نشریه علمی پژوهشی مهندسی فناوری اطلاعات مکانی

سال هفتم، شماره چهارم، زمستان ۱۳۹۸

Vol.7, No.4, Winter 2020

۱۰۱ -۱۱٤

مقاله پژوهشی

# ارائه روشی برای استخراج الگوهای پر بازدید در خطوط سیر بر اساس توالی فعالیت های کاربران

**امین حسین پور میل آغاردان[۱]، رحیم علی عباسپور[۲]*، علیرضا چهرقان[۳]**

۱- دانشجوی دکتری *GIS*، دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی، پردیس دانشکده‌های فنی، دانشگاه تهران

۲- استادیار دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی، پردیس دانشکده‌های فنی، دانشگاه تهران

۳- استادیار دانشکده مهندسی معدن، دانشگاه صنعتی سهند

تاریخ دریافت مقاله: ۱۳۹۷/۰۸/۲۶ تاریخ پذیرش مقاله: ۱۳۹۸/۰۴/۱۸

## چکیده

در سال‌های اخیر فراگیر شدن استفاده از ابزارهای ثبت موقعیت مکانی، امکان تولید خطوط سیر هندسی با استفاده از مسیرهای حرکت کاربران را فراهم ساخته است. بدین ترتیب می‌توان علاوه بر هندسه و شکل مسیر، هدف کاربران از انجام سفر و فعالیت‌های مرتبط با آن را مورد توجه قرار داد. در این راستا خط سیر فعالیت کاربر که بیانگر توالی فعالیت‌های بازدید شده می‌باشد، در سال‌های اخیر مورد توجه بسیاری از محققین بوده است. از جمله مهمترین موضوعات مرتبط را می‌توان شناسایی الگوهای پرتکرار کاربر به منظور پیش‌بینی فعالیت بعـدی ذکر کرد. عمـده مطالعات پیشین تنها با تمرکز بر داده‌هـای یک کاربر است که این دسته از مطالعات، الگوهای پرتکرار ارائه شده فعالیت‌های قبلی را در نظر نمی‌گیرند. برای این منظور در این مقاله روشی پیشنهاد شده است تا پس از کدگذاری فعالیت‌ها و تشکیل ماتریس توالی آن‌ها، الگوهای پرتکرار را با استفاده از خطوط سیر تمام کاربران و در نظر گرفتن فعالیت‌های قبلی شناسایی نماید. همچنین روش پیشنهادی، قابلیت شناسایی الگوهای پرتکرار را برای مبدأ، مقصد و یا یک فعالیت پربازدید ارائه می‌نماید. پس از پیاده‌سازی بر روی داده‌های اخذ شده از تعداد ۱۰۶ کاربر در دانشگاه ام آی تی، نتایج روش پیشنهادی با دو روش بر مبنای ساختار درختی مورد مقایسه قرار گرفت. نتایج بیانگر میانگین ۶۰ درصدی کاهش زمان محاسبات برای تشکیل پایگاه داده مرتبط و افزایش ۱۷٫۵۰ درصدی در پیش‌بینی صحیح می‌باشد.

کلید واژه‌ها : خط سیر فعالیت، الگوی پر بازدید، ماتریس توالی.

* نویسنده مکاتبه کننده: تهران، خیابان کارگر شمالی، بعد از تقاطع جلال آل احمد، پردیس دانشکده‌های فنی دانشگاه تهران، دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی

تلفن: ۸۸۰۰۸۸۴۱

*Email: abaspour@ut.ac.ir*

[ Downloaded from jgit.kntu.ac.ir on 2025-07-04 ]

[ DOI: 10.29252/jgit.7.4.101 ]

## ۱- مقدمه

امروزه با توجه به ظهور فناوری‌های نوین در حوزه اطلاعات و ارتباطات مانند سیستم تعیین موقعیت جهانی (*GPS*[1])، موقعیت مکانی افراد در طی سفرهای درون و یا برون شهری توسط دستگاه‌های مربوط و تلفن همراه قابل ثبت و ذخیره است. موقعیت مکانی افراد را می‌توان به صورت خطوط سیر در نظر گرفت که زمینه بسیاری از تحقیقات در سال‌های اخیر می‌باشد. از جمله موضوعات مهم مورد توجه محققین، شناسایی الگوی هندسی حرکت کاربران در سفرهای درون شهری می‌باشد. در حقیقت جابجایی افراد در سفرهای درون شهری به منظور رسیدن به مقصدی خاص و انجام فعالیت مرتبط با موقعیت هدف می‌باشد. به عبارت دیگر کاربر از یک موقعیت مکانی با فعالیت مشخصی، سفر خود را شروع کرده و در طی سفر پس از توقف و انجام یک یا چندین فعالیت (مانند خرید، سینما و باشگاه) به مقصد (آخرین فعالیت) می‌رسد. معمولاً این نقاط هدف، با فعالیت مرتبط به عنوان نقاط مورد علاقه شناخته می‌شوند. نقاط مورد علاقه به مکان‌هایی گفته می‌شود که توسط فرد یا افرادی بطور مکرّر بازدید شده و یا این افراد زمانی را در آنها توقف کرده اند[۱].. بطور کلی شناسایی دقیق هدف فعالیت کاربر در موقعیت‌های مکـانی مرتبط به دو روش پس-پردازش[2] و پیش‌پردازش[3] می‌باشد. روش پس‌پردازش از تکنیک‌های غنی‌سازی داده‌های خط‌سیر استفاده می-کند و در نهایت فعالیت کاربر در هر نقطه استخراج می‌شود. طبق بررسی روش‌های ارائه شده، بهترین راه حل برای این منظور استخراج نقاط توقف از خط سیر هندسی می‌باشد. اما در بسیاری موارد نقاط توقف شناسایی شده از خط سیر هندسی، بیانگر فعالیت خاصی نبوده و بطور مثال نقاط توقف در تقاطع‌ها را شامل می‌شود. این موضوع چالشی است که در بسیاری از مطالعات استخراج خط سیر فعالیت به آن پرداخته شده است [۲و۳]. در روش پس‌پردازش نام و فعالیت مورد نظر توسط کاربر ذخیره می‌شود که عمدتاً زمان انجام فعالیت نیز در این روش ذخیره می‌شود.

توالی فعالیت‌های کاربر در طی سفر به عنوان خط سیر فعالیت در نظر گرفته می‌شود. به همین ترتیب الگوی فعالیت شامل زنجیره‌ای از فعالیت‌ها می‌شود که بیشترین تعداد تکرار را در میان سایر دنباله‌های فعالیت در طول یک بازه زمانی دارد [۴]. امروزه استخراج الگوی فعالیت در بسیاری از خدمات مکان مبنا (*LBS*[4]) اهمیت ویژه‌ای دارد. از آن جمله می‌توان به سرویس‌های تبلیغات هدفمند و شبکه‌های اجتماعی مکان مبنا[5] اشاره نمود. در این سرویس‌ها با توجه به الگوی فعالیت افراد و همچنین پیش‌بینی موقعیت بعدی آنها بر اساس الگوی مورد نظر، سرویس‌های تبلیغاتی و اجتماعی ارائه می‌شوند [۵و۶]. روش‌هایی که برای پیش‌بینی موقعیت ارائه شده‌اند بطور کلی با در نظر گرفتن الگوی حرکتی فرد به بررسی ارتباط بین نقاط فعلی و قبلی مورد علاقه و فاصله آنها پرداخته و در نهایت، قرارگرفتن فرد در موقعیت بعدی را بصورت احتمالی بیان می‌کنند [۷]. روش‌هایی که بدین منظور استفاده می‌شوند را می‌توان در دو گروه دسته بندی نمود. گروه اول روش‌های بر مبنای تئوری‌های احتمال است که مهمترین آنها مدل مارکوف می‌باشد [۷،۸و۹].

در این رویکرد ابتدا الگوی حرکتی مشابه در یک خوشه شناسایی شده و سپس برای یادگیری هر یک از این الگوها، مدل مارکوف بطور جداگانه آموزش داده شده است [۹]. در این راستا در سال ۲۰۰۳ از مدل مارکوف برای پیش‌بینی مقصد بعدی استفاده شد. در این مطالعه ابتدا نقاطی که در آنها سیگنال *GPS* دریافت نشده است به عنوان نقاطی که توقف بوده و احتمال

[1] *Global Positioning System*
[2] *Post-Processing*
[3] *Preprocessing*
[4] *Location-based Services*
[5] *Location-based social Networks*



[ Downloaded from jgit.kntu.ac.ir on 2025-07-04 ]
[ DOI: 10.29252/jgit.7.4.101 ]

تردد در آنها زیاد است شناسایی شده‌اند. در قدم بعدی از طریق الگوریتم *K-means* و با استفاده از شعاع حدآستانه، این نقاط به عنوان مراکز خوشه در نظر گرفته شده است. سپس با استفاده از الگوی حرکتی بین این نقاط و روش مارکوف، با توجه به تعداد بازدید بالا، نقاطی که بیشترین احتمال در میان نقاط هر خوشه را دارند شناسایی شده و مراکز خوشه به آن نقاط جابجا می‌شود. همزمان شعاع خوشه کاهش می‌یابد تا جایی که تغییر مراکز خوشه متوقف شود [۱۰]. همچنین در تحقیقی دیگر در سال ۲۰۱۵ با استفاده از مدل مارکوف و در نظر گرفتن طول خطوط - سیر نقاطی را به عنوان کاندید معرفی کردند [۱۱]. گروه دوم روش‌های بر مبنای درخت تصمیم‌گیری هستند. در این روش‌ها ابتدا موقعیت‌های پر بازدید کاربر شناسایی شده و کدگذاری می‌شوند. در مرحله بعد و با استخراج الگوهای حرکتی، درخت تصمیم شکل می‌گیرد که چندین الگوی حرکتی را در لایه‌های مختلف در خود ذخیره می‌کند. سپس در مرحله پیش‌بینی با توجه به الگوهای مشابه با رفتار فعلی از شاخه‌های این درخت برای پیش‌بینی موقعیت بعدی استفاده می‌شود [۱۲, ۱۳]. بطور مثال در پژوهش [۱۴] که در سال ۲۰۰۹ ارائه گردید، ابتدا نقاط با فرکانس بازدید زیاد شناسایی شده‌اند، سپس با ایجاد یک ساختار درختی موقعیت نقطه بعدی پیش‌بینی می‌شود؛ بدین ترتیب که برای پیش‌بینی نقطه بعد، مسیر طی شده توسط فرد با درخت مورد نظر تطبیق داده و نقطه بعدی در صورت وجود در داده‌های درخت ارائه می‌شود [۱۴]. با توجه به تعداد تکرار در الگوهای مختلف، در برخی از روش‌های ارائه شده برای هر یک از شاخه‌های درخت تصمیم، وزن نیز در نظر گرفته شده است [۱۵].

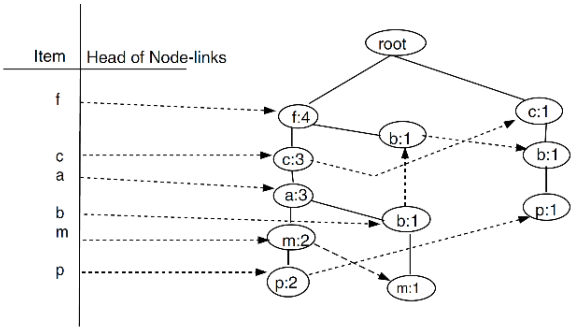
آنچه در زمینه پیش‌بینی موقعیت بعدی اهمیت دارد شناسایی الگوی فعالیت می‌باشد. همچنین نحوه شناسایی خطوط سیر فعالیت پربازدید و مشابهت آنها با یکدیگر چالش اصلی بسیاری از تحقیقاتی است که به آنها اشاره شد. در این راستا [۱۶] روشی بر مبنای ترکیب توابع خطی و غیر خطی حرکت کاربر ارائه داند که هندسه کاربر را با استفاده از تکنیک اپری‌وری[1] به‌صورت مدل‌های ریاضی ارائه می‌نماید. از جمله تحقیق مشابه می‌توان [۱۷] را نام برد. همچنین [۱۸] با بررسی رفتار حرکتی یک کاربر، روابط معنایی[2] را شناسایی کرده و در نهایت با استفاده از پارامترهای پشتیبانی[3] و اطمینان[4] پیش‌بینی موقعیت بعدی را انجام داده است. [۱۹] الگویی با نام الگوی زندگی شخصی ارائه دادند که به استخراج رفتارهای دوره‌ای می‌پردازد. [۱۲] با ارائه ساختاری جدید موارد الگوهای پرتکرار را برای هر کاربر شناسایی نموده است. این الگوریتم بر اساس مفهوم پیشوند و پسوند، الگوهای ترتیبی را در دنباله‌ها جستجو می‌کند. در این الگوریتم بعد از یک مرحله جستجوی تمامی‌اعضای دنباله‌ها، الگوهایی که بیشتر از یک مقدار آستانه تکرار شده‌اند را به عنوان الگوهای ترتیبی تک عضوی شناسایی می‌کند. سپس تمام الگوهای ترتیبی موجود در دنباله‌ها به زیرمجموعه‌هایی افراز می‌شوند. در هر یک از این زیرمجموعه‌ها یکی از الگوهای بدست آمده پیشوند می‌باشد. بر اساس این افرازها پایگاه داده انعکاسی هر یک از پیشوندها بدست می‌آید و الگوریتم در هر یک از زیرمجموعه‌ها، الگوهای ترتیبی تک عضوی را پیدا کرده و پیشوندهای جدید را محاسبه می‌کند. در پایان بصورت بازگشتی تمام الگوها را تا رسیدن به زیرمجموعه‌های تهی محاسبه می‌نماید [۱۲]. بطور مشابه در بسیاری از تحقیقات، از ساختار درختی برای استخراج الگوهای پرتکرار استفاده شده است که می‌توان به [۲۰] اشاره نمود که ساختاری را با نام *FP-Tree* برای این منظور ارائه کرده است. در این روش از یک درخت برای ذخیره اطلاعات مرتبط با مجموعه

[1] *Apriori*
[2] *Association Rules*
[3] *Support*
[4] *Confidence*



[ Downloaded from jgit.kntu.ac.ir on 2025-07-04 ]
[ DOI: 10.29252/jgit.7.4.101 ]

آیتم‌های متناوب استفاده می‌کند. هر مسیری که از ریشه این درخت آغاز شود و به یکی از برگ‌های آن ختم شود؛ نشان‌دهنده یکی از تراکنش‌های موجود در پایگاه تراکنش خواهد بود. شاخص ها به ترتیب نزولی تعداد تکرارشان در پایگاه تراکنش درخت سازمان می‌یابند؛ به‌طوری‌که شاخص هایی که تعداد تکرارشان بیشتر باشد نزدیک‌تر به ریشه قرار می‌گیرند.*FP-tree* ، یک جدول سرآیند دارد که همه آیتم‌های متناوب موجود در پایگاه تراکنش (به همراه تعداد تکرار هر یک) را در خود دارد. هر رکورد موجود در جدول سرآیند نقطه آغاز لیستی است که همه گره‌های *FP-tree* را که حاوی شاخص این رکورد جدول سرآیند هستند را به هم پیوند می‌دهد. نمونه‌ای از نحوه تشکیل این درخت در شکل (۱) ارائه شده است.




**شکل ۱: نحوه تشکیل درخت *FP* [۲۰].**

همچنین [۲۱] برای شناسایی و خوشه بندی الگوی حرکتی یک کاربر به منظور پیش‌بینی موقعیت بعدی روشی را با نام پیش‌بینی بر مبنای خوشه‌بندی پویا[1] ارائه کرده است.

با بررسی تحقیقات انجام شده چند نکته قابل توجه وجود دارد. اول اینکه تقریباً در تمامی این مطالعات، شناسایی الگوهای پرتکرار و یا روابط معنایی با استفاده از داده‌های یک کاربر صورت گرفته است. این در حالی است که تعداد خطوط سیر فعالیت مشابه یک کاربر محدود بوده و با توجه به وجود بازدید غیرمنتظره[2] در انتخاب موقعیت بعدی توسط کاربر، پیش‌بینی موقعیت بعدی دقت بالایی نخواهد داشت. ثانیاً روش‌هایی که بر مبنای استخراج روابط معنایی هستند، ویژگی توالی بین فعالیت‌ها را بطور کامل در نظر نمی‌گیرند که البته این موضوع از ضعف‌های روش اپری‌وری نیز می‌باشد. بطور دقیق‌تر باید گفت که در این موارد توالی بین فعالیت‌های در نظر گرفته نمی‌شود و این فعالیت‌ها بصورت دسته‌ای[3] مدنظر قرار می‌گیرد. این در حالی است که برای شناسایی دقیق الگوهای پرتکرار، توالی بین فعالیت‌ها از جمله موارد ضروری می‌باشد. از طرفی روش‌هایی که بر مبنای درخت داده کاوی هستند؛ هر خط سیر فعالیت را بصورت جداگانه به این درخت اضافه می‌کنند. همچنین با در نظر گرفتن یک حداقل پشتیبانی توالی، گره‌هایی که بیشترین تعداد بازدید را داشته باشند به عنوان الگوهای پرتکرار شناسایی می‌شوند. این در حالی است که با این روش شناسایی دقیق الگوهای پرتکرار با تعداد فعالیت مشخص که همگی تعداد بازدید برابر داشته باشند؛ ممکن نمی‌باشد. همچنین اضافه نمودن خط‌سیر فعالیت بصورت جداگانه به درخت زمان بر بوده و در صورتی که حجم داده زیاد باشد و یا تعداد بازدید غیرمنتظره در آن‌ها بالا باشد، حجم و پیچیدگی درخت موردنظر بسیار زیاد خواهد بود.

[1] *Dynamic clustering based prediction*
[2] *Drift*
[3] *Itemset*

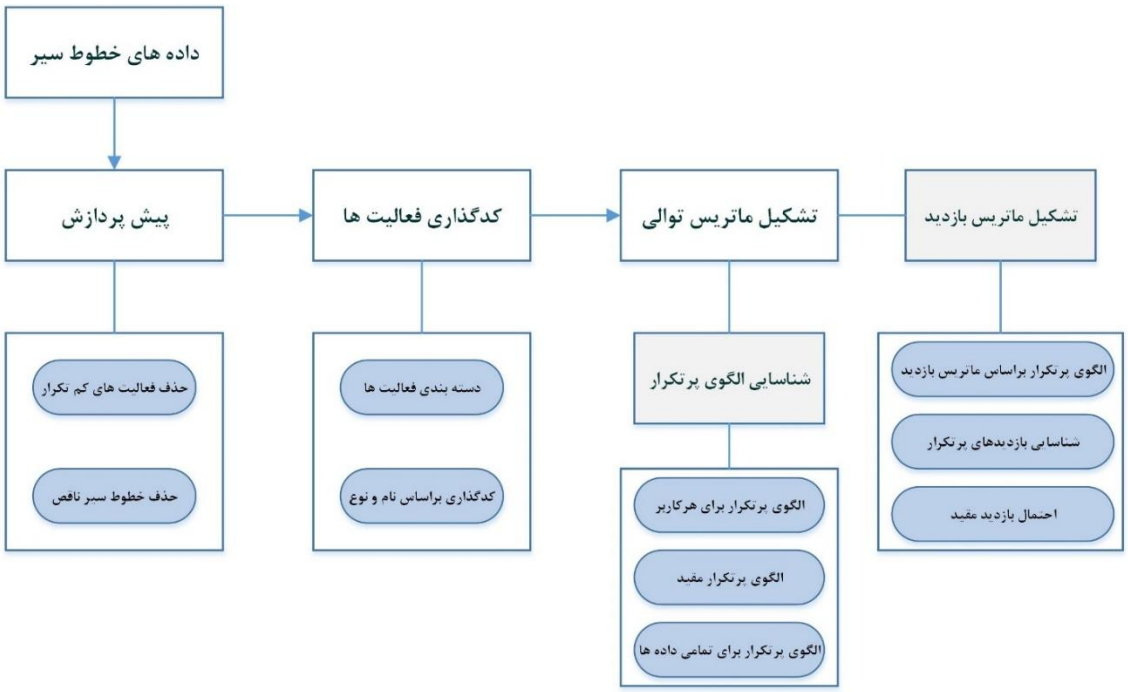


[ Downloaded from jgit.kntu.ac.ir on 2025-07-04 ]
[ DOI: 10.29252/jgit.7.4.101 ]

در این تحقیق روشی برای شناسایی الگوهای پرتکرار ارائه شده است تا داده‌های تمام کاربران برای استخراج الگو به کار گرفته شود. همچنین مهمترین ویژگی روش پیشنهادی در نظر گرفتن توالی بین فعالیت‌ها در شناسایی الگوهای پرتکرار می‌باشد. همچنین در این تحقیق برای پیش‌بینی موقعیت بعدی تنها به مقدار حداکثر بدست آمده در ماتریس انتقال و انتخاب آن به عنوان گزینه محتمل اکتفا نمی‌شود. بدین معنی که علاوه بر شماره موقعیت فعلی در خط سیر فعالیت طی شده، فعالیت‌های قبلی نیز برای پیش‌بینی در نظر گرفته می‌شود؛ چرا که مسلماً احتمال انتخاب یک گزینه ارتباط مستقیمی با تعداد، نوع و توالی فعالیت‌های پیشین دارد و این موضوعی است که در مطالعات پیشین به آن پرداخته نشده است. همچنین در روش پیشنهادی شناسایی الگوهای پرتکرار نیازمند اسکن کامل تمام داده‌ها نمی‌باشد؛ بطوری که برای شناسایی الگوهای پرتکرار، خطوط سیر فعالیت غیرمشترک، در هر مرحله از پروسه شناسایی حذف می‌شوند. همچنین روش پیشنهادی نیازمند ساختار درختی نبوده و استخراج الگوهای پرتکرار توسط ماتریس جدیدی به نام ماتریس توالی فعالیت صورت می‌گیرد. موارد مذکور در نهایت باعث کاهش حجم محاسبات و زمان می‌باشد که در ادامه به آن پرداخته خواهد شد. بخش دوم این مقاله به روش پیشنهادی پرداخته و جزئیات آن تشریح می‌شود. بخش سوم به بررسی و ارزیابی نتایج بدست آمده از پیاده‌سازی روش پیشنهادی می‌پردازد. همچنین نتیجه‌گیری و پیشنهادات برای تحقیقات آتی در بخش چهارم مقاله گنجانده شده است.

## ۲- روش پیشنهادی

ساختار کلی روش پیشنهادی در شکل (۲) ارائه شده است. در این تحقیق فرض بر این است که فعالیت‌های صورت گرفته توسط کاربران و یا از طریق روش‌های غنی‌سازی استخراج شده و به صورت خطوط سیر فعالیت در دسترس می‌باشد.




**شکل۲: مراحل انجام روش پیشنهادی**



[ Downloaded from jgit.kntu.ac.ir on 2025-07-04 ]

[ DOI: 10.29252/jgit.7.4.101 ]

**پیش پردازش:** در اولین مرحله از روش پیشنهادی، پیش پردازش بر روی خطوط سیر انجام می‌پذیرد. در این مرحله خطوط سیر با تعداد تکرار کم و همچنین داده‌های پرت (داده‌های با ثبت اشتباه موقعیت مکانی و یا نداشتن اطلاعات فعالیت) شناسایی و حذف می‌شوند.

**کدگذاری فعالیت‌ها**: در مرحله دوم آماده‌سازی داده‌های فعالیت جهت تکمیل ماتریس توالی فعالیت‌ها انجام می‌گیرد. برای این منظور فعالیت‌های ارائه شده دسته بندی شده و کدگذاری می‌شوند. این کدگذاری بصورت ترکیبی از اعداد و حروف می‌باشد. به طور مثال فعالیت‌های خانه، کار، رستوران و باشگاه بصورت ۱ تا ۴ کدگذاری شده و در صورتی که داده‌های فعالیت ارائه شده شامل چندین نوع مختلف از یک نوع فعالیت باشند از حروف برای تمییز دادن آنها استفاده می‌شود. بطور مثال سه نوع رستوران مختلف با مشخصه‌های $3A$، $3B$ و $3C$ از یکدیگر تشخیص داده می‌شوند.

**تشکیل ماتریس توالی:** در مرحله بعد، ماتریس توالی فعالیت، بر اساس کدگذاری انجام شده در مرحله قبل تشکیل می‌شود؛ بطوری که ستون‌های این ماتریس بیانگر انواع فعالیت‌های ارائه شده در داده‌های مورد مطالعه بوده و رکوردهای آن در برگیرنده خطوط‌سیر فعالیت کاربران می‌باشد. به عبارت دیگر برای هر خط سیر فعالیت یک رکورد در نظر گرفته می‌شود که ترتیب بازدید کاربر از فعالیت‌های انجام شده بصورت ترتیبی با اعداد زیر ستون مربوط به هر فعالیت درج می‌شود. نمونه‌ای از ماتریس توالی فعالیت در جدول(۱) ارائه شده است.

**جدول۱: نمونه‌ای از یک ماتریس توالی فعالیت**

| خط‌سیر- فعالیت | ۱ (محل سکونت) | ۲ (محل کار) | ۳- الف (تفریح-ورزش) | ۳-ب (تفریح- سینما) |
|---|---|---|---|---|
| ۱۳۶-۱ | ۱ | ۳ | ۲ | ۴ |
| ۱۰۵-۲ | ۱ | ۲ | ۳ | - |
| ۲۹-۳ | ۴ | ۱ | ۳ | ۲ |

پس از تشکیل این ماتریس یکی از مهمترین نتایج قابل استخراج، الگوهای پرتکرار می‌باشد که در واقع هدف اصلی از این مقاله می‌باشد. نکته قابل ذکر این است که منظور از الگوهای پرتکرار خط سیری است که تمام گره‌های آن توسط کاربران با رعایت توالی بیشترین بازدید را داشته است؛ چرا که در مطالعات پیشین الگوهای پرتکرار به الگویی اطلاق شده است که رعایت توالی بین گره‌های آن مهم نبوده و تنها احتمال بازدید بین هر دو گره بیشترین مقدار را داشته است. برای مثال برای یک خط‌سیر فعالیت با سه گره $A$، $B$ و $C$ بیشترین مقدار احتمال انتقال[1] از نقطه $A$ به $B$ و از $B$ به $C$ بوده است. بدین ترتیب به این معنی نیست که خط سیر $ABC$ بیشترین بازدید را بین کاربران داشته است. نحوه شناسایی الگوهای پرتکرار در روش پیشنهادی بصورت مرحله‌ای می‌باشد، بدین ترتیب که پس از هر مرحله رکوردهایی که در شرط ارائه شده صدق نمی‌کنند از مرحله بعد حذف می‌شوند. برای این منظور در مرحله اول پربازدیدترین مبدأ شناسایی می‌شود، بطوری که تعداد عدد ۱ در تمام ستون‌ها شمرده شده و ستونی که بیشترین تعداد را داشته باشد به عنوان پربازدیدترین مبدأ معرفی می‌شود. از این رو رکوردهای مربوط به این پرس‌وجو شناسایی شده و بقیه رکوردها برای مرحله بعد حذف شده و در واقع یک ماتریس با رکورد کمتر مورد استفاده قرار می‌گیرد. در مرحله بعد رکوردهایی که بیشترین تعداد عدد ۲ را داشته باشند؛ شناسایی می‌شوند و این کار تا یافتن الگوهای پرتکرار

[1] *Transition Probability*



[ Downloaded from jgit.kntu.ac.ir on 2025-07-04 ]

[ DOI: 10.29252/jgit.7.4.101 ]

بسته به تعداد گره معرفی شده ادامه پیدا می‌کند. راهکار دوم برای شرط توقف مراحل، تعریف یک حدآستانه برای تعداد رکوردها می‌باشد، بدین معنی که تعداد رکوردهای الگوهای پرتکرار از یک حدآستانه کمتر نباشد.

الگوی پرتکرار مقید: یکی از ویژگیهای مرتبط با خط‌سیر الگوهای پرتکرار شناسایی شده، یکنوابودن[1] می‌باشد. بدین ترتیب که برای هر خط سیری که به عنوان الگوی پرتکرار شناسایی شده، زیر مجموعه از این خط سیر نیز، به عنوان الگوهای پرتکرار مطرح می‌باشد. یکی از مسایل مطرح در ارتباط با الگوی پرتکرار، شناسایی این الگو بصورت مقیّد می باشد. بدین صورت که یک یا چند قید برای الگوی پرتکرار تعریف می شود که می تواند مبدا، مقصد و یا یک موقعیت یا فعالیت خاص در مسیر الگو باشد. به منظور یافتن الگوهای پرتکرار از مبدأ خاص، مرحله اول حذف شده و مراحل دوم از ستون مربوط به همان مبدأ ادامه پیدا می‌کند. اما این موضوع برای یافتن الگوهای پرتکرار با مقصد خاص متفاوت می‌باشد. بطوری که پرس‌وجوی مورد نظر با در نظر گرفتن ستون مربوط به مقصد مورد نظر، رکوردهایی را بیابد که اولا دارای بیشترین مقدار در این ستون بوده و مقدار موردنظر بیشترین تکرار را داشته است. از دیگر ویژگی‌های موردنظر این تحقیق یافتن الگوهای پرتکرار بصورتی است که فعالیتی خاص را شامل شود. برای این منظور یک پرس‌وجو به پردازش اضافه می‌شود. بدین ترتیب که رکوردهایی را بیابد که ستون مربوط به فعالیت موردنظر بیشترین تکرار را داشته و جزء فعالیت‌های مبدا یا مقصد نباشد. در نهایت خروجی این پرس‌وجو به عنوان ماتریس ورودی برای یافتن الگوهای پرتکرارمورد استفاده قرار می‌گیرد. همچنین با استفاده از ماتریس توالی مورد اشاره می‌توان الگوی پربازدید و همچنین الگوی مقید را مانند آنچه برای تمام داده بیان شد، برای یک کاربر خاص در نظر گرفت؛ بدین صورت که تنها از خطوط سیر فعالیت کاربر موردنظر برای شناسایی الگو استفاده می شود.

**تشکیل ماتریس بازدید:** اما یکی دیگر از مهمترین مراحل روش پیشنهادی استخراج ماتریس بازدید از ماتریس توالی فعالیت می‌باشد. این ماتریس در مطالعات بر مبنای درخت برای تشکیل آن مورد استفاده قرار می‌گیرد و بیانگر احتمال انجام هر یک از فعالیت‌های مورد مطالعه بعد از فعالیتی خاص می‌باشد. در مطالعات پیشین این ماتریس با یک خط سیر فعالیت تشکیل شده و با اضافه شدن هر خط‌سیر به‌روزرسانی می‌شود [۲۲]، که حجم بسیار زیادی از محاسبات و تغییر اندازه ماتریس مورد نظر را پس از اضافه شدن فعالیت‌های جدید برای خط سیرهای جدید در پی خواهد داشت. این در حالی است که ماتریس مورد نظر در روش پیشنهادی بعد از تشکیل ماتریس توالی فعالیت با تعداد فعالیت‌های مشخص و تنها یک بار محاسبه خواهد شد. در این رابطه پرس‌وجوی مورد نظر بین دو فعالیت، شامل تعداد رکوردهایی می‌گردد که توالی بین این دو فعالیت را داشته باشند. این پرس‌وجو بین هر دو زوج فعالیت اجرا می‌شود تا در نهایت ماتریس بازدید تشکیل شود.

ماتریس بازدید در بیشتر مطالعات پیشین به عنوان پایگاه داده برای تشکیل ساختار درختی خط سیر کاربران مورد استفاده قرار گرفته است که در نهایت در مراحل بعدی برای پیش‌بینی فعالیت یا موقعیت بعدی به کار گرفته شده است. با وجود این، تاکنون الگوی پیشینه[2] فعالیت‌های قبلی کاربر در این راستا در نظر گرفته نشده است و تنها بیشترین مقدار احتمال محاسبه شده توسط ماتریس بازدید برای موقعیت فعلی به عنوان گزینه محتمل بعدی در نظر گرفته شده است. در روش پیشنهادی در نظر گرفتن فعالیت‌های قبلی کاربر برای این موضوع ارائه می‌گردد. به عبارت دیگر الگوی پیشنهادی برای پیش‌بینی فعالیت بعدی از نقطه

[1] *Monotonocity*

[2] *History*



[ Downloaded from jgit.kntu.ac.ir on 2025-07-04 ]

[ DOI: 10.29252/jgit.7.4.101 ]

$P$ شامل خطوط سیری می‌شود که تعداد و توالی فعالیت بازدید مشترک با کاربر مورد نظر دارند. بطور مثال کاربر مورد نظر سه فعالیت خانه، محل کار و سینما را انجام داده است و فعالیت بعدی کاربر به عنوان موضوع پیش‌بینی می‌باشد. برای این منظور روش پیشنهادی خطوط‌سیری را برای پیش‌بینی در نظر می‌گیرد که هر سه فعالیت مذکور و توالی آنها را بازدید کرده‌اند. در نهایت با محاسبه احتمال بازدید بعدی پس از سینما می‌توان بیشترین مقدار آن را به عنوان پیش‌بینی محتمل ارائه نمود. در این راستا مقدار احتمال بازدید فعالیت بعدی $P_i(U_{a_n})$ بصورت رابطه(۱) معرفی می‌شود.

$$P_i(U_{a_n}) = \frac{N_i}{N_a} \ , \ 1 < i < n \ , \quad n = \text{Number of possible activiyies} \qquad \text{رابطه(۱)}$$

در این رابطه پارامترهای $N_i$ و $N_a$ به ترتیب بیانگر تعداد بازدید فعالیت $i$ ام و تعداد کل بازدید فعالیت‌های بعدی می‌باشد.

به همین ترتیب با استفاده از ماتریس بازدید می‌توان الگوی پرتکرار را استخراج نمود. بدین ترتیب که در الگوی مستخرج از این الگو، توالی فعالیت‌های قبلی کاربر در نظر گرفته نمی شود و بیشترین بازدید بین دو فعالیت به عنوان بازدید پرتکرار شناسایی شده و در ادامه با توالی این بازدیدهای پرتکرار، الگوی پرتکرار شناسایی می‌گردد. به همین ترتیب برای شناسایی الگوی پرتکرار مقید توسط ماتریس بازدید، قید موردنظر به عنوان اولویت درنظر گرفته شده و با شناسایی بازدیدهای پرتکرار قبل و بعد از قید مورد انتظار، الگوی فعالیت پرتکرار مقید شناسایی می شود.

## ۳- پیاده‌سازی و نتایج

در این تحقیق به منظور پیاده‌سازی روش پیشنهادی از داده‌های واقعی[1] استفاده می‌شود. این داده‌ها بین سالهای ۲۰۰۴ تا ۲۰۰۵ توسط تعداد ۱۰۶ کاربر در دانشگاه ام آی تی جمع آوری شده است. این داده‌ها در طول یک سال تحصیلی و شامل حدودا ۳۵۰ هزار ساعت داده فعالیت دانش آموزان می‌باشد. پارامترهای موقعیت، فعالیت و وضعیت ارتباط تلفن همراه از جمله مشخصات ثبت شده در این مجموعه داده می‌باشد.

در این تحقیق داده‌های فعالیت کاربران از این مجموعه داده استخراج شده تا برای پیاده‌سازی از آنها استفاده شود. از میان ۱۰۶ کاربر تعداد ۱۷ کاربر فاقد اطلاعات خط سیر فعالیت بودند که از میان داده‌ها حذف شدند. بنابراین در نهایت خطوط سیر فعالیت ۸۹ کاربر برای پیاده‌سازی انتخاب شد. همچنین تعدادی از خطوط سیر فعالیت کاربران با تعداد فعالیت بیشتر از ۲۵ یا کمتر از ۴، از مجموعه داده حذف گردید؛ زیرا پس از کدگذاری فعالیت‌های مربوط به داده‌ها، تعداد این خطوط سیر نسبت به مجموعه داده بسیار اندک می‌باشد. همچنین تعدادی از فعالیت‌های ثبت شده در این مجموعه داده، نام موقعیت مکانی میادین، خیابان‌ها و یا اسامی مخفف مانند *ML* می‌باشد که با توجه به عدم توضیح در فایل کمکی و در نظر گرفتن هدف تحقیق کنار گذاشته شدند. در نهایت، با توجه به حجم بسیار بالای داده مورد اشاره، تعداد ۸۹۰۰ خط سیر فعالیت انتخاب شد، بطوریکه ۸۹ کاربر و برای هریک ۱۰۰ خط سیر فعالیت در نظر گرفته شد. هر خط سیر داده‌های مربوط به یک کاربر برای یک روز می‌باشد که با توجه به مقادیر زمانی ارائه شده مجموعه داده برای هر کاربر تفکیک شده است.

در ادامه با توجه به اینکه داده‌های فعالیت ارائه شده در مجموعه داده مورد استفاده با در نظر گرفتن نزدیکی به دکل‌های مخابراتی و همچنین سنسورهای بی سیم ثبت شده اند، کدگذاری آنها طبق توضیحات ارائه شده در بخش قبلی انجام شد. پس از پیش پردازش و حذف

[1] -*http://reality.media.mit.edu*



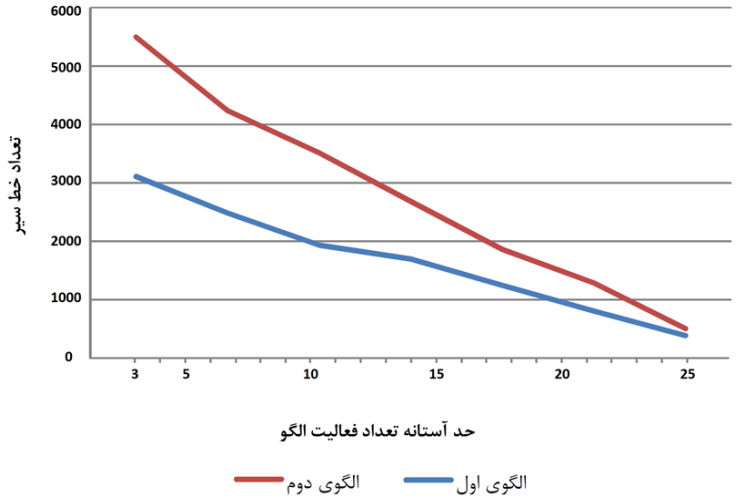
[ Downloaded from jgit.kntu.ac.ir on 2025-07-04 ]

[ DOI: 10.29252/jgit.7.4.101 ]

موقعیت‌های مکانی، تعداد ۳۸ فعالیت معتبر برای تشکیل ماتریس توالی فعالیت‌ها استخراج شد. در ادامه ماتریس موردنظر تشکیل شده و الگوهای پرتکرار موجود در داده‌ها برای تمام کاربران استخراج گردید. در این راستا به منظور تحلیل وابستگی تعداد الگوهای پرتکرار اول و دوم به حد آستانه، تعداد فعالیت‌ها به صورت نموداری تهیه شده و در شکل (۳) ارائه شده است. توضیح اینکه الگوی پرتکرار اول به عنوان الگویی هست که بیشترین تعداد خط سیر را شامل می شود. به همین ترتیب الگوی دوم بیانگر دومین الگو از نظر تعداد خط سیر می باشد.

اما با توجه به ساختار روش پیشنهادی می‌توان حالاتی را در نظر گرفت که الگوهای پرتکرارمورد نظر دارای بیشترین فعالیت بازدید شده در ترتیب $n$ام باشد. برای این منظور الگوهای پرتکرارهای مورد اشاره با در نظر گرفتن مقدار $n$ از ۲ تا ۶ در شکل (۴) نمایش داده شده اند. این موضوع در زمینه‌های پیش‌بینی موقعیت بعدی بسیار پرکاربرد می‌باشد. چرا که یافتن بیشترین احتمال فعالیت بعدی با مقایسه الگوهای پرتکرار در حالات مختلف از $n$ می‌تواند نتایج کامل و دقیق‌تری را ارائه نماید.

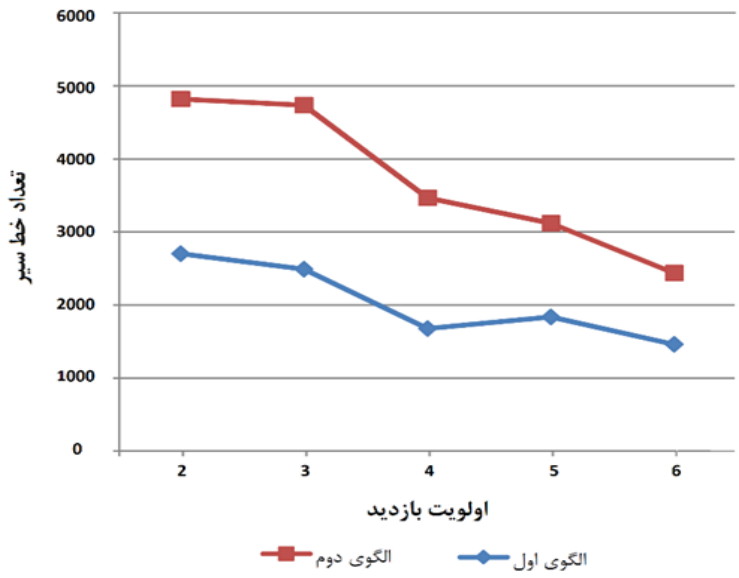
همان طور که در شکل (۴) ملاحظه می‌شود با افزایش شماره توالی فعالیت، تعداد خطوط سیر الگوها برای هر دو اولویت اول و دوم کاهش می‌یابد. با وجود این، بیشترین تعداد خطوط‌سیر در اولویت اول این شکل از تعداد خطوط سیر اشاره شده در شکل (۳) برای اولویت اول کمتر است که این موضوع به دلیل شروع فعالیت بیشتر کاربران از مبدا مشترک می‌باشد که عموما خانه یا باشگاه است. همان طور که در بخش قبل اشاره شد، یکی از قابلیت‌های روش پیشنهادی امکان شناسایی الگوهای پرتکرار با قید بازدید از یک فعالیت خاص می‌باشد. برای این منظور، در مجموعه داده مورد استفاده، الگوهای پربازدید شامل فعالیت کافه شناسایی شده و مشخصات آنها در جدول (۲) ارائه شده است.




**شکل ۳: تغییرات تعداد خطوط‌سیر الگوهای اول و دوم با توجه به حد آستانه تعداد فعالیت الگو**

[ Downloaded from jgit.kntu.ac.ir on 2025-07-04 ]

[ DOI: 10.29252/jgit.7.4.101 ]




**شکل ۴: تغییرات تعداد خطوط‌سیر فعالیت نسبت به تغییر اولویت بازدید کاربر**

**جدول ۲: شناسایی الگوهای پربازدید با قید بازدید از یک *AOI***

| تعداد فعالیت‌های قبلی | کمترین احتمال بازدید فعالیت بعدی(٪) | بیشترین احتمال بازدید فعالیت بعدی(٪) | تعداد خطوط‌سیر الگو | اولویت الگوی پربازدید |
|---|---|---|---|---|
| ۴ | ۶٫۷ | ۳۸٫۹ | ۷۳ | اول |
| ۷ | ۴٫۲۵ | ۴۰٫۳ | ۴۶ | دوم |
| ۵ | ۹٫۸ | ۳۲٫۴۱ | ۳۹ | سوم |

قابل ذکر است که نتایج ارائه شده در جدول (۲) برای فعالیت کاربران و فارغ از نوع آن ارائه شده است.

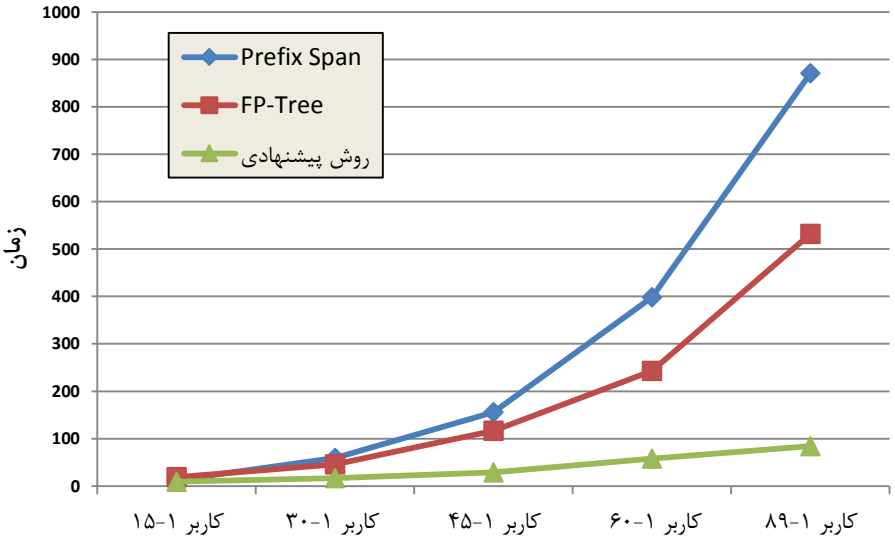
در ادامه به منظور ارزیابی روش پیشنهادی زمان تشکیل ماتریس توالی فعالیت‌ها در روش پیشنهادی و ماتریس انتقال با استفاده از دو روش *Prefix span* و *FP-Tree* برای تعداد خطوط سیر مختلف ارائه شده است (شکل (۵)).

آنچه در شکل (۵) نمایان است اختلاف قابل توجه در زمان تشکیل ماتریس توالی فعالیت در روش پیشنهادی و ساختار درختی در دو روش مذکور می‌باشد. دلیل این اختلاف را می‌توان در نحوه تشکیل پایگاه داده‌های مرتبط بررسی نمود؛ زیرا ساختارهای درختی با دو مورد اصلی همراه هستند. اول اینکه با اضافه شده هر خط سیر بایستی بروزرسانی شود که باعث افزایش حجم محاسبات می‌شود. همچنین با افزایش و متنوع بودن خطوط سیر تعداد شاخه‌های این ساختار به طور گسترده افزایش یافته که در نهایت زمان تشکیل ماتریس انتقال را نیز افزایش خواهد داد. این موضوع در روش پیشنهادی کاملا متفاوت می‌باشد؛ زیرا در این روش پس از کد گذاری فعالیت‌ها برای هر خط سیر به طور مجزا یک رکورد در ساختار ماتریس توالی فعالیت در نظر گرفته می‌شود و افزایش حجم خطوط‌سیر تاثیر زیادی بر پیچیدگی محاسبات نداشته و تنها تعداد رکوردهای این ماتریس را افزایش می‌دهد. این در حالی است که در ساختار درختی ممکن است با اضافه شدن یک خط‌سیر چندین شاخه، خصوصا خط سیر متفاوت،



[ Downloaded from jgit.kntu.ac.ir on 2025-07-04 ]
[ DOI: 10.29252/jgit.7.4.101 ]

به آن اضافه گردد؛ زیرا ساختار درختی بازدید بین هر دو فعالیت را بصورت جداگانه در نظر گرفته و به مجموعه پایگاه داده خود اضافه می‌نماید. بطور کلی روش پیشنهادی توانسته است پایگاه داده موردنظر خود را با زمان تقریبا ۶۵ درصد کمتر نسبت به پایگاه داده ساختار درختی تشکیل دهد. این بهبود در زمان در رابطه با حجم خطوط‌سیر زیاد، بیشتر بوده و قابل توجه می‌باشد.




**شکل ۵: مقایسه روش پیشنهادی و دو روش *FP-Tree* و *Prefix Span* در تشکیل پایگاه داده موردنظر**

در ادامه به منظور ارزیابی دقیق تر روش پیشنهادی از احتمال وقوع فعالیت ارائه شده در رابطه (۱) برای پیش بینی فعالیت چهارم پس از سه فعالیت مشترک و پیش بینی فعالیت پنجم پس از چهار فعالیت استفاده شده و نتایج آن در جدول شماره (۳) ارائه شده است.

**جدول ۳: نتایج پیش بینی فعالیت بعدی توسط روش پیشنهادی و روش *Fp-Tree***

| روش موردنظر | پیش بینی فعالیت پنجم | | پیش بینی فعالیت چهارم | |
|---|---|---|---|---|
| | پیش بینی صحیح | پیش بینی غلط | پیش بینی صحیح | پیش بینی غلط |
| روش پیشنهادی | ۷۵/۰۴ | ۲۴/۹۶ | ۶۹/۶۰ | ۳۰/۴۰ |
| روش FP-Tree | ۵۸/۳۰ | ۴۱/۷ | ۵۱/۲۵ | ۴۸/۷۵ |

به همین ترتیب نتایج پیش‌بینی موردنظر با استفاده از روش *FP-Tree* برای مقایسه با روش پیشنهادی در این جدول ارائه شده است. در این رابطه، الگوی پرتکرار با درنظر گرفتن فعالیت بعدی و قبلی برای مقایسه با روش*FP-Tree* در نظر گرفته شده است. همچنین برای این ارزیابی تعداد ۲۰۰ خط سیر بصورت اتفاقی از میان کاربران مختلف انتخاب شده اند که دارای شش فعالیت یکسان متوالی می باشند.

یکی از موارد قابل توجه در این جدول افزایش درصد پیش‌بینی درست توسط روش پیشنهادی با افزایش فعالیت‌های قبلی می‌باشد؛ بطوریکه درصد پیش‌بینی درست برای فعالیت چهارم ۶۹/۶۰ بوده و برای فعالیت



[ Downloaded from jgit.kntu.ac.ir on 2025-07-04 ]
[ DOI: 10.29252/jgit.7.4.101 ]

پنجم ۷۵/۰۴ می‌باشد. در توجیه این موضوع می‌توان گفت که در الگوی پرتکرار مستخرج از روش پیشنهادی با افزایش تعداد فعالیت‌های قبلی، تعداد خطوط سیری که از الگوی موردنظر پیروی می‌کنند افزایش می‌یابد. از طرفی عدم تغییر زیاد در تشخیص فعالیت بعدی صحیح توسط روش *Fp-Tree* می‌تواند به دلیل در نظر نگرفتن فعالیت‌های قبلی باشد؛ زیرا این روش تنها توالی فعالیت‌های پربازدید را بصورت جمعی در نظر می‌گیرد.

## ۴- نتیجه‌گیری

تاکنون تحقیقات بسیاری برای آنالیز و شناسایی الگوی هندسی کاربران ارائه شده است. در سال‌های اخیر علاوه بر الگوی هندسی حرکت، خط سیر فعالیت کاربران و شناسایی الگوهای فعالیت پربازدید به منظور پیش‌بینی فعالیت و موقعیت بعدی مورد توجه بوده است. مطالعات صورت گرفته در این زمینه تنها از داده‌های یک کاربر استفاده نموده و همچنین اطلاعات فعالیت‌های قبلی کاربران را برای پیش‌بینی موقعیت بعدی به کار نگرفته اند. در این راستا ساختارهای درختی و روش‌های بر مبنای روابط معنایی برای شناسایی الگوهای بازدید ارائه شده اند.

از این رو در این تحقیق روشی برای شناسایی الگوهای پرتکرار ارائه شد تا داده‌های تمام کاربران برای استخراج الگو به کار گرفته شود. همچنین مهمترین ویژگی روش پیشنهادی در نظر گرفتن توالی بین فعالیت‌ها در شناسایی الگوهای پرتکرار می‌باشد. همچنین در این تحقیق برای پیش‌بینی موقعیت بعدی تنها به مقدار حداکثر بدست آمده در ماتریس انتقال و انتخاب آن به عنوان گزینه محتمل اکتفا نمی‌شود؛ بدین معنی که علاوه بر شماره موقعیت فعلی در خط سیر فعالیت طی شده، فعالیت‌های قبلی نیز برای پیش‌بینی در نظر گرفته می‌شود؛ زیرا مسلماً احتمال انتخاب یک گزینه ارتباط مستقیمی با تعداد، نوع و توالی فعالیت‌های پیشین دارد و این موضوعی است که در مطالعات پیشین به آن پرداخته نشده است. همچنین در روش پیشنهادی شناسایی الگوهای پرتکرار نیازمند اسکن کامل تمام داده‌ها نمی‌باشد، بطوری که برای شناسایی الگوهای پرتکرار خطوط سیر فعالیت غیر مشترک در هر مرحله از پروسه شناسایی حذف می‌گردند. همچنین روش پیشنهادی نیازمند ساختار درختی نبوده و استخراج الگوهای پرتکرار توسط ماتریس جدیدی به نام ماتریس توالی فعالیت صورت می‌گیرد.

در این راستا برای ارزیابی روش پیشنهادی، پیاده‌سازی بر روی داده‌های اخذ شده از تعداد ۱۰۶ کاربر در دانشگاه ام آی تی صورت پذیرفت. نتایج روش پیشنهادی با دو روش بر مبنای ساختار درختی مورد مقایسه قرار گرفت. نتایج بیانگر میانگین ۶۰ درصدی کاهش زمان محاسبات برای تشکیل پایگاه داده مرتبط و افزایش ۱۷/۵۰ درصدی در پیش‌بینی صحیح می‌باشد.

در روش پیشنهادی معیارهای شخصی کاربران مانند سن، شغل، جنسیت و تحصیلات به کار گرفته نشده است که می‌تواند برای تحقیقات بعدی مورد استفاده قرار گیرند. به همین ترتیب معیار زمان حرکت و همچنین شرایط محیطی از دیگر معیارهای موردنظر مولف می‌باشد که می‌تواند در بهبود نتایج و تفسیر دقیقتر الگوهای مورداشاره تاثیر بسزایی داشته باشد. بررسی ارتباط گره‌های فعالیت، زمان توقف برای فعالیت، روز هفته، و همچنین فواصل زمانی مکانی بین نقاط توقف فعالیت از جمله مواردی است که از دید مولف می‌تواند برای بهبود دقت مدل پیشنهادی جهت بررسی معنایی خطوط سیر به کار گرفته شود. داده-های فعالیت مورد استفاده در این تحقیق محدود بوده و بسیاری از فعالیت های مورد اشاره از مجموعه داده‌ها حذف و یا یکپارچه شده‌اند. مسلماً استفاده از داده‌هایی که همزمان مولفه‌های معنایی و هندسی با دقت مناسبی را در بر داشته باشد می‌تواند نتایج تحقیق را بهبود بخشد.



[ Downloaded from jgit.kntu.ac.ir on 2025-07-04 ]

[ DOI: 10.29252/jgit.7.4.101 ]

## مراجع

[1] A. Y. Xue, R. Zhang, Y. Zheng, X. Xie, J. Yu, and Y. Tang, "Desteller: A system for destination prediction based on trajectories with privacy protection," Proceedings of the VLDB Endowment, vol. 6, pp. 1198-1201, 2013.

[2] V. Bogorny, B. Kuijpers, and L. O. Alvares, "ST - DMQL: a semantic trajectory data mining query language," International Journal of Geographical Information Science, vol. 23, pp. 1245-1276, 2009.

[3] N. Eagle, A. S. Pentland, and D. Lazer, "Inferring friendship network structure by using mobile phone data," Proceedings of the national academy of sciences, vol. 106, pp. 15274-15278, 2009.

[4] R. Monclar, A. Tecla, J. Oliveira, and J. M. de Souza, "MEK: Using spatial–temporal information to improve social networks and knowledge dissemination," Information Sciences, vol. 179, pp. 2524-2537, 2009.

[5] S. S. Banerjee and R. R. Dholakia, "Mobile advertising: Does location based advertising work?," 2008.

[6] C.-C. Chan, Y.-C. Lin, and M.-S. Chen, "Recommendation for advertising messages on mobile devices," in Proceedings of the 23rd International Conference on World Wide Web, 2014, pp. 235-236.

[7] B. Wang, Y. Hu, G. Shou, and Z. Guo, "Trajectory Prediction in Campus Based on Markov Chains," in International Conference on Big Data Computing and Communications, 2016, pp. 145-154.

[8] R. Hariharan and K. Toyama, "Project Lachesis: parsing and modeling location histories," in International Conference on Geographic Information Science, 2004, pp. 106-124.

[9] M. Chen, Y. Liu, and X. Yu, "Predicting next locations with object clustering and trajectory clustering," in Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2015, pp. 344-356.

[10] D. Ashbrook and T. Starner, "Using GPS to learn significant locations and predict movement across multiple users," Personal and Ubiquitous computing, vol. 7, pp. 275-286, 2003.

[11] M. Chen, X. Yu, and Y. Liu, "Mining moving patterns for predicting next location," Information Systems, vol. 54, pp. 156-168, 2015.

[12] M. Morzy, "Mining frequent trajectories of moving objects for location prediction," Machine Learning and Data Mining in Pattern Recognition, pp. 667-680, 2007.

[13] A. Y. Xue, R. Zhang, Y. Zheng, X. Xie, J. Huang, and Z. Xu, "Destination prediction by sub-trajectory synthesis and privacy protection against such prediction," in Data Engineering (ICDE), 2013 IEEE 29th International Conference on, 2013, pp. 254-265.

[14] A. Monreale, F. Pinelli, R. Trasarti, and F. Giannotti, "Wherenext: a location predictor on trajectory pattern mining," in Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, 2009, pp. 637-646.

[15] M. Boukhechba, A. Bouzouane, B. Bouchard, C. Gouin-Vallerand, and S. Giroux, "Online prediction of people's next Point-of-Interest: Concept drift support," in Human Behavior Understanding, ed: Springer, 2015, pp. 97-116.

[16] H. Jeung, Q. Liu, H. T. Shen, and X. Zhou, "A hybrid prediction model for moving objects," in Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on, 2008, pp. 70-79.

[17] M. Morzy, "Prediction of moving object location based on frequent trajectories," in International Symposium on Computer and Information Sciences, 2006, pp. 583-592.

[18] G. Yavaş, D. Katsaros, Ö. Ulusoy, and Y. Manolopoulos, "A data mining approach



[ Downloaded from jgit.kntu.ac.ir on 2025-07-04 ]

[ DOI: 10.29252/jgit.7.4.101 ]

*for location prediction in mobile environments," Data & Knowledge Engineering, vol. 54, pp. 121-146, 2005.*

*[19] Y. Ye, Y. Zheng, Y. Chen, J. Feng, and X. Xie, "Mining individual life pattern based on location history," in Mobile Data Management: Systems, Services and Middleware, 2009. MDM'09. Tenth International Conference on, 2009, pp. 1-10.*

*[20] C. I. Ezeife and Y. Su, "Mining incremental association rules with generalized FP-tree," in Conference of the Canadian Society for Computational Studies of Intelligence, 2002, pp. 147-160.*

*[21] D. Katsaros, A. Nanopoulos, M. Karakaya, G. Yavas, Ö. Ulusoy, and Y. Manolopoulos, "Clustering mobile trajectories for resource allocation in mobile environments," in International Symposium on Intelligent Data Analysis, 2003, pp. 319-329.*

*[22] C. Yu, Y. Liu, D. Yao, L. T. Yang, H. Jin, H. Chen, and Q. Ding, "Modeling user activity patterns for next-place prediction," IEEE Systems Journal, vol. 11, pp. 1060-1071, 2017.*

[ Downloaded from jgit.kntu.ac.ir on 2025-07-04 ]

[ DOI: 10.29252/jgit.7.4.101 ]

*K. N. TOOSI UNIVERSITY OF TECHNOLOGY*
*FACULTY OF GEODESY AND GEOMATICS ENGINEERING*

*Journal of Geospatial Information Technology*
**Vol.7 No.4, Winter 2020**

***Research Paper***

# *A Framework for Exploring the Frequent Patterns based on Activities Sequence*

***Amin Hosseinpoor Milaghardan[1], Rahim Ali Abbaspour[2]*, Alireza Chehreghan[3]***

*1- School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran*
*2- School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran*
*3- Mining Engineering Faculty, Sahand University of Technology, Tabriz, Iran*

***Abstract***

*In recent years, the development of the use of location-based tools has made it possible to produce geometric trajectories from the user's movement paths. In this way, users' goal of traveling and related activities can be considered in addition to the geometry and route shape. the user activity trajectory represents the sequence of the visited activities and its related analysis as presented in many studies. In the meantime, the most important analysis is the identification of frequent patterns to predict future activities. In previous studies, only one user's data was used, as well as previous activities of the user were not taken into account in presented frequent patterns. For this purpose, this paper presents a framework such as After encoding the activities and forming a sequence matrix, the frequent patterns using all users' trajectories are identified considering previous activities. The proposed method also offers the ability to identify frequent patterns for the origin, destination, or interesting activity. The results of comparing the proposed method with two methods based on tree structure represent a mean of 60% reduction in computing time to form the database.*

***Key words :*** *activity trajectory, frequent pattern, sequence matrix.*

[ Downloaded from jgit.kntu.ac.ir on 2025-07-04 ]

[ DOI: 10.29252/jgit.7.4.101 ]

***Correspondence Address****: School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran.*
***Tel****: : 021-61114251.*
***Email :*** *abaspour@ut.ac.ir*