

ایجاد خطوط سیر مکانی ساختار یافته بر اساس مدل توقف-حرکت به منظور غنی سازی معنایی

سمیه عاقل شاه نشین^۱، رحیم علی عباسپور^{۲*}

^۱ دانشجوی کارشناسی ارشد سیستم های اطلاعات مکانی - دانشکده مهندسی نقشه برداری و اطلاعات مکانی -

پردیس دانشکده های فنی - دانشگاه تهران

s_aghel@ut.ac.ir

^۲ استادیار دانشکده مهندسی نقشه برداری و اطلاعات مکانی - پردیس دانشکده های فنی - دانشگاه تهران

abaspour@ut.ac.ir

(تاریخ دریافت دی ۱۳۹۴، تاریخ تصویب اردیبهشت ۱۳۹۵)

چکیده

امروزه به واسطه توسعه فناوری، استفاده از سیستم های تعیین موقعیت افزایش چشمگیری داشته است. این ابزار داده های موقعیت را به صورت رشته ای از نقاط مکانی-زمانی به نام خط سیر تولید می کنند. در سال های اخیر تحقیقات مربوط به مدیریت داده های خط سیر، اغلب با تمرکز بر تکنیک های ذخیره سازی/بازیابی، مدلسازی، داده کاوی و کشف دانش از داده های خط سیر بوده است. این مطالعات عمدتاً بر جنبه های هندسی داده های خام تاکید داشته اند. با این حال، ما امروزه شاهد گسترش تقاضا در بخش های کاربردی مختلف از ردیابی کالا توسط کشتی های تجاری تا شبکه های اجتماعی، با فهم رفتار (الگوی) معنایی اشیاء در حال حرکت هستیم. الگوهای معنایی نه تنها شامل الگوهای هندسی هستند، بلکه دانش استخراج شده از داده های حرکت و اطلاعات مربوط به حوزه جغرافیایی و کاربردی را نیز شامل می شوند. فرآیند کشف دانش از داده های حرکت و غنی سازی معنایی خطوط سیر مکانی معمولاً شامل چند گام پردازش است تا داده های ورودی خام به اپیزودهای معنی دار تبدیل شوند. این مقاله، ضمن ارائه چارچوب محاسباتی خط سیر معنایی، به بررسی و پیاده سازی هر یک از مراحل برای آماده سازی داده ها به منظور کشف دانش و اضافه کردن معنا به داده ها می پردازد. در این کار، فرآیند پردازش داده ها در هر گام چکیده سازی به ترتیب با داده های خام پاکسازی شده، خط سیر مکانی-زمانی و خط سیر ساختاریافته مورد بررسی قرار گرفته است و الگوریتم های ارائه شده روی داده های واقعی خط سیر پیاده سازی و ارزیابی شده است.

واژگان کلیدی: خط سیر مکانی، داده کاوی و کشف دانش، غنی سازی معنایی، فشرده سازی

* نویسنده رابط

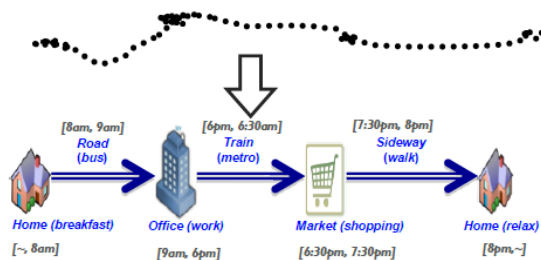
۱- مقدمه

امروزه با ظهور و توسعه دستگاه‌هایی مانند تلفن‌های هوشمند، PDA و سیستم‌های ناوبری در خودروها که مجهز به سیستم‌های تعیین موقعیت نظیر GPS هستند، امکان گردآوری داده‌های موقعیت تولید شده توسط اشیای متحرک افزایش چشمگیری داشته است. علاوه بر این، تکنیک‌های ردیابی حسگر مانند ماهواره‌ها، رادار و RFID، جمع‌آوری و پردازش حجم عظیمی از داده‌های حرکت را انجام می‌دهند [۱]. عموماً این ابزار داده‌های خام حرکت را با یک شناسه از شی و موقعیت آن در یک لحظه از زمان تولید می‌کنند. این داده‌ها که بیشتر به صورت رشته‌ای از نقاط مکانی و زمانی ذخیره و مدلسازی می‌شوند، اصطلاحاً خط سیر^۱ نامیده می‌شوند [۲].

در سالهای اخیر موضوع خط سیر به واسطه اهمیت روزافزون آن، مورد توجه پژوهشگران بوده است. تحقیقات مربوط به خط سیر در حوزه‌های گوناگون توسعه یافته است که از آن جمله می‌توان به ایجاد روش‌های پایه برای بازسازی خط سیر خام [۳]، توسعه نوع داده‌های جدید و پیچیده و عملگرها و زبان‌هایی برای پایگاه داده‌های اشیاء متحرک [۴، ۵] و ایجاد الگوریتم‌های داده‌کاوی جدید برای داده‌های خط سیر [۶، ۷، ۸] اشاره کرد. مطالعات انجام شده در حوزه خط سیر، عموماً با تمرکز بر خط سیر خام انجام شده‌اند و اطلاعات زمینه را که می‌تواند توصیف کننده دانش معنایی درباره حرکت باشد نادیده می‌گیرند؛ در نتیجه یک توصیف کلی‌نگر که هم جنبه هندسی و هم جنبه معنایی را دربرگیرد، به سختی امکانپذیر است [۹].

غنی‌سازی خطوط سیر معنایی^۲ یک مفهوم نسبتاً جدید در حوزه مطالعات مربوط به خط سیر است که در سال‌های اخیر مورد توجه محققین قرار گرفته است. در اصطلاحات مربوط به خط سیر معنایی، حرکت به عنوان یک رشته از بخش‌های معنی‌دار تعریف می‌شود که شی متحرک در آن یا متوقف است (به عنوان مثال محل کار، مرکز خرید، خانه) و یا حرکت می‌کند (رانندگی، پیاده‌روی و غیره). شکل ۱ مثالی از خط سیر هندسی (شامل نقاط خام جمع‌آوری شده با GPS) و خط سیر معنایی متناظر با آن را نشان می‌دهد [۱۰]. خط سیر هندسی، نمایش

مسیر حرکت شی متحرک به صورت رشته‌ای از نقاط خام به شکل $\langle (P_1, t_1), (P_2, t_2), \dots, (P_N, t_N) \rangle$ است که در آن $P_i = \langle x_i, y_i \rangle$ موقعیت نقطه و $t_1 < t_2 < \dots < t_n$ زمان ثبت هر موقعیت است و خط سیر معنایی متناظر، دنباله‌ای از بخش‌های همگن شناسایی شده و غنی‌سازی شده با اطلاعات معنایی بر اساس کاربرد موردنظر است.



شکل ۱- خط سیر هندسی (بالا) و خط سیر معنایی متناظر آن (پایین) [۱۰]

اضافه کردن معنا به داده‌های خام خط سیر، به منظور ایجاد یک مدل داده از حرکت شی متحرک است که قادر به نمایش معانی سطح بالا از داده‌های حرکت باشد [۹]. غنی‌سازی معنایی خط سیر، باعث بهبود مدلسازی و تحلیل داده‌های شی متحرک می‌شود. در این حوزه، یک شی متحرک یک موجودیت^۳ است که موقعیتش با زمان تغییر می‌کند. این شی متحرک می‌تواند یک شخص با حالات مختلف حمل و نقل (پیاده‌روی، با اتومبیل، با اتوبوس و...)، یک تاکسی یا اتوبوس در حال حرکت در یک شهر و یا حتی کشتی در حال حرکت در دریا باشد [۱۱].

غنی‌سازی خطوط سیر مکانی با معنا، فرآیندی با چندین گام پردازش است. در این فرآیند، داده‌های جمع‌آوری شده با دستگاه‌های تعیین موقعیت (مثل GPS) در پایین‌ترین سطح انتزاعی^۴ پردازش می‌شود و خطوط سیر غنی شده با اطلاعات معنایی در بالاترین سطح تولید می‌شود. بر همین اساس، این فرآیند را می‌توان بر اساس خروجی هر مرحله از چکیده‌سازی در سه گام خلاصه کرد:

۱. پیش‌پردازش و تولید داده‌های خام پاکسازی شده،
۲. شناسایی اپیزودها و تولید خطوط سیر ساختاریافته،
۳. غنی‌سازی خطوط سیر مکانی با معنا و تولید خط سیرهای معنایی [۱۲].

^۳ Entity

^۴ Abstraction

^۱ Trajectory

^۲ Semantic Enrichment of Trajectories

باشند. در این مدل، کاندیداهای توقف توسط کاربر به الگوریتم معرفی می‌شود و احتمال اینکه توقفی از دید کاربر دور بماند وجود دارد.

Rocha و همکاران [۱۵]، شناسایی مناطق توقف و حرکت در خط سیر و غنی‌سازی معنایی را بر اساس تغییر جهت نقاط تعریف کردند. خوشه‌هایی از نقاط که تغییر جهت بیشتری نسبت به نقاط دیگر داشته باشند، مناطق مهم خط سیر هستند. در روش پیشنهادی که روی داده‌های خط سیر کشتی‌های ماهیگیری (به عنوان شی متحرک) پیاده‌سازی شد، هدف یافتن مناطقی است که فعالیت‌های ماهیگیری در آن بیشتر است.

Alvares و همکاران [۱۶]، گرچه مدلی معنایی ارائه ندادند، اما چارچوبی برای افزودن اتوماتیک معنا به داده‌های خط سیر پیشنهاد کردند. چارچوب ارائه شده، فرآیند کشف دانش از داده‌های خط سیر را تحت پوشش قرار می‌دهد و شامل سه سطح است: انبار داده، آماده‌سازی داده و داده‌کاوی. در چارچوب پیشنهادی، از نرم‌افزار Weka به عنوان ابزار داده‌کاوی و PostGIS به عنوان انبار داده استفاده شد. در این کار برای اضافه کردن معنا به داده‌های خط سیر، در Weka، واسط اتصال به پایگاه داده توسعه یافت. در فرآیند اضافه کردن معنا به داده‌های خط سیر از مدل معنایی ارائه شده در مدل SMoT استفاده شده است.

Yan و همکاران [۱۲] سیستم SeMiTri را در قالب معماری سه لایه پیشنهاد دادند. لایه اول، محاسبه خط سیر که شامل پاکسازی داده از خطا و شناسایی اپیزودهای خط سیر است. لایه دوم شامل تکنیک‌هایی برای غنی‌سازی داده‌های خط سیر است. لایه آخر سیستم نیز لایه کاربرد است که واسط بین سیستم و کاربر است و شامل تکنیک‌هایی برای داده‌کاوی خطوط سیر غنی‌سازی شده و بصری‌سازی نتایج است.

Cammosi و همکاران [۱۱] فرآیند کشف الگوهای معنایی از خط سیرها را براساس آنتولوژی رویدادهای معنایی انجام دادند. روش ارائه شده می‌تواند در کاربردهایی که رویدادهای معنایی می‌تواند به صراحت بر اساس توقف‌ها و حرکت‌ها توصیف شود، مورد استفاده قرار گیرد. روش ارائه شده در کشف الگوهای مشکوک حمل-ونقل کانتینرهای دریایی مورد آزمایش قرار گرفته است. در این روش، یک پایگاه دانش برای کانتینرهای دریایی

از آنجا که در فرآیند غنی‌سازی خط سیر، هدف پردازش صرف مسیرهای جغرافیایی برای آنالیزهای مرسوم در سیستم‌های اطلاعات مکانی نیست، هر یک از مراحل کار در فرآیند غنی‌سازی داده‌ها با رویکردی متفاوت در مقایسه با خط سیر هندسی به تحلیل داده‌ها می‌پردازد. این مقاله ضمن ارائه و تحلیل چارچوب غنی‌سازی معنایی به ارائه و پیاده‌سازی روش‌های مؤثر در پاکسازی داده‌های خام و شناسایی اپیزودها با حفظ خصوصیات یک خط سیر برای ورود به مرحله غنی‌سازی معنایی می‌پردازد. ساختار کلی مقاله به این شکل است: بخش ۲ به بررسی و مرور مطالعات انجام شده در حوزه خط سیر معنایی می‌پردازد، بخش ۳ به معرفی چارچوب محاسبه خط سیر برای غنی‌سازی معنایی اختصاص یافته است، در بخش ۴ پیاده سازی و ارزیابی نتایج روی داده‌های واقعی ارائه می‌گردد و بخش آخر نیز به نتیجه‌گیری و پیشنهادات می‌پردازد.

۲- مروری بر مطالعات انجام شده

اولین تلاش در زمینه خط سیر معنایی، ارائه مدل مفهومی توقف - حرکت^۱ بود [۱۳]. در مدل ارائه شده، توقف مکان‌های مهم یک خط سیر از نقطه نظر کاربرد هستند که شی متحرک برای یک بازه زمانی مشخص در آن بخش متوقف است، در حالیکه حرکت‌ها زیررشته‌های بین توقف‌های متوالی هستند. در این مدل، خط سیر معنایی توسط یک رشته از توقف‌ها و حرکت‌ها نمایش می‌یابد و اطلاعات معنایی مربوط به خط سیر مجموعه‌ای از مکان‌هایی است که شی متحرک در آن برای یک حداقل بازه زمانی متوقف باقی مانده است.

با ارائه مدل توقف - حرکت، بسیاری از مطالعات بر همین اساس توسعه یافته است. مدل SMoT^۲ [۱۴] یک الگوریتم تقاطع‌مبنا برای غنی‌سازی معنایی خط سیر است. در این روش، مناطق توقف بر اساس تقاطع خطوط سیر مکانی با مناطق جغرافیایی از پیش تعریف شده توسط کاربر (به عنوان کاندیدای توقف) شناسایی می‌شوند. تعیین مناطق کاندیدای توقف بر اساس کاربرد انجام می‌شود. به عنوان مثال، در کاربرد گردشگری، کاندیداهای توقف می‌تواند فرودگاه، هتل، اماکن تفریحی و گردشگری

^۱ Stop & Move

^۲ Stop & Move of Trajectories

تعریف و براساس آن الگوهای غیرعادی توصیف کننده حرکات و رفتار مشکوک می‌تواند مدلسازی شود.

Bogorny و همکاران در [۲] به ارائه مدل مفهومی CONSTAnT در غنی‌سازی خط سیر پرداختند. مدل ارائه شده در چارچوب پروژه SEEK^۱ انجام شد که تمرکزش روی مدلسازی، نمایش، آنالیز و کشف دانش از داده‌های خط سیر بود. در مدل مفهومی ارائه شده، فرض این بود که خط سیر معنایی مجموعه‌ای از زیررشته‌های معنایی است که با توجه به دامنه کاربردی موردنظر تعیین می‌شوند. زیررشته‌های معنایی می‌توانند بر اساس هدف شیء متحرک، ابزار حمل و نقل و یا الگوی حرکت باشند. در مدل ارائه شده، استنتاج جنبه‌های معنایی مختلف به راحتی از روی داده‌ها امکانپذیر نیست و بر اساس الگوریتم‌های داده‌کاوی انجام می‌شود.

Ritcher و همکاران [۱۷]، فرآیند غنی‌سازی معنایی را با تمرکز بر خط سیرهای مقید شبکه‌ای پیاده‌سازی کردند. در این پژوهش از یک مدل گراف برای نمایش شبکه راه‌های شهری استفاده شده است. در روش ارائه شده، الگوریتم STC به شناسایی نقاط مرجع در خط سیر می‌پردازد. فرآیند شناسایی نقاط مرجع بر اساس الگوریتم‌های تطبیق نقشه انجام می‌شود. سپس الگوریتم برای همه نقاط شناسایی شده تمام توصیفات ممکن از نحوه ادامه حرکت از هر نقطه را شناسایی می‌کند. در نهایت بر اساس توصیفات شناسایی شده، نقاط مرجع همسان متوالی ترکیب شده و فرآیند اضافه کردن معنا به داده‌ها انجام می‌شود. دقت روش ارائه شده به دقت الگوریتم تطبیق نقشه پیاده‌سازی شده بستگی دارد.

در تحقیقات انجام شده در حوزه غنی‌سازی معنایی خطوط سیر مکانی، اغلب مطالعات جنبه خاصی از کاربرد را در نظر می‌گیرند و با تمرکز بر اضافه کردن معنا به داده‌ها بوده‌اند [۱۵، ۱۷، ۶، ۱۱] و بخش دیگر نیز تقریباً مستقل از کاربرد به پیاده‌سازی الگوریتم‌ها روی داده‌های خط سیر اشیاء مختلف می‌پردازند [۱۲، ۲، ۶]. این در حالی است که مطالعات کمی پیرامون ارائه یک چارچوب جامع، که فرآیند غنی‌سازی معنایی را بررسی کند صورت گرفته است. در واقع اضافه کردن معنا به داده‌های حرکت، یک فرآیند پردازشی است که در چند گام انجام می‌شود. در

این فرآیند پردازشی داده‌های خام GPS وارد فرآیند محاسباتی شده و خطوط سیر غنی شده با اطلاعات معنایی به عنوان خروجی در اختیار کاربردهای مختلف قرار می‌گیرد. در این مقاله سعی بر این است تا مراحل آماده‌سازی داده‌ها از ورود داده‌های خام حرکت تا تولید خطوط سیر ساختاریافته مورد تحلیل قرار گرفته و در هر گام یک روش مؤثر پیاده‌سازی و ارزیابی شود.

۳- محاسبه خط سیر معنایی

در فرآیند تبدیل داده‌های جمع‌آوری شده با GPS به خطوط سیر غنی‌شده با معنا چندین مفهوم کاربردی وجود دارد. در این بخش ضمن معرفی این مفاهیم، چارچوب غنی‌سازی معنایی بررسی می‌شود.

۳-۱- تعاریف

تعریف ۱- خط سیر مکانی (خام، هندسی): یک خط سیر مکانی، معمولاً به شکل لیست منظمی از نقاط مکانی و زمانی به شکل $\langle P_0, P_1, \dots, P_N \rangle$ نمایش داده می‌شود که در آن، $P_i = \langle x_i, y_i, t_i \rangle$ که $x_i, y_i \in \mathbb{R}$ مختصات نقاط در یک سیستم مختصات تعریف شده و $t_i \in \mathbb{R}^t$ و $i = 0, 1, \dots, N$ مولفه زمان این لیست بوده و شرط $t_1 < t_2 < \dots < t_n$ بر مولفه‌های زمانی حاکم است.

تعریف ۲- توقف: یک توقف S از خط سیر T به صورت چندتایی $\langle Tid, Sid, ST, t_{start}, t_{end} \rangle$ تعریف می‌شود که در آن، Tid شناسه خط سیر، Sid شناسه توقف، ST زیررشته شناسایی شده از خط سیر به عنوان توقف، t_{start}, t_{end} به ترتیب زمان شروع و زمان پایان توقف در خط سیر است که در آن $|t_{end} - t_{start}| \geq \Delta_{threshold}$ و حداقل بازه زمانی تعریف شده بر اساس کاربرد موردنظر است برای اینکه این بخش از خط سیر با زمان شروع t_{start} و زمان پایان t_{end} به عنوان توقف در نظر گرفته شود.

تعریف ۳- حرکت: حرکت M در خط سیر T، یک چندتایی $\langle Tid, Mid, MT, t_{start}, t_{end} \rangle$ است که Tid شناسه خط سیر، Mid شناسه حرکت، MT زیررشته شناسایی شده به عنوان حرکت، t_{start}, t_{end} به ترتیب زمان شروع و پایان حرکت در خط سیر است. به عبارت دیگر، هر نقطه از خط سیر که توقف نباشد، حرکت است.

^۱ SEmantic Enrichment of trajectory Knowledge discovery

با توجه به اینکه فرآیند غنی سازی معنایی خط سیر در این مقاله از پیش پردازش داده ها تا شناسایی مناطق توقف در خط سیر را دربرمی گیرد، در ادامه، این مراحل مورد تحلیل و بررسی قرار می گیرد.

۳-۲-۱- پیش پردازش داده ها

فرآیند پیش پردازش داده ها با تمرکز بر جنبه هندسی داده های خام GPS است و معمولاً شامل تکنیک هایی برای پاکسازی داده های خط سیر از خطا و فشرده سازی است. مجموعه داده های جمع آوری شده توسط دستگاه های تعیین موقعیت اغلب غیردقیق و به دلیل وجود داده های پرت می تواند نادرست هم باشد [۱۹]. این خطاها به دلیل محدودیت در تکنیک های موقعیت یابی (به عنوان مثال اندازه گیری های نادرست GPS و یا از دست رفتن سیگنال در فضای بسته) و یا جنبه های مربوط به حریم خصوصی کاربر است [۲۰]. تمرکز اصلی تکنیک های پاکسازی داده های خط سیر، حذف یا کاهش این خطاها است. خطاهای GPS معمولاً خطاهای مربوط به بعد مکانی نقاط خط سیر را در نظر می گیرد و اطلاعات مربوط به زمان، دقیق فرض می شود که به دلیل کالیبراسیون دقیق ساعت هایی است که ماهواره ها به آن مجهز هستند [۲۱]. فرآیند کشف دانش از داده های خط سیر معنایی، علاوه بر خصوصیات هندسی خط سیر، با تکیه بر اطلاعات معنایی خطوط سیر نیز انجام می شود. از این رو شناسایی و حذف داده های پرت مهمترین گام در پاکسازی داده های خط سیر است. برای این کار، یک راه شناسایی داده های پرت، به صورت بصری و حذف داده های خطا دار است. واضح است که این روش برای مجموعه داده های بزرگ امکان پذیر نیست [۲۱]. معمولاً یکی از روش ها استفاده از تکنیک های فیلتر کردن داده هاست. برای این کار، باید هر دو بعد مکانی و زمانی نقاط در نظر گرفته شود. یک راه، استفاده از سری زمانی سرعت نقاط است. به این ترتیب هر خط سیر به صورت یک لیست منظم از سرعت نقاط نمایش می یابد که سرعت لحظه ای هر نقطه به صورت تقریبی با کمک نقطه قبل و بعد محاسبه می شود:

$$T = \langle V_1, V_2, \dots, V_N \rangle \quad (1)$$

$$V_i = \sqrt{\left(\frac{x_{i+1} - x_{i-1}}{t_{i+1} - t_{i-1}}\right)^2 + \left(\frac{y_{i+1} - y_{i-1}}{t_{i+1} - t_{i-1}}\right)^2} \quad (2)$$

$i = 2, \dots, n - 1$

تعریف ۴- اپیزود: اپیزود e در خط سیر T بخش های همگن در خط سیرهای مکانی هستند. شناسایی اپیزودها در خطوط سیر مکانی بر اساس مدل مفهومی استفاده شده در فرآیند غنی سازی معنایی انجام می شود. مثلاً در مدل مفهومی توقف-حرکت، اپیزودهای خط سیر توقف ها و حرکت های شناسایی شده هستند.

تعریف ۵- خط سیر ساختاریافته T' : خط سیر ساختاریافته T' به صورت چندتایی $\langle e_0, e_1, \dots, e_N \rangle$ تعریف می شود که در آن، e_i اپیزودهای شناسایی شده در فرآیند شناسایی بخش های همگن خط سیر هستند.

تعریف ۶- خط سیر معنایی T'' : خط سیر معنایی T'' به صورت رشته $\langle Q_0, Q_1, \dots, Q_N \rangle$ تعریف می شود که در آن، $Q_i = \langle x_i, y_i, t_i, A_i \rangle$ معانی اضافه شده به خط سیر براساس کاربرد موردنظر هستند. سه منبع اطلاعاتی می تواند برای افزودن معنا به داده های خام حرکت استفاده شود: (۱) معنا از داده های خام حرکت استخراج شود (مثلاً فاصله از نقطه شروع)؛ (۲) از داده های ضمنی به دست آید (مثلاً داده های مربوط به آب و هوا)؛ (۳) با استنتاج از داده ها حاصل شوند (مثلاً حالات مختلف حمل و نقل می تواند از داده های سرعت و شتاب استنتاج شود) [۱۸].

غنی سازی معنایی خطوط سیر مکانی را می توان به عنوان یک فرآیند محاسباتی در نظر گرفت که داده های خام حرکت را به داده های خط سیر غنی شده با اطلاعات معنایی تبدیل می کند. بر همین اساس این فرآیند محاسباتی در چندین گام پیاده سازی می شود. ادامه این بخش به ارائه این مراحل پردازشی می پردازد.

۳-۲- چارچوب محاسباتی خط سیر معنایی

فرآیند غنی سازی معنایی خط سیرهای مکانی معمولاً شامل چندین گام پردازش است. در این روند، داده های خام جمع آوری شده توسط دستگاه تعیین موقعیت (مثل GPS) وارد فرآیند محاسباتی شده و خط سیر غنی شده معنایی در اختیار کاربردهای مختلف قرار می گیرد. شکل ۲ فرآیند غنی سازی معنایی خط سیر را در چهارگام محاسباتی نشان می دهد:

(۱) پیش پردازش داده های خام GPS.

(۲) تقسیم بندی داده ها،

(۳) شناسایی اپیزودهای حرکت،

(۴) غنی سازی معنایی

در رابطه (۱)، خط سیر به صورت لیستی از مولفه های سرعت لحظه ای V_i است که هر مولفه به صورت تقریبی با رابطه (۲) به دست می آید. سرعت هر نقطه از خط سیر با استفاده از نقطه قبل، $P_{i-1} = (x_{i-1}, y_{i-1}, t_{i-1})$ و نقطه بعد $P_{i+1} = (x_{i+1}, y_{i+1}, t_{i+1})$ محاسبه می شود.



شکل ۲- فرآیند غنی سازی معنایی خط سیر

با محاسبه سرعت لحظه ای نقاط، از قانون ۳-سیگما برای حذف داده های پرت استفاده شده است. اما تعریف حد آستانه سرعت صرفاً بر اساس قانون ۳-سیگما ممکن است باعث خطا در شناسایی شود، که این موضوع به دلیل ناهمگن بودن داده ها در تمام خطوط سیر است. بر اساس راه حل پیشنهادی، فرآیند شناسایی و حذف داده های پرت یک فرآیند دو مرحله ای است: در گام نخست، نقاطی که در بازه تعریف شده بر اساس قانون ۳-سیگما باشند به عنوان نقطه اصلی خط سیر در نظر گرفته می شوند و نقاط خارج از بازه به عنوان کاندیدای نقاط پرت ذخیره سازی می شوند. پس از شناسایی کاندیداهای نقاط پرت، الگوریتم به گروه بندی نقاط پرت متوالی شناسایی شده می پردازد. با محاسبه بازه زمانی هر یک از گروه های کاندیدا، هر یک از گروه ها ارزیابی می شوند. اگر بازه زمانی نقاط متوالی شناسایی شده از حد آستانه زمانی تعریف شده بیشتر باشد این نقاط به مجموعه داده اصلی خط سیر بازگردانده می شوند. به عنوان مثال این

داده ها ممکن است مربوط به بخشی از خط سیر باشند که شی متحرک وسیله نقلیه خود را تغییر داده است.

علاوه بر پاکسازی داده ها از خطا، گام دیگر در پیش پردازش، حذف داده های اضافی از مجموعه داده های خط سیر است. در واقع، به دلیل محدودیت در سازوکار جمع آوری و ذخیره سازی داده ها، حرکت پیوسته اشیاء در حال حرکت، با نمونه های گسسته از موقعیت مکانی-زمانی نقاط نمایش می یابد. واضح است که هر چه تعداد نقاط نمونه بیشتر باشد، دقت نمایش خط سیر نیز بیشتر است. با وجود این، با بیشتر شدن تعداد نقاط، حجم پردازش نیز افزایش می یابد. از این رو، کاربرد تکنیک های کاهش داده خط سیر به منظور مدیریت مؤثر داده ها ضروری است [۲۲].

اغلب تکنیک های ارائه شده در فشرده سازی، بر اساس ساده سازی خطی هستند و از توابع فاصله برای تقریب خط سیر استفاده می کنند. شکل ۳ نمونه ای از توابع فاصله استاندارد و ارتباط آنها با هم را نشان می دهد. تابع E_2 فاصله در الگوریتم داگلاس-پوکر است که در ساده سازی خطی تنها بعد مکانی نقاط را در نظر می گیرد. در تابع فاصله E_3 زمان به صورت بعد سوم مکانی در نظر گرفته می شود. در تابع فاصله E_u زمان به صورت متفاوت عمل می کند و هنگام محاسبه فاصله با درونیایی خطی، از بعد زمان نیز برای محاسبه فاصله استفاده می شود. در اصطلاح این تابع، تابع فاصله اقلیدسی همزمان^۱ نیز نام دارد. تابع E_t نیز مشابه تابع E_u است اما به جای تعیین اختلاف مکانی نقاط همزمان، از اختلاف زمان نقاط نزدیک به هم استفاده می کند [۲۳]. واضح است که استفاده از توابع E_2 و E_t برای کاهش داده خط سیر کافی نیست، چون به ترتیب بعد زمانی و مکانی خط سیر نادیده گرفته می شود. Cao و همکاران [۲۳] یک تابع خطا معرفی کرده اند که بر اساس ترکیب E_t و E_u است و از روش ساده در نظر گرفتن بزرگترین مقدار استفاده می کند. با این حال کاملاً مشخص نیست که اختلاف در مقیاس دو تابع خطا چگونه باید در نظر گرفته شود.

در فنون فشرده سازی خطوط سیر معنایی، علاوه بر بعد زمانی و مکانی خط سیر، توجه به یک موضوع دیگر نیز ضروری است و آن حفظ مناطق توقف و حرکت خط سیر همزمان با کاهش تعداد نقاط خط سیر است. یک راه در نظر گرفتن تابع فاصله بر اساس سرعت نقاط است. در

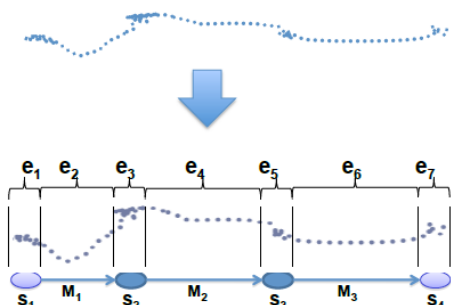
^۱ Synchronized Euclidean distance

۳-۲-۲- تقسیم‌بندی خط سیر

فرآیند تقسیم‌بندی خط سیر از داده‌های خام پاکسازی شده استفاده می‌کند و خطوط سیر مکانی - زمانی بدون همپوشانی تولید می‌کند. ایده اصلی در این بخش، معرفی روش‌هایی برای شناسایی نقاط تقسیم است که نقاط خام خط سیر را به خط سیرهای متوالی تبدیل می‌کند [۲۲]. شناسایی خطوط سیر می‌تواند بر اساس شکاف مکانی^۱ (بیشتر از حد آستانه تعریف شده)، فاصله زمانی از پیش تعریف شده (مثل خط سیرهای ساعتی، روزانه، هفتگی، ماهانه و...) و یا نقاط مکانی از پیش تعریف شده (فاصله ثابت، فاصله بین نقاط از پیش تعریف شده) باشد [۹]. انتخاب روش مناسب برای شناسایی خطوط سیر مکانی- زمانی بستگی به کاربرد موردنظر و ویژگی‌های داده دارد. به عنوان مثال زمانی که داده‌های خط سیر با استفاده از گوشی‌های هوشمند توسط افراد جمع‌آوری شده باشد و هدف غنی‌سازی معنایی خطوط سیر برای برنامه‌ریزی شهری است، خطوط سیر مکانی-زمانی می‌تواند به صورت روزانه شناسایی شوند. در کاربرد دیگر، داده‌های جمع‌آوری شده توسط تاکسی‌ها می‌تواند از نقاط از پیش تعریف شده مکانی به عنوان نقاط تقسیم استفاده کند.

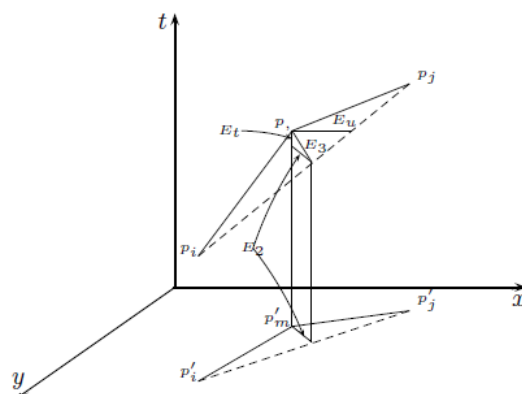
۳-۲-۳- شناسایی مناطق توقف

بعد از شناسایی خطوط سیر مکانی-زمانی، فرآیند شناسایی مناطق توقف و حرکت روی خطوط سیر شناسایی شده انجام می‌شود، که خروجی آن خط سیرهای ساختاریافته به شکل دنباله‌ای از اپیزودهاست. به عبارت دیگر، ایده اصلی در این مرحله، گروه‌بندی نقاط همگن در هر خط سیر است. شکل ۴ خط سیر مکانی-زمانی و خط سیر ساختاریافته متناظر با آن را نشان می‌دهد.



شکل ۴- خط سیر مکانی-زمانی و خط سیر ساختاریافته به شکل دنباله‌ای از اپیزودها [۱۲]

واقع ایده اصلی این است که رفتار توقف و حرکت نقاط بیشتر در مشتق آن (سرعت) ظاهر می‌شود. بر همین اساس در الگوریتم ارائه شده برای کاهش داده، خط سیر به صورت رشته‌ای از سرعت نقاط در نظر گرفته می‌شود و تابع فاصله سرعت بر اساس رابطه (۳) تعریف می‌شود.



شکل ۳- رابطه بین توابع مختلف فاصله [۲۳]

$$E_V(\langle P_i \rangle, \langle P_1 \rangle, \langle P_N \rangle) = \sqrt{(v_i - v'_i)^2},$$

$$P_i = \langle x_i, y_i, v_i, t_i \rangle, \quad (3)$$

$$v'_i = v_s + a_{se}(t_i - t_s);$$

$$a_{se} = (v_e - v_s)/(t_e - t_s) \quad (4)$$

در رابطه (۳)، E_V تابع فاصله بر اساس سرعت است. در تابع فاصله E_V هدف یافتن نقطه همزمان متناظر با نقطه P_i یعنی $P'_i = \langle x'_i, y'_i, v'_i, t_i \rangle$ است. تابع فاصله سرعت از فرض ثابت بودن شتاب هنگام درونیابی استفاده می‌کند. رابطه (۴) درونیابی بر اساس سرعت با فرض ثابت بودن شتاب را نشان می‌دهد. در اینجا، v'_i سرعت نقطه متناظر روی خط سیر تقریبی، v_s و v_e به ترتیب سرعت نقاط ابتدا و انتهای خط سیر و t_s و t_e به ترتیب، زمان ابتدا و انتهای خط سیر است.

برای کاهش تعداد نقاط از ترکیب توابع فاصله سرعت و فاصله اقلیدسی قائم استفاده شده است. با این حال، دو تابع از دو فرض متفاوت در درونیابی برای تقریب خط سیر استفاده می‌کنند. تابع E_2 از فرض ثابت بودن سرعت هنگام درونیابی استفاده می‌کند و از سوی دیگر، تابع E_V بر اساس فرض شتاب ثابت به درونیابی می‌پردازد. بنابراین برای استفاده از ترکیب دو تابع فاصله، تابع E_2 نیز به جای ثابت بودن سرعت از ثابت بودن شتاب استفاده می‌کند. پس خط واصل دو نقطه در درونیابی در فضای مکانی-زمانی تبدیل به سهمی درجه دو (با شتاب ثابت) می‌شود.

^۱ Spatial Gap

شناسایی مناطق توقف در خطوط سیر مکانی بر اساس مدل توقف-حرکت به دو شکل می‌تواند انجام شود [۲۴]:
(۱) فرآیند شناسایی صرفاً بر اساس داده‌های خام خط سیر باشد؛

(۲) در شناسایی مناطق توقف و حرکت از ترکیب داده‌های خام همراه با اطلاعات جغرافیایی و اطلاعات مرتبط با کاربرد استفاده شود.

اطلاعات خام خط سیر می‌تواند سرعت، شتاب، جهت و اطلاعات مربوط به موقعیت نقاط مثل تراکم باشد. اطلاعات جغرافیایی و کاربردی نیز شامل اطلاعات مربوط به محیطی که شیء متحرک در آن حرکت می‌کند است. به عنوان مثال، در کاربرد گردشگری، مکان‌های توقف می‌توانند هتل، موزه، اماکن تاریخی و گردشگری باشد.

روش مناسب برای شناسایی مناطق توقف بر اساس کاربرد موردنظر انتخاب می‌شود. در این کار، بر اساس اطلاعات مربوط به زمینه شیء متحرک و محیطی که شیء متحرک در آن حرکت می‌کند از سرعت شیء متحرک برای شناسایی مناطق توقف استفاده شده است. به این ترتیب، حد آستانه سرعت به صورت (۵) تعریف می‌شود [۱۲]:

$$\Delta_v = \min\{\delta_1 \times \overline{objAvgVlc}, \delta_2 \times \overline{PsnAvgVlc}\} \quad (5)$$

در این رابطه $\overline{objAvgVlc}$ سرعت میانگین شیء متحرک در هر خط سیر است و $\overline{PsnAvgVlc}$ نیز سرعت بر اساس محیطی است که شیء متحرک در آن حرکت می‌کند. در این کار با توجه به قید مکانی داده‌های خط سیر، یعنی خطوط سیر اتوبوس در شبکه راه شهری، $\overline{PsnAvgVlc}$ میانگین سرعت حرکت وسایط نقلیه در شبکه راه تعریف شده است. با تعریف حد آستانه سرعت، بخش‌هایی از خط سیر که سرعت بیشتری از حد آستانه تعریف شده دارند، به عنوان مناطق توقف در خط سیر شناسایی می‌شوند. در این تعریف، مقدار حداقل زمان توقف نیز به الگوریتم شناسایی مناطق توقف و حرکت معرفی می‌شود تا از ثبت توقف‌های اشتباه جلوگیری شود. به این ترتیب، مناطقی از خط سیر که بازه زمانی آن‌ها کمتر از حد آستانه تعریف شده باشد، مناطق حرکت در خط سیر خواهند بود.

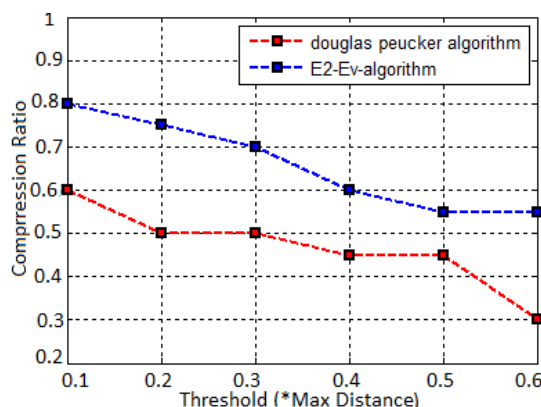
۴- پیاده سازی و ارزیابی نتایج

به منظور ارزیابی کارایی روند پیشنهادی، الگوریتم پیشنهادی بر روی مجموعه‌ای از داده‌های خط سیر پیاده

سازی و سپس نتایج حاصل مورد ارزیابی قرار گرفت. مجموعه داده مورد استفاده، از مجموعه داده GeoLife انتخاب شده است [۲۸، ۲۷، ۲۶]. مجموعه داده GeoLife شامل ۱۷۶۲۱ خط سیر است که توسط ۱۸۲ کاربر در طول ۵ سال (از آوریل ۲۰۰۷ تا آگوست ۲۰۱۲) در محدوده شهر پکن جمع‌آوری شده است. مجموعه داده انتخابی، شامل ۹ خط سیر است. در این مجموعه داده، بزرگترین خط سیر طولی حدود ۳ کیلومتر و کوتاهترین خط سیر طولی حدود ۵۰۰ متر دارد.

فرآیند پیش‌پردازش شامل حذف خطا از داده‌های خط سیر و کاهش تعداد نقاط است. در فرآیند شناسایی و حذف خطا، از مجموعه داده‌های خط سیر باید هر دو بعد مکانی و زمانی نقاط در نظر گرفته شود. برای این کار، ابتدا از روابط (۱) و (۲)، برای شناسایی و حذف داده‌های پرت استفاده شده است. برای این کار، داده‌های پرت از هر خط سیر بر اساس حد آستانه تعریف شده شناسایی و به طور موقت حذف می‌شوند، سپس داده‌های شناسایی شده بر اساس حد آستانه زمانی تعریف شده ارزیابی می‌شوند. شکل ۵ نسبت داده‌های پرت به نقاط خط سیر را قبل و بعد از بازیابی داده‌ها نشان می‌دهد. در این شکل محور افقی بیانگر شماره خط سیر و محور قائم نسبت تعداد نقاط پرت شناسایی شده به تعداد کل نقاط است. همانطور که مشاهده می‌شود نسبت تعداد نقاط پرت به تعداد کل نقاط در خط سیر ششم بعد از ارزیابی داده‌های پرت شناسایی شده به شدت کاهش یافته است. علت این مسئله را می‌توان در نمودار سرعت نقاط هنگامی که نقاط پرت اشتباه به داده‌ها بازگردانده می‌شود، مشاهده کرد. شکل ۶ نمودار سرعت نقاط خط سیر را قبل و بعد از شناسایی و حذف داده‌های پرت نشان می‌دهد. در این شکل محور افقی بیانگر شماره نقاط در خط سیر ششم و محور قائم نمایانگر سرعت در هر نقطه است. در شکل ۶ (الف) نقاط در محدوده شماره ۳۰۰ تا ۴۰۰ به عنوان داده پرت شناسایی شده و از داده‌ها حذف شده‌اند. شکل ۶ (ب) نمودار سرعت داده‌ها را بعد از ارزیابی نقاط شناسایی شده به عنوان داده‌های پرت، نشان می‌دهد. معمولاً این بخش از خط سیر مربوط به زمانی است که شیء متحرک وسیله نقلیه خود را تغییر داده است. صحت این موضوع را می‌توان در مجموعه اطلاعات جمع‌آوری شده توسط کاربر نیز مشاهده کرد. در این مجموعه داده علاوه بر داده‌های موقعیت خط سیر، حالات مختلف حرکت نیز توسط کاربر جمع‌آوری شده است. بر

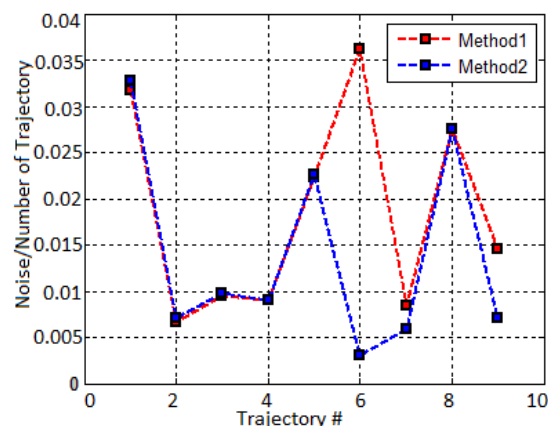
پس از شناسایی و حذف داده‌های پرت از نقاط خام GPS، فرآیند فشرده‌سازی روی داده‌ها انجام می‌شود. در شکل ۷، نرخ فشرده‌سازی الگوریتم E_v-E_2 با روش داگلاس-پوکر در حد آستانه‌های مختلف مقایسه شده است. همانطور که دیده می‌شود، نرخ فشرده‌سازی الگوریتم پیاده‌سازی شده نسبت به الگوریتم داگلاس-پوکر بالاتر است، که البته این موضوع به دلیل حفظ نقاط بیشتر خط سیر است. بنابراین حفظ نقاط توقف به بهای افزایش نرخ فشرده‌سازی است. با ارزیابی نتایج مربوط به پیاده‌سازی الگوریتم پیشنهادی E_v-E_2 مشاهده شد که مناطق توقف در خط سیر نیز که عمدتاً به شکل مناطق متراکم در خط سیر نمایان می‌شوند در مجموعه داده خط سیر به عنوان نقاط مهم حفظ شده‌اند. این در حالی است که الگوریتم‌های کاهش داده که بر اساس ساده‌سازی خطی هستند، قادر به حفظ این نقاط در خط سیر نبوده‌اند.



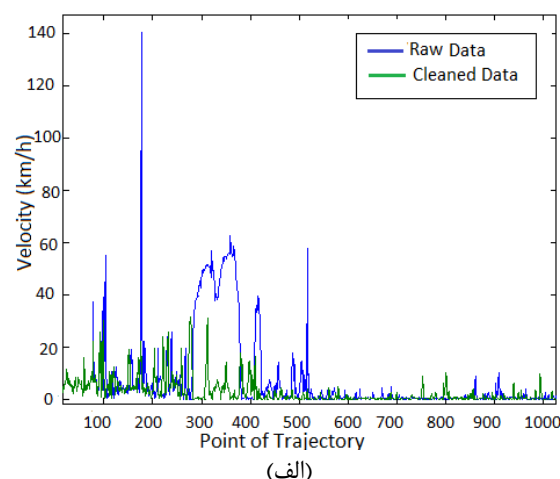
شکل ۷- مقایسه نرخ فشرده الگوریتم داگلاس-پوکر با الگوریتم E_2-E_v

بعد از پیش‌پردازش داده‌های خام حرکت، فرآیند شناسایی خطوط سیر مکانی-زمانی روی داده‌های پاکسازی شده انجام می‌شود. بر اساس کاربرد، که با تمرکز بر مدیریت حمل و نقل و ترافیک شهری است، فرآیند تقسیم‌بندی خطوط سیر به صورت خطوط سیر روزانه انجام شده است. در خطوط سیر شناسایی شده، کمینه نرخ نمونه‌برداری در داده‌های جمع‌آوری شده ۱ ثانیه، میانگین نمونه‌برداری حدود ۱۵ ثانیه و بیشینه شکاف زمانی بین دو نقطه از خط سیرهای شناسایی شده حدود ۱۲ ساعت است. این شکاف زمانی مربوط به زمانی است که کاربر GPS خود را خاموش کرده است. جدول ۱ مشخصات خطوط سیر روزانه شناسایی شده را نشان می‌دهد.

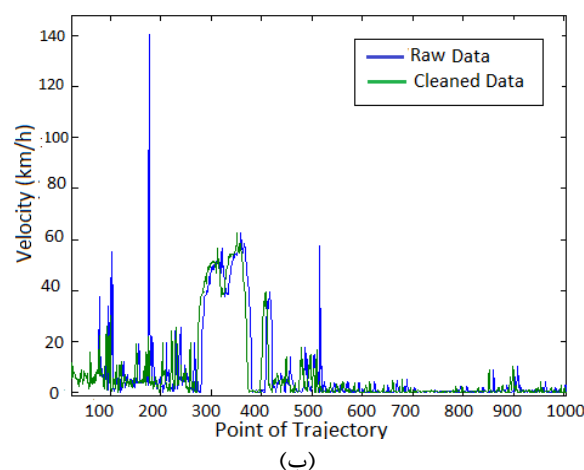
اساس این اطلاعات جمع‌آوری شده در خط سیر ششم در محدوده نقاط ۳۰۰ تا ۴۰۰ حالت حرکت کاربر از پیاده‌روی به حرکت با تاکسی تغییر کرده است.



شکل ۵- نسبت تعداد نقاط پرت به تعداد کل نقاط: روش ۱ با قانون ۳-سیگما و روش ۲ الگوریتم پیشنهادی



(الف)



(ب)

شکل ۶- نمودار سرعت نقاط در خط سیر ششم، (الف) حذف داده‌های پرت با قانون ۳-سیگما (ب) حذف داده‌های پرت با الگوریتم پیشنهادی

جدول ۱- خصوصیات خطوط سیر روزانه شناسایی شده

بیشینه شکاف زمانی (S)	میانگین نرخ نمونه برداری (S)	شناسه خط سیر
۸۷۳	۹/۴۶	۲۰۱۱۰۸۰۷
۱۱۲۳	۶/۱۲	۲۰۱۱۰۸۰۸
۴۳۸۰۸	۵۵/۹۲	۲۰۱۱۰۸۰۹
۲۹۹۸	۷/۵۳	۲۰۱۱۰۸۲۴
۴۷۹	۶/۳۴	۲۰۱۱۰۸۲۷
۲۶۰۴۴	۱۸/۱۷	۲۰۱۱۰۸۲۸
۱۶۰۲	۷/۱۸	۲۰۱۱۰۸۲۹
۱۰۰۳	۹/۴۱	۲۰۱۱۰۸۳۰
۶۶۶۲	۹/۴۱	۲۰۱۱۰۹۰۲

گام بعد در چارچوب محاسبه خط سیر معنایی، شناسایی اپیزودهای حرکت است. بر اساس کاربرد موردنظر، فرآیند شناسایی اپیزودهای توقف و حرکت بر اساس سرعت نقاط انجام شده است. برای شناسایی اپیزودها در داده‌های خط سیر، سرعت لحظه‌ای هر یک از نقاط محاسبه شده است. سپس با توجه به حد آستانه تعریف شده، هر یک از نقاط خط سیر بر اساس سرعت محاسبه شده با S و M برچسب گذاری می‌شوند. در این مرحله، S برچسب نقاطی از خط سیر است که ممکن است مناطق توقف در خط سیر باشند. در واقع این احتمال وجود دارد که این نقطه بعد از یکپارچه سازی توقف‌های متوالی و تولید منطقه توقف دارای حداقل حد آستانه زمانی لازم نباشد. M نیز برچسب نقاط حرکت است. سپس نقاط توقف و حرکت با یکپارچه سازی و ادغام توقف‌ها و حرکت‌های متوالی با اعمال شرط کمترین زمان توقف شناسایی شده‌اند. بر اساس کاربرد موردنظر حداقل زمان توقف در هر منطقه از خط سیر ۱۲۰ ثانیه تعریف شده است. مقدار δ نیز به صورت تجربی برای δ_1 و δ_2 برابر $\frac{0}{3}$ در نظر گرفته شده است. بر اساس تحلیل‌های انجام شده تغییر مقدار δ به مقادیر بیشتر از $\frac{0}{3}$ باعث شناسایی مناطق توقف با طول بیشتر می‌شود. جدول ۲ توقف‌های شناسایی شده را روی داده‌های خط سیر نشان می‌دهد.

جدول ۲- توقف‌های شناسایی شده خطوط سیر مکانی - زمانی

میانگین زمان توقف (S)	کمینه زمان توقف (S)	بیشینه زمان توقف (S)	تعداد مناطق توقف	شناسه خط سیر
۱۵۰	۱۵۰	۱۵۰	۱	۲۰۱۱۰۸۰۷
۶۰۷/۲	۲۵۸	۱۱۸۷	۵	۲۰۱۱۰۸۰۸
۱۶۶۳۵	۱۱۴۲	۴۳۸۳۳	۳	۲۰۱۱۰۸۰۹
۱۷۳۵/۷	۱۵۱	۶۰۷۶	۱۸	۲۰۱۱۰۸۲۴
۱۰۶۲/۵	۴۱۹	۱۷۰۶	۲	۲۰۱۱۰۸۲۷
۱۷۵۲۱	۴۵۱	۲۶۱۰۰	۳	۲۰۱۱۰۸۲۸
۱۳۳۵	۱۷۳	۷۹۴۷	۹	۲۰۱۱۰۸۲۹
۱۲۳۳/۱	۱۳۱	۱۰۴۴	۲	۲۰۱۱۰۸۳۰
۲۵۱۵۷	۱۶۳۲۶	۳۳۹۸۸	۲	۲۰۱۱۰۹۰۲

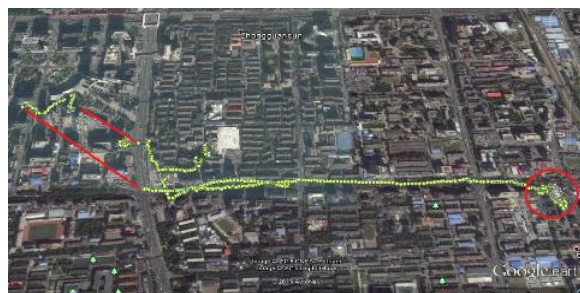
شکل ۸ چند نمونه از توقف‌های شناسایی شده در خطوط سیر را نشان می‌دهد. همانطور که دیده می‌شود مناطق توقف شناسایی شده توسط الگوریتم، در اطراف تقاطع‌ها و ساختمان‌ها به شکل تجمع نقاط و شکاف مکانی هستند. این موضوع را می‌توان به دلیل کاهش سرعت حرکت شی متحرک در این مناطق و ورود شی متحرک به فضای بسته و از دست رفتن سیگنال GPS دانست. این مناطق همان مکان‌های مهمی هستند که شی متحرک برای حداقل بازه زمانی تعریف شده در آن‌ها حضور داشته است. در شکل ۸ (الف) دو توقف به شکل شکاف مکانی در اطراف ساختمان و توقف دیگر در تقاطع شناسایی شده است. توقف در تقاطع می‌تواند بیانگر کاهش سرعت شی متحرک در پشت چراغ‌های راهنمایی باشد که در آن شی متحرک بیش از حداقل بازه زمانی تعریف شده در آن حضور داشته است. این روند را می‌توان در توقف‌های شناسایی شده در شکل ۸ (ب) و ۸ (ج) نیز مشاهده کرد. با شناسایی مناطق توقف و حرکت، خط سیر ساختار یافته به صورت لیستی از اپیزودها، آماده غنی سازی می‌شود. بر این اساس، اپیزودهای شناسایی شده را می‌توان در دو گروه تحلیل کرد. گروه اول اپیزودهایی هستند که بر اساس اطلاعات معنایی موجود به عنوان مناطق توقف غنی سازی می‌شوند. دسته دیگر مناطقی هستند که الگوریتم شناسایی توقف آن‌ها را به عنوان مناطق توقف شناسایی کرده، اما براساس اطلاعات معنایی موجود در هیچ دسته‌ای جای ندارند. این مناطق را می‌توان به عنوان مناطق توقف ناشناخته مورد تحلیل قرار داد تا الگوهای معنایی از این توقف‌های ناشناخته نیز کشف و استخراج شود.

۵- نتیجه گیری

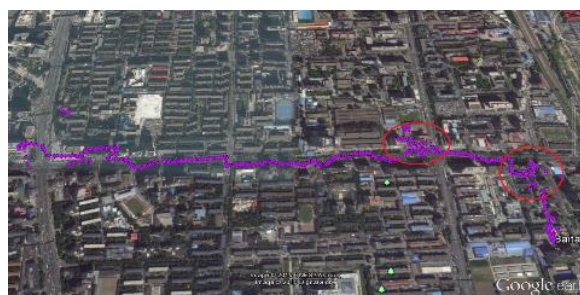
در مباحث مربوط به خط سیر، فرآیند اضافه کردن معنا به داده‌های خط سیر مکانی در چندین سال اخیر موردتوجه محققین این حوزه بوده است. اضافه کردن معنا که با هدف چکیده‌سازی حجم عظیم داده‌های خط سیر و انجام آنالیزهای معنایی متنوع انجام می‌شود، شامل چندین گام پردازش است تا داده‌های خام GPS به سطح چکیده سازی موردنیاز برای غنی‌سازی معنایی تبدیل شود. در این مقاله، با ارائه چارچوب غنی‌سازی معنایی، سه گام موردنیاز برای آماده‌سازی داده‌ها برای اضافه کردن معنا به داده‌های خط سیر مورد بررسی قرار گرفته است. در گام نخست چکیده‌سازی، الگوریتم پیاده‌سازی شده ضمن شناسایی و حذف داده‌های پرت، داده‌های شناسایی شده را ارزیابی و در صورت نیاز آن‌ها را حذف می‌کند. در فشرده‌سازی داده‌ها نیز که گام دیگر مرحله پاکسازی است، روش پیشنهادی ضمن شناسایی و حذف داده‌های اضافی، مناطق توقف و حرکت را که اجزای مهم خط سیر در فرآیند غنی‌سازی هستند حفظ کرده است. شناسایی خطوط سیر مکانی-زمانی از روی داده‌های پاکسازی شده گام بعدی چکیده‌سازی خط سیر است. در این مرحله بر اساس کاربرد موردنظر خطوط سیر روزانه شناسایی شده‌اند. خطوط سیر مکانی-زمانی در مرحله بعد که شامل شناسایی مناطق همگن در خط سیر است به صورت اپیزودهای توقف و حرکت شناسایی می‌شوند. در الگوریتم پیاده‌سازی شده بر اساس کاربرد مدیریت حمل و نقل، از روش سرعت-مبنا استفاده شده است. در این روش از اطلاعات زمینه شیء متحرک و محیطی که در آن حرکت می‌کند برای تعیین حدآستانه توقف استفاده شده است.

مراجع

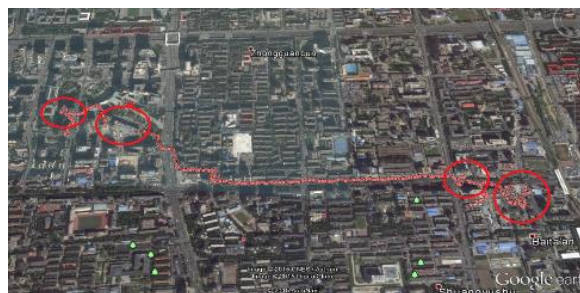
- [1] Z. Yan, "Towards semantic trajectory data analysis: a conceptual and computational approach," in VLDB, Lyon, France, 2009.
- [2] Bogorny, V., Renso, C., Aquino, A. R., Lucca Siqueira, F., and Alvares, L. O., "CONSTANT – A Conceptual Data Model for Semantic," Transactions in GIS, vol. 18, no. 1, p. 66–88, 2014.
- [3] Marketos G, Frentzos E, Ntoutsi I, Pelekis N, Raffaeta A, and Theodoridis Y, "Building real-world trajectory warehouses," in Proceedings of the Seventh ACM International Workshop on Data Engineering for Wireless and Mobile Access (MobiDE 2008), Vancouver, British Columbia, 2008.
- [4] Guting, R. H. And Schneider, M., Moving Objects Databases, Morgan Kaufmann, 2005.
- [5] Pelekis, N., Theodoridis, Y., Vosinakis, S., Panayiotopoulos T., "HERMES - A Framework for Location-Based Data Management," in Advances in Database Technology - EDBT 2006, Springer Berlin Heidelberg, 2006, pp. 1130-1134.



(الف)



(ب)



(ج)

شکل ۸- توقف‌های شناسایی شده در خطوط سیر

غنی‌سازی خطوط سیر مکانی یک موضوع نسبتاً جدید در مطالعات مربوط به خط سیر است که در چند سال اخیر موردتوجه محققین این حوزه بوده است. با توجه به تنوع در خطوط سیر اشیاء متحرک مختلف مطالعه پردازش‌های خط سیر برای غنی‌سازی معنایی می‌تواند بر اساس مدل‌های مفهومی دیگر نیز گسترش یابد.

- [6] Li, Z., Han, J., Ji, M., Tang, L. A., Yu, Y., Ding, B., Lee, J.-G., And Kays, R, "MoveMine: Mining moving object data for discovery of animal movement patterns," TIST, vol. 2, no. 4, 2011.
- [7] Giannotti F, Nanni M, Pinelli F, and Pedreschi D, "Trajectory pattern mining," in In Proceedings of the Thirteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, California, 2007.
- [8] Cao H, Mamoulis N, and Cheung D W, "Discovery of periodic patterns in spatiotemporal sequences," IEEE Transactions on Knowledge and Data Engineering (TKDE) , vol. 19, no. 1, pp. 453 - 467 , 2007.
- [9] Yan, Z., Spaccapietra, S., Aberer, K., Parent, C., Chakraborty, D. , "Semantic Trajectories: Mobility Data Computation and Annotation," ACM Transactions on Intelligent Systems and Technology, vol. 4, no. 3, 2013.
- [10] Pelekis, N., Theodoridis, Y., Janssens, D., "On the Management and Analysis of Our LifeSteps," ACM SIGKDD Explorations Newsletter, vol. 15, no. 1, pp. 23-32, 2013.
- [11] Camossi, E., Villa, P., and Mazzola, L., "Semantic-based Anomalous Pattern Discovery in Moving Object Trajectories," arXiv:1305.1946, 2013.
- [12] Yan, Z., Chakraborty, D., Parent C., Spaccapietra, S., Aberer K, "SeMiTri: a framework for semantic annotation of heterogeneous trajectories," in Proceedings of the 14th International Conference on Extending Database Technology, New York, 2011.
- [13] Spaccapietra, S., Parent, C., Damiani, M L., Macedo, J A., Porto, F., and Vangenot, C., "A conceptual view on trajectories," Data & Knowledge Engineering , vol. 65, no. 1, pp. 126-146 , 2008.
- [14] Alvares, L.O., Bogorny, V., Kuijpers, B., DE Macedo, J.A.F., Moelans, B., And Vaisman, A, "A model for enriching trajectories with semantic geographical information," in ACM-GIS, New York, NY, USA, 2007.
- [15] Rocha, J.A.M., Times, V.C., Oliveira, G., Alvares, L.O., And Bogorny, V., "DB-SMoT: A directionbased spatio-temporal clustering method," in Conf. of Intelligent Systems, London , 2010.
- [16] Alvares, L. O., Oliveira, G., Heuser, C. A., and Bogorny, V., "A Framework for Trajectory Data Preprocessing for Data Mining," in Conf. on Software Engineering and Knowledge Engineering, 2009.
- [17] Richter KF, Schmid F, Laube P, "Semantic trajectory compression:Representing urban movement in a nutshell," SPATIAL INFORMATION SCIENCE, vol. 4, no. 1, pp. 3-30, 2012.
- [18] Guc, B., May, M., Saying Y., And Korner, "Semantic annotation of GPS trajectories," in 11th AGILE Int. Conf. on Geographic Information Science, Gerona, Spain, 2008.
- [19] Bamba,B., Liu,L , Pesti,P., and Wang, T., "Supporting Anonymous Location Queries in Mobile Environments With Privacygrid," in Proceedings of the17th International Conference on World Wide Web, 2008.
- [20] Zhang J and Goodchild M.F , Uncertainty in Geographical Information, CRC, 2002.
- [21] Jun,J., Guensler, R., and Ogle J., "Smoothing Methods to Minimize Impact of Global Positioning System Random Error on Travel Distance,Speed, and Acceleration Profile Estimates," Transportation Research Record: Journal of the Transportation Research Board, p. 141–150, 2006.
- [22] Zheng, Y., , Zhou, X., Computing with Spatial Trajectories, New York: Springer-Verlag New York, 2011.
- [23] Cao, H.,Wolfson, O., And Trajcevski, G., "Spatio-temporal data reduction with deterministic error bounds," in In Proceedings of the DIALM-POMC Joint Workshop on Foundations of Mobile Computing, 2003.
- [24] Parent, C., Spaccapietra, S., Renso, C., Andrienko, G., Andrienko, N., Bogorny, V., Damiani, M. L., Gkoulalas-Divanis, A., Macedo, J. A., Pelekis, N., Theodoridis, Y., and Yan, Z., "Semantic trajectories modeling and analysis," in ACM Computing Surveys, 2013.
- [25] Zheng, Y., Zhang, L., Xie, X., & Ma, W. Y., "Mining interesting locations and travel sequences from GPS trajectories," in In Proceedings of the 18th international conference on World wide, Madrid Spain, 2009.
- [26] Zheng, Y., Li, Q., Chen, Y., Xie, X., & Ma, W. Y., "Understanding mobility based on GPS data," in In Proceedings of the 10th international conference on Ubiquitous computing, Seoul, Korea, 2008.
- [27] Zheng, Y., Xie, X., & Ma, W. Y., "GeoLife: A Collaborative Social Networking Service among User, location and trajectory," in IEEE Data Engineering Bulletin, 2010.
- [28] de Vries, G.K.D., "Kernel methods for vessel trajectories," in Ph.D. thesis, Informatics Institute, University of Amsterdam., 2012.