر اساس نسبت تعداد پیش‌بینی‌های صحیح به تعداد کل پیش‌بینی‌ها به دست می‌آید که طبق رابطه (۱۹) محاسبه می‌شود [۱].

$$precision = \frac{\text{number of items that are relevant}}{\text{number of all recommended items}} \quad (19)$$

معیار متوسط مطلق دقت نیز بر اساس رابطه ۲۰ محاسبه می‌شود [۱]:

$$MAP = \frac{1}{Nq} \sum_{i=1}^{Nq} AP_i \quad (20)$$

در این رابطه  $Nq$  تعداد پرس‌وجوی مورد آزمایش و  $AP_i$  متوسط دقت برای پرس‌وجوی  $i$  ام می‌باشد.  $AP_i$  بر اساس رابطه  $\frac{1}{r}$  محاسبه می‌شود که  $r$  نشان‌دهنده رتبه آیتم مورد نظر در لیست مرتب شده آیتم‌ها است. در واقع معیار دقت و متوسط مطلق دقت نشان‌دهنده کیفیت پیشنهاداتی است که ارائه شده هستند و هر چه مقدار آن‌ها بیشتر باشد، نشان دهنده توانایی بیشتر سیستم توصیه‌گر برای ارائه پیشنهاداتی درست است.

یکی دیگر از روش‌های ارزیابی نتایج سیستم‌های پیشنهاددهنده، بررسی میزان خطای پیشنهادات ارائه شده به کاربران است. منابع مختلف، روابط متفاوتی را برای محاسبه میزان خطا ارائه داده‌اند که یکی از ساده‌ترین و پایه‌ای‌ترین رابطه‌ها که در [۴۶] مطرح شده است به جای در نظر گرفتن مقادیر عددی رتبه‌های پیش‌بینی‌شده و رتبه‌های واقعی، مقادیر آن‌ها را به صورت باینری در نظر گرفته و هر پیشنهاد را در دو حالت مفید و غیر مفید مورد بررسی قرار می‌دهد. در این مقاله، برای مکان‌های پیشنهاد داده شده، دو حالت بازدیدشده یا بازدیدنشده در نظر گرفته شده است. معیار میانگین قدر مطلق خطا<sup>۲۲</sup> بر اساس رابطه (۲۱) محاسبه می‌شود که در آن، مقدار Accuracy بر اساس تعداد پیشنهادات

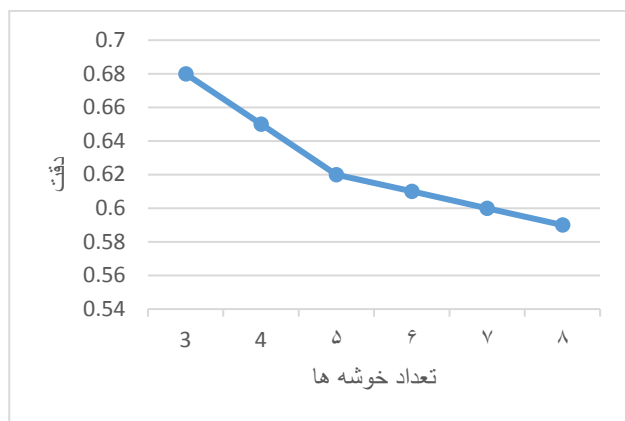
<sup>22</sup> Mean Absolute Error

<sup>20</sup> Precision

<sup>21</sup> Mean Absolute Precision (MAP)

جدول ۵: مقادیر دقت با در نظر گرفتن تعداد خوشه‌های مختلف برای الگوریتم K-Means

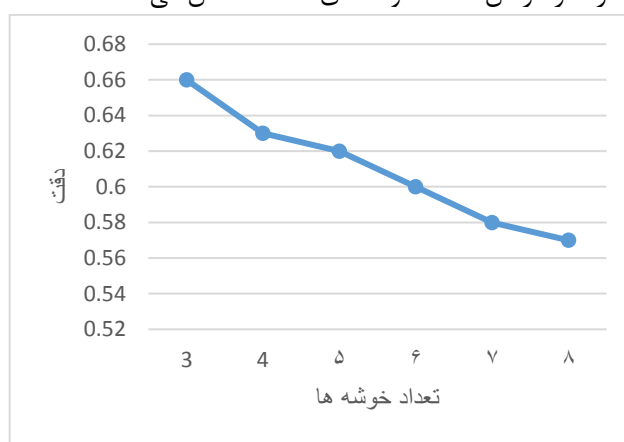
تعداد خوشه	دقت
۳	۰.۶۶
۴	۰.۶۳
۵	۰.۶۲
۶	۰.۶
۷	۰.۵۸
۸	۰.۵۷



شکل ۵: مقادیر معیار دقت الگوریتم خوشه بندی C-Means فازی را با در نظر گرفتن تعداد خوشه‌های مختلف

برای الگوریتم K-Means همانند الگوریتم C-Means فازی بالاترین میزان دقت با تعداد ۳ خوشه به دست می‌آید. در مورد معیارهای موثر در الگوریتم رقابت استعماری نتایج ارزیابی نشان می‌دهند که معیار DB نسبت به معیار CS مناسب‌تر بوده و برای تعیین تعداد و مراکز خوشه‌ها بهتر عمل می‌کند. معیار CS نسبت به DB پیچیدگی محاسباتی نسبتاً بالاتری دارد و نتایج ارزیابی‌ها در [۴۷] نیز نشان می‌دهد که معیار DB نسبت به دیگر معیارهای ارزیابی خوشه‌بندی عملکرد بهینه‌تری را ارائه می‌دهد. شکل ۶ تغییرات این دو معیار را در تکرارهای مختلف نشان می‌دهد.

شکل ۴: مقادیر معیار دقت الگوریتم خوشه بندی K-Means را با در نظر گرفتن تعداد خوشه‌های مختلف نشان می‌دهد.

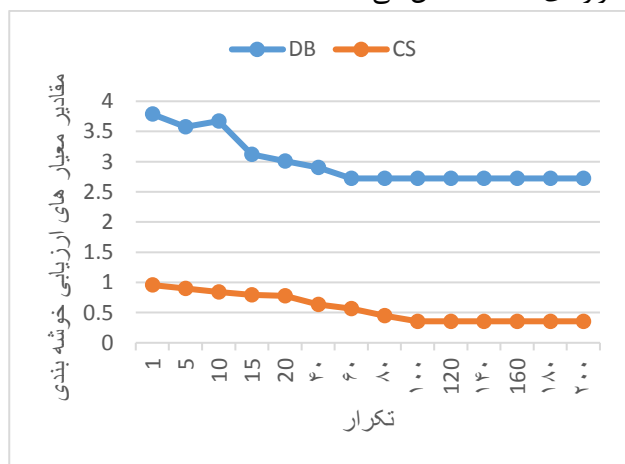


شکل ۴: مقادیر معیار دقت الگوریتم خوشه بندی K-Means را با در نظر گرفتن تعداد خوشه‌های مختلف

معیار دقت با در نظر گرفتن تعداد خوشه‌های مختلف برای الگوریتم خوشه‌بندی C-Means فازی در جدول ۶ و نمودار مقادیر آن در شکل ۵ نشان داده شده است.

جدول ۶: مقادیر دقت با در نظر گرفتن تعداد خوشه‌های مختلف برای الگوریتم FCM

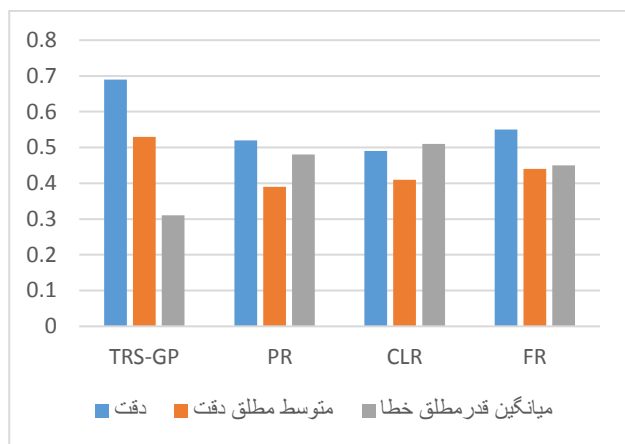
تعداد خوشه	دقت
۳	۰.۶۸
۴	۰.۶۵
۵	۰.۶۲
۶	۰.۶۱
۷	۰.۶
۸	۰.۵۹



شکل ۶: مقایسه معیارهای ارزیابی خوشه‌بندی DB و CS

#### ۵-۵- مقایسه TRS-GP با روش‌های موجود

در این بخش، TRS-GP را با برخی از روش‌های موجود در سیستم توصیه‌گر زمینه آگاه سفر مقایسه می‌کنیم تا جایگاه روش پیشنهادی در پژوهش‌های موجود در این حوزه مشخص شود.



شکل ۷: میزان دقت، متوسط مطلق دقت و میانگین قدرمطلق خطا TRS-GP در مقایسه با روش‌های پایه‌ای

با هدف بررسی تاثیرگذاری روش پیشنهادی، نتایج آن نسبت به چهار روش‌های پایه که در این قسمت توضیح داده شده‌اند، مقایسه شده است.

- **Popularity Rank (PR):** در این روش مکان‌ها بر اساس میزان محبوبیت عمومی رتبه بندی می‌شوند. میزان محبوبیت هر مکان بر اساس تعداد بازدیدهای غیرتکراری از مکان موردنظر است [۴۸].
- **Classic Rank (CLR):** این روش، همه کاربرها را یکسان در نظر نمی‌گیرد و بر اساس استنتاج مبتنی بر HITS<sup>۲۳</sup>، جذابیت‌های توریستی مکان‌ها و تجربه‌های سفر کاربران را برحسب امتیاز اعتباری برای آن‌ها محاسبه می‌کند و همچنین امتیاز محوری را از داده‌های مربوط به ارتباطات کاربران و مکان‌ها استخراج می‌کند [۴۹].
- **Frequent Rank (FR):** در این روش برای تعیین مکان‌های پرتکرار و الگوهای ترتیبی که فراوانی‌های آن‌ها کوچک‌تر از مقدار حدآستانه پشتیبان کمینه نباشد، از روش PrefixSpan استفاده می‌شود [۵۰]. نتایج نشان می‌دهد که عملکرد TRS-GP نسبت به دیگر روش‌های پایه بهتر است. دلیل آن، استفاده از داده‌های زمینه‌ای مبتنی بر زمان، مکان، شرایط آب و هوا، دما، حالت روحی کاربر، میزان ترافیک و همچنین تحلیل و خوشه‌بندی دقیق داده‌ها است که با بررسی تاریخچه مربوط به بازدیدهای گذشته کاربر، پیشنهادات دقیقی را به کاربر ارائه می‌دهد.

برای مقایسه روش پیشنهادی با سایر پژوهش‌های موجود در این زمینه از دو رویکرد زیر استفاده می‌شود تا نتایج به دست آمده در یک چهارچوب استاندارد صحت‌گذاری مناسبی گردد.

در ابتدا، TRS-GP با یکسری روش‌های پایه‌ای استاندارد که در بسیاری از مقالات برای سنجش دقت از آن‌ها استفاده می‌شود مقایسه می‌گردد، نظیر Popularity Rank (PR) [48]، Classic Rank (CLR) [49]، Frequent Rank (FR) [50]. لازم به ذکر است که تمامی این روش‌های پایه‌ای با مجموعه داده ارائه شده در این مقاله پیاده‌سازی و ارزیابی شده‌اند. سپس، تعدادی از پژوهش‌های موجود در زمینه سیستم‌های توصیه‌گر سفر مبتنی بر عکس‌های برچسب‌دار جغرافیایی استخراج شده از فلیکر (۲۴، ۳۲، ۳۳، ۳۴) برای مقایسه با روش پیشنهادی ارزیابی می‌شود. معیارهای ارزیابی در تمام مقالات پارامترهای مرتبط با کیفیت توصیه سیستم نظیر دقت، متوسط مطلق دقت و میانگین قدرمطلق خطا است.

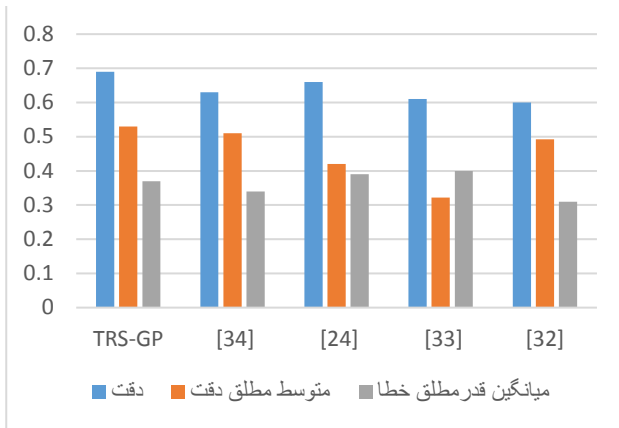
جدول ۷ مقادیر دقت، متوسط مطلق دقت و میانگین قدرمطلق خطا اندازه‌گیری شده TRS-GP را در مقایسه با روش‌های پایه‌ای دیگر نشان می‌دهد. برای ارزیابی تمام روش‌ها، از مجموعه داده استفاده شده در این پژوهش تعداد ۲۵۰۰ رکورد از دیتاست فلیکر برای ۱۰۰ کاربر استخراج شده است. شکل ۷ نیز این ارزیابی را نشان می‌دهد.

جدول ۷: میزان دقت، متوسط مطلق دقت و میانگین قدرمطلق خطا TRS-GP در مقایسه با روش‌های پایه‌ای

روش	دقت	متوسط مطلق دقت	میانگین قدرمطلق خطا
TRS-GP	۰.۶۹	۰.۵۳	۰.۳۱
PR	۰.۵۲	۰.۳۹	۰.۴۸
CLR	۰.۴۹	۰.۴۱	۰.۵۱
FR	۰.۵۵	۰.۴۴	۰.۴۵

<sup>23</sup> Hypertext Induced Topic Search (HITS)





شکل ۸: میزان دقت، متوسط مطلق دقت و میانگین قدرمطلق خطا TRS-GP در مقایسه با مقالات دیگر

### ۶- نتیجه گیری

این مقاله بر اساس تاریخچه سفرها و بازدیدهای کاربر از مکان‌های مختلف، بر اساس داده‌های زمینه‌ای، مکان‌های جدیدی را برای سفر به کاربر توصیه می‌کند. روش پیشنهادی، از تصاویر با برجسب‌های جغرافیایی در رسانه‌های اجتماعی، برای تولید توصیه‌های مکانی مطابق با علاقه‌مندی‌های کاربران و موقعیت‌های بازدید، استفاده می‌کند و در این راستا قادر به درک موقعیت، زمان، تاریخ، آب و هوا، دما و حالات روحی کاربر نیز می‌باشد. این روش، مبتنی بر ترکیب دو الگوریتم رقابت استعماری و خوشه‌بندی C-Means فازی است که از آن‌ها، برای شناسایی خوشه‌های مکان‌های توریستی استفاده می‌شود و سوابق سفرها از تصاویری با برجسب‌های جغرافیایی استخراج می‌شوند. روش پیشنهادی به بعضی از روش‌های توصیه مکان سفر با استفاده از مجموعه تصاویر فلیکر که در دسترس عموم قرار دارد ارزیابی و مقایسه شده است. نتایج ارزیابی بیان‌گر آن است که این روش قادر به فراهم کردن توصیه‌های مکانی مطابق با علاقه‌مندی‌های کاربر و موقعیت فعلی بازدید او می‌باشد و با در نظر گرفتن شباهت‌های میان کاربران (مدل تشابه کاربر-کاربر) بر اساس سفرهایشان برای توصیه‌های سفر مناسب است. بررسی نتایج ارزیابی‌های روش پیشنهادی نشان می‌دهند که بهره‌گیری از توان نفوذ اجتماعی و استفاده از داده‌های زمینه‌ای برای آیت‌های حساس به مکان موثر است. در واقع، روش پیشنهادی با ترکیب دو الگوریتم رقابت استعماری و خوشه‌بندی C-Means فازی موجب افزایش دقت پیشنهادها شده و استفاده از سرویس‌های شخصی‌سازی شده و همچنین داده‌های زمینه‌ای بیشتر (مکان، تعداد بازدید، وضعیت آب و هوا، موقعیت زمانی در روز، حالت روحی کاربر، میزان ترافیک، فصل و دما)

در جدول ۸ TRS-GP با مقالات مشابه بر اساس پارامترهای دقت، متوسط مطلق دقت و میانگین قدرمطلق خطا مقایسه شده است. همانطور که قبلاً نیز ذکر شد تمام روش‌ها از مجموعه داده‌های فلیکر که در مکان‌های مختلف جمع‌آوری شده‌اند استفاده کرده‌اند. شکل ۸ نیز این ارزیابی را نشان می‌دهد.

جدول ۸: میزان دقت، متوسط مطلق دقت و میانگین قدرمطلق خطا TRS-GP در مقایسه با مقالات دیگر

مقاله	دقت	متوسط مطلق دقت	میانگین قدرمطلق خطا
[۳۴]	۰.۶۳	۰.۵۱	۰.۳۷
[۲۴]	۰.۶۶	۰.۴۲	۰.۳۴
[۳۳]	۰.۶۱	۰.۳۲	۰.۳۹
[۳۲]	۰.۶۰	۰.۴۹	۰.۴۰
TRS-GP	۰.۶۹	۰.۵۳	۰.۳۱

همانطور که در جدول ۸ و شکل ۸ مشخص شده است میزان دقت و متوسط مطلق دقت TRS-GP نسبت به دیگر روش‌ها بالاتر و میانگین قدرمطلق خطای آن نسبت به روش‌های دیگر پایین‌تر است. استفاده از چندین زمینه مختلف و ترکیب دو الگوریتم خوشه‌بندی فازی و بهینه‌سازی رقابت استعماری قبل از ارایه توصیه نهایی در TRS-GP در مقایسه با [۲۴] که از تعداد زمینه‌های کمتری استفاده می‌کند و تنها مبتنی بر الگوریتم خوشه‌بندی K-Means است، دقت را افزایش و خطا را کاهش داده است. همچنین در مقایسه با روش مقالات [۳۴] و [۳۳] که بر اساس اولویت‌های سفر کاربران از سفرهای قبلی آن‌ها و همچنین رویکرد فیلترکننده مشارکتی پیشنهاد سفر می‌دهند TRS-GP از دقت بالاتری برخوردار است زیرا TRS-GP اطلاعات کامل‌تر و همه‌جانبه‌تری از یک بازدید را مورد پردازش قرار داده است و بر اساس آن‌ها خوشه‌بندی و ارایه پیشنهادات دقیق‌تری را انجام می‌دهد. در روش [۳۲] مکان‌ها بر اساس نوع جنسیت گردشگران دسته‌بندی می‌شوند و اطلاعات زمینه‌ای کمتری نسبت به TRS-GP به کار گرفته شده است. آنچه موجب تمایز اصلی میان TRS-GP در این مقاله و روش‌های ارایه شده در مقالات دیگر شده است افزودن تعداد زمینه‌های بیشتری (وضعیت آب و هوا، وضعیت روحی کاربر، میزان ترافیک و ...) است. در حالی که در بسیاری از مقالات مشابه فقط از زمینه‌های زمانی و مکانی استفاده شده است.

۶. صابری، نفیسه. منتظر، غلامعلی. (۱۳۸۹). «شخصی سازی محیط یادگیری الکترونیکی به کمک توصیه گر فازی مبتنی بر تلفیق سبک یادگیری و سبک شناختی» نشریه فناوری اطلاعات و ارتباطات ایران ۲، شماره ۳ (۱۳۸۹): ۹۱-۱۰۹.

7. K. Choi, D. Yoo, G. Kim and Y. Suh, "A hybrid online-product recommendation system: Combining implicit Rating based collaborative filtering and sequential pattern analysis," *Electronic Commerce Research and Applications*, Vol.11, pp. 309-317, 2012.

8. G. Adomavicius and A. Tuzhilin, "Context-Aware Recommender Systems," In: *Recommender Systems Handbook*, F. Ricci, L. Rokach, B. Shapira, Ed. Springer, Boston, MA, Springer US, pp. 217-253. 2011.

9. K. Verbert, N. Manouselis, X. Ochoa, M. Wolpers, H. Drachsler, I. Bosnic and E. Duval, "Context-aware recommender systems for learning: a survey and future challenges," *IEEE Transactions on Learning Technologies*, Vol. 5, no. 4, pp. 318-335, 2012.

10. F. Ricci, "Mobile recommender systems," *Information Technology & Tourism journal*, vol. 12, no. 3, pp.205-231., 2010.

11. R. Beale and P. Lonsdale, "Mobile context aware systems: The intelligence to support tasks and effectively utilise resources," In: *Mobile Human-Computer Interaction-Mobile HCI 2004*. S. Brewster, M. Dunlop, Ed. Springer Berlin Heidelberg. pp. 240-251. 2004.

12. D. Weib, M. Duchon, F. Fuchs and C. Linnhoff-Popien, "Context-aware personalization for mobile multimedia services," *Proceedings of the 6th International Conference on Advances in Mobile Computing and Multimedia*. ACM, 2008.

13. N. Manouselis, H. Drachsler, R. Vuorikari, H. Hummel and R. Koper, "Recommender systems in technology enhanced learning," *Recommender systems handbook*. Springer US. pp. 387-415. 2011.

14. A. K. Dey, A. Gregory and D. Salber, "A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications," *International Journal of Human-Computer Studies*, Vol. 16, no. 2, pp. 97-166. 2001.

15. P. Brusilovsky and E. Millán, "User models for adaptive hypermedia and adaptive

باعث بهبود عملکرد سیستم توصیه‌گر و ارائه پیشنهادهاى دقیق‌تر با خطای کمتر می‌شود.

برای بهبود کارایی سیستم توصیه‌گر مبتنی بر زمینه می‌توان پژوهش‌های زیر را برای کارهای آینده بیشتر مورد بررسی قرار داد:

✓ ارائه راهکارهایی برای حفظ محرمانگی و امنیت داده‌ها.

✓ توسعه واسط‌های کاربر باکیفیت بالا و پیچیدگی کم که تعامل بین کاربران نهایی و سیستم توصیه‌گر را ساده کنند.

✓ بررسی برخی محدودیت‌های زمانی- مکانی برای نتایج توصیه‌ای (مثلا چه مدت زمان برای بازدید نیاز است و کاربر چقدر زمان دارد؟)

✓ شخصی‌سازی در ترکیب اطلاعات به دست آمده از موقعیت در سفرها (پیگیری مکان‌ها) برای توصیه‌های توریستی

✓ ارائه یک مجموعه داده منسجم و کاربردی

✓ استخراج موقعیت جغرافیایی (طول و عرض جغرافیایی) و همچنین امکان افزودن برجسب به صورت دستی با استفاده از سیستم‌های موبایل

✓ وزن‌دهی به اطلاعات زمینه‌ای مختلف

#### مراجع

1. J. Bobadilla, F. Ortega, A. Hernando and A. Gutiérrez, "Recommender systems survey," *Know.-Based Syst.*, vol. 46, pp.109-132, 2013.

۲. بهشتی نژاد، راحله. سمیع، محمد ابراهیم. حمزه، علی. (۱۳۹۸). «بهبود سیستم‌های توصیه‌گر با کمک وب معنایی». نشریه فناوری اطلاعات و ارتباطات ایران ۹، شماره ۳۱ (۱۳۹۸): ۴۵-۵۶.

3. M. Slehat, "Evaluation of potential tourism resources for developing different forms of tourism : case study of Iraq Al-Amir and its surrounding areas – Jordan," PhD thesis, Catholic University of Eichstätt-Ingolstadt. 2018.

۴. منتظر، غلامعلی. فتاحی، وحید. (۱۳۹۴). «تخصیص بهینه درس پار به کمک الگوریتم بهینه سازی گروه ذرات». نشریه فناوری اطلاعات و ارتباطات ایران ۶، شماره ۲۱ (۱۳۹۴): ۱۵-۲۶.

5. N. Henze, P. Dolog and N. Wolfgang, "Reasoning and Ontologies for Personalized Elearning in the Semantic Web," *Educational Technology & Society*, Vol. 7, 82-97.

26. Alrasheed, H., Alzeer, A., Alhowimel, A. and Althyabi, A., 2020. A Multi-Level Tourism Destination Recommender System. *Procedia Computer Science*, 170, pp.333-340.
27. H. Khallouki, A. Abatal and M. Bahaj, "An ontology-based context awareness for smart tourism recommendation system," In: *Proceedings of the Inter-national Conference on Learning and Optimization Algorithms: Theory and Applications, LOPAL 2018, Rabat, Morocco, May 2-5, 2018*, pp.43:1-43:5.
28. E. Pantano, C-V. Priporas, N. Stylos and C. Dennis, "Facilitating tourists' decision making through open data analyses: A novel recommender system," *Tourism Management Perspectives*, Vol. 31. pp. 323-331, 2019.
29. B. Kaya, "A hotel recommendation system based on customer location: a link prediction approach," *Multimedia Tools and Applications*, Vol. 79, pp. 1745-1758, 2020.
30. L. Esmaili, S. Mardani, A. Hashemi Golpayegani and Z. Zanganeh Madar. "A Novel Tourism Recommender System in the Context of Social Commerce," *Expert Systems With Applications*, Vol.149, 113301, July 2020.
31. BM. Veloso, F. Leal, B. Malheiro and JC. Burguillo, "On-line guest profiling and hotel recommendation," *Electronic Commerce Research and Applications*, Vol. 34, 100832, 2019.
32. Z. Xu, L. Chen, H. Guo, M. Lv and G. Chen, "User similarity-based gender-aware travel location recommendation by mining geotagged photos," *International Journal of Embedded Systems*, Vol. 10, no. 5, 356, 2018.
33. A. Majid, L. Chen, G. Chen, H. Mirza, I. Hussain and J. Woodward, "A context-aware personalized travel recommendation system based on Geotagged social media data mining," *International Journal of Geographical Information Science*, Vol. 27, no. 4, pp. 662-684, 2013.
34. I. Memon, L. Chen, A. Majid, M. Lv, I. Hussain and G. Chen, "Travel recommendation using geo-tagged photos in social media for educational systems," *The adaptive web*. Springer-Verlag, 2007.
16. K. Xu, Clustering. In: Dubitzky W., Wolkenhauer O., Cho KH., Yokota H. (eds) *Encyclopedia of Systems Biology*. 2013, Springer, New York, NY.
17. S. Chen, Z. Xu and Y. Tang, "A Hybrid Clustering Algorithm Based on Fuzzy c-Means and Improved Particle Swarm Optimization," *Arabian Journal for Science and Engineering*, Vol. 39, pp.8875-8887, 2014.
18. KP. Soman, S. Diwakar and V. Ajay, "Data mining: theory and practice," PHI Learning Pvt. Ltd.; 2006.
19. J. Nayak, B. Naik and HS. Behera, "Fuzzy C-means (FCM) clustering algorithm: a decade review from 2000 to 2014," In *Computational intelligence in data mining-Vol. 2*, pp. 133-149. 2015, Springer, New Delhi.
20. A. Stetco, X. Zeng and J. Keane, "Fuzzy C-means++: Fuzzy C-means with effective seeding initialization," *Expert Systems with Applications*, Vol. 42, no. 21, pp. 7541-7548, 2015.
21. A. Gargari and E. Lucas, "Imperialist competitive algorithm: an algorithm for optimization inspired by imperialistic competition," *IEEE Congress on Evolutionary Computation*, pp. 4661-4667, 2007.
22. N.M. Villegas, C. Sanchez, J. Dõaz-Cely and G. Tamura, "Characterizing Context -Aware Recommender Systems: A Systematic Literature Review," *Knowledge-Based Systems*. Vol. 140, pp. 173-200, 2017.
23. G. Chen and L. Chen, "Augmenting service recommender systems by incorporating contextual opinions from user reviews," *User Modeling and User-Adapted Interaction*, Vol. 25, no. 3, pp. 295-329, 2015.
24. Z. Xu, C. Ling and C. Gencai, "Topic based context-aware travel recommendation method exploiting geotagged photos," *Neurocomputing*, Vol. 155, pp. 99-107, 2015.
25. Shafaqat, W. and Byun, Y.C., 2020. A Recommendation Mechanism For Under-Emphasized Tourist Spots Using Topic Modeling And Sentiment Analysis. *Sustainability*, 12(1), p.320.

44. H. Chou, M. C. Su, and E. Lai, "A new cluster validity measure and its application to image compression," *Pattern Analysis and Applications* vol. 7, no. 2, pp. 205–220, Jul. 2004.
45. "Flickr 10K Dataset" <https://www.kaggle.com>, [accessed: Sep 2020]
46. Del Olmo, F.H. and Gaudioso, E., 2008. Evaluation of recommender systems: A new approach. *Expert Systems with Applications*, 35(3), pp.790-804.
47. E. Rendón, I. M. Abundez, C. Gutierrez, S. D. Zagal, A. Arizmendi, E. M. Quiroz and H. E. Arzate, "A comparison of internal and external cluster validation indexes," In *Proceedings of the 2011 American conference on applied mathematics and the 5th WSEAS international conference on Computer engineering and applications (AMERICAN-MATH'11/CEA'11)*. World Scientific and Engineering Academy and Society (WSEAS), Stevens Point, Wisconsin, USA, 158–163, 2011.
48. A. Majid, L. Chen, G. Chen, H. T. Mirza and I. Hussain, "GoThere: Travel suggestions using geotagged photos," In *WWW 2012 companion*, April 16–20, 2012. Lyon, France: ACM.
49. Y. Zheng, L. Zhang, X. Xie and W.Y. Ma, "Mining interesting locations and travel sequences from GPS trajectories," in: *Proceedings of the 18th International Conference on World Wide Web, WWW '09*, ACM, New York, NY, USA, 2009, pp. 791–800.
50. Z. Yin, L. Cao, J. Han, J. Luo and T.S. Huang, "Diversified trajectory pattern ranking in geo-tagged social media," in: *Proceedings of SIAM International Conference on Data Mining*, 2011, pp. 980–991.
- tourist," *Wireless Personal Communications*. vol. 80, no. 4, pp. 1347-1362, 2015.
35. Z. Xu, L. Chen, Y. Dai and G. Chen. "A Dynamic Topic Model and Matrix Factorization based Travel Recommendation Method Exploiting Ubiquitous Data," *IEEE Transactions on Multimedia*. Vol. 19, no. 8, pp. 1933-1945, 2017.
36. S. Ojagh, M. Malek, S. Saeedi, and S. Liag, "A location-based orientation-aware recommender system using IoT smart devices and Social Networks," *Future Generation Computer Systems*, Vol. 108, pp. 97-118, 2020.
37. G. Zhao, P. Lou, X. Qian and X. Hou, "Personalized location recommendation by fusing sentimental and spatial context," *Knowledge-Based Systems*, Vol. 196, 2020.
38. M. Memarzadeh and A. Kamandi, "Model-Based Location Recommender System Using Geotagged Photos On Instagram," *2020 6th International Conference on Web Research (ICWR)*, Tehran, Iran, 2020, pp. 203-208.
39. L.W. Dietz, A. Sen, R. Roy and W. Worndl, "Mining trips from location-based social networks for clustering travelers and destinations," *Information Technology & Tourism*, Vol. 22, pp. 131–166, 2020.
40. D. Lyu, L. Chen, Z. Xu and S. Yu, "Weighted multi-information constrained matrix factorization for personalized travel location recommendation based on geo-tagged photos," *Applied Intelligence*, Vol. 50, pp. 924–938, 2020.
41. A. Chaghari and M. Feizi-Derakhshi, "Automatic Clustering Using Improved Imperialist Competitive Algorithm," *JSDP*. 2017; 14 (2) :159-169.
۴۲. امیری، مریم. ختن لو، حسن. (۱۳۹۲). «خوشه بندی اسناد، مبتنی بر آنتولوژی و رویکرد فازی» *نشریه فناوری اطلاعات و ارتباطات ایران* ۵، شماره ۱۷ (1392): ۷۳-۹۶.
43. D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence.*, Vol. 1, no. 2, pp. 224–227, 1979