

توسعه الگوریتم توصیه گر مکان احتمالاتی بر اساس دسته‌بندی مکان در شبکه‌های اجتماعی مکان‌مبنا مبتنی بر تأثیر زمانی، تأثیر مکانی و تأثیر روابط اجتماعی

صدیقه گودرزی^۱، علی اصغر آل شیخ^{۲*}، سپهر هنرپرور^۳

^۱ دانشجوی کارشناسی ارشد سیستم‌های اطلاعات مکانی - دانشکده مهندسی نقشه‌برداری - دانشگاه صنعتی

خواجه‌نصیرالدین طوسی

sgudarzi@email.kntu.ac.ir

^۲ استاد دانشکده مهندسی نقشه‌برداری - دانشگاه صنعتی خواجه‌نصیرالدین طوسی

alesheikh@kntu.ac.ir

^۳ دانشجوی دکتری سیستم اطلاعات مکانی - دانشکده مهندسی نقشه‌برداری - دانشگاه صنعتی خواجه‌نصیرالدین طوسی

sep.honarparvar@mail.kntu.ac.ir

(تاریخ دریافت اردیبهشت ۱۳۹۶، تاریخ تصویب شهریور ۱۳۹۶)

چکیده

یکی از سرویس‌هایی که در بین کاربران تلفن‌های همراه هوشمند با استقبال زیادی روبه‌رو شده است، سرویس‌های مبتنی بر شبکه‌های اجتماعی است. توصیه مکان یکی از سرویس‌های محبوب برای شبکه‌های اجتماعی مکان‌مبناست. این سرویس بر اساس پیشینه رفتار کاربر و اطلاعات مکان مانند دسته‌بندی مکان، مکان‌های بازدید نشده را به کاربران پیشنهاد می‌دهد. روش‌های موجود که از اطلاعات ثبت موقعیت کاربر و اطلاعات مکان، استفاده می‌کنند، تنها تأثیر مکانی و تأثیر زمانی را در توصیه مکان در نظر گرفته‌اند. با توجه به اینکه روابط اجتماعی کاربر می‌تواند بر عملکرد بهتر توصیه مکان تأثیر بگذارد، این تأثیر نیز می‌تواند نقش مهمی در توصیه مکان ایفا کند. در این پژوهش، الگوریتم توصیه مکان PCLRTGS با در نظر گرفتن تأثیر مکانی، تأثیر زمانی و تأثیر روابط اجتماعی توسعه داده شده است. در این الگوریتم، تأثیر مکانی با استفاده از تابع توزیع مکانی مدل‌سازی شده است. همچنین برای مدل‌سازی تأثیر زمانی، منحنی زمانی هر کاربر در هر دسته ایجاد شده و شباهت این منحنی با منحنی‌های زمانی کاربران دیگر در آن دسته به‌دست‌آمده و این شباهت برای مدل‌سازی تأثیر زمانی بکار گرفته شده است. تأثیر روابط اجتماعی نیز با استفاده از سه معیار شباهت ارتباطات اجتماعی کاربر و دوستانش، شباهت فعالیت‌های ثبت موقعیت کاربر و دوستانش و شباهت زمانی رفتار ثبت موقعیت بین کاربر و دوستانش، مدل‌سازی شده است. الگوریتم توسعه داده شده با استفاده از دو معیار دقت و جامعیت ارزیابی شده و با دو الگوریتم PCLR و SPCLR در پنج گروه با تعداد توصیه‌های مختلف مقایسه شده است. نتایج نشان می‌دهد که الگوریتم پیشنهادی PCLRTGS از نظر دقت و جامعیت، حدود ۱۵-۱۰ درصد عملکرد بهتری نسبت به دو الگوریتم PCLR و SPCLR دارد.

واژگان کلیدی: شبکه‌های اجتماعی مکان‌مبنا، توصیه مکان، منحنی زمانی، تأثیر روابط اجتماعی

* نویسنده رابط

۱- مقدمه

رشد روزافزون کاربران شبکه‌های اجتماعی مکان مبنا و وجود سرویس‌های مختلف در این شبکه‌ها باعث شده است که تأثیر قابل توجهی بر زندگی افراد داشته باشد. شبکه‌های اجتماعی مکان مبنا این امکان را به کاربران می‌دهند که بتوانند تجربیات مکانی خود را با دیگران به اشتراک بگذارند، همچنین امکان ذخیره پیشینه مکانی آن‌ها در این شبکه‌ها وجود دارد. سرویس توصیه مکان در شبکه‌های اجتماعی مکان مبنا از جمله سرویس‌های بسیار محبوب این شبکه‌ها محسوب می‌شود؛ بنابراین توصیه مکان یک زمینه تحقیقاتی بسیار مهم هم در حوزه دانشگاه و هم در عرصه صنعت است. با توجه به تحقیقات انجام شده در زمینه توصیه مکان در شبکه‌های اجتماعی مکان مبنا، افراد در این شبکه‌ها، به احتمال زیاد فعالیت‌های یکسانی را در زمان‌های تقریباً یکسانی از روز انجام می‌دهند [۱]. این الگوی زمانی برای هر فرد نسبت به دیگری متفاوت است. از طرفی مکان‌ها می‌توانند به کلاس‌های مختلفی به عنوان دسته‌بندی مکان^۱ با توجه به ویژگی‌های مشترکشان تقسیم شوند [۱]. از آنجاکه دسته مکان فعالیت‌های انجام شده در آن مکان را منعکس می‌کند، پژوهشگران بر اساس آنالیز داده‌های ثبت موقعیت به این نتیجه رسیده‌اند که الگوهای زمانی جداگانه‌ای برای هر دسته مکان وجود دارد [۱]. پژوهش‌های متعددی در زمینه توصیه مکان در شبکه‌های اجتماعی مکان مبنا انجام شده است. برخی از این پژوهش‌ها بدون در نظر گرفتن اطلاعات دسته‌بندی مکان، بر تأثیرات زمانی، مکانی و یا تأثیر روابط اجتماعی بر رفتار ثبت موقعیت کاربر تمرکز داشته‌اند [۲، ۳]. در عین حال تعداد کمی از پژوهش‌ها نیز با توجه به این اطلاعات، تأثیرات زمانی، مکانی و اثرات تشابه الگوهای زمانی دو کاربر را در توصیه مکان در نظر گرفته‌اند [۱، ۴]. این در حالی است که هیچ‌یک از این پژوهش‌ها، تأثیر روابط اجتماعی را همزمان با تأثیرات زمانی و مکانی ثبت موقعیت‌ها به کار نگرفته‌اند. روابط اجتماعی بین کاربران بر رفتار ثبت موقعیت آن‌ها تأثیر می‌گذارد و حدود ۳۰٪-۱۰٪ از ثبت موقعیت‌ها از روابط اجتماعی تأثیر می‌پذیرد [۱].

در مدل‌سازی تأثیر مکانی، توصیه گر احتمال اینکه یک کاربر مکان‌هایی را بازدید کند را بر اساس فاصله آن مکان‌ها تا خانه، یا فاصله تا مکان‌های بازدید شده قبلی کاربر، پیدا می‌کند. در صورتی که تأثیر زمانی بر رفتار ثبت موقعیت افراد در نظر گرفته نشود ممکن است توصیه گر مکان‌هایی را با توجه به پیشینه ثبت موقعیت کاربر، به او پیشنهاد دهد که به این موضوع که در چه ساعتی از روز است، توجهی نداشته باشد. به عنوان مثال یک کاربر در ساعت ۸ صبح به کافی‌شاپ و در ساعت ۶ عصر به باشگاه می‌رود؛ بنابراین، اگر زمان توصیه در عصر باشد، بهتر است باشگاه به او پیشنهاد شود. از طرف دیگر باید شباهت زمانی بین کاربران در نظر گرفته شود زیرا ممکن است کاربری تنها تعداد کمی از دسته‌های مکان را بازدید کرده باشد که در پیشینه ثبت موقعیت او وجود دارد و در نتیجه مکان‌های بالقوه‌ای وجود داشته باشد که به آن کاربر پیشنهاد نشود. در نظر گرفتن روابط اجتماعی بین کاربران نیز می‌تواند به یافتن احتمال ثبت موقعیت کاربر کمک کند؛ زیرا دوستان رفتارهای مشابهی از نظر فعالیت‌های ثبت موقعیت دارند؛ به دلیل اینکه علایق مشترک زیادی دارند. برای مثال دو دوست ممکن است باهم به کافی‌شاپ رفته و قهوه بخورند، یا ممکن است یک کاربر به تماشای فیلمی برود که دوستش به او توصیه کرده است [۳].

در این پژوهش، رفتار دوره‌ای کاربر، با در نظر گرفتن شباهت زمانی دو کاربر، تأثیر مکانی و تأثیر اجتماعی، در توصیه مکان به کار گرفته شده است. لذا ابتدا توابع توزیع احتمال مکانی و زمانی برای مدل کردن رفتار ثبت موقعیت کاربران، ایجاد می‌شود. تابع توزیع احتمال زمانی، الگوهای متناوب ثبت موقعیت‌های کاربر را مدل‌سازی می‌کند. پس از آن شباهت زمانی بین کاربران نیز برای به دست آوردن تشابه رفتارهای ثبت موقعیت آن‌ها، به دست می‌آید. این شباهت با توجه به منحنی‌های زمانی هر کاربر و اتصال این منحنی‌ها به دست می‌آید. تابع توزیع مکانی نیز برای به دست آوردن احتمال ثبت موقعیت در یک مکان به عنوان تابعی از فاصله آن مکان تا خانه کاربر مورداستفاده قرار می‌گیرد. سپس مدل تأثیر اجتماعی به منظور مدل‌سازی تأثیری که دوستان بر رفتار ثبت موقعیت کاربر دارند، ارائه می‌شود و روشی برای ترکیب این سه تأثیر و به دست آوردن احتمال ثبت موقعیت کاربر بیان می‌گردد. نهایتاً ارزیابی الگوریتم با آزمایش روی مجموعه واقعی از داده‌های

^۱ Location Category

ثبت موقعیت یک شبکه اجتماعی مکان مبنا، برای ارزیابی الگوریتم PCLRTGS^۱ در مقایسه با دو الگوریتم PCLR^۲ و sPCLR^۳ انجام می شود.

این پژوهش در شش بخش تدوین شده است. در بخش دوم، مروری بر ادبیات موجود در الگوریتم های توصیه مکان برای شبکه های اجتماعی مکان مبنا ارائه شده است. در بخش سوم، مجموعه داده مورد استفاده در این پژوهش معرفی شده و سپس به آنالیز زمانی، مکانی و اجتماعی این داده ها پرداخته می شود. در بخش چهارم الگوریتم PCLRTGS ارائه می شود. ارزیابی نتایج در بخش پنجم ارائه می شود. در نهایت، در بخش ششم نتیجه گیری حاصل از این پژوهش بیان گردیده است.

۲- پیشینه تحقیق

توصیه مکان یک زمینه مطالعاتی بسیار مهم است که در حال حاضر پژوهش های زیادی در این زمینه صورت می گیرد. Beeharee و همکاران در سال ۲۰۰۷ و Simon و همکاران نیز در سال ۲۰۰۷ بر اساس مکان لحظه ای کاربر در سیستم راهنمای تور همراه، توصیه مکان را ارائه دادند [۵، ۶]. Park و همکاران در سال ۲۰۰۷ یک سیستم توصیه شخصی را ارائه دادند که ترجیحات کاربر را بر اساس مدل شبکه Bayesian مدل می کند [۷]. Zheng و همکاران در سال ۲۰۱۰ از پیشینه GPS^۴ و نظرات کاربران برای توصیه مکان و فعالیت ها به آن ها استفاده کردند [۸]. Zhou و همکاران در سال ۲۰۱۲ روشی به نام PLSA^۵ را همراه با روش های پالایش گروهی آیت مینا و کاربر مینا، ارائه دادند که در آن ابتدا یک مدل معنایی پنهان^۶ را آموزش داده و سپس از آن مدل برای یافتن احتمال ثبت موقعیت هر کاربر در هر مکان مشخص استفاده می کند [۹]. پژوهش های فوق، این واقعیت که رفتار ثبت موقعیت افراد، با توجه به فاصله مکان ثبت موقعیت کاربر تا خانه او، تحت تأثیر قرار می گیرد را در نظر نگرفته اند. Cheng و همکاران در سال ۲۰۱۲، یک

مدل گوسی چندمرکزی^۷ برای به دست آوردن تأثیر مکانی در توصیه مکان ارائه دادند [۱۰]. Zhao و همکاران در سال ۲۰۱۳، یک مدل ترکیبی گوسی مبتنی بر ژنتیک^۸ را برای به دست آوردن تأثیر مکانی ارائه دادند که این روش عملکرد بهتری نسبت به مدل گوسی چندمرکزی دارد [۱۱]. Zhang و همکاران در سال ۲۰۱۳، در پژوهش خود به این نتیجه دست یافتند که تأثیر مکانی به جای اینکه با یک توزیع مشترک مانند توزیع قانون توانی مدل سازی شود، باید به صورت شخصی برای هر کاربر مدل سازی شود. به همین دلیل آن ها از برآورد چگالی کرنل^۹ با استفاده از توزیع فاصله شخصی برای هر کاربر، استفاده کردند [۱۲]. روش های فوق نمی تواند منجر به یک توصیه مکان خوب شود، زیرا آن ها مکان ها را نیز مانند فیلم ها در نظر می گیرند. در توصیه فیلم اگر ترجیح فیلم کاربر مشخص باشد، بدون توجه به این مسئله که کاربر کجا زندگی می کند و در چه زمانی توصیه را درخواست می کند، سیستم می تواند بهترین فیلم را بر اساس ترجیحاتش در اختیار او قرار دهد. باین حال پژوهش های فوق از الگوهای زمانی^{۱۰} رفتار ثبت موقعیت افراد برای توصیه مکان استفاده نکرده اند؛ بنابراین در توصیه مکان معیار زمان و محدوده مکانی بسیار اهمیت دارد.

مطالعات اخیر ارتباطات اجتماعی بین کاربران را نیز برای توصیه مکان در شبکه های اجتماعی مکان مبنا در نظر گرفته اند. برخی از این پژوهش ها با استفاده از روش های پالایش گروهی حافظه مینا و مدل مینا انجام شده است. در این تحقیقات شباهت کاربران به روش پالایش گروهی کاربر مینا به دست آمده است [۱۰، ۱۲-۱۶]. یکی از این مطالعات که قیود مکانی و ارتباطات اجتماعی را نیز در نظر می گیرد روش ارائه شده توسط Ye و همکاران در سال ۲۰۱۱ است [۱۷]. آن ها یک چارچوب ترکیبی به نام USG^{۱۱} را برای توصیه مکان در شبکه های اجتماعی مکان مبنا ارائه دادند. این توصیه گر احتمال ثبت موقعیت کاربر در یک مکان را با استفاده از احتمالات به دست آمده از سه مدل به دست می آورد. این سه مدل عبارتند از: مدل پالایش گروهی کاربر مینا، مدل تأثیر اجتماعی و مدل تأثیر مکانی [۱۷].

^۷ Multi center Gaussian Model (MGM)

^۸ Genetic-based Gaussian Mixture Model

^۹ Kernel density estimation

^{۱۰} Temporal patterns

^{۱۱} user preference (U), social influence from friends (S), geographical influence from POIs (G)

^۱ Probabilistic Category-based Location Recommendation Utilizing Temporal Influence, Geographical Influence and Social Influence

^۲ Probabilistic Category-based Location Recommendation

^۳ Probabilistic Category-based Location Recommendation Utilizing Temporal Influence and Geographical Influence

^۴ Global Positioning System

^۵ Probabilistic Latent Semantic Analysis

^۶ Latent Semantic Model

مدل پالایش گروهی کاربر مبنا، ترجیح ضمنی^۱ یک کاربر برای یک مکان را با ترکیب رفتار کاربران مشابهش برآورد می‌کند. مدل تأثیر اجتماعی از یک روش پالایش گروهی دوست مبنا^۲ استفاده می‌کند؛ بنابراین در این مدل ترجیحات کاربر نسبت به یک مکان با استفاده از ترکیب ترجیحات دوستانش برآورد می‌شود. مدل تأثیر مکانی از توزیع قانون توانی^۳ برای مدل کردن احتمال ثبت موقعیت در فاصله مشخصی از مکان‌های مشاهده‌شده قبلی کاربر، استفاده می‌کند. درنهایت یک چهارچوب تلفیقی برای ترکیب این مدل‌ها ارائه می‌دهد [۱۷]. Ye و همکاران در سال ۲۰۱۱ همچنین بیان کرده‌اند که ویژگی‌های مکانی نقش مهمی در شبکه‌های اجتماعی مکان مبنا ایفا می‌کنند [۱۸]. با توجه به مطالب فوق، مشاهده می‌شود که پژوهشگر در مدل USG مدلی برای اطلاعات زمانی موجود در داده‌های ثبت موقعیت در نظر نگرفته است. ویژگی زمانی، می‌تواند با استفاده از یک مدل حرکتی دوره‌ای^۴ برای کاربران در نظر گرفته شود. Berjani و همکاران دو استراتژی استنتاج برای تفسیر داده‌های ثبت موقعیت به‌عنوان ترجیحات کاربر ارائه دادند [۱۹]. اولین استراتژی از یک تعریف ترجیح باینری ساده^۵ و استراتژی دوم از روش فواصل عرض برابر^۶ استفاده کرده است. یک مدل فاکتور پنهان برای پیش‌بینی رده‌بندی‌های نامشخص آموزش داده‌شده است [۱۹]. Zhou و همکاران در سال ۲۰۱۲ روشی ارائه دادند که در آن از فرکانس‌های درون داده‌های ثبت موقعیت استفاده‌شده است و میزان معنادار بودن بازدید کاربر نسبت به یک مکان و همچنین میزان اهمیت یک مکان برای یک کاربر را نشان می‌دهد [۹].

رفتارهای تغییر مکان دوره‌ای افراد و همچنین الگوهای زمانی موجود در داده‌های ثبت موقعیت اخیراً مورد بررسی قرار گرفته است [۲۰-۲۳]. برخی از این پژوهش‌ها تنها الگوی زمانی دوره‌ای در بازدید مکان را یافته و آن را نمایش داده‌اند اما از این الگو برای توصیه مکان استفاده نکرده‌اند [۲۴-۲۶]. Eagle و همکاران در سال ۲۰۰۹ رفتار یک شخص را با استفاده از جمع

وزن دار مجموعه‌ای از بردارهای ویژگی به نام رفتارهای ویژه مدل کردند [۲۰]. Li و همکاران در سال ۲۰۱۰ رفتار دوره‌ای یک شی متحرک را به‌صورت «فعالیت‌های تکراری در مکان‌های مشخص بافاصله‌های زمانی منظم» تعریف کردند و سپس حرکات دوره‌ای اشیا را استخراج کردند [۲۱]. Cho و همکاران در سال ۲۰۱۱ یک روش توصیه مکان را بر اساس حرکت دوره‌ای افراد ارائه دادند [۲]. آن‌ها همچنین دو روش PMM و PSMM^۷ را ارائه دادند که در این الگوریتم‌ها بر دو جنبه تعداد ثبت موقعیت‌ها و روابط اجتماعی تأکید داشته‌اند و اطلاعات دسته‌بندی مکان درون داده‌های ثبت موقعیت را در نظر نگرفته‌اند [۲]. Gao و همکاران در سال ۲۰۱۳، الگوی زمانی رفتار ثبت موقعیت کاربران را به واحدهای هفته، ماه و سال تقسیم کردند و برای محاسبه شباهت رفتاری بین کاربران از روش فاکتورگیری ماتریس و ترکیب مدل زمانی و مدل حرکت کاربر استفاده کردند [۲۷].

دسته هر مکان فعالیت‌های انجام‌شده در آن مکان را منعکس می‌کند [۴]. در حال حاضر پژوهش‌های کمی از این اطلاعات برای توصیه مکان استفاده کرده‌اند. Bao و همکاران در سال ۲۰۱۲، از اطلاعات دسته‌بندی مکان برای محاسبه شباهت کاربران به روش پالایش گروهی کاربر مبنا استفاده کرده‌اند [۲۸]. Liu و همکاران در سال ۲۰۱۳، مکان‌ها را بر اساس دسته آن‌ها گروه‌بندی کرده و سپس با استفاده از پیشینه ثبت موقعیت کاربران، ماتریس انتقال کاربر-دسته^۸ را به‌جای ماتریس ثبت موقعیت کاربر-مکان^۹ ایجاد کردند. درنهایت با استفاده از روش فاکتورگیری ماتریس^{۱۰} برای یافتن مکان‌هایی که کاربر تمایل بیشتری به بازدید از آن‌ها دارد، استفاده کردند [۲۹]. Rahimi و همکاران در سال ۲۰۱۳ دو الگوریتم توصیه گر به نام‌های PCR^{۱۱} و PCLR را با استفاده از اطلاعات دسته‌بندی مکان^{۱۲} درون داده‌های ثبت موقعیت ارائه دادند [۱]. در این الگوریتم‌ها رفتارهای ثبت موقعیت مکانی و زمانی کاربران با استفاده از توابع توزیع احتمال مکانی و زمانی مدل شده است. PCLR مدل دسته‌بندی

^۷ Periodic Social Mobility Model

^۸ User-category transition matrix

^۹ User-location checkins matrix

^{۱۰} Matrix Factorization(MF)

^{۱۱} Probabilistic Category Recommender

^{۱۲} Location Category Information

^۱ Implicitly preference

^۲ friend-based collaborative filtering

^۳ power law distribution

^۴ periodic model of human movements

^۵ simple binary preference

^۶ equal width intervals (EWI) obility Model

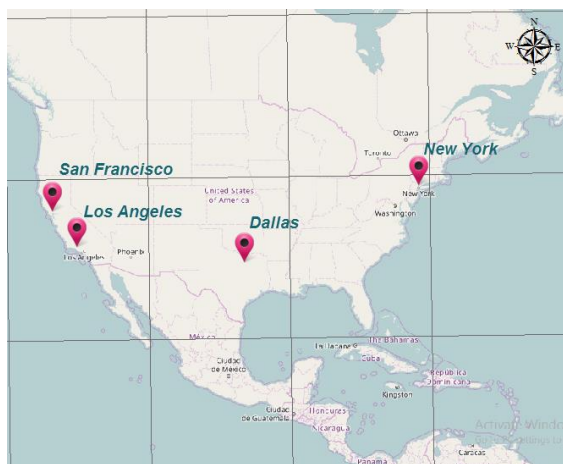
از ارزیابی الگوریتم با الگوریتم‌های PCLR و sPCLR مقایسه گردد و میزان بهبود توصیه را در الگوریتم پیشنهادی نسبت به این دو الگوریتم سنجیده شود.

۳- مواد و روش‌ها

در این بخش چهار شهر ایالات متحده آمریکا به عنوان منطقه مورد مطالعه معرفی گردیده و داده‌های مورد استفاده در این پژوهش توضیح داده شده است. همچنین به تشریح روش‌های مورد استفاده در سه بخش مکانی، زمانی و اجتماعی پرداخته شده است.

۳-۱- منطقه مورد مطالعه

منطقه مورد مطالعه در این پژوهش شامل چهار شهر پرجمعیت ایالات متحده آمریکا به نام‌های نیویورک، لس‌آنجلس، سانفرانسیسکو و دالاس است. این چهار شهر به دلیل تراکم مناسب داده‌ها، در این پژوهش انتخاب شدند. شکل ۱ موقعیت این شهرها را در ایالات متحده آمریکا نشان می‌دهد.



شکل ۱- مناطق مورد مطالعه

۳-۲- داده‌های مورد استفاده

در این پژوهش از داده‌های شبکه اجتماعی مکان‌مبنا Gowalla استفاده شده است. Gowalla یکی از بزرگ‌ترین شبکه‌های اجتماعی مکان‌مبناست که در سال ۲۰۰۷ راه‌اندازی شد و در سال ۲۰۱۲ بسته شد. مجموعه داده‌های استفاده شده در این پژوهش شامل داده‌های مربوط به ۱۰۰۰ کاربر و ۱۵۹۰۵ رکورد ثبت موقعیت مربوط به چهار شهر نیویورک، لس‌آنجلس، سانفرانسیسکو

زمانی^۱ استفاده شده در PCR را با یک مدل تأثیر مکانی^۲ که در تابع توزیع مکانی ایجاد شده ترکیب می‌کند. در این الگوریتم‌ها مدل زمانی تنها تناوب رفتار ثبت موقعیت یک کاربر را در نظر می‌گیرد؛ در صورتی که تشابه بین رفتارهای ثبت موقعیت کاربران را در نظر نمی‌گیرد [۱]. Ying و همکاران در سال ۲۰۱۴، ابتدا دسته مکان‌های مربوط به هر مکان را با توجه به برجسب‌های^۳ آن‌ها استخراج کرده و سپس ترجیحات شخصی هر کاربر را با استفاده از تعداد ثبت موقعیت‌ها در مکان‌ها با دسته یکسان، به دست آوردند [۱۵]. Zhao و همکاران در سال ۲۰۱۴، کاربران را به گروه‌های مختلف خوشه‌بندی کرده و هر خوشه را با یک بردار وزن‌دار نشان دادند. هر بردار وزن‌دار نشان‌دهنده یک بازدید از یک دسته مکان خاص و در یک دوره زمانی مشخص است. آن‌ها برای محاسبه شباهت کاربران بر اساس بردار وزن‌دار، از روش پالایش گروهی کاربر مبنا^۴ استفاده کردند [۳۰]. Hu و همکاران در سال ۲۰۱۴ از روش فاکتورگیری ماتریس برای ارتباط هر دسته با هر بردار پنهان^۵ استفاده کردند. سپس امتیاز هر کاربر به هر مکان را بر اساس بردارهای پنهان هر دسته مکان به دست آوردند [۳۱].

Zhou و همکاران در سال ۲۰۱۴ الگوریتم sPCLR را با استفاده از اطلاعات دسته‌بندی مکان درون داده‌های ثبت موقعیت ارائه دادند [۴]. در این الگوریتم از توابع توزیع احتمال مکانی و زمانی ارائه شده توسط Rahimi استفاده شده است [۱]. علاوه بر این، این الگوریتم شباهت بین رفتارهای ثبت موقعیت کاربران را نیز در نظر می‌گیرد. این شباهت با استفاده از مفهومی به نام منحنی‌های زمانی^۶ برای هر کاربر و اتصال منحنی^۷ به دست آمده است [۴]. با اینکه این الگوریتم عملکرد بهتری نسبت به الگوریتم‌های پیشین دارد اما روابط اجتماعی بین کاربران را در نظر نمی‌گیرد.

هدف در این پژوهش این است که روابط اجتماعی نیز در توصیه مکان با استفاده از اطلاعات دسته‌بندی مکان در داده‌های ثبت موقعیت در نظر گرفته شود و نتایج حاصل

^۱ temporal category model

^۲ geographical influence model

^۳ Tag

^۴ user-based collaborative filtering

^۵ Latent factor

^۶ Temporal curves

^۷ Curves Coupling

و دالاس است. یک رکورد ثبت موقعیت نشان‌دهنده بازدید یک کاربر از یک مکان در زمان مشخص است که شامل شناسه کاربر، شناسه مکان، تاریخ و زمان ثبت موقعیت و دسته مکان است. این داده‌ها در چهار بخش قرار می‌گیرد. بخش اول شامل رکوردهای ثبت موقعیت است که شامل شناسه ثبت موقعیت، شناسه کاربر، شناسه مکان و تاریخ و زمان است. بخش دوم مربوط به اطلاعات مکان است. این اطلاعات شامل شناسه مکان، تعداد ثبت موقعیت، تاریخ و زمان، طول و عرض جغرافیایی و دسته مکان است. بخش سوم از داده‌ها اطلاعات مربوط به کاربران است که شامل شناسه کاربر، تعداد دوستان و تعداد ثبت موقعیت‌هاست. بخش سوم داده‌ها مربوط به روابط اجتماعی کاربران است که شامل شناسه کاربر و شناسه دوستانش است.

۳-۳- روش‌های مورد استفاده در توسعه الگوریتم

روش‌هایی که برای پیاده‌سازی الگوریتم توسعه داده‌شده در این پژوهش، به کار گرفته شده است، در سه بخش به شرح زیر بررسی شده است.

۳-۳-۱- تأثیر مکانی

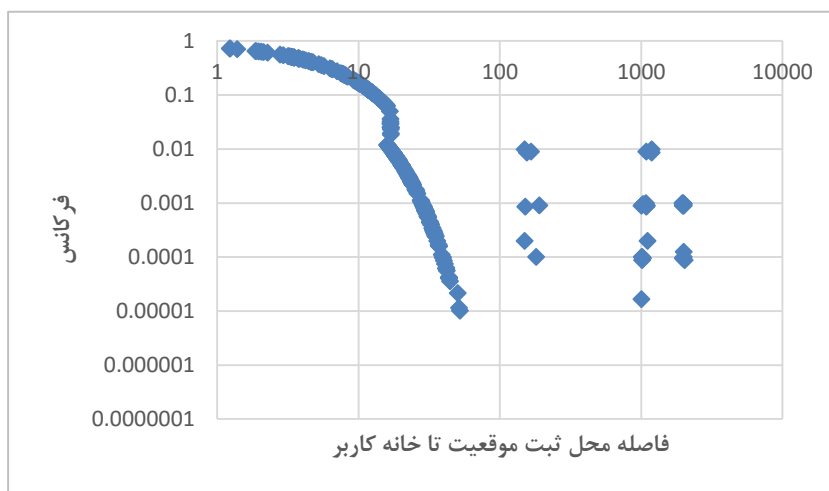
نزدیکی و دوری یک مکان به محل زندگی کاربر رفتار ثبت موقعیت او را تحت تأثیر قرار می‌دهد [۱۷]. به‌طوری‌که احتمال اینکه یک کاربر مکانی را در نزدیکی محل زندگی خود، انتخاب کند بالاتر از احتمال انتخاب مکانی است که در فاصله دورتری از محل زندگی‌اش قرار دارد. در این پژوهش با توجه به روش به کار گرفته‌شده توسط Rahimi و همکاران در سال ۲۰۱۳ [۱]، یک نمودار

لگاریتمی از فرکانس ثبت موقعیت کاربر بر اساس فاصله مکان از محل زندگی‌اش با استفاده از مجموعه داده‌های به کار گرفته‌شده در این پژوهش ایجاد شده است. با توجه به شکل ۲ نتیجه حاصل از این نمودار به این صورت است که زمانی که فاصله مکان تا محل زندگی کاربر کمتر از ۵۰ کیلومتر است بین فرکانس ثبت موقعیت با فاصله مکان از خانه کاربر ارتباطی وجود دارد؛ اما زمانی که این فاصله بیشتر از ۵۰ کیلومتر است، فرکانس ثبت موقعیت به‌صورت تصادفی تغییر می‌کند که می‌تواند نشان‌دهنده این موضوع باشد که کاربر در سفر بوده است. فاصله کمتر از ۵۰ کیلومتر را نیز می‌توان به دو بخش در مرز ۱۶ کیلومتر تقسیم کرد. بر این اساس، احتمال در بخش کمتر از ۱۶ کیلومتر به آهستگی کاهش می‌یابد. می‌توان فرض کرد که درون منطقه با شعاع ۱۶ کیلومتر، مکان خانه کاربر قرار دارد و خارج از این شعاع، برای کاربر منطقه دوری محسوب می‌شود؛ بنابراین تابع احتمال مکانی به‌صورت زیر تعریف می‌شود:

$$SP(l, h) = p(d = \text{dis} \tan ce(l, h)) =$$

$$\begin{cases} 0.0886e^{-0.166*d}, d \leq 16km \\ 0.3122e^{-0.204*d}, 16km \leq d \leq 50km \\ 0, d > 50km \end{cases} \quad (1)$$

در معادله ۱، l مکانی است که هدف، یافتن احتمال ثبت موقعیت برای آن است، h مکان خانه کاربر، d فاصله بین l و h است.



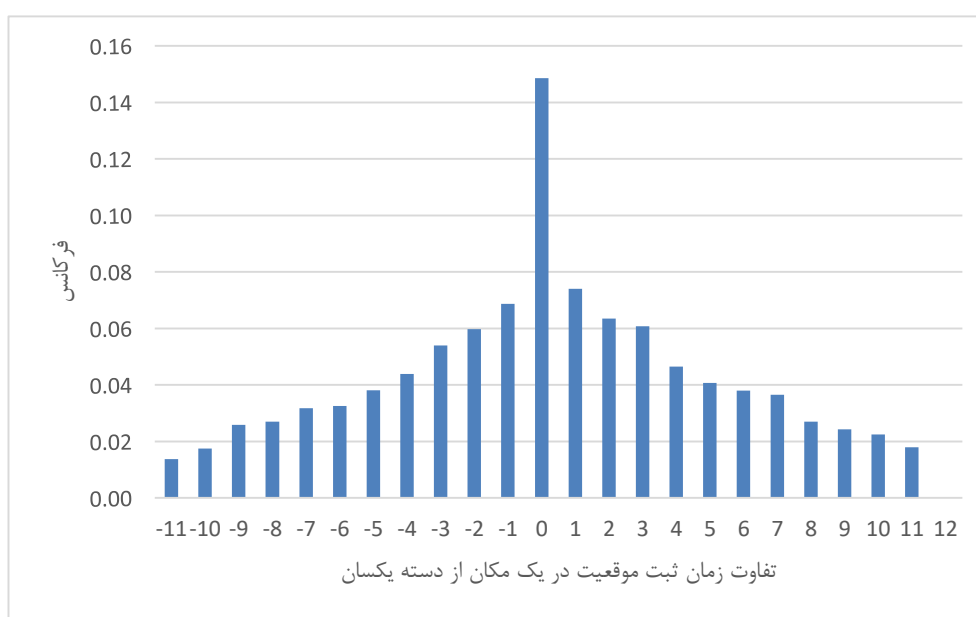
شکل ۲- نمودار لگاریتمی فرکانس ثبت موقعیت کاربر بر اساس فاصله مکان موردنظر تا محل خانه کاربر

۳-۳-۲- تأثیر زمانی

برخی کاربران تعداد کمی ثبت موقعیت در پیشینه ثبت موقعیت خود، دارند، بنابراین در زمان توصیه اگر تنها بر اساس پیشینه ثبت موقعیت، مکان‌هایی به آن‌ها توصیه شود ممکن است توصیه مناسبی نباشد. دلیل این امر این است که مکان‌هایی ممکن است وجود داشته باشد که کاربر به آن‌ها علاقه داشته باشد اما چون بر اساس پیشینه کاربر توصیه انجام شده است، آن مکان‌ها به او توصیه نشود؛ بنابراین به دست آوردن شباهت بین کاربران می‌تواند باعث ایجاد توصیه‌های مناسب‌تر شود. روش پالایش گروهی در سیستم توصیه گر، ترجیحات ضمنی کاربر را با استفاده از ترکیب رفتارهای کاربران مشابه به دست می‌آورد. شباهت رفتار ثبت موقعیت دوره‌ای کاربر، در دسته‌بندی‌های مختلف را با توجه به مراحل زیر، می‌توان به دست آورد [۴].

در مرحله اول، از روش تابع توزیع احتمال زمانی ارائه شده توسط Rahimi و همکاران استفاده شده است [۱]. این تابع احتمال ثبت موقعیت کاربر در دسته‌های مختلف و در زمان‌های مختلف از روز را به دست می‌آورد. برای ایجاد تابع توزیع احتمال زمانی، فرکانس‌های ثبت موقعیت هر کاربر در هر دسته بر اساس فاصله زمانی مشخص با پنجره زمانی یک ساعت، به دست آمده و به صورت یک نمودار لگاریتمی رسم شده است. نمونه‌ای از این نمودارها در شکل ۳ نشان داده شده است.

با توجه به اینکه مکان کاربر به طور مستقیم در داده‌های ثبت موقعیت وجود ندارد، بنابراین لازم است با توجه به داده‌های ثبت موقعیت برای هر کاربر به دست آید. روش‌های مختلفی برای یافتن محل زندگی کاربر وجود دارد. به عنوان مثال در نظر گرفتن میانگین ثبت موقعیت‌ها در زمانی از شب که کاربر در خانه حضور داشته باشد؛ اما با توجه به اینکه این روش استثنائات زیادی را دربرمی گیرد از جمله اینکه زمان اشتغال همه افراد در طول روز نبوده و بنابراین کاربر در طول شب در خانه ثبت موقعیت نمی‌کند؛ بنابراین این روش در مورد همه کاربران صادق نیست. از سوی دیگر با توجه به اینکه هدف نهایی این پژوهش پس از توسعه الگوریتم، ارزیابی و مقایسه آن با دو الگوریتم موجود است از مدل‌های مشابه در بخش مکانی سه الگوریتم استفاده شده است. بر این اساس برای به دست آوردن مکان خانه کاربر، فرض شده است که مرکز ثبت موقعیت‌های هر کاربر در محل خانه‌اش قرار می‌گیرد؛ بنابراین میانگین ثبت موقعیت‌های هر کاربر می‌تواند بیانگر مکان خانه او باشد، اما به دلیل اینکه ممکن است در مجموعه ثبت موقعیت‌های کاربر ثبت موقعیت‌های مربوط به سفرهای او نیز وجود داشته باشد پس این میانگین به‌ویژه زمانی که تعداد ثبت موقعیت‌های کاربر کم است تحت تأثیر قرار می‌گیرد. برای حل این مشکل ابتدا سطح زمین به مناطق کوچک بدون هم‌پوشانی تقسیم می‌شود [۲]. سپس منطقه با بیشترین ثبت موقعیت برای هر کاربر به دست آمده و میانگین ثبت موقعیت‌های درون آن منطقه محاسبه می‌شود.



شکل ۳- نمودار فرکانس داده‌های ثبت موقعیت مکان‌ها با دسته یکسان بر اساس اختلاف زمان آن‌ها

$$TP(t, \mu) = F(\lfloor t - \mu \rfloor) \quad (3)$$

در معادله ۳، t زمان محاسبه احتمال، μ میانگین زمان ثبت موقعیت‌های مجموعه داده است [۱].

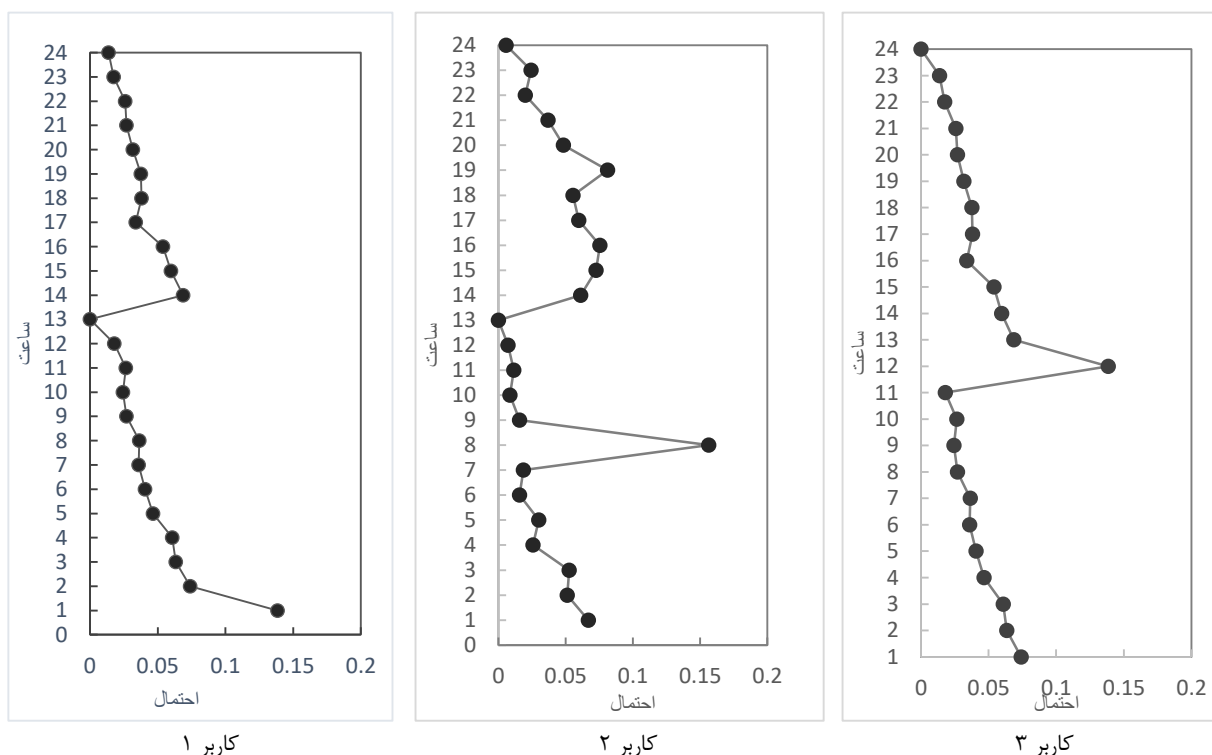
با توجه به تابع توزیع احتمال زمانی، احتمال ثبت موقعیت برای هر کاربر در ۲۴ ساعت شبانه‌روز به‌دست‌آمده است؛ بنابراین هر کاربر دارای یک دنباله احتمالاتی در ۲۴ ساعت شبانه‌روز در هر دسته مکان است. با توجه به دنباله‌هایی که برای هر کاربر به‌دست‌آمده، منحنی‌های زمانی کاربر در هر دسته ترسیم‌شده است. منحنی زمانی کاربر U در دسته مکان j شامل دنباله‌ای از مقادیر احتمالاتی است که به‌صورت $U^j = (u_1^j, u_2^j, \dots, u_m^j)$ نشان داده می‌شود و در آن u_m^j احتمال ثبت موقعیت کاربر در ساعت m و دسته j است [۱۸]. شکل ۴ نمونه‌هایی از منحنی‌های زمانی سه کاربر مختلف در یک دسته مکان را نشان می‌دهد که محور عمودی آن ۲۴ ساعت شبانه‌روز و محور افقی آن احتمال ثبت موقعیت در هر ساعت در آن دسته مکان است.

برای به دست آوردن تابع توزیع احتمال زمانی برای مجموعه مشخصی از ثبت موقعیت‌ها ابتدا باید تابع F را با توجه به شکل ۳، به‌صورت معادله ۲ تعریف کرد.

$$F(\Delta t) = \begin{cases} 0.15, \lfloor \Delta t \rfloor = 0 \\ 0.075, \lfloor \Delta t \rfloor = 1 \\ 0.067, \lfloor \Delta t \rfloor = -1, \\ \dots \\ 0.018, \lfloor \Delta t \rfloor = 12 \end{cases} \quad (2)$$

در معادله ۲، $F(\Delta t)$ فرکانس ثبت موقعیت در اختلاف‌زمان Δt است. همان‌طور که در معادله ۲ مشاهده می‌شود این تابع دارای ۲۴ خروجی بر اساس میزان فرکانس در اختلاف‌زمان پنجره زمانی ۱ ساعت است.

در نهایت تابع توزیع احتمال زمانی به‌صورت زیر، اختلاف‌زمان را بر اساس نمودار به‌دست‌آمده، به‌احتمال ثبت موقعیت تبدیل می‌کند؛ بنابراین احتمال ثبت موقعیت برای یک مکان از یک دسته مشخص با استفاده از معادله ۳ به دست می‌آید:



شکل ۴- منحنی‌های زمانی سه کاربر مختلف در یک دسته مکان

کاربر U و کاربر V ، به‌صورت $C(U^j, V^j)$ نشان داده می‌شود که به‌صورت $(u_{a_1}^j, v_{b_1}^j), (u_{a_2}^j, v_{b_2}^j), \dots, (u_{a_n}^j, v_{b_n}^j)$ از زوج‌های منطبق شده جداگانه از $U^j \times V^j$ است که

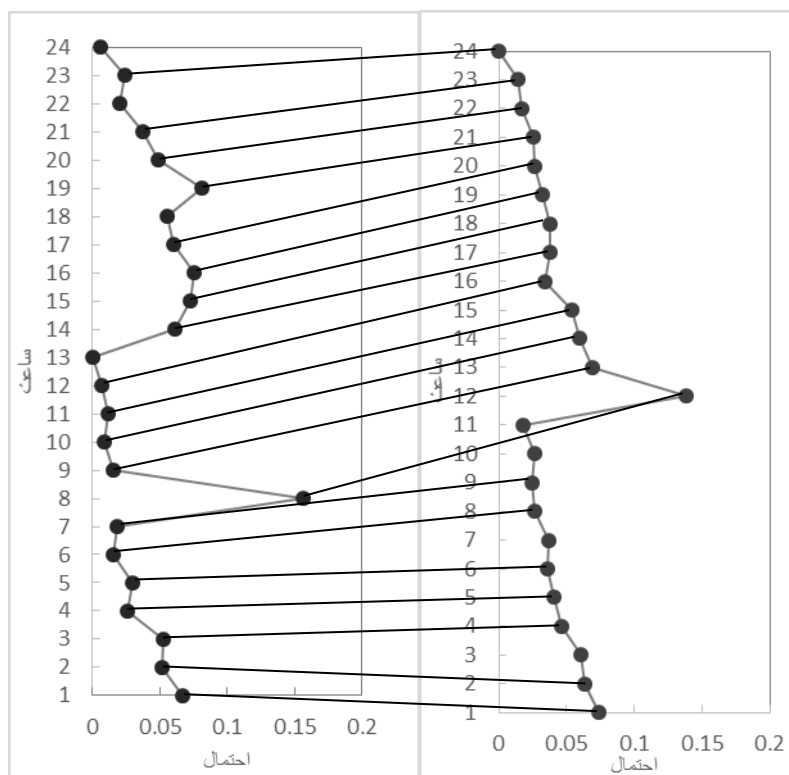
در مرحله بعد، منحنی اتصال^۱ بین منحنی‌های هر دو کاربر، به‌دست‌آمده است. منحنی اتصال بین دو منحنی

^۱ Curve Coupling

V برای دسته z که به صورت $MC(U^j, V^j)$ نشان داده می‌شود، یک منحنی اتصال است که از دو شرط زیر پیروی می‌کند [۳۲]:

$$\begin{aligned} 1) & \arg \max(C(U^j, V^j)) \\ 2) & \arg \min(\sum_{(p,q) \in C(U^j, V^j)} dist(p, q)) \end{aligned} \quad (۴)$$

شرط اول بیانگر این است که تعداد زوج‌های منطبق شده بیشینه باشد و شرط دوم کمینه بودن فاصله کلی بین هر زوج را نشان می‌دهد. شکل ۵ یک نمونه از بهترین منحنی اتصال بین دو منحنی زمانی U و V را نشان می‌دهد.



شکل ۵- نمونه از بهترین منحنی اتصال بین دو منحنی

است. کلاس NP-hard تنها به مسائل تصمیم‌گیری محدود نمی‌شود. نمونه‌ای از مسائل NP-hard «مسئله توقف‌پذیری»^۲ است. در این مسئله هدف ساختن الگوریتمی است که مشخص کند آیا یک برنامه کامپیوتری داده‌شده هیچ‌وقت پایان خواهد یافت (اجزای آن تکمیل خواهد شد و توقف خواهد کرد و یا وارد حلقه نامتناهی

^۲ Halting problem

برای $a_1 \geq 1$ و $b_1 \geq 1$ و $a_n \leq 24$ و $b_n \leq 24$ است و برای همه $i=1, \dots, n$ و $a_{i-1} < a_i$ و $b_{i-1} < b_i$ است که در آن $u_{a_i}^j$ احتمال ثبت موقعیت برای دسته z در ساعت a_1 است و $n = |C(U^j, V^j)|$ تعداد زوج‌های منطبق شده در دسته z است [۳۲]. با توجه به مطالب فوق هر زوج منطبق شده زمانی به صورت $(u_{a_i}^j, v_{b_i}^j)$ تعریف می‌شود که در آن $u_{a_i}^j$ احتمال ثبت موقعیت کاربر u برای دسته z در ساعت a_i است و $v_{b_i}^j$ احتمال ثبت موقعیت کاربر V برای دسته z در ساعت b_i است.

از آنجاکه تعداد منحنی‌های اتصال بین دو منحنی بسیار زیاد است، بهترین منحنی اتصال در نظر گرفته می‌شود. بهترین منحنی اتصال بین دو منحنی زمانی U و

یافتن بهترین منحنی اتصال یک مسئله NP-hard^۱ است. مسئله NP-hard به معنای حل نشدن مسئله در زمان چندجمله‌ای برحسب اندازه ورودی مسئله (زمان معقول) است. مجموعه NP-hard شامل چند هزار مسئله مختلف با کاربردهای فراوان است که تاکنون برای آن‌ها راه‌حل سریع و قابل انجام در زمان معقول پیدا نشده

^۱ Non-deterministic Polynomial-time hard

در معادله ۶، C مجموعه دسته‌های بازدید شده توسط هر دو کاربر u و v ، $cdist'(U^j, V^j)$ فاصله اتصال میانگین بین U^j و V^j که نرمال شده است، می‌باشد؛ بنابراین پس‌ازاینکه کاربران مشابه پیدا شدند، رفتار دوره‌ای ثبت موقعیت یک کاربر مشخص می‌تواند با استفاده از جمع وزن‌دار کاربران مشابهش به دست آید؛ بنابراین احتمال ثبت موقعیت کاربر در بازدید از مکان دسته C در زمان t از رابطه ۷ به دست می‌آید [۴]:

$$\hat{T}(u, c|t) = \frac{\sum_{v \in U_c} tsim(u, v) * T(v, c|t)}{\sum_{v \in U_c} tsim(u, v)} \quad (7)$$

در معادله ۷، $T(v, c|t)$ نشان‌دهنده احتمال این است که کاربر v در مکان دسته C در زمان t ثبت موقعیت می‌کند، U_c مجموعه کاربرانی است که از دسته c بازدید کرده‌اند و $tsim(u, v)$ تشابه زمانی بین کاربر u و v است؛ و به‌این‌ترتیب تأثیر زمانی ثبت موقعیت محاسبه می‌شود.

۳-۳-۳- تأثیر روابط اجتماعی

دوستان به دلیل داشتن علایق مشترک بسیار، تمایل به داشتن رفتارهای مشابه دارند و این امر منجر به رفتارهای ثبت موقعیت مشابه می‌شود [۳۴، ۳۳]. به‌عنوان مثال دوستان ممکن است برای تماشای فیلم باهم به سینما بروند و یا یک کاربر ممکن است به رستورانی برود که دوستش آن را بسیار توصیه کرده است [۱۷].

تأثیر تعاملات اجتماعی در انتخاب مکان‌ها برای توصیه مکان استفاده می‌شود [۳، ۱۸]؛ بنابراین زمانی که کاربر قصد دارد مکان جدیدی را برای بازدید انتخاب کند، رفتار مجموعه دوستان کاربر فاکتور مهمی خواهد بود. از سوی دیگر دوستانی که روابط اجتماعی نزدیک‌تر و صمیمی‌تری دارند، به توصیه‌های یکدیگر اعتماد بیشتری دارند. به‌عبارت‌دیگر دوستانی که رفتارهای ثبت موقعیت شبیه‌تری دارند، باید سلاقی مشترک بیشتری باهم داشته باشند و به همین دلیل پیشنهادهایی که از طرف آن دوستان ارائه می‌شود بهتر خواهد بود. بر این اساس در این پژوهش شباهت بین کاربر و دوستانش از سه دیدگاه موردبررسی قرار گرفته که به شرح زیر است:

خواهد گردید). برای حل این مسئله Zhou و همکاران یک روش ابتکاری ارائه داده‌اند [۳۲]. در این روش تطبیق نقاط از نقطه‌ای شروع می‌شود که در یک ساعت مشخص بیشترین احتمال را دارد؛ بنابراین نقطه شروع اتصال نقاطی خواهند بود که بیشترین مقدار احتمال را در هر دو منحنی کاربر داشته باشند. نتایج اتصال کاندید به‌صورت لیستی به دست می‌آید و بهترین نتیجه اتصال با استفاده از تعریف بهترین منحنی اتصال به دست می‌آید. ۲۴ ساعت شبانه‌روز به‌عنوان شمارنده دنباله در نظر گرفته شده است. هر نتیجه اتصال شامل تعدادی زوج مرتب است که برای انجام این اتصال دو درخت رو به بالا (با افزایش شمارنده دنباله) و درخت رو به پایین (با کاهش شمارنده دنباله) ایجاد می‌شود. هر مسیر روی درخت از گره ریشه که اولین زوج منطبق شده است و تا گره برگ که آخرین زوج منطبق شده است، یک نتیجه اتصال یک‌طرفه است؛ بنابراین ترکیب دو درخت رو به بالا و رو به پایین یک نتیجه اتصال دوطرفه را ایجاد می‌کند که با توجه به دو شرط بهترین منحنی اتصال می‌توان بهترین منحنی اتصال را از این روش به دست آورد. پس از به دست آوردن بهترین منحنی اتصال، شباهت زمانی را می‌توان با استفاده از اندازه فاصله بین دو منحنی زمانی کاربر به دست آورد.

بنابراین، با توجه به دو منحنی زمانی کاربران U و V و یافتن بهترین منحنی اتصال بین آن‌ها، فاصله اتصال میانگین بین دو منحنی به‌صورت زیر به دست می‌آید [۴]:

$$cdist(U^j, V^j) = \frac{\sum_{(p,q) \in MC(U^j, V^j)} dist(p, q)}{|MC(U^j, V^j)|} \quad (5)$$

در معادله ۵، U^j و V^j منحنی زمانی دو کاربر در دسته j است، $dist$ یک تابع فاصله متقارن و $MC(U^j, V^j)$ بهترین منحنی اتصال بین U^j و V^j است. چون فاصله میانگین فاصله کلی به‌دست‌آمده از زوج‌های منطبق شده است، هر چه این فاصله کم‌تر باشد، دو کاربر شباهت بیشتری به هم دارند.

شباهت بین کاربران بر اساس بهترین منحنی اتصال از معادله ۶ به دست می‌آید [۴]:

$$tsim = \frac{\sum_{j \in C} (1 - cdist'(U^j, V^j))}{|C|} \quad (6)$$

در معادله ۹، C_v دسته مکان‌هایی است که دوست کاربر بازدید کرده است و C_u دسته مکان‌هایی است که کاربر بازدید کرده است.

به‌منظور به دست آوردن میزان تأثیر روابط اجتماعی بر ثبت موقعیت کاربر در هر دسته روشی که به‌کاربرده شده است ترکیب سه معیار فوق و به دست آوردن وزن تأثیر اجتماعی است.

$$SI_{u,v} = \alpha FSI + \beta CSI + \gamma TSIM \quad (10)$$

هر سه معیار FSI ، CSI و $TSIM$ با توجه به اینکه نشان‌دهنده شباهت بین کاربر و دوستانش هستند اعدادی در بازه $[0,1]$ هستند. ضرایب α ، β و γ پارامترهای وزندهی هستند که به ترتیب میزان اهمیت FSI ، CSI و $TSIM$ را نشان می‌دهند. بین این سه ضریب وزندهی رابطه $\alpha + \beta + \gamma = 1$ برقرار است و هر سه پارامتر در بازه $[0,1]$ قرار دارند. برای به دست آوردن این ضرایب از روش سرشکنی ترکیبی استفاده شده است.

پس از محاسبه وزن تأثیر اجتماعی می‌توان احتمال ثبت موقعیت هر کاربر در هر دسته را بر اساس ثبت موقعیت دوستانش در آن دسته به دست آورد. طبق معادله زیر این احتمال محاسبه می‌شود:

$$\hat{c}(u, c|t) = \frac{\sum_{v \in F_u} SI_{u,v} \cdot c(v, c|t)}{\sum_{v \in F_u} SI_{u,v}} \quad (11)$$

۳-۴- توسعه الگوریتم

در این بخش، الگوریتم توصیه مکان جدیدی به نام PCLRTGS با استفاده از تأثیر زمانی، تأثیر مکانی و تأثیر روابط اجتماعی ارائه شده است. جزء زمانی از تأثیر زمانی رفتارهای ثبت موقعیت استفاده کرده و احتمال ثبت موقعیت را با استفاده از احتمال ثبت موقعیت کاربران مشابه به دست می‌آورد. جزء مکانی از تأثیر مکانی مربوط به هر مکان استفاده می‌کند. احتمال ثبت موقعیت کاربر در یک مکان را با توجه به فاصله مکان تا خانه کاربر مدل‌سازی می‌کند، به‌طوری‌که هر چه مکان به خانه کاربر نزدیک‌تر باشد، دارای احتمال ثبت موقعیت بیش‌تری خواهد بود. جزء روابط اجتماعی، از تأثیر دوستان بر روی رفتارهای ثبت موقعیت کاربر استفاده می‌کند. احتمال ثبت

۱- شباهت زمانی رفتار ثبت موقعیت بین کاربر و دوستانش در هر دسته مکان. از آنجاکه کاربران دارای علائق مشابه هستند، رفتار دوره‌ای ثبت موقعیت مشابهی نیز خواهند داشت؛ بنابراین شباهت زمانی رفتارهای ثبت موقعیت دوستان نیز می‌تواند در توصیه مکان مناسب‌تر به کاربر، مؤثر باشد. این امر بخصوص زمانی که کاربر در یک دسته تعداد کمی ثبت موقعیت دارد، مؤثر خواهد بود.

با توجه به مطالبی که در بخش زمانی بیان شده است، می‌توان با استفاده از روش اتصال منحنی‌های زمانی کاربر و هر یک از دوستانش، شباهت زمانی بین آن‌ها را به دست آورد؛ بنابراین ابتدا منحنی‌های زمانی هر کاربر و دوستانش در هر دسته را ایجاد کرده و سپس بهترین منحنی اتصال بین منحنی زمانی کاربر و هریک از دوستانش با توجه به دو شرط بیان‌شده در معادله ۴، به دست می‌آید. درنهایت با استفاده از معادله‌های ۵ و ۶، شباهت زمانی بین هر کاربر و هریک از دوستانش محاسبه می‌شود.

۲- شباهت ارتباطات اجتماعی کاربر و دوستانش. تقریباً ۳۰ درصد از مکان‌های جدید بازدید شده توسط یک کاربر، قبلاً توسط یکی از دوستانش و یا دوستان دوستانش بازدید می‌شود [۱۴]؛ بنابراین به‌منظور به دست آوردن این معیار از ضریب جاکارد^۱ برای محاسبه شباهت بین کاربر و دوستش از نظر تعداد دوستان مشترکی که دارند، استفاده شده است. معادله ۸ این مطلب را بیان می‌کند:

$$FSI = \frac{|F_u \cap F_v|}{|F_u \cup F_v|} \quad (8)$$

در معادله ۸، F_v مجموعه دوستان دوست کاربر و F_u مجموعه دوستان کاربر است.

۳- شباهت فعالیت‌های ثبت موقعیت کاربر و دوستانش در هر دسته مکان‌های بازدید شده توسط آن‌ها. در این قسمت با استفاده از ضریب جاکارد، شباهت بین کاربر و دوستش، از نظر دسته مکان‌های مشترکی که بازدید کرده‌اند، محاسبه شده است. هر چه مقدار این شباهت بیشتر باشد، توصیه‌های دوست کاربر قابل‌اعتمادتر است. معادله ۹ این شباهت را محاسبه می‌کند:

$$CSI = \frac{|C_u \cap C_v|}{|C_u \cup C_v|} \quad (9)$$

^۱ Jaccard coefficient

$$p(u, l|t) = SP(l, h) * (\hat{T}(u, c|t) * C(u, c|t)) \quad (12)$$

در معادله ۱۲، h مکان خانه کاربر، c دسته مکان مربوط به مکان، SP احتمال مکانی بازدید از مکان با توجه به مکان خانه کاربر، $\hat{T}(u, c|t)$ احتمال ثبت موقعیت در دسته مکان در زمان مشخص بر اساس شباهت زمانی و $C(u, c|t)$ احتمال ثبت موقعیت در یک دسته مکان در زمان مشخص بر اساس مجموعه دوستان کاربر است. در شکل ۶، مراحل اجرای الگوریتم به صورت مختصر بیان شده است.

موقعیت کاربر در هر دسته را بر اساس روابط اجتماعی بین کاربر و دوستانش، شباهت رفتارهای ثبت موقعیت دوستانش و شباهت کاربر و دوستانش از نظر زمانی، به دست می‌آورد.

با توجه به مستقل بودن هر سه جزء، احتمال مکانی، زمانی و اجتماعی بر اساس قانون احتمالات ترکیب سه احتمال با ضرب آن‌ها در هم امکان‌پذیر است؛ بنابراین با ترکیب سه جزء زمانی، مکانی و روابط اجتماعی احتمال ثبت موقعیت کاربر u در مکان l در زمان مشخص t به صورت معادله ۱۲ تعریف می‌شود:



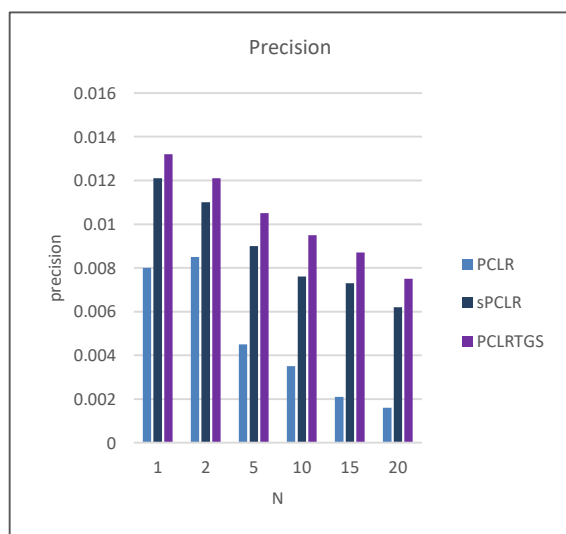
شکل ۶- فلوچارت مراحل اجرای الگوریتم

که در بخش ۳-۳ بیان شد، با استفاده از داده‌های مذکور در بخش ۳-۲ به ارزیابی الگوریتم توسعه داده شده پرداخته شده است. به منظور ارزیابی الگوریتم توصیه گر، داده‌ها به مجموعه داده‌های آموزشی و آزمایشی تقسیم شدند. این کار به این صورت انجام شد که به صورت

۴- نتایج

توسعه الگوریتم از طریق روش‌های بیان شده در بخش ۳-۳، در محیط برنامه‌نویسی جاوا و همچنین با استفاده از نرم‌افزار مدیریت پایگاه داده PostgreSQL پیاده‌سازی شد. در این بخش، پس از توسعه الگوریتم از طریق روش‌هایی

می‌تواند دو دلیل باشد. اول اینکه تعداد توصیه‌های صحیح کاهش می‌یابد در صورتی که تعداد توصیه‌ها افزایش می‌یابد و دوم اینکه تعداد توصیه‌های صحیح با سرعت کم‌تری نسبت به تعداد توصیه‌ها افزایش می‌یابد. برای مشخص کردن اینکه کدام دلیل باعث ایجاد این اتفاق شده است باید نمودار جامعیت نیز بررسی شود. شکل ۸ مقادیر جامعیت روش‌های مختلف را نشان می‌دهد. با مقایسه مقادیر جامعیت روش‌های توصیه گر، مشاهده می‌شود که مقادیر جامعیت، زمانی که تعداد توصیه‌ها افزایش می‌یابد، بیشتر می‌شود. با توجه به اینکه تعداد جواب‌های صحیح یک مقدار ثابت است، می‌توان نتیجه گرفت که تعداد توصیه‌های صحیح، زمانی که تعداد توصیه‌ها افزایش می‌یابد، بیشتر می‌شود. با توجه به شکل ۸ روند رو به رشد نمودار جامعیت نشان می‌دهد که تعداد توصیه‌های صحیح افزایش می‌یابد، اما سرعت افزایش آن کم‌تر از تعداد توصیه‌هاست. به کارگیری روابط اجتماعی در توصیه مکان باعث شده است که الگوریتم PCLRTGS نسبت به هر دو الگوریتم PCLR و sPCLR حدود ۱۰ تا ۱۵ درصد افزایش عملکرد داشته باشد؛ اما یکی از محدودیت‌های این پژوهش در مورد پیچیدگی زمانی الگوریتم در بخش زمانی است. با توجه به اینکه زمان توصیه برای کاربر و سیستم توصیه گر اهمیت بسیاری دارد بنابراین حل پیچیدگی زمانی که در یافتن بهترین منحنی اتصال وجود دارد بسیار مهم است. محدودیت دیگر حجم داده‌هایی است که برای توسعه و ارزیابی الگوریتم استفاده شده است. با توجه به اینکه الگوریتم برای توصیه مکان در مناطق مختلف نیاز به بررسی بیش‌تری دارد بنابراین نیاز به داده‌های وسیع‌تری خواهد بود.



شکل ۷- نمودار دقت

تصادفی یکی از ثبت موقعیت‌های هر کاربر برای ایجاد مجموعه آزمایشی انتخاب شد؛ و بقیه رکوردهای ثبت موقعیت در مجموعه داده‌های آموزشی قرار گرفت. به صورت تصادفی ۵ گروه از مجموعه داده‌های آموزشی و آزمایشی ایجاد شد. میانگین عملکرد پنج اجرا به عنوان عملکرد نهایی گزارش شده است.

عملکرد الگوریتم توصیه گر مکان به وسیله دو معیار دقت^۱ و جامعیت^۲ ارزیابی شد. این دو معیار در سطح وسیعی برای ارزیابی عملکرد سیستم‌های توصیه گر پذیرفته شده است. دقت، نسبت تعداد موارد مرتبط به تعداد موارد بازبازی شده است، در حالی که جامعیت، نسبت تعداد موارد بازبازی شده مرتبط به تعداد موارد مرتبط است [۳۵]. این دو معیار به ترتیب به صورت معادله‌های ۱۳ و ۱۴ در زیر تعریف می‌شوند [۳۵].

$$\text{Precision} = \frac{N_{rs}}{N_s} \quad (13)$$

در معادله ۱۳، N_{rs} تعداد آیتم‌های مرتبط مرتبط بازبازی شده، N_s تعداد آیتم‌های بازبازی شده است.

$$\text{Recall} = \frac{N_{rs}}{N_r} \quad (14)$$

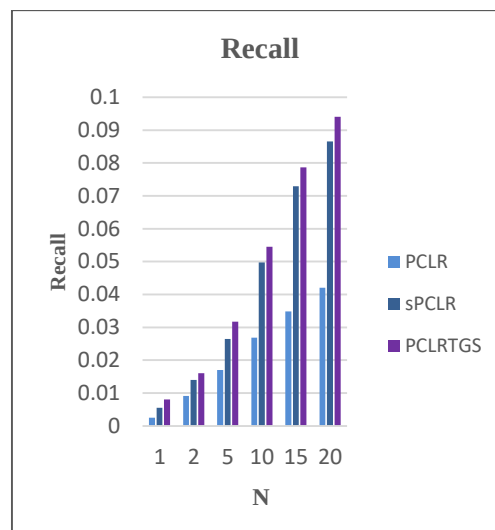
در معادله ۱۴، N_{rs} تعداد آیتم‌های مرتبط مرتبط بازبازی شده و N_r تعداد کل آیتم‌های مرتبط در دسترس است. الگوریتم توصیه گر مکان PCLRTGS از نظر عملکرد با دو الگوریتم موجود PCLR و sPCLR مقایسه شده است. شکل‌های ۷ و ۸ نشان می‌دهد که الگوریتم PCLRTGS، دارای عملکرد بهتری نسبت به هر دو الگوریتم از نظر دقت و جامعیت است. با توجه به اینکه الگوریتم PCLRTGS از روابط اجتماعی نیز در توصیه مکان استفاده کرده است، باعث بهبود توصیه مکان شده است. در شکل ۷ و ۸ که دقت و جامعیت را نشان می‌دهد، محور افقی تعداد توصیه‌ها و محور عمودی به ترتیب دقت و جامعیت توصیه گر را نشان می‌دهد. تعداد توصیه‌ها ۱، ۲، ۵، ۱۰، ۱۵ و ۲۰ در نظر گرفته شده است. شکل ۷ همچنین نشان می‌دهد که مقادیر دقت با افزایش تعداد توصیه‌ها کاهش می‌یابد. علت این اتفاق

^۱ Precision
^۲ Recall

در یک مکان را با توجه به تأثیر مکانی فاصله خانه کاربر تا مکان موردنظر، به دست می‌آورد. سپس مکان‌هایی که دارای احتمال بیشتری هستند را به‌عنوان مکان‌های موردعلاقه کاربر در نظر می‌گیرد. در بخش اجتماعی، احتمال ثبت موقعیت کاربر در یک دسته با توجه به مجموعه دوستان کاربر به دست می‌آید. وزن تأثیر اجتماعی به کار گرفته‌شده در این احتمال، بر اساس سه معیار روابط اجتماعی بین دو دوست، شباهت رفتارهای ثبت موقعیت آن‌ها و همچنین شباهت زمانی بین آن‌ها در یک دسته به دست می‌آید. درنهایت با توجه به احتمال ثبت موقعیت دوستان کاربر و وزن تأثیر اجتماعی، میزان احتمالی که هر کاربر در هر مکان ثبت موقعیت کند، محاسبه می‌شود.

الگوریتم PCLRTGS با دو الگوریتم PCLR و sPCLR به دلیل استفاده از مدل‌های مشابه در توصیه مکان با استفاده از دسته‌بندی مکان، مقایسه شده است. نتایج نشان می‌دهد که الگوریتم PCLRTGS از نظر دقت و جامعیت عملکرد بهتری نسبت به دو الگوریتم داشته است و این به دلیل در نظر گرفتن تأثیر روابط اجتماعی در توصیه مکان است.

در تحقیقات آینده به ارائه روشی برای اتصال منحنی‌های زمانی پرداخته می‌شود. با توجه به اینکه اتصال منحنی‌های زمانی دارای پیچیدگی زمانی زیادی است، نیاز به ارائه روشی که به کاهش این پیچیدگی‌ها کمک کند، وجود دارد. ارزیابی الگوریتم با استفاده از مجموعه داده‌های بزرگ‌تر، نیز یکی از اهداف آینده این تحقیق خواهد بود.



شکل ۸- نمودار جامعیت

۵- نتیجه‌گیری

در این تحقیق، الگوریتم توصیه مکان PCLRTGS ارائه‌شده است که در زمان مشخص از روز و بر اساس دسته‌بندی مکان، مکان‌هایی را به کاربر توصیه می‌کند. الگوریتم PCLRTGS علاوه بر در نظر گرفتن دو بخش زمانی و مکانی، بخش روابط اجتماعی را نیز در نظر می‌گیرد. بخش زمانی، از منحنی‌های زمانی که از رفتار ثبت موقعیت کاربر در طول زمان‌های مختلف در هر دسته به دست می‌آید، استفاده می‌کند. سپس با توجه به منحنی‌های زمانی، شباهت زمانی کاربر با کاربران دیگر را محاسبه می‌کند. درنهایت بر اساس این شباهت، احتمال ثبت موقعیت زمانی یک کاربر در یک دسته مکان را به دست می‌آورد. بخش مکانی نیز احتمال ثبت موقعیت یک کاربر

مراجع

- [1] Rahimi, S. M., & Wang, X. (2013). Location recommendation based on periodicity of human activities and location categories. Paper presented at the Pacific-Asia Conference on Knowledge Discovery and Data Mining.
- [2] Cho, E., Myers, S. A., & Leskovec, J. (2011). Friendship and mobility: user movement in location-based social networks. Paper presented at the Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining.
- [3] Ye, M., Yin, P., & Lee, W.-C. (2010). Location recommendation for location-based social networks. Paper presented at the Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems.
- [4] Zhou, D., & Wang, X. (2014). Probabilistic category-based location recommendation utilizing temporal influence and geographical influence. Paper presented at the Data Science and Advanced Analytics (DSAA), 2014 International Conference on.
- [5] Beeharee, A., & Steed, A. (2007). Exploiting real world knowledge in ubiquitous applications. Personal and Ubiquitous Computing, 11(6), 429-437.

- [6] Simon, R., & Fröhlich, P. (2007). A mobile application framework for the geospatial web. Paper presented at the Proceedings of the 16th international conference on World Wide Web.
- [7] Park, M.-H., Hong, J.-H., & Cho, S.-B. (2007). Location-based recommendation system using bayesian user's preference model in mobile devices. Paper presented at the International Conference on Ubiquitous Intelligence and Computing.
- [8] Zheng, V. W., Cao, B., Zheng, Y., Xie, X., & Yang, Q. (2010). Collaborative Filtering Meets Mobile Recommendation: A User-Centered Approach. Paper presented at the AAAI.
- [9] Zhou, D., Wang, B., Rahimi, S. M., & Wang, X. (2012). A study of recommending locations on location-based social network by collaborative filtering. Paper presented at the Canadian Conference on Artificial Intelligence.
- [10] Cheng, C., Yang, H., King, I., & Lyu, M. R. (2012). Fused Matrix Factorization with Geographical and Social Influence in Location-Based Social Networks. Paper presented at the Aaai.
- [11] Zhao, S., King, I., & Lyu, M. R. (2013). Capturing geographical influence in POI recommendations. Paper presented at the International Conference on Neural Information Processing.
- [12] Zhang, J.D., & Chow, C.-Y. (2013). iGSLR: personalized geo-social location recommendation: a kernel density estimation approach. Paper presented at the Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems.
- [13] Ference, G., Ye, M., & Lee, W.-C. (2013). Location recommendation for out-of-town users in location-based social networks. Paper presented at the Proceedings of the 22nd ACM international conference on Information & Knowledge Management.
- [14] Wang, H., Terrovitis, M., & Mamoulis, N. (2013). Location recommendation in location-based social networks using user check-in data. Paper presented at the Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems.
- [15] Ying, J. J.-C., Kuo, W.-N., Tseng, V. S., & Lu, E. H.-C. (2014). Mining user check-in behavior with a random walk for urban point-of-interest recommendations. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(3), 40.
- [16] Zhang, J.-D., Chow, C.-Y., & Li, Y. (2014). Lore: Exploiting sequential influence for location recommendations. Paper presented at the Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems.
- [17] Ye, M., Yin, P., Lee, W.-C., & Lee, D.-L. (2011). Exploiting geographical influence for collaborative point-of-interest recommendation. Paper presented at the Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval.
- [18] Ye, M., Janowicz, K., Mülligann, C., & Lee, W.-C. (2011). What you are is when you are: the temporal dimension of feature types in location-based social networks. Paper presented at the Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems.
- [19] Berjani, B., & Strufe, T. (2011). A recommendation system for spots in location-based online social networks. Paper presented at the Proceedings of the 4th Workshop on Social Network Systems.
- [20] Eagle, N., & Pentland, A. S. (2009). Eigenbehaviors: Identifying structure in routine. *Behavioral Ecology and Sociobiology*, 63(7), 1057-1066.
- [21] Li, Z., Ding, B., Han, J., Kays, R., & Nye, P. (2010). Mining periodic behaviors for moving objects. Paper presented at the Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining.
- [22] Li, Z., Wang, J., & Han, J. (2012). Mining event periodicity from incomplete observations. Paper presented at the Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining.
- [23] Wang, J., & Prabhala, B. (2012). Periodicity Based Next Place Prediction. the Proceedings of Mobile Data Challenge by Nokia Workshop at the Tenth International Conference on Pervasive Computing. University of Illinois at Urbana.
- [24] Bannur, S., & Alonso, O. (2014). Analyzing temporal characteristics of check-in data. Paper presented at the Proceedings of the 23rd International Conference on World Wide Web.
- [25] Gao, H., Tang, J., Hu, X., & Liu, H. (2013a). Exploring temporal effects for location recommendation on location-based social networks. Paper presented at the Proceedings of the 7th ACM conference on Recommender systems.
- [26] Zhang, K., Jin, Q., Pelechrinis, K., & Lappas, T. (2013). On the importance of temporal dynamics in modeling urban activity. Paper presented at the Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing.

- [27] Gao, H., Tang, J., Hu, X., & Liu, H. (2013a). Exploring temporal effects for location recommendation on location-based social networks. Paper presented at the Proceedings of the 7th ACM conference on Recommender systems.
- [28] Bao, J., Zheng, Y., & Mokbel, M. F. (2012). Location-based and preference-aware recommendation using sparse geo-social networking data. Paper presented at the Proceedings of the 20th international conference on advances in geographic information systems.
- [29] Liu, X., Liu, Y., Aberer, K., & Miao, C. (2013). Personalized point-of-interest recommendation by mining users' preference transition. Paper presented at the Proceedings of the 22nd ACM international conference on Information & Knowledge Management.
- [30] Zhao, Y.-L., Nie, L., Wang, X., & Chua, T.-S. (2014). Personalized recommendations of locally interesting venues to tourists via cross-region community matching. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(3), 50.
- [31] Hu, L., Sun, A., & Liu, Y. (2014). Your neighbors affect your ratings: on geographical neighborhood influence to rating prediction. Paper presented at the Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval.
- [32] Zhou, D., Rahimi, S. M., & Wang, X. (2016). Similarity-based probabilistic category-based location recommendation utilizing temporal and geographical influence. *International Journal of Data Science and Analytics*, 1(2), 111-121.
- [33] Ma, H., King, I., & Lyu, M. R. (2009). Learning to recommend with social trust ensemble. Paper presented at the Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval.
- [34] Ma, H., Yang, H., Lyu, M. R., & King, I. (2008). Sorec: social recommendation using probabilistic matrix factorization. Paper presented at the Proceedings of the 17th ACM conference on Information and knowledge management.
- [35] Toleva-Stoimenova, S. (2010). Evaluation of web based information systems: Users' informing criteria. *Issues in Informing Science and Information Technology*, 7, 297-310.