

مقایسه توابع فاصله اندازه‌گیری میزان شباهت در خطوط‌سیر مکانی

محمد شاعری^۱، رحیم علی‌عباسپور^{۲*}

^۱دانشجوی کارشناسی ارشد سیستم‌های اطلاعات مکانی - گروه مهندسی نقشه‌برداری - پردیس دانشکده‌های فنی - دانشگاه تهران
mshaeri@ut.ac.ir

^۲استادیار گروه مهندسی نقشه‌برداری - پردیس دانشکده‌های فنی - دانشگاه تهران
abaspour@ut.ac.ir

(تاریخ دریافت خرداد ۱۳۹۳، تاریخ تصویب مهر ۱۳۹۳)

چکیده

خط‌سیر مکانی (spatial trajectory)، تغییرات مکانی یک شی در طی زمان را ثبت کرده و به صورت رشته‌ای از نقاط گسسته که دارای مولفه‌های مکانی-زمانی هستند، مدل می‌شود. امروزه افزایش روز افزون دستگاه‌های متحرک و تکنولوژی‌های موقعیت‌یابی منجر به حجم انبوه و بیشماری از اطلاعات مکانی-زمانی و در پی آن نیازمند کاربردها و تحلیل‌های زیادی شده است. یکی از تحلیل‌های مهم در خطوط‌سیر مکانی استخراج خطوط‌سیر مکانی متشابه است. تاکنون توابع فاصله متنوعی برای اندازه‌گیری میزان شباهت خطوط‌سیر مکانی ارائه شده‌اند که هر یک، شباهت خطوط‌سیر را از جنبه‌ی خاصی بررسی کرده و برای داده‌ها با ویژگی‌های متنوع کارآیی متفاوتی دارند. در نتیجه کارآیی توابع برای همه‌ی داده‌ها و کاربردها یکسان نبوده و برای انتخاب تابع مناسب نیازمند شناخت قابلیت‌ها و ویژگی‌های توابع هستیم. در این مقاله به بررسی ۷ تابع بنیادی و متداول که اساس بسیاری از توابع فاصله دیگر هستند پرداخته و برتری‌ها و نواقص هر یک از روشها ضمن روند ارزیابی مشخص می‌گردند.

واژگان کلیدی: خطوط‌سیر مکانی، توابع فاصله، شباهت

* نویسنده رابط

۱- مقدمه

داده، تابع مورد استفاده برای مقایسه و کاربرد مورد نظر باشد.

در این مقاله با اعمال تبدیلات گوناگون مکانی و زمانی بر روی داده اولیه، مجموعه داده از جنبه‌های گوناگون جهت کمک به تصمیم‌گیری صحیح در انتخاب تابع مناسب و انجام مقایسه جامع و هدفمند آماده می‌گردد. در این تحقیق توابع از جهات زیر بررسی می‌شوند:

- نويز
- تبدیلات هندسی (انتقال و دوران)
- تغییر در نرخ برداشت (کاهش و افزایش)
- انتقال زمان محلی
- تغییر سرعت (کشیدگی زمانی)

سپس برای بررسی جنبه‌های یاد شده، مستقل از انتقال و دوران خطوط سیر نسبت به هم، در توابع فاصله از روش‌های پیش‌پردازش استفاده می‌شود. علاوه بر پیش‌پردازش نرمال، جهت بررسی دقیق‌تر جنبه‌های مختلف در خطوط سیر، روش بردار میانگین ارائه شده است. در انتها، مجموعه داده‌های تبدیل یافته با مجموعه داده اولیه توسط ۷ تابع اندازه‌گیری شباهت خطوط سیر مقایسه و به بررسی رفتار توابع از لحاظ تبدیلات گوناگون پرداخته می‌شود.

۲- پیشینه پژوهش

با توجه به کاربردهای متعدد فاصله، از اوایل قرن بیستم تاکنون در زمینه ارائه توابع فاصله شباهت توسط محققان مختلف تلاش‌های بسیاری صورت گرفته است. تبدیل سری فوری به صورت گسسته از اولین روش‌هایی بود که برای تحلیل شباهت سری‌های زمانی مورد استفاده قرار گرفت [۱، ۲].

فواصل اقلیدسی و منهن از قدیمی‌ترین روش‌ها برای اندازه‌گیری فاصله هستند که توابع متعددی بر اساس آن‌ها تعریف شده‌اند [۳، ۴، ۵]. فاصله فرشت^۳ در سال ۱۹۰۶ مطرح شده و در اواخر قرن بیستم انواع گسسته و قابل استفاده آن برای اندازه‌گیری شباهت در خط سیر مکانی ارائه شده است [۶، ۷، ۸]. پیچش زمانی پویا^۴ (DTW) [۹] که توسط سونگ و روزنبرگ در سال ۱۹۸۸ و طولانی‌ترین

خط سیر مکانی، تغییرات مکانی یک شیء در طی زمان را ثبت کرده و به صورت نقاط گسسته و متوالی که دارای مؤلفه‌های مکانی-زمانی هستند، مدل می‌شود. امروزه به خاطر افزایش روزافزون دستگاه‌های متحرک و فناوری‌های تعیین موقعیت مانند سامانه موقعیت‌یابی جهانی (GPS) و شناسایی امواج رادیویی^۱ منجر به حجم انبوه و بیشماری از اطلاعات مکانی-زمانی شده و در پی آن کاربردها و تحلیل‌های زیادی در زمینه‌های مختلف مطرح گردیده‌اند. یکی از تحلیل‌های متداول و مهم، الگوکاوی^۲ خطوط سیر مکانی است که موضوع آن مطالعه الگوهای رفتاری خطوط سیر مکانی نسبت به یکدیگر است. مهم‌ترین بخش الگوکاوی خطوط سیر مکانی، استخراج خطوط سیر مکانی متشابه است. تحلیل خوشه‌بندی، یکی از روش‌های متداول در استخراج خطوط سیر مکانی متشابه است که برای اجرا، به اندازه‌گیری میزان شباهت دو خط سیر نیاز دارد. علاوه بر این کاربرد اساسی، اندازه‌گیری میزان شباهت دو خط سیر کاربردهای متعددی در شاخه‌های مختلف علمی مانند بهینه‌سازی حمل و نقل، انتخاب مسیر در راه‌سازی، تشخیص فعالیت و شناسایی آنومالی، پیش‌بینی ترافیک قبل از ساخت راه، پیش‌بینی مسیر در سرویس‌های مکان-مبنا و تشخیص حرکت حیوانات در مطالعات جانورشناسی اشاره کرد.

مسأله مهم در مطالعات مرتبط با اندازه‌گیری میزان شباهت بین خطوط سیر، پیدا کردن معیار مناسب برای شناخت میزان شباهت است که این امر توسط توابع فاصله محقق می‌شود. تاکنون توابع فاصله متنوعی ارائه شده‌اند که هر یک شباهت خطوط سیر را از جنبه خاصی بررسی کرده‌اند. اما برخلاف ماهیت متفاوت خطوط سیر، بر اساس ویژگی‌های گوناگون و کاربردهای مختلف، توابع فاصله، شباهت را بر اساس کاربرد و نیاز خود تعریف می‌کنند. برای مثال، شباهت برای یک کاربرد صرفاً نزدیکی مکانی است در حالی که برای کاربرد دیگر نزدیکی زمانی مبنای شباهت است. همین امر سبب می‌شود که همه توابع برای داده‌ها و کاربردهای گوناگون مناسب نبوده و تعیین میزان شباهت بین خطوط سیر مکانی نیازمند شناخت درست از

^۳ Frechet

^۴ Dynamic Time Warping

^۱ Radio-frequency identification (RFID)

^۲ Pattern mining

نگردیده و تنها در نتیجه نهایی به یک یا دو شاخص یا چند تبدیل مکانی-زمانی بسنده شده و مقایسه توسط آن-ها انجام شده است.

۳- پیش پردازش

برای بررسی جنبه‌های مختلف در خطوط سیر به صورت مستقل، لازمست تا مقایسه‌ای مستقل از مکان، مقیاس و دوران انجام گردد. به این منظور ضروری است که در مرحله پیش پردازش بر روی خطسیر پردازش‌هایی انجام گیرد. روش نرمال [۱۱] معرفی شده در تابع EDR عمده‌ترین راهکار استفاده شده تاکنون است. به علت ناتوانی روش نرمال در رفع دوران و عدم مقیاس یکسان در دو راستای افقی و عمودی، پیش پردازشی دیگر به نام بردار میانگین ارائه شده است.

۳-۱- روش بردار میانگین

روش بردار میانگین برای حل دوران و مقیاس یکسان در هر دو مؤلفه مسطحاتی پیشنهاد می‌شود. μ_x و μ_y به ترتیب میانگین مولفه‌های x و y نقاط خط سیر هستند. در این روش، بردار \overrightarrow{MV} (مطابق رابطه ۱) محور y سیستم مختصات خطسیر تبدیل یافته خواهد شد و مبدأ، نقطه (μ_x, μ_y) خواهد بود و بردار $\overrightarrow{P_iV}$ (مطابق رابطه ۲) بردار نقطه از مبدأ جدید می‌باشد. در نهایت x و y از روابط (۳) تا (۴) به دست می‌آیند. مثبت یا منفی بودن x با توجه به زوایای دو بردار تعیین می‌شود. نقاط در انتها بر مقدار $\max\{\max(x), \max(y), -\min(x), -\min(y)\}$ تقسیم خواهند شد.

$$\overrightarrow{MV} = (\mu_x - x_1, \mu_y - y_1) \quad (1)$$

$$\overrightarrow{P_iV} = (x_i - \mu_x, y_i - \mu_y) \quad (2)$$

$$y'_i = \frac{\overrightarrow{MV} \cdot \overrightarrow{P_iV}}{|\overrightarrow{MV}|} \quad (3)$$

$$x'_i = \sqrt{|\overrightarrow{P_iV}|^2 - y'^2_i} \quad (4)$$

x_i و y_i مولفه افقی و عمودی نقطه i ام خط سیر و متغیرهای x'_i و y'_i مولفه‌های نقاط خط سیر تبدیل یافته هستند.

زیر رشته مشترک^۱ (LCSS) [۱۰] که توسط کرنی و هسن در سال ۱۹۹۰ پیشنهاد شدند، دو روش غیرمتریک هستند که مشکل نابرابری تعداد نقاط و شیفت زمانی را حل کرده‌اند. تصحیح فاصله روی رشته حقیقی^۲ (EDR) [۱۱] توسط چن در سال ۲۰۰۵ از دیگر روش‌های غیرمتریک است که به مشکل تعداد نقاط و شیفت زمانی پرداخته است. تصحیح فاصله با پنالتی حقیقی^۳ (ERP) [۴] توسط چن و نگ در سال ۲۰۰۴ از ترکیب دو روش Lp-norm و تصحیح فاصله^۴ به وجود آمد. یکی دیگر از توابع فاصله متریک که خطسیر مکانی را مجموعه‌ای از پاره‌خط‌های گسسته در نظر می‌گیرد، هاسدورف^۵ است [۱۲].

برخلاف توابع فاصله پیشنهادی متعدد، تاکنون مطالعات بسیار اندکی در زمینه مقایسه این توابع صورت گرفته است. در مقایسه صورت گرفته توسط ژانگ، هوانگ و تان [۱۳] به مقایسه کارآمدی توابع در خوشه‌بندی اکتفا کرده و خروجی آن اطلاعات کاملی در رابطه با ویژگی‌های توابع به خواننده نمی‌دهد.

وانگ و همکارانش [۱۴] در مقایسه خود از تبدیلات کمی استفاده کرده و تنها به افزودن نویز، انتقال نقاط و تغییر نرخ برداشت بسنده کرده‌اند. سومرن و وریز [۱۵] در مقایسه خود دو دسته از توابع را با هم مقایسه کرده‌اند و نتیجه با مقایسه نحوه خوشه‌بندی پایان یافته است.

تحقیقات انجام شده از لحاظ تنوع در تبدیلات و ویژگی‌های مقایسه کامل نبوده و کمک چندانی به تصمیم‌گیری در رابطه با انتخاب تابع مناسب نمی‌کنند. در این مقاله به بررسی و مقایسه هفت تابع اساسی و متداول اقلیدسی، فرشت، هاسدورف، DTW، LCSS، EDR و ERP می‌پردازیم. توابع از جنبه‌های گوناگونی بررسی می‌شوند: نویز، انتقال، دوران، تغییر نرخ برداشت به صورت کاهشی و افزایشی، انتقال زمان محلی^۶ و کشیدگی زمانی.

تاکنون در هیچ یک از مقایسه‌های انجام شده از روش پیش‌پردازش مناسب برای تعیین شباهت در مقایسه استفاده نشده و طیف متنوعی از توابع در آن‌ها بررسی

^۱ Longest Common Subsequence

^۲ Edit distance on real sequence

^۳ Edit distance with real penalty

^۴ Edit distance

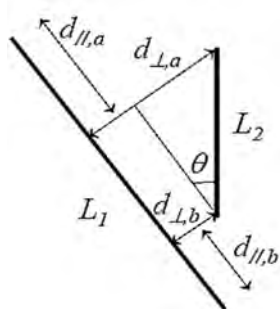
^۵ Hausdorff

^۶ Local time shifting

۴- توابع فاصله

دهی به هر یک، بسته به کاربرد خاص، فاصله نهایی را بر اساس مجموعه روابط (۶) تا (۹) محاسبه می‌کند.

- فاصله موازی دو پاره‌خط نسبت به هم: d_{\perp}
- زاویه دو پاره‌خط نسبت به هم: d_{θ}
- تفاوت اندازه دو پاره‌خط نسبت به هم: d_{\parallel}



شکل ۱- فواصل هاسدورف

در شکل (۱) دو پاره‌خط از دو خط سیر، نشان داده شده‌اند که سه فاصله بیان شده توسط فواصل نمایش داده شده در شکل (۱) محاسبه می‌شوند.

$$d_{\perp} = \frac{d_{\perp,a}^2 + d_{\perp,b}^2}{d_{\perp,a} + d_{\perp,b}} \quad (۶)$$

$$d_{\theta} = \begin{cases} \|L_2\| \times \sin\theta & \theta < 90 \\ 2 - \|L_2\| \times \sin\theta & \text{otherwise} \end{cases} \quad (۷)$$

$$d_{\parallel} = \min\{d_{\parallel,a}, d_{\parallel,b}\} \quad (۸)$$

$$D_{Hausdorff} = w_{\perp} \cdot d_{\perp} + w_{\parallel} \cdot d_{\parallel} + w_{\theta} \cdot d_{\theta} \quad (۹)$$

مقدار شباهت، میانگین فواصل پاره‌خطها خواهد بود.

۴-۳- تابع فرشت

توابع ارائه شده بر اساس این فاصله در چندخطی‌ها کاربرد داشته و اخیراً نوع گسسته آن نیز مطرح شده است. مقدار شباهت $C(n, m)$ از رابطه زیر بدست می‌آید.

$$C(i, j) = \begin{cases} \text{dist}(\alpha_{x,y(i)}, \beta_{x,y(j)}) & \text{if } i \text{ and } j = \\ \max\{C(i-1, 1), \text{dist}(\alpha_{x,y(i)}, \beta_{x,y(1)})\} & \text{if } i > 1 \text{ and } j = \\ \max\{C(1, j-1), \text{dist}(\alpha_{x,y(1)}, \beta_{x,y(j)})\} & \text{if } i = 1 \text{ and } j > \\ \max\{\min\{C(i-1, j), C(i-1, j-1), C(i, j-1)\}, \\ \text{dist}(\alpha_{x,y(i)}, \beta_{x,y(j)})\} & \text{if } i \text{ and } j > 1 \end{cases} \quad (۱۰)$$

۴-۴- DTW

این تابع برای اندازه‌گیری شباهت دو خطسیر، مستقل از تغییرات غیر خطی زمانی کاربرد دارد. در حقیقت دو

دو خطسیر A به صورت $\{(\alpha_{x,1}, \alpha_{y,1}, \alpha_{t,1}), \dots, (\alpha_{x,n}, \alpha_{y,n}, \alpha_{t,n})\}$ و B به صورت $\{(\beta_{x,1}, \beta_{y,1}, \beta_{t,1}), \dots, (\beta_{x,m}, \beta_{y,m}, \beta_{t,m})\}$ را در نظر بگیرید. بررسی میزان شباهت خطوطسیر نسبت به هم توسط تابعی انجام می‌گیرد که دو خطسیر مورد مقایسه، متغیر-های ورودی آن هستند. تاکنون توابع متفاوتی با در نظر گرفتن جنبه خاصی از شباهت مطرح شده‌اند که متداول-ترین و بنیادی‌ترین‌های آنها که پایه بسیاری از توابع دیگر هستند به صورت توابع اقلیدسی، فرشت، هاسدورف، DTW، LCSS، EDR و ERP می‌باشند.

۴-۱- تابع اقلیدسی

تاکنون توابع متعددی از فاصله اقلیدسی استفاده نموده‌اند. علت این امر، بنیادی بودن این اندازه‌گیری در مقایسه دو خطسیر است. در مقایسه انجام شده، تابع اقلیدسی بین دو نقطه را به صورت زیر تعریف می‌کنیم:

$$\text{Euclidean}(A, B) = \frac{\text{dist}(\alpha_{x,y(i)}, \beta_{x,y(i)}) + |\alpha_{t(i)} - \beta_{t(i)}|}{3} + \sqrt{\frac{\text{dist}(\alpha_{x,y(i)}, \beta_{x,y(i)}) * |\alpha_{t(i)} - \beta_{t(i)}|}{3}} \quad (۵)$$

فاصله بین دو نقطه، میانگینی از سه فاصله مکانی، زمانی و مکانی-زمانی است. در صورتی که تعداد نقاط دو خط سیر به صورت $(n > m)$ باشد، نقاط نظیر $\beta_{(i)}$ $\alpha_{((i-1)*n/m)}$ تا $\alpha_{((i)*n/m)}$ می‌باشند. مقدار شباهت میانگین فواصل نقاط خواهد بود.

۴-۲- تابع هاسدورف

در روش‌های متداول خوشه‌بندی، شباهت کل خطوط سیر محاسبه می‌شود. در برخی روشهای خوشه‌بندی مانند TRACCLUS [۱۲] برای بررسی زیرخطوسیر^۱ و یافتن قطعات مشابه خطوطسیر، آن‌ها را به مجموعه پاره‌خطها تقسیم نموده و توسط روش هاسدورف شباهت محاسبه می‌شود. تابع هاسدورف برای محاسبه شباهت، سه فاصله نشان داده شده در شکل (۱) را محاسبه کرده و با وزن-

^۱ Subtrajectory

نمی‌کنند و گروه دوم انتقال زمان محلی را پشتیبانی کرده، ولی متریک نیستند. برای بهره بردن از قابلیت‌های هر دو گروه، یعنی متریک بودن و پشتیبانی از انتقال زمان محلی، تابع ERP با ترکیب روش Lp-norm و تصحیح فاصله ارائه شده است. در مطالعات صورت گرفته، مقدار مناسب برای شکاف با در نظر گرفتن انتقال مبداء به نقطه (μ_x, μ_y) در مرحله پیش‌پردازش، $(0,0)$ پیشنهاد شده است [۴].

$$\text{dist}(\alpha_{x,y,(i)}, \beta_{x,y,(j)}) = \begin{cases} \text{euclidean}(\alpha_{x,y,(i)}, \beta_{x,y,(j)}) & \text{if } |\alpha_{t,(i)} - \beta_{t,(j)}| \leq 5 \\ \text{euclidean}(\alpha_{x,y,(i)}, \text{gap}) & \text{if } \beta_{(j)} \text{ is gap} \\ \text{euclidean}(\text{gap}, \beta_{x,y,(j)}) & \text{if } \alpha_{(i)} \text{ is gap} \end{cases}$$

$$\text{ERP}(A, B) = \begin{cases} \sum_1^n \text{euclidean}(\alpha_{x,y,(i)}, \text{gap}) & \text{if } m = 0 \\ \sum_1^m \text{euclidean}(\text{gap}, \beta_{x,y,(j)}) & \text{if } n = 0 \\ \min\{\text{ERP}(\text{Rest}(A), \text{Rest}(B)) + \text{dist}(\alpha_{x,y,(1)}, \beta_{x,y,(1)}), \\ \text{ERP}(A, \text{Rest}(B)) + \text{dist}(\text{gap}, \beta_{x,y,(1)}), \\ \text{ERP}(\text{Rest}(A), B) + \text{dist}(\alpha_{x,y,(1)}, \text{gap})\} & \text{otherwise} \end{cases} \quad (13)$$

مقدار شباهت $\frac{\text{ERP}(A,B)}{n}$ خواهد بود.

۴-۷- EDR

روش‌های گوناگون در زمینه بررسی شباهت خطوط سیر از ۳ مشکل بزرگ نویز، انتقال مکانی و مقیاس داده‌ها رنج می‌برند. روش EDR با ارائه پیش‌پردازش نرمال مقایسه خطوط سیر را از جنبه انتقال مکانی و اختلاف مقیاس بین آن‌ها مستقل کرده است. این تابع نیز همچون تابع LCSS نیاز به یک حدآستانه (δ) برای فاصله اقلیدسی دو نقطه دارد. در صورتی که فاصله دو نقطه بیشتر از حدآستانه باشد یا نقطه معادل زمانی در خط سیر دوم برای خط سیر اول وجود نداشته باشد نیاز به تصحیح وجود دارد.

۳ نوع تصحیح عبارتند از:

- افزودن نقطه (به علت وجود شکاف)
- تصحیح نقطه (تصحیح مکانی)
- حذف نقطه

$$\text{dist}(\alpha_{x,y,(i)}, \beta_{x,y,(j)}, \delta) = \begin{cases} 0 & \text{if euclidean}(\alpha_{x,y,(i)}, \beta_{x,y,(j)}) \leq \delta \\ & \text{and } |\alpha_{t,(i)} - \beta_{t,(j)}| \leq 5 \\ 1 & \text{otherwise} \end{cases}$$

$$\text{EDR}(A, B, \delta) = \begin{cases} n & \text{if } m = 0 \\ m & \text{if } n = 0 \\ \min\{\text{EDR}(\text{Rest}(A), \text{Rest}(B)) + \text{dist}(\alpha_{x,y,(1)}, \beta_{x,y,(1)}, \delta), \\ \text{EDR}(A, \text{Rest}(B)) + \text{dist}(\text{gap}, \beta_{x,y,(1)}, \delta), \\ \text{EDR}(\text{Rest}(A), B) + \text{dist}(\alpha_{x,y,(1)}, \text{gap}, \delta)\} & \text{otherwise} \end{cases} \quad (14)$$

مقدار شباهت $\frac{\text{EDR}(A,B)}{n}$ خواهد بود.

خط سیر، مستقل از انتقال زمان محلی و کشیدگی زمانی (تفاوت سرعت) بررسی می‌شوند. این تابع دو مجموعه از نقاط دو خط سیر را طوری پیدا می‌کند که دو خط سیر کمترین فاصله را با هم داشته باشند و خروجی تابع، فاصله دو خط سیر نماینده از خطوط سیر واقعی است. در صورتی که در یک زمان مشخص با فقدان داده روبه‌رو شویم آن را به عنوان شکاف^۱ می‌شناسیم.

$$\text{dist}(\alpha_{x,y,(i)}, \beta_{x,y,(j)}) = \begin{cases} \text{euclidean}(\alpha_{x,y,(i)}, \beta_{x,y,(j)}) & \text{if } |\alpha_{t,(i)} - \beta_{t,(j)}| \leq 5 \\ \text{euclidean}(\alpha_{x,y,(i)}, \beta_{x,y,(j-1)}) & \text{if } \beta_{(j)} \text{ is gap} \\ \text{euclidean}(\alpha_{x,y,(i-1)}, \beta_{x,y,(j)}) & \text{if } \alpha_{(i)} \text{ is gap} \end{cases}$$

$$\text{DTW}(A, B) = \begin{cases} 0 & \text{if } m = n = 0 \\ \infty & \text{if } m = 0 \text{ or } n = 0 \\ \text{dist}(\alpha_{x,y,(1)}, \beta_{x,y,(1)}) + \min\{\text{DTW}(\text{Rest}(A), \text{Rest}(B)), \\ \text{DTW}(A, \text{Rest}(B)), \text{DTW}(\text{Rest}(A), B)\} & \text{otherwise} \end{cases} \quad (11)$$

$\text{Rest}(A)$ به معنای خط سیر A به جز نقطه اول آن است. مقدار شباهت $\frac{\text{DTW}(A,B)}{n}$ خواهد بود.

۴-۵- LCSS

همان‌طور که از نام تابع نیز پیداست، این تابع اندازه طولانی‌ترین زیر رشته‌های متشابه از دو خط سیر را محاسبه می‌کند. این روش برای حل مشکل نویز مطرح شده است و هدف آن پایداری و مقایسه شباهت دو خط سیر مستقل از وجود نویز است. مقایسه این تابع مستقل از کشیدگی در زمان است. در رابطه (۱۲) تعداد نقاطی از دو خط سیر که فاصله آن‌ها کمتر از حدآستانه (δ) تعیین شده باشند محاسبه می‌شود.

$$\text{LCSS}(A, B, \delta) = \begin{cases} 0 & \text{if } m = 0 \text{ or } n = 0 \\ \text{LCSS}(\text{Rest}(A), \text{Rest}(B)) + 1 & \text{if } |\alpha_{x,(1)} - \beta_{x,(1)}| \leq \delta \text{ and } \\ & |\alpha_{y,(1)} - \beta_{y,(1)}| \leq \delta \text{ and } |\alpha_{t,(1)} - \beta_{t,(1)}| \leq 5 \\ \max\{\text{LCSS}(A, \text{Rest}(B)), \text{LCSS}(\text{Rest}(A), B)\} & \text{otherwise} \end{cases} \quad (12)$$

مقدار شباهت $1 - \frac{\text{LCSS}(A,B)}{n}$ خواهد بود.

۴-۶- ERP

مطالعات انجام گرفته بر روی سری‌های زمانی بر مبنای دو گروه از توابع فاصله قرار دارند. گروه نخست بر مبنای Lp-norm بوده و متریک هستند، ولی انتقال زمان محلی را پشتیبانی

^۱ Gap

۵- پیاده‌سازی و ارزیابی

۵-۱- داده‌ها

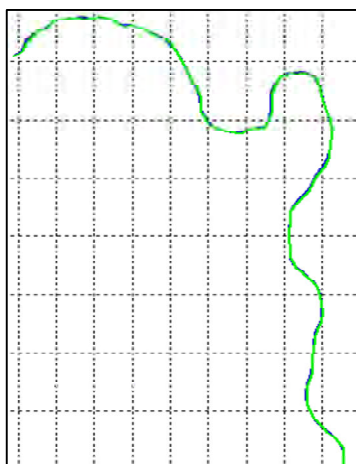
مجموعه داده با ۱۷۶۲۱ خطسیر (شکل ۲) توسط ۱۸۲ کاربر در بیش از ۳۰ شهر چین و شهرهایی از اروپا و آمریکا به دست آمده است [۱۶، ۱۷، ۱۸].



شکل ۲- مجموعه داده اولیه مورد استفاده در ارزیابی

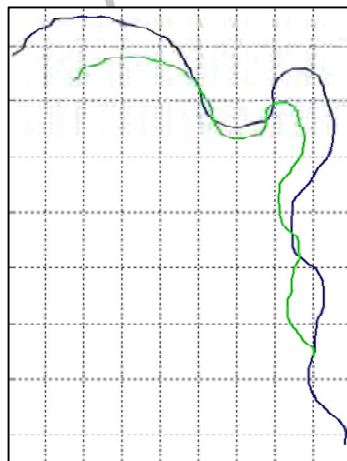
فقط از جنبه تبدیل انجام شده متفاوت خواهد بود. برای مثال، با دوران خطسیر، به خطسیر نويز اضافه نمی‌شود. در نتیجه مقدار تابع شباهت تنها ناشی از تفاوت دوران در آن خطسیر است و این کار بررسی جنبه‌های گوناگون مستقل از هم را امکان‌پذیر می‌کند.

اولین تبدیل، ایجاد نويز روی داده می‌باشد. به صورت اتفاقی حدود نیمی از نقاط بین صفر تا ۵ متر نويزدار می‌شوند. نمونه ای از این تبدیل بر روی یک خط سیر در شکل (۳) نشان داده شده است.



شکل ۳- خطسیر (آبی) و نويز دار شده (سبز)

تبدیل بعدی ایجاد اعوجاج در خطسیر به صورتی است که شباهت آن تا حدود زیادی حفظ شود. نمونه‌ای از این تبدیل بر روی یک خط سیر در شکل (۴) نشان داده شده است. سپس خطسیر اعوجاج‌داده شده به اندازه‌های ۳۰، ۶۰، ۹۰، ۱۲۰ و ۱۵۰ درجه دوران می‌یابد تا تأثیر زاویه در تعیین شباهت، در هریک از مراحل پیش‌پردازش و توابع بررسی شود.



شکل ۴- خط سیر (آبی) و اعوجاج داده شده (سبز)

با اصلاح و آماده‌سازی داده‌ها بر اساس مشخصات خطوطسیر، ۴۶۸ خطسیر که حداقل دارای ۱۰۰ و حداکثر ۱۵۳۶ نقطه برداشت‌شده و طولی بین ۲۵ تا ۳۵۰۰۰ متر دارند، برای مقایسه انتخاب شده‌اند که مشخصات آن‌ها در جدول (۱) نشان داده شده است.

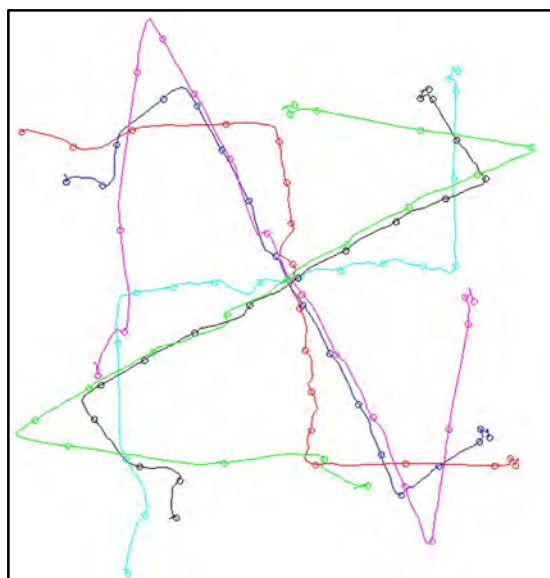
جدول ۱- مشخصات مجموعه داده

بازه زمانی برداشت (s)	سرعت (km/h)	شتاب (km/h/s)	
۵۹	۶۹/۹۴۹۷	۴/۹۹	بیشینه
۴/۴۳۵۷	۸/۸۵۶۸	۲/۶۸۵	متوسط
۱	۰/۰۹۲۴	-۴/۹۸	کمینه
۱/۵۹۴۷	۶/۶۵۱۰	۱/۱۳۳۷	انحراف

۵-۱-۱- تبدیلات خطوطسیر

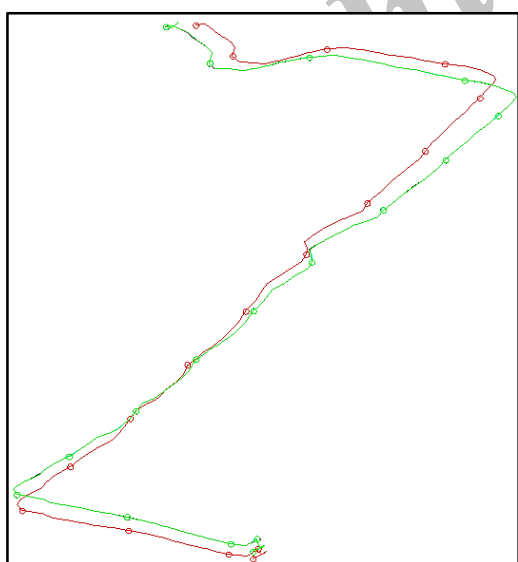
برای مقایسه توابع از جنبه‌های مختلف نیازمند اعمال تبدیلاتی روی خطوطسیر اولیه و ایجاد خطوطسیر تبدیل‌یافته هستیم. با اعمال یک تبدیل روی خطسیر، دیگر ویژگی‌های خطسیر ثابت نگه داشته شده و خط سیر

در شکل (۶) و (۷) یک خطسیر با ۵ خطسیر اعوجاج و دوران داده شده به ترتیب توسط روش نرمال و بردار میانگین نمایش داده شده‌اند.



شکل ۶- خطسیر (قرمز) و ۵ خطسیر اعوجاج و دوران داده

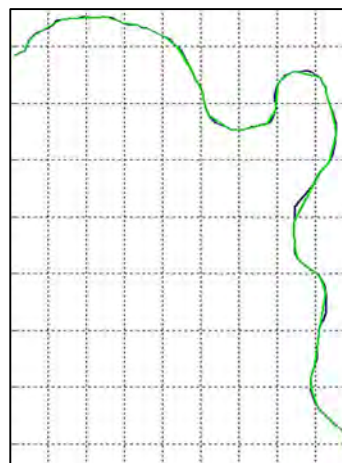
در روش نرمال دوران حذف نشده ولی در روش بردار میانگین دوران کاملاً حذف گردیده است و تنها دو خط-سیر، که یکی خطسیر اصلی و دیگری ۵ خطسیر دوران و اعوجاج داده شده هستند که بر هم منطبق شده‌اند، قابل مشاهده‌اند.



شکل ۷- خطسیر(قرمز) و ۵ خطسیر اعوجاج و دوران داده

به علاوه توابع را در ۵ تبدیل دوران توسط دو روش نرمال و بردار میانگین بررسی می‌کنیم. در مقایسه انجام شده با استفاده از روش نرمال با افزایش دوران مقادیر

در تبدیل بعدی، ۴۰ درصد از نقاط خطسیر به صورت اتفاقی حذف شده تا در ارزیابی، تغییر نرخ برداشت بررسی گردد. نمونه ای از این تبدیل بر روی یک خط سیر در شکل (۵) نشان داده شده است.



شکل ۵- خط سیر(آبی) و کاهش نرخ برداشت شده (سبز)

در تبدیل بعدی، به اندازه ۴۰ درصد از نقاط خطسیر به صورت اتفاقی به خطسیر افزوده می‌شود تا در ارزیابی، تغییر نرخ برداشت بررسی گردد. سه مولفه نقطه افزوده شده بین دو نقطه با درون یابی خطی معین می‌شوند. واضح است که خطسیر تبدیل یافته کاملاً بر خطسیر اولیه منطبق است.

در نهایت، خطوطسیر تبدیل یافته هر کدام به صورت اتفاقی بین صفر تا ۱۰۰۰ متر انتقال می‌یابند تا تبدیلات از لحاظ انتقال نیز متفاوت باشند.

برای بررسی انتقال زمان محلی، زمان نقاط به اندازه ۵، ۱۰، ۱۵، ۲۰ و ۲۵ ثانیه افزایش داده شده تا توابع از این حیث نیز بررسی گردند. تبدیل آخر مربوط به کشیدگی زمانی و تغییر سرعت است که با ضرب بازه‌های زمانی در ۲، ۳، ۴، ۵ و ۱۰، سرعت آن‌ها به ترتیب به یک‌دوم، یک-سوم، یک‌چهارم، یک‌پنجم و یک‌دهم کاهش می‌یابد.

۵-۲- پیاده‌سازی

۵-۲-۱- پیش‌پردازش

در مرحله پیش‌پردازش روش بردار میانگین برای حل مشکل روش نرمال در دوران و مقیاس یکسان در دو جهت ارائه شد.

توابع افزایش می‌یابد که قابل پیش‌بینی بود ولی با استفاده از روش بردار میانگین مقادیر توابع کاملاً ثابت باقی می‌ماند که این به معنای موفقیت روش بردار میانگین در حل دوران است. پس از نمایش برتری روش بردار میانگین بر روش نرمال در ادامه تحقیق تنها از این روش برای مرحله پیش‌پردازش استفاده شده است.

۵-۲-۲- توابع

در این مرحله به مقایسه توابع بر روی خطوط سیر تبدیل یافته می‌پردازیم. مقایسه ما روی ۷ تابع می‌باشد که سه تابع نیازمند ورودی اولیه می‌باشند. تابع هاسدورف سه وزن برای فواصل خود می‌گیرد. برای این تابع هر بار یکی از وزن‌ها را یک و بقیه را صفر قرار می‌دهیم. برای توابع LCSS و EDR حد آستانه را ۱۰، ۱۵ و ۳۰ قرار می‌دهیم.

تبدیلات شامل یک نویز، پنج دوران خط سیر اعوجاج- یافته، یک تبدیل کاهش و یک تبدیل افزایش نرخ برداشت، پنج تبدیل انتقال زمان محلی و پنج تبدیل کشیدگی زمانی (کاهش سرعت) هستند که در مجموع ۱۸ تبدیل خواهیم داشت اما چون جواب‌های ۵ دوران به علت استفاده از روش بردار میانگین یکسان می‌باشند در بررسی‌ها فقط از واژه دوران استفاده می‌شود.

۵-۲-۱- مقایسه تبدیلات مکانی-زمانی

این تبدیلات شامل ۱. نویز ۲. دوران ۳. افزایش نرخ برداشت ۴. کاهش نرخ برداشت می‌باشند. دو معیار میانگین و سطح اطمینان ۹۵٪ برای مقایسه انتخاب شده‌اند.

جدول ۲- تبدیلات زمانی- مکانی

تبدیلات	نویز	دوران	افزایش نرخ	کاهش نرخ	نویز	دوران	افزایش نرخ	کاهش نرخ
تابع	LCSS- حد آستانه ۱۰ متر				EDR- حد آستانه ۱۰ متر			
میانگین	۰/۰۰۱۲	۰/۴۶۰۱	۰/۶۶۹۲	۰/۷۴۳۶	۰/۰۰۱۱	۰/۴۲۹۵	۰/۲۶۴۰	۰/۷۳۵۱
سطح اطمینان	۰/۰۳۳۳	۰/۷۳۲۲	۰/۶۵۲۰	۰/۴۵۳۷	۰/۰۳۳۰	۰/۷۲۹۰	۰/۶۳۱۸	۰/۴۶۰۷
تابع	LCSS- حد آستانه ۱۵ متر				EDR- حد آستانه ۱۵ متر			
میانگین	۰/۰۰۰۲	۰/۳۳۶۰	۰/۵۸۹۲	۰/۶۸۲۸	۰/۰۰۰۲	۰/۳۰۹۶	۰/۱۸۷۰	۰/۶۷۳۱
سطح اطمینان	۰/۰۱۰۸	۰/۷۰۷۶	۰/۵۷۹۲	۰/۴۵۹۲	۰/۰۱۱۱	۰/۶۹۵۳	۰/۵۵۹۳	۰/۴۶۱۳
تابع	LCSS- حد آستانه ۳۰ متر				EDR- حد آستانه ۳۰ متر			
میانگین	۰	۰/۱۷۲۶	۰/۴۶۹۳	۰/۵۷۳۹	۰	۰/۱۵۷۲	۰/۰۷۰۸	۰/۵۶۶۳
سطح اطمینان	۰/۰	۰/۵۸۳۶	۰/۳۸۶۷	۰/۴۱۶۵	۰/۰	۰/۵۵۷۸	۰/۳۶۳۳	۰/۴۱۰۱
تابع	افلیدسی				فرشت			
میانگین	۰/۰۰۲۹	۰/۰۱۱۱	۲/۲۸۶۷	۵/۶۶۷۰	۰/۰۱۹۷	۰/۰۶۸۸	۰/۰۲۹۴	۰/۰۷۹۸
سطح اطمینان	۰/۰۱۲۸	۰/۰۱۵۲	۲/۹۵۹۶	۶/۸۹۸۹	۰/۰۸۹۸	۰/۱۰۴۱	۰/۰۳۷۴	۰/۱۰۳۲
تابع	DTW				ERP			
میانگین	۰/۰۰۸۴	۰/۰۲۵۶	۰/۰۱۳۶	۰/۰۲۶۳	۰/۰۰۸۷	۰/۰۳۳۳	۰/۲۱۸۷	۰/۲۳۴۸
سطح اطمینان	۰/۰۳۵۶	۰/۰۳۸۶	۰/۰۱۷۸	۰/۰۳۹۷	۰/۰۳۵۹	۰/۰۴۵۹	۰/۰۹۳۶	۰/۰۹۵۹

با توجه به نتایج حاصل از تبدیلات مکانی- زمانی که در جدول (۲) نشان داده شده رفتار دو تابع LCSS و EDR بسیار نزدیک به هم بوده و با انتخاب حد آستانه مناسب می‌توان نویز را حذف کرد. هر دو تابع به تغییر نرخ برداشت به خصوص کاهش آن حساس بوده و مقدار آنها همواره از ۰/۴ بیشتر است که با مقدار ۰/۴۰ نقاطی

که از خط سیر در مرحله تبدیل حذف شده همخوانی دارد. اما تابع LCSS در افزایش نرخ برداشت حساسیت کمتری نسبت به EDR دارد. در مقایسه ابتدایی، نویز در توابع افلیدسی، فرشت، DTW و ERP از دوران کمتر است، اما در سطح اطمینان ۹۵٪ مقدار نویز از دوران بیشتر نیز می‌تواند باشد که در عمل هم در برخی موارد

جدول ۳- تبدیلات زمانی- مکانی

تبدیلات	نویز	دوران	افزایش نرخ	کاهش نرخ
تابع		هاسدورف w_{\perp}		
میانگین	۰/۰۰۳۶	۰/۰۱۹۷	۰/۰۱۵۸	۰/۰۳۶۸
سطح اطمینان	۰/۰۱۵۵	۰/۰۲۷۴	۰/۰۱۶۳	۰/۰۴۳۲
تابع		هاسدورف w_{\parallel}		
میانگین	۰/۰۱۰۸	۰/۰۲۰۵	۰/۰۱۰۸	۰/۰۲۹۷
سطح اطمینان	۰/۰۵۰۳	۰/۰۳۳۰	۰/۰۱۳۷	۰/۰۴۴۷
تابع		هاسدورف w_{θ}		
میانگین	۰/۰۰۶۷	۰/۰۰۱۰	۰/۰۰۲۷	۰/۰۰۸۶
سطح اطمینان	۰/۰۳۳۳	۰/۰۰۱۷	۰/۰۰۳۵	۰/۰۱۱۱

نویز از دوران بیشتر بود. در نتیجه توابع نسبت به نویز حساس هستند. تابع ERP رفتار یکسانی در برابر کاهش و افزایش نرخ برداشت نشان می‌دهد، در حالی که ۳ تابع دیگر به کاهش حساس‌تر از افزایش بوده‌اند. توابع اقلیدسی و ERP به تغییر نرخ برداشت بسیار حساس بوده و مقادیر به دست آمده همواره از نویز بزرگتر بوده و غالباً از دوران نیز مقدار بیشتری دارند. در دو تابع فرشت و DTW، حساسیت به تغییر نرخ برداشت با نویز تقریباً به یک اندازه است.

با توجه به جدول (۳)، در تابع هاسدورف حساسیت به تغییر در نرخ برداشت و نویز و دوران در دو فاصله اول یکسان بوده ولی فاصله سوم به نویز بسیار حساس بوده و تغییر نرخ برداشت مقادیر بیشتری نسبت به دوران دارد.

جدول ۴- تبدیلات زمانی

تبدیلات	انتقال ۵	انتقال ۱۰	انتقال ۱۵	انتقال ۲۰	انتقال ۲۵	کشیدگی ۲	کشیدگی ۳	کشیدگی ۴	کشیدگی ۵	کشیدگی ۱۰
تابع	اقلیدسی									
میانگین	۱/۶۶۶۶	۳/۳۳۳۳	۵	۶/۶۶۶۶	۸/۳۳۳۳	۱۴۴/۳۰	۲۸۸/۶۰	۴۳۲/۹۰	۵۷۷/۲۰	۱۲۹۸/۷۰
سطح اطمینان	۲۰۲/۸۱	۴۰۵/۶۳	۶۰۸/۴۴	۸۱۱/۲۶	۱۸۲۵/۳۴
تابع	DTW									
میانگین	۰/۰۰۰۱	۰/۰۰۰۱	۰/۰۰۰۱	۰/۰۰۰۱	۰/۰۰۰۱	۰/۰۰۰۱	۰/۰۰۰۱	۰/۰۰۰۱	۰/۰۰۰۱	۰/۰۰۰۱
سطح اطمینان	۰/۰۰۰۶	۰/۰۰۰۶	۰/۰۰۰۶	۰/۰۰۰۶	۰/۰۰۰۶	۰/۰۰۰۶	۰/۰۰۰۶	۰/۰۰۰۶	۰/۰۰۰۶	۰/۰۰۰۶
تابع	LCSS									
میانگین	۰/۲۵۶۴	۰/۴۹۶۳	۰/۶۴۰۵	۰/۷۵۰۱	۰/۸۱۷۵	۰/۹۴۷۸	۰/۹۷۰۸	۰/۹۷۸۹	۰/۹۸۲۹	۰/۹۸۹۸
سطح اطمینان	۰/۵۵۱۱	۰/۶۸۲۷	۰/۵۸۷۴	۰/۴۸۲۱	۰/۴۰۳۱	۰/۱۳۷۴	۰/۰۸۱۶	۰/۰۵۷۳	۰/۰۴۵۰	۰/۰۲۱۰
تابع	EDR									
میانگین	۰/۳۰۹۴	۰/۵۳۵۰	۰/۷۰۶۷	۰/۸۰۹۹	۰/۸۶۹۴	۰/۹۸۳۲	۰/۹۸۶۱	۰/۹۸۷۲	۰/۹۸۷۲	۰/۹۸۷۴
سطح اطمینان	۰/۵۸۹۷	۰/۶۷۶۶	۰/۵۴۲۶	۰/۴۳۷۰	۰/۳۵۶۰	۰/۰۱۸۱	۰/۰۱۱۱	۰/۰۰۹۶	۰/۰۰۹۶	۰/۰۰۹۰
تابع	ERP									
میانگین	۰/۰۴۰۶	۰/۰۷۷۸	۰/۱۲۱۹	۰/۱۴۸۳	۰/۱۸۰۷	۰/۴۶۴۷	۰/۴۹۹۸	۰/۵۱۲۲	۰/۵۱۷۶	۰/۵۲۴۲
سطح اطمینان	۰/۰۸۳۳	۰/۰۵۹۳	۰/۰۹۸۸	۰/۱۱۴۹	۰/۱۴۳۳	۰/۱۶۹۶	۰/۱۸۱۶	۰/۱۸۴۸	۰/۱۸۵۷	۰/۱۸۵۷

۵-۲-۲-۲-مقایسه تبدیلات زمانی

تابع هاسدورف و فرشت به علت نداشتن بعد زمانی مقایسه نمی‌شوند. از تابع EDR و LCSS به علت رفتار تقریباً یکسان حدآستانه، تنها حدآستانه ۱۵ متر مقایسه

تبدیلات زمانی، شامل انتقال زمان محلی به اندازه‌های ۵، ۱۰، ۱۵، ۲۰ و ۲۵ ثانیه (تبدیلات ۱-۵) و کشیدگی زمانی ۲، ۳، ۴، ۵ و ۱۰ برابر (تبدیلات ۶-۱۰) هستند.

نداشته باشد مشاهده نشده است. این در حالیست که در اکثر کاربردها توابع از چندین جنبه با هم تفاوت دارند. برای مثال ممکن است دو خطسیر مقایسه شونده دارای نویز، تفاوت نرخ برداشت و انتقال زمان محلی باشند. در چنین وضعیتی پیشنهاد تابع مناسب کار دشوار یا غیرممکنی است. توابع ارائه شده به جز هاسدورف که بر بعد مکانی تاکید دارد و EDR که برای حل انتقال و مقیاس، پیش‌پردازش نرمال را افزوده است، ارائه توابع دیگر صرفاً برای بررسی شباهت سری‌های زمانی بوده و کاربرد عمومی دارند و این به معنای عدم توجه به مشخصات و جنبه‌های خطسیر مکانی در تعریف توابع است. به علاوه خروجی توابع صرفاً با یک مقدار عددی بیان می‌شود که اطلاعات مفیدی درباره علت عدم شباهت دو خطسیر (مانند تفاوت مکانی، انتقال زمان محلی و ...) بیان نمی‌کند. از طرف دیگر بیان میزان شباهت به صورت عددی روش مناسبی برای تعیین میزان شباهت نیست. برای مثال مقدار $0/32$ در تابع DTW نمی‌تواند بیان‌کننده شبیه یا غیر شبیه بودن آن باشد ولی در مقایسه با عدد $0/97$ می‌توان آن را شبیه‌تر دانست. توابع متعدد دیگری نیز مطرح شده‌اند که اساس کار آن‌ها بر پایه‌ی روش‌های محاسباتی این توابع است. با توجه به نقاط ضعف توابع در بیان شباهت و مجهز نبودن آن‌ها به ابزار-های تحلیل مکانی-زمانی در حوضه خطوطسیر مکانی، ارائه تابعی که با توجه به نیاز کاربر شباهت را با تحلیل-های مناسب بررسی کرده و نتایج را به صورت ملموس و مفیدتر بیان کند پیشنهاد می‌شود.

می‌شود. در نتیجه، در مقایسه تبدیلات زمانی، توابع اقلیدسی، DTW، LCSS، EDR و ERP مقایسه می‌شوند. در جدول (۴)، نتایج حاصل از تبدیلات زمانی آورده شده است. برتری تابع DTW در انتقال زمان محلی و کشیدگی زمانی بر طبق ادعایی که مطرح شده کاملاً واضح است و با مقایسه مقادیر آن با تبدیلات زمانی-مکانی توانایی آن در مواجهه با این دو تبدیل زمانی کاملاً آشکار می‌شود.

دو تابع LCSS و EDR نیز رفتار شبیه به یکدیگر داشته و حساسیت آن‌ها به کشیدگی زمانی بیشتر از انتقال زمان محلی است. تابع اقلیدسی نیز بیشترین حساسیت را به تبدیلات زمانی-مکانی دارد. تابع ERP به تبدیلات زمانی-مکانی به خصوص کشیدگی زمانی حساسیت بالایی دارد.

۶- نتیجه‌گیری و پیشنهادها

توابع EDR و LCSS در نویز نسبت به دیگر توابع برتری دارند اما نیازمند حدآستانه مناسب برای رفع نویز هستند. تابع DTW خطوط سیر را مستقل از انتقال زمان محلی و کشیدگی زمانی بررسی کرده، در نتیجه برای محاسبه فاصله شباهت حاصل از اختلاف زمانی نقاط، این تابع پیشنهاد نمی‌شود. اما برای بررسی دو خط سیر مستقل از این دو تبدیل زمانی، این تابع پیشنهاد می‌شود. همه توابع به جز ERP به کاهش نرخ برداشت بیش از افزایش آن حساسیت داشته‌اند که یکی از دلایل آن می‌تواند ایجاد تغییر شکل در خطسیر باشد. در مجموع، تابعی که در همه‌ی تبدیلات موفق بوده و حساسیت به آن

مراجع

- [1] Agrawal, R., Faloutsos, C., & Swami, A. (1993). Efficient similarity search in sequence databases (pp. 69-84). Springer Berlin Heidelberg.
- [2] Rafiei, D., Mendelzon, A. (1998). Efficient retrieval of similar time sequences using DFT. arXiv preprint cs/9809033.
- [3] Laurinen, P., Siirtola, P., & Röning, J. (2006, February). Efficient Algorithm for Calculating Similarity between Trajectories Containing an Increasing Dimension. In Artificial Intelligence and Applications (pp. 392-399).
- [4] Chen, L., & Ng, R. (2004, August). On the marriage of lp-norms and edit distance. In Proceedings of the Thirtieth international conference on Very large data bases-Volume 30 (pp. 792-803).

- [5] Sherkat, R., & Rafiei, D. (2008). On efficiently searching trajectories and archival data for historical similarities. *Proceedings of the VLDB Endowment*, 1(1), 896-908.
- [6] Agarwal, P. K., Avraham, R. B., Kaplan, H., & Sharir, M. (2013, February). Computing the Discrete Fréchet Distance in Subquadratic Time. In *SODA* (pp. 156-167).
- [7] Driemel, A., & Har-Peled, S. (2013). Jaywalking your dog: computing the Fréchet distance with shortcuts. *SIAM Journal on Computing*, 42(5), 1830-1866.
- [8] Alt, H., & Godau, M. (1995). Computing the Fréchet distance between two polygonal curves. *International Journal of Computational Geometry & Applications*, 5(01n02), 75-91.
- [9] Yi, B. K., Jagadish, H. V., & Faloutsos, C. (1998, February). Efficient retrieval of similar time sequences under time warping. In *Data Engineering, 1998. Proceedings., 14th International Conference on* (pp. 201-208).
- [10] Vlachos, M., Kollios, G., & Gunopulos, D. (2002). Discovering similar multidimensional trajectories. In *Data Engineering, 2002. Proceedings. 18th International Conference on* (pp. 673-684).
- [11] Chen, L., Özsü, M. T., & Oria, V. (2005, June). Robust and fast similarity search for moving object trajectories. In *Proceedings of the 2005 ACM SIGMOD international conference on Management of data* (pp. 491-502).
- [12] Lee, J. G., Han, J., & Whang, K. Y. (2007, June). Trajectory clustering: a partition-and-group framework. In *Proceedings of the 2007 ACM SIGMOD international conference on Management of data* (pp. 593-604).
- [13] Zhang, Z., Huang, K., & Tan, T. (2006, August). Comparison of similarity measures for trajectory clustering in outdoor surveillance scenes. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on* (Vol. 3, pp. 1135-1138).
- [14] Wang, H., Su, H., Zheng, K., Sadiq, S., & Zhou, X. (2013, January). An effectiveness study on trajectory similarity measures. In *Proceedings of the Twenty-Fourth Australasian Database Conference-Volume 137* (pp. 13-22).
- [15] de Vries, G. K. D., & van Someren, M. W. (2014). A Comparison Between Alignment and Integral Based Kernels for Vessel Trajectories.
- [16] Zheng, Y., Zhang, L., Xie, X., & Ma, W. Y. (2009, April). Mining interesting locations and travel sequences from GPS trajectories. In *Proceedings of the 18th international conference on World wide web* (pp. 791-800).
- [17] Zheng, Y., Li, Q., Chen, Y., Xie, X., & Ma, W. Y. (2008, September). Understanding mobility based on GPS data. In *Proceedings of the 10th international conference on Ubiquitous computing* (pp. 312-321).
- [18] Zheng, Y., Xie, X., & Ma, W. Y. (2010). GeoLife: A Collaborative Social Networking Service among User, Location and Trajectory. *IEEE Data Eng. Bull.*, 33(2), 32-39.