

# بازسازی خطوط سیر اشیاء متحرک با استفاده از روش تشابه پیچش زمانی پویا زمینه-مبنا

میلاد جمالی<sup>۱</sup>، علی اصغر آل‌شیخ<sup>۲</sup>، محمد شریف<sup>۳\*</sup>

<sup>۱</sup> دانشجوی کارشناسی ارشد سیستم اطلاعات مکانی، دانشکده مهندسی نقشه‌برداری، دانشگاه صنعتی خواجه

نصیرالدین طوسی

miladjamali88@yahoo.com

<sup>۲</sup> استاد گروه سیستم اطلاعات مکانی، دانشکده مهندسی نقشه‌برداری، دانشگاه صنعتی خواجه نصیرالدین طوسی

alesheikh@kntu.ac.ir

<sup>۳</sup> استادیار گروه جغرافیا، دانشکده علوم انسانی، دانشگاه هرمزگان

m.sharif@hormozgan.ac.ir

(تاریخ دریافت: تیر ۱۴۰۲، تاریخ تصویب: شهریور ۱۴۰۲)

## چکیده

با رشد روزافزون فناوری‌های تعیین موقعیت و استفاده از سامانه‌های ناوبری، امروزه حجم وسیعی از داده‌های حرکتی اشیاء متحرک مانند افراد، خودروها، کشتی‌ها و حیوانات در اختیار می‌باشد. با این وجود عدم تمامیت و ناقص بودن این داده‌ها به دلایل دستگاهی، انسانی و محیطی، تجزیه و تحلیل خطوط سیر و بکارگیری موثر آن‌ها را در زمینه‌های مختلف به چالش می‌کشد. از اینرو بازتولید داده‌های از دست رفته نقش مهمی در استفاده حداکثری از ظرفیت داده‌های حرکتی بویژه در حوزه ناوبری و ردیابی مسیر دارد. در این پژوهش با بهره‌گیری از روش اندازه‌گیری تشابه خطوط سیر اشیاء، به بازسازی خطوط سیر حاوی گپ پرداخته شده‌است. در این راستا از روش پیچش زمانی پویا زمینه-مبنا (CDTW) به همراه پارامترهای حرکتی سرعت و جهت برای اندازه‌گیری شباهت و بازسازی خطوط سیر کشتی‌ها در دو منطقه از اقیانوس اطلس و آرام استفاده شده‌است. دو سازوکار استفاده از تعداد خطوط سیر مشابه مشخص و تعداد خطوط سیر از حد آستانه مشخص برای بازسازی در نظر گرفته شده‌است. نتایج نشان‌دهنده آن است که استفاده از تعداد خطوط سیر مشابه مشخص در مقایسه با تعریف حد آستانه، باعث کاهش خطای مجذور میانگین مربعات و میانگین مطلق از ۱/۵ و ۱/۴ درجه به ۰/۵ و ۰/۴ درجه می‌شود. همچنین، افزایش طول خطوط سیر باعث بهبود مقادیر مجذور میانگین مربعات و میانگین مطلق از ۰/۵ به ۰/۱ درجه در حالت تعداد خطوط سیر مشخص و ۱/۵ به ۰/۳ درجه در حالت حد آستانه شده‌است.

**واژگان کلیدی:** خط سیر، گپ، اندازه‌گیری تشابه، داده از دست رفته، سامانه شناسایی خودکار (AIS)

## ۱- مقدمه

داده‌های اشیاء متحرک نقطه‌ای مانند انسان، وسایل نقلیه و حیوانات به شکل مختصات‌های مکانی در طول زمان ذخیره می‌شوند. خطوط مسیر این اشیاء با دنباله‌ای از نقاط متوالی مرتب شده نمایش داده می‌شوند که از مختصات‌های مکانی به همراه بُعد زمان تشکیل شده‌اند [۱]. به منظور تجزیه و تحلیل دقیق خطوط مسیر، کامل بودن آن‌ها از اهمیت بالایی برخوردار است. منظور از کامل بودن یا تمامیت داده، عدم وجود شکاف یا گپ در خط‌سیر است. دلایل متعددی موجب ایجاد گپ در داده‌های حرکتی می‌شوند که از جمله آن‌ها می‌توان به از دست رفتن سیگنال، نوفه موجود در سیگنال‌ها یا خاموش بودن عمدی یا سهوی دستگاه‌ها و حسگرهای ردیابی مسیر اشاره نمود. بسته به مدت زمان در دسترس نبودن سیگنال (که می‌تواند از چند دقیقه تا یک روز کامل باشد)، داده‌های از دست رفته تأثیرات متفاوتی بر روی نتایج تحلیل‌ها یا فرآیندهای اکتشاف الگو دارند [۲].

روش‌های متعددی برای بازسازی گپ در خطوط مسیر اشیاء از جمله روش‌های آماری [۳]، درونیابی [۴]، یادگیری ماشین [۵]، شبکه عصبی [۶] و مبتنی بر اندازه‌گیری شباهت خطوط مسیر [۷] ارائه شده‌است که در ادامه به آن‌ها اشاره شده‌است.

یکی از روش‌های برخورد با داده‌های از دست رفته و بازسازی آن‌ها، تحلیل‌های آماری برای شناسایی و تشخیص وابستگی بین داده‌های از دست رفته و مجموعه داده‌ها بوده‌است. در یک نمونه از این روش، میزان وابستگی ارتباط داده‌های از دست رفته با ویژگی‌های جمعیتی مجموعه داده‌ها از جمله سن و جنسیت و درآمد و غیره مورد بررسی قرار گرفته بود [۳]. در حالتی دیگر از روش‌های آماری برای برخورد با داده‌های از دست رفته، بحث جایگزینی گپ با مقادیر میانگین محاسبه شده از داده‌های موجود از همان شی متحرک یا مجموعه‌ای از داده‌ها مطرح بوده‌است [۸]. با این حال روش‌های آماری برای شناسایی وابستگی بین ویژگی‌های شی متحرک با داده‌های از دست رفته، علاوه بر هزینه‌های محاسباتی سنگین مستلزم دسترسی به حجم عظیمی از داده‌های شی متحرک به همراه ویژگی‌های آن‌ها می‌باشد و قابل تعمیم به تمامی اشیاء متحرک نمی‌باشد.

یکی دیگر از روش‌های متداول بازسازی خطوط مسیر اشیاء متحرک، روش درونیابی می‌باشد. روش‌های درونیابی به طور گسترده در پیش‌بینی خطوط مسیر اشیاء متحرک مورد استفاده قرار گرفته‌اند. با این حال از آنجائیکه گپ نیز دنباله‌ای از موقعیت‌های مشخص مکانی شی متحرک را پیش از خود دارد، این روش در بازسازی خطوط مسیر اشیاء نیز مورد توجه قرار گرفته‌است. درونیابی حالات مختلفی از جمله درونیابی خطی، کینماتیک، مکعبی، منحنی و گام تصادفی مقید و غیره را شامل می‌شود [۴]. بسته به نوع داده‌ها و هندسه خط‌سیر شی متحرک می‌توان از حالات مختلف درونیابی برای بازسازی خط‌سیر شی متحرک استفاده نمود. به عنوان نمونه در صورتی که حرکت شی متحرک دارای سرعت و شتاب نباشد، درونیابی خطی مورد استفاده قرار می‌گیرد. در بسیاری از موارد، عملکرد روش درونیابی خطی به خوبی روش‌های پیچیده یا حتی بهتر از آن‌ها بوده‌است. به عنوان مثال در مورد داده‌های مسیریابی حیوانات، مخصوصاً پستانداران خشکی که حرکات آن‌ها به طور مداوم توسط گام‌های تصادفی مدل شده [۹]، درونیابی خطی روشی مناسب بوده‌است. در پژوهشی دیگر درونیابی منحنی برای پستانداران دریایی عملکرد بهتری از درونیابی خطی را نمایش می‌دهد [۱۰]. اما در مورد اشیائی که حرکات سریع دارند و داده‌های حرکتی آن‌ها در بازه زمانی نسبتاً بالایی جمع‌آوری شده‌است (مانند دوچرخه‌سواران، موتورسواران و ورزشکاران)، روش کینماتیک روش مناسبی برای درونیابی موقعیت اشیاء می‌باشد. در یک نمونه لانگ [۴] از درونیابی کینماتیک برای بازیابی مسیر اشیاء متحرک استفاده کرد. درونیابی کینماتیک برای ردیابی اشیاء متحرک با وسعت بزرگ همچون طوفان‌ها که تأثیرات کینماتیک را در محدوده وسیع مکانی-زمانی به نمایش گذاشته‌اند، مناسب بوده‌است. از نقاط ضعف روش کینماتیک می‌توان به عملکرد ضعیف‌تر آن در خطوط مسیر با سرعت ثابت نسبت به سایر حالات درونیابی اشاره کرد. با این حال تمامی روش‌های مذکور درونیابی تنها از خود داده‌های خط‌سیر برای بازسازی و درونیابی استفاده می‌کنند و عدم استفاده از الگوهای حرکتی مشابه و شباهت خطوط مسیر را می‌توان از نقاط ضعف این روش‌ها برشمرد.

از سایر روش‌های مورد استفاده برای بازسازی خطوط مسیر، می‌توان به روش‌های یادگیری ماشین اشاره

نمود. به عنوان نمونه لیانگ و همکاران [۶] روشی را برای بازسازی خودکار مسیر حرکت کشتی‌ها با استفاده از تلفیق روش‌های یادگیری ماشین ارائه کردند. نتایج نشان داده که مسیر بازیابی شده توسط این روش در مقایسه با روش درونیایی مکعبی شباهت بیشتری به مسیر اصلی داشته‌است. از محدودیت‌های این روش می‌توان به عدم در نظر گرفتن عوامل محیطی از قبیل آب و هوا و جزر و مد اشاره کرد. بطور کلی روش‌های یادگیری ماشین برای بازسازی خطوط‌سیر، نیازمند حجم زیادی از داده‌ها برای آموزش مدل می‌باشند که این مساله هزینه محاسباتی سنگینی را ایجاد می‌کند و این نکته را می‌توان از نقاط ضعف این روش‌ها عنوان نمود.

در برخی روش‌ها سعی شده تا از تلفیق روش‌های درونیایی و شبکه‌عصبی برای بازسازی استفاده شود [۱۱]. از نقاط قوت این روش می‌توان به حذف داده‌های اضافی اولیه به منظور رسیدن به بیشترین همبستگی اشاره نمود. با این حال به دلیل پیچیدگی‌ها و هزینه‌های محاسباتی روش‌های ترکیبی، تعمیم این نوع روش‌ها برای سایر داده‌های خطوط‌سیر مرسوم با ساختار کاملاً متفاوت (افراد پیاده و وسایل نقلیه) مشخص نیست. همچنین، در این روش بدلیل بهره‌گیری از درونیایی، تنها از یک مجموعه داده برای بازسازی استفاده شده و امکان استفاده از تشابه خطوط‌سیر و الگوهای تکراری در داده‌های گذشته وجود ندارد.

روش دیگری که برای بازسازی خطوط‌سیر مورد استفاده قرار گرفته، روش‌های مبتنی بر شباهت خطوط‌سیر و الگوهای موجود در آن‌ها می‌باشد. به عنوان مثال پاتاک و همکاران [۷] روشی را برای بازسازی خطوط‌سیر کشتی‌ها با استفاده از اندازه‌گیری تشابه خطوط‌سیر ارائه کردند. مزیت این روش استفاده از مجموعه‌ای از خطوط‌سیر همسایه‌ها در فرآیند بازسازی بوده تا برخلاف روش‌های دیگر از رسیدن به جواب واحد جلوگیری و عدم قطعیت نیز در فرآیند بازسازی گنجانده شود. روش‌های شباهت‌یابی به نسبت روش‌های درونیایی خطی و مکعبی از دقت بالاتری برخوردار می‌باشند. با این حال می‌توان در کنار بهره‌گیری از شباهت‌یابی خطوط‌سیر از پارامترهای حرکتی همچون سرعت و جهت نیز در فرآیند بازسازی استفاده نمود.

استفاده از الگوهای حرکتی مشابه در منطقه و بکارگیری شباهت خطوط‌سیر در کنار پارامترهای حرکتی، نقش بسزایی در فرآیند بازسازی گپ موجود در خطوط‌سیر

دارند، که به ندرت این مساله مورد توجه قرار گرفته‌است. رفتارهای حرکتی اغلب توسط خود شی یا سایر اشیاء متحرک در طول بازه زمانی و مکان مشخص تکرار می‌شوند. به عنوان مثال خودروها تنها می‌توانند از بزرگراه‌ها و سایر مسیرهای قابل تردد استفاده کنند؛ یا افراد پیاده از معابر عمومی و پیاده‌روها برای عبور و مرور استفاده می‌کنند؛ یا حتی کشتی‌ها نیز با وجود مقید نبودن در آب‌ها، از مسیرهای مشخصی برای تردد استفاده می‌کنند. به همین دلیل اغلب، خطوط‌سیر مشابهی از شی متحرک در بازه‌های زمانی یا مکانی مختلف موجود است و می‌توان از آن‌ها برای محاسبه میزان شباهت استفاده نمود.

شباهت خطوط‌سیر عموماً به عنوان شباهت در ابعاد مکانی و زمانی و بصورت مطلق یا نسبی تعریف می‌شود. در بیشتر کاربردهای اندازه‌گیری شباهت خطوط‌سیر، بعد زمان از اهمیت کمتری نسبت به بعد مکان برخوردار است. در کاربردهایی که هدف، اندازه‌گیری شباهت مکانی و استخراج الگوهای حرکتی خطوط‌سیر می‌باشند، مجاورت مکانی خطوط‌سیر اهمیت بیشتری دارد [۱۲]. درک شباهت مکانی بین خطوط‌سیر اشیاء به بازسازی هر چه بهتر خطوط‌سیر و داده‌های از دست رفته کمک شایانی می‌کند. با این حال شباهت خطوط‌سیر به اندازه‌گیری شباهت مکانی یا زمانی محدود نمی‌شود و می‌توان در کنار ابعاد مکانی و زمانی از پارامترهای حرکتی نیز برای مقایسه و ارزیابی میزان شباهت خطوط‌سیر استفاده نمود. بنابراین، در این پژوهش از پارامترهای حرکتی سرعت و جهت برای اندازه‌گیری شباهت خطوط‌سیر استفاده شده‌است. همچنین، تاثیر طول خطوط‌سیر و مکان و طول گپ بر دقت بازسازی مورد بررسی قرار گرفته‌است.

روش‌های مختلفی برای اندازه‌گیری میزان شباهت خطوط‌سیر بر اساس سری‌های زمانی وجود دارد، مانند فاصله اقلیدسی<sup>۱</sup>، فاصله ویرایش<sup>۲</sup>، طولانی‌ترین دنباله مشترک<sup>۳</sup> و روش پیچش زمانی پویا<sup>۴</sup> (DTW) که هر کدام دارای مزایا و معایب خاص خود هستند [۱۲-۱۴]. در این پژوهش از روش پیچش زمانی پویا زمینه‌مبنای<sup>۵</sup> (CDTW)

۱ Euclidean Distance

۲ Edit Distance

۳ Longest Common Subsequence

۴ Dynamic Time Warping

۵ Context-based Dynamic Time Warping

که از توابع قدرتمند در زمینه اندازه‌گیری شباهت خطوط مسیر می‌باشد، استفاده شده است [۱۴]. این روش از هیچ حد آستانه‌ای برای تناظریابی بین نقاط خطوط مسیر استفاده نمی‌کند و برخلاف روش طولانی‌ترین دنباله مشترک که تناظریابی برای برخی نقاط صورت نمی‌گیرد، هر نقطه از خط مسیر با یک یا چند نقطه از خط مسیر دیگر متناظر می‌شود. روش CDTW قادر است تا در کنار بعد مکان از انواع پارامترهای حرکتی و اطلاعات زمینه‌ای ثبت شده حین حرکت نیز در فرآیند شباهت‌یابی استفاده نماید. هدف این پژوهش بازسازی گپ موجود در خطوط مسیر کشتی‌ها با استفاده از روش اندازه‌گیری تشابه و پارامترهای حرکتی است. مشارکت این پژوهش شامل بهره‌گیری از پارامترهای حرکتی سرعت و جهت حرکت کشتی‌ها در فرآیند شباهت‌یابی و بازسازی و همچنین ارزیابی تاثیر طول و مکان گپ در دقت بازسازی می‌باشد. روش مذکور روی خطوط مسیر کشتی‌ها در دو منطقه در اقیانوس اطلس و آرام پیاده‌سازی و ارزیابی می‌شوند.

## ۲- مواد و روش

### ۲-۱- منطقه مطالعاتی

در این پژوهش از داده‌های سامانه شناسایی خودکار<sup>۱</sup> (AIS) کشتی‌ها در سال ۲۰۱۷ در دو منطقه از اقیانوس اطلس و آرام واقع در زون‌های ۱۰ و ۱۸ استفاده شده است (شکل ۱). منطقه اول در عرض جغرافیایی ۲۶ درجه شمالی تا ۳۴ درجه شمالی و طول جغرافیایی ۱۲۹- درجه غربی تا ۱۱۹- درجه غربی به ابعاد ۵۷۰ در ۹۰۰ کیلومتر مربع و منطقه دوم در عرض جغرافیایی ۳۷ درجه شمالی تا ۳۹ درجه شمالی و طول جغرافیایی ۷۴- درجه غربی تا ۷۲- درجه غربی به ابعاد ۱۹۲ در ۱۰۰ کیلومتر مربع واقع شده‌اند. مناطق مورد مطالعه بدلیل مجاورت به سواحل آمریکای شمالی، مسیر عبور و مرور کشتی‌های زیادی می‌باشند و از تراکم خطوط مسیر بالایی برخوردار هستند، بطوری که مسیرهای تردد توسط خود کشتی و یا کشتی‌های دیگر تکرار شده و می‌توان از این تاریخچه برای شناسایی خطوط مسیر مشابه در منطقه استفاده نمود. از دیگر مزایای مجاورت مناطق مورد استفاده به سواحل آمریکای

شمالی، در اختیار داشتن ابتدا و یا انتهای خطوط مسیر مورد مطالعه در دو منطقه می‌باشد. با این حال با وجود نزدیکی مناطق مورد مطالعه به ساحل و تراکم بالای خطوط مسیر در هر دو منطقه، خطوط مسیر موجود در این مناطق ساختار متفاوتی از لحاظ تراکم نقاط نسبت به هم دارند که این مساله منجر به تولید خطوط مسیر با طول‌های متفاوت در دو منطقه گردیده که در ادامه به آن اشاره می‌گردد.

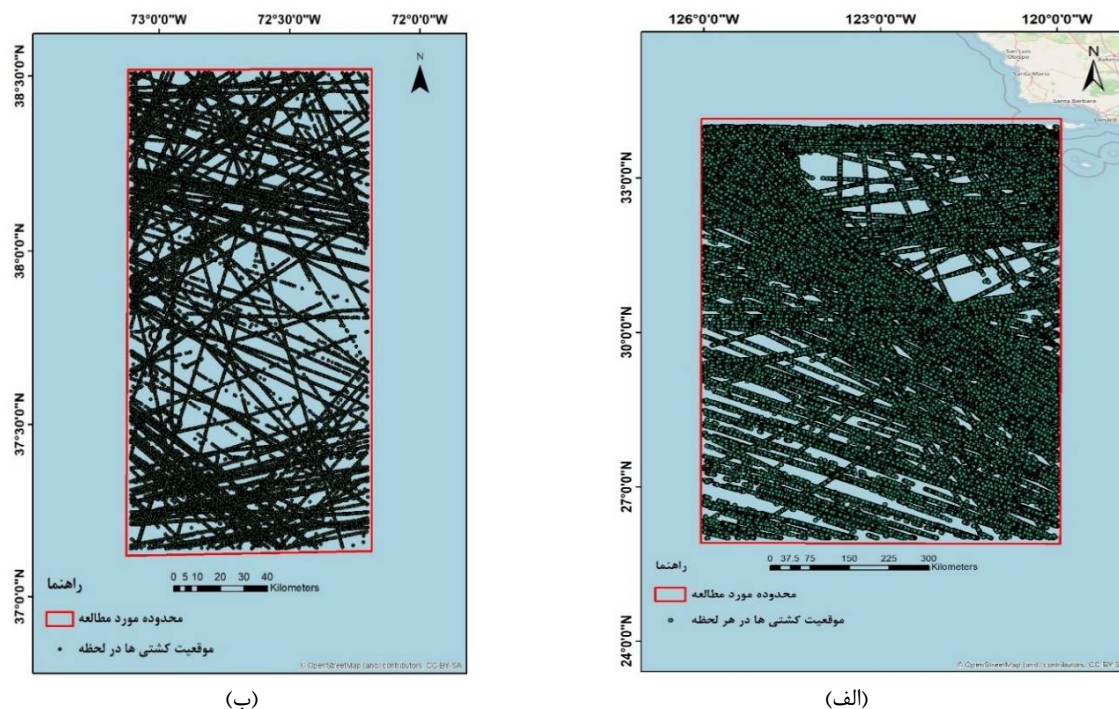
### ۲-۲- داده

AIS سامانه‌ای است که برای اهداف گوناگونی همچون مسیر یابی خودکار، نظارت بر ترافیک دریایی و امنیت در حمل و نقل دریایی کشتی‌ها مورد استفاده قرار می‌گیرد. AIS توسط کمیته‌هایی فنی، به عنوان فناوری برای جلوگیری از برخورد کشتی‌های بزرگ در دریا که در محدوده سامانه‌های مستقر در ساحل نیستند، توسعه داده شد. این فناوری هر کشتی را به صورت جداگانه به همراه موقعیت و حرکات خاص آن شناسایی می‌کند و امکان ایجاد یک تصویر مجازی در زمان واقعی را فراهم می‌کند. کشتی‌های مجهز به گیرنده‌های AIS را می‌توان توسط ایستگاه‌های پایه AIS واقع در امتداد خطوط ساحلی یا زمانی که خارج از محدوده شبکه‌های زمینی هستند، از طریق تعداد فزاینده‌ای از ماهواره‌هایی که به گیرنده‌های AIS ویژه مجهز شده‌اند و قادر به رفع تعارض تعداد زیادی از شناسه‌ها هستند، ردیابی کرد.

داده‌های AIS حاوی اطلاعات مختلف ایستا و پویا مانند موقعیت، سرعت، وضعیت بارگیری، نوع کشتی، جهت و غیره هستند. در این پژوهش از نمونه داده‌های نمایش داده شده در جدول ۱ استفاده شده است. پس از پیش‌پردازش داده‌ها با توجه به تراکم متفاوت خطوط مسیر در هر منطقه، در منطقه اول تعداد ۱۰۴۴ خط مسیر و در منطقه دوم ۲۸۸ خط مسیر باقی ماندند. همچنین، متوسط نرخ جمع‌آوری داده‌ها یا به عبارت دیگر فاصله زمانی که هر کشتی موقعیت خود را برای دستگاه‌های ردیابی مسیر ارسال می‌کند، در هر منطقه ۶۰ تا ۷۰ ثانیه می‌باشد. همان‌طور که در جدول ۱ نشان داده شده است، پارامترهای حرکتی موجود در داده‌های AIS شامل سرعت حرکت کشتی به نسبت زمین<sup>۲</sup>

<sup>۲</sup> Speed Over Ground

<sup>۱</sup> Automatic Identification System



(ب)

(الف)

شکل ۱- خطوط سیر کشتی‌ها واقع در (الف) اقیانوس اطلس و (ب) اقیانوس آرام

جدول ۱- نمونه‌ای از داده‌های AIS

MMSI	Base Data Time	Lat	Lon	(degree) COG	(knot) SOG	Heading	Vessel Name
۳۶۷۰۱۵۸۸۰	۲۰۱۷-۰۷-۰۱T۰۰:۰۰:۰۳	۴۰/۶۳۸۹۴	-۷۴/۱۵۵۷۹	-۱۰۸/۱	۰/۷	۵۱۱	ROBERT IV
۳۶۷۷۷۸۶۷۰	۲۰۱۷-۰۷-۰۱T۰۰:۰۱:۵۴	۴۵/۵۱۰۴۵	-۷۳/۵۴۸۷۹	-۴۹/۶	۱/۶	۵۱۱	H201

پیش‌پردازش و پاکسازی داده‌ها، اولین مرحله در استفاده از این داده‌ها می‌باشد. پیش‌پردازش داده‌های AIS بسته به هدف استفاده از این داده‌ها می‌تواند متفاوت باشد. در مطالعاتی که تجزیه و تحلیل خطوط سیر و حرکت کشتی‌ها مدنظر می‌باشند، حذف نقاطی که در آن‌ها کشتی در حالت سکون یا نزدیک به سکون قرار دارد، یکی از مراحل پیش‌پردازش است. به منظور تشکیل خطوط سیر، در ابتدا نقاطی که اطلاعات ناقص دارند و شناسه منحصر بفرد کشتی<sup>۲</sup> (MMSI) آن‌ها مشخص نمی‌باشد، از پایگاه داده حذف می‌شوند و نقاطی که در آن‌ها کشتی‌ها در حالت سکون هستند و سرعت آن‌ها از مقدار مجاز تعیین شده (یک نات) پایین‌تر باشند، حذف می‌شوند [۱۵]. پس از آن خطوط سیر هر کشتی با استفاده از MMSI تشکیل می‌شوند. لازم به ذکر است که این خطوط سیر، تمامی

(SOG) و زاویه حرکت کشتی با زمین<sup>۱</sup> (COG) می‌باشند. اطلاعات دیگر شامل نوع کشتی، نام کشتی، وضعیت بارگیری و دیگر داده‌های ایستا و توصیفی نیز در فرآیند تجزیه و تحلیل خطوط سیر کشتی بی‌تاثیر و در ادامه این پژوهش از آن‌ها صرف‌نظر شده‌است.

## ۲-۳- روش

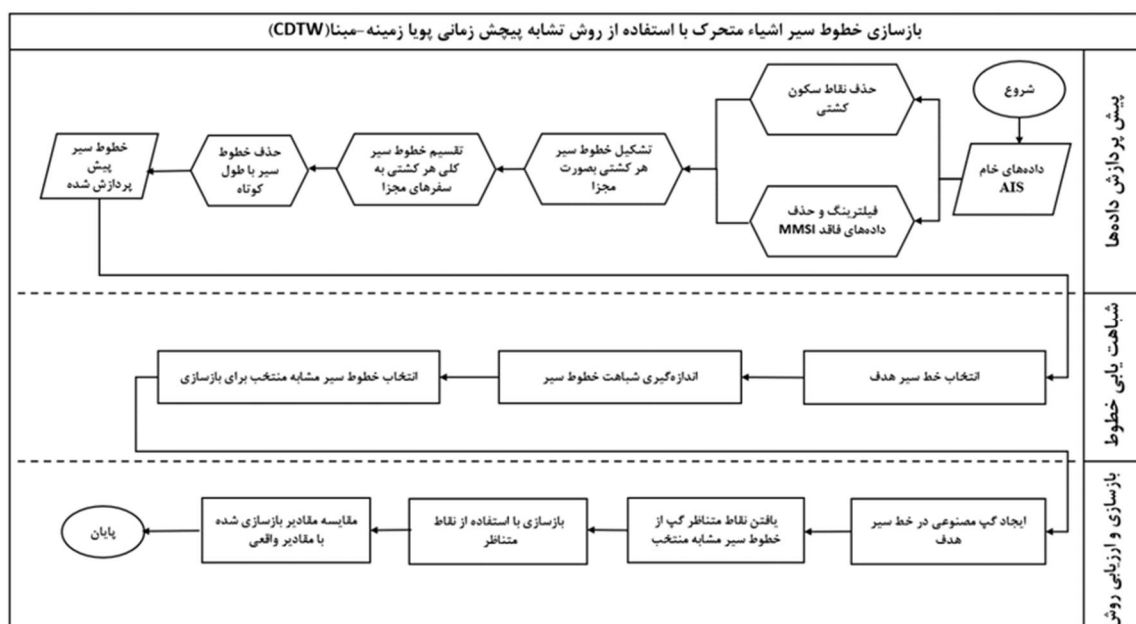
بازسازی خطوط سیر در سه مرحله پیش‌پردازش داده، شباهت‌یابی خطوط سیر و بازسازی و ارزیابی انجام می‌پذیرد که جزئیات آن در شکل ۲ نمایش داده شده‌است.

## ۲-۳-۱- پیش‌پردازش داده‌ها و تشکیل خطوط سیر

از آنجایی که داده‌های AIS بصورت خام از ماهواره‌های ردیابی مسیر بدست می‌آیند، اغلب دارای نوفه می‌باشند و

<sup>۲</sup> Maritime Mobile Service Identity

<sup>۱</sup> Course Over Ground



شکل ۲- روندنمای بازسازی خطوط سیر مبتنی بر اندازه‌گیری شباهت

کرده‌است (شکل ۳). در این پژوهش از این روش برای اندازه‌گیری تشابه خطوط سیر استفاده شده‌است. CDTW امان‌هایی از خطوط سیر که اطلاعات زمینه‌ای مشابه اما در بعد زمانی با یکدیگر تفاوت دارند را به یکدیگر متناظر می‌کند. برای اندازه‌گیری شباهت چندبُعدی بین خطوط سیر، خط سیر  $S$  دارای  $n$  نقطه و خط سیر  $T$  دارای  $m$  نقطه را فرض نمایید (روابط ۱ و ۲).

$$S[s_1, s_2, \dots, s_n] \quad (1)$$

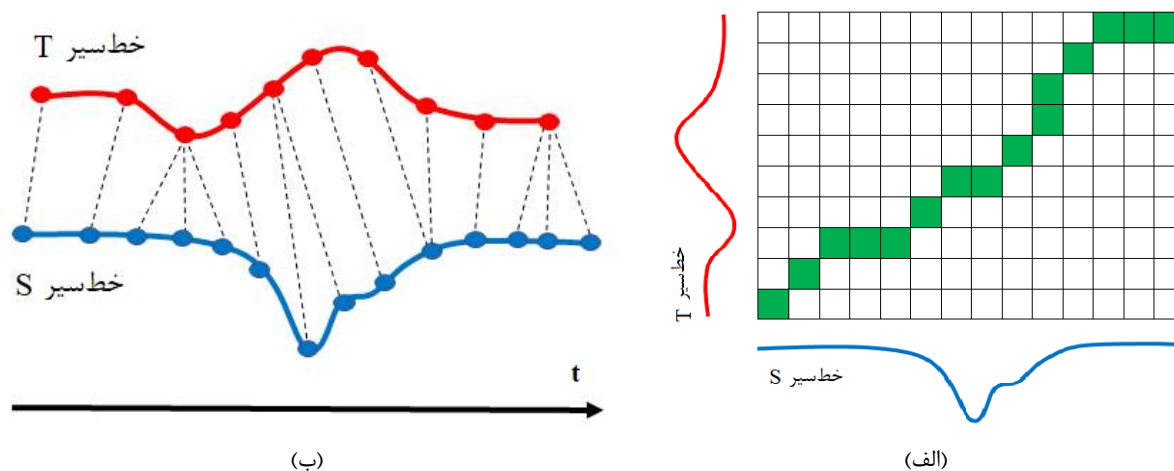
$$T[t_1, t_2, \dots, t_m] \quad (2)$$

در این خطوط سیر هر امان و نقطه می‌توانند شامل  $k$  بُعد باشند و هر کدام از این ابعاد با  $C$  تعریف می‌شوند. به منظور از بین بردن اثر مقیاس و همچنین سهولت در نمایش و تفسیر نتایج، پیش از اندازه‌گیری شباهت مقادیر خطوط سیر بین صفر و یک نرمال می‌شوند. در ادامه مراحل زیر برای محاسبه مقدار CDTW بین دو خط سیر اجرا می‌شود. در ابتدا برای دو خط سیر  $S$  و  $T$  با طول‌های به ترتیب  $n$  و  $m$  نقطه، یک ماتریس تهی ( $M$ ) با اندازه  $n$  در  $m$  ایجاد می‌شود. در ادامه از پایین سمت چپ ماتریس ( $M_{0,0}$ ) تا بالا سمت راست آن ( $M_{n,m}$ )، فاصله محلی بین امان‌های خطوط سیر با یکی از روش‌های فاصله مینوسکی یا فاصله اقلیدسی و یا مجذور فاصله اقلیدسی که در رابطه (۳) آورده شده‌اند، محاسبه می‌شوند.

سفرهای کشتی در طول یک ماه را شامل می‌شود و هر خط سیر شامل چندین سفر مجزا در بازه‌های زمانی مختلف می‌باشد و می‌بایست این خطوط به سفرهای مجزا تقسیم شوند. بدین منظور چنانچه اختلاف زمانی میان دو نقطه متوالی از خط سیر یک کشتی، بیش از ۳۰ دقیقه باشد، سفرها از یکدیگر مجزا شده‌اند. در ادامه این پژوهش منظور از خطوط سیر، همین سفرهای مجزا شده‌است. همچنین، به منظور افزایش دقت شباهت‌یابی و در نظر گرفتن حداقل ۱۰ نقطه گپ مصنوعی، خطوط سیری که کمتر از ۱۰ نقطه داشته باشند از کل فرآیند شباهت‌یابی و بازسازی خارج گردیده‌اند.

## ۲-۳-۲ روش تشابه پیچش زمانی پویا زمینه-مبنا (CDTW)

روش DTW یک تابع پویا است که در ابتدا برای تشخیص صوت استفاده می‌شد. با این حال روشی متداول در اندازه‌گیری شباهت بین دنباله‌های زمانی با اندازه‌های متفاوت است. این روش هر نقطه از خط سیر را با یک یا بیش از چند نقطه از خط سیر دیگر تطابق می‌دهد و بهینه‌ترین جفت را انتخاب می‌کند. حالت‌های مختلفی از روش DTW توسعه و مورد استفاده قرار گرفته‌است. روش CDTW [۱۴] در کنار داده‌های مکانی از پارامترهای حرکتی و اطلاعات زمینه‌ای در فرآیند اندازه‌گیری تشابه خطوط سیر استفاده



شکل ۳- عملکرد روش CDTW: الف) تطبیق الاستیک به دست آمده و ب) همترازی خطوط مسیر [۱۴]

تجمعی از مجموع مقادیر مسیر پیچشی خواهد بود که از طریق رابطه (۴) محاسبه می‌شود.

$$G = \sum d(s_i, t_j) \quad (4)$$

با این حال الزاماً  $G$  را نمی‌توان به عنوان مسیر بهینه بین دو خط‌سیر تلقی کرد. به منظور محاسبه مسیر بهینه دو خط‌سیر (فاصله CDTW) در مرحله بعدی برنامه‌نویسی پویا بر روی ماتریس  $M$  طبق رابطه (۵) پیاده‌سازی می‌شود.

$$\forall i, j, k: i = 1, \dots, n, \quad j = 1, \dots, m, \quad k = 1, \dots, x$$

$$CDTW(\varphi, \varphi) = 0$$

$$CDTW(s_i, \varphi) = CDTW(\varphi, t_j) = \infty$$

$$CDTW_k(s_i, t_j) = (C_k(s_i) - C_k(t_j))^2 + \min \begin{cases} CDTW_k(s_i, t_{j-1}) \\ CDTW_k(s_{i-1}, t_j) \\ CDTW_k(s_{i-1}, t_{j-1}) \end{cases} \quad (5)$$

وزن مربوط به هر پارامتر ( $w_k$ ) در فاصله CDTW ضرب می‌شود که در آن طبق رابطه (۶) تعریف می‌شود.

$$w_k: S \times T \rightarrow [0, 1] \quad (6)$$

مقدار صفر بیانگر عدم وزندهی یا عدم حضور پارامتر زمینه‌ای و مقدار ۱ بیانگر بیشترین مقدار وزندهی به پارامتر مورد نظر می‌باشد. در نهایت  $w$  به عنوان شباهت نسبی بین خطوط‌سیر  $S$  و  $T$  خواهد بود که طبق رابطه (۷) تعریف می‌شود.

$$M_{i,j}^k = d(s_i, t_j) = L_p = (|s_i - t_j|^p)^{1/p} \quad (3)$$

با پُر کردن درایه‌های ماتریس، مسیرهای متعددی از  $M_{0,0}$  تا  $M_{i,j}$  که مسیرهای پیچشی اطلاق می‌شوند، ایجاد می‌شوند. هر کدام از مسیرهای پیچشی خطوط‌سیر  $S$  و  $T$  را به گونه‌ای متناظر می‌کنند که در آن تمامی المان‌ها به حداقل یک المان از خط‌سیر دیگر جفت می‌شوند. بر همین اساس، فاصله کلی بین خطوط‌سیر  $S$  و  $T$  یک مقدار

که در رابطه (۵)  $\varphi$  بیانگر خط‌سیر تهی است. نتایج این معادلات به یک ماتریس جدید وارد می‌شود که ماتریس CDTW نامیده می‌شود. در زمان تکمیل این ماتریس، مسیر بهینه پیچشی می‌تواند از  $CDTW_{m,n}$  تا  $CDTW_{0,0}$  بازیابی شود. در ادامه، بهترین فاصله  $CDTW$  بین دو خط‌سیر  $S$  و  $T$  مقدار درایه بالا سمت راست ماتریس CDTW خواهد بود. با این حال در فرآیند اندازه‌گیری شباهت خطوط‌سیر در ابعاد و کاربردهای مختلف، یک بعد می‌تواند نسبت به بعدهای دیگر از اهمیت بالاتری برخوردار باشد. به همین دلیل در مرحله بعدی،

$$W_{S,T} = \sum_{k=1}^x CDTW_k(s_i, t_j) * w_k \quad (7)$$

خروجی روش CDTW یک ماتریس فاصله می باشد که از یکسری المان های W تشکیل شده که شباهت نسبی میان تمام خطوط سیر را در تمامی ابعاد (X) نمایش می دهد. ماتریس فاصله یک ماتریس مجذور متقارن می باشد که ابعاد آن بستگی به تعداد خطوط سیر دارد [۱۴]. با این حال در این پژوهش وزن های در نظر گرفته شده برای موقعیت مکانی برابر با ۰/۶ و برای پارامترهای حرکتی سرعت و جهت ۰/۲ می باشند [۱۵].

### ۲-۳-۳- بازسازی خطوط سیر

استفاده از تمامی خطوط سیر موجود در پایگاه داده در فرآیند بازسازی، علاوه بر کاهش دقت، هزینه محاسباتی زیادی نیز به همراه دارد. به همین دلیل پس از محاسبه میزان شباهت تمامی خطوط سیر به خط سیر دارای گپ، بسته به روش مورد استفاده می توان تنها از تعداد محدودی خط سیر برای فرآیند بازسازی استفاده نمود. در این پژوهش از دو روش حد آستانه و تعداد خطوط سیر ثابت برای انتخاب خطوط سیر مشابه استفاده شده که در ادامه هر یک از این روش ها توضیح داده خواهند شد.

### ۲-۳-۳-۱- انتخاب خطوط سیر مشابه منتخب با روش حد آستانه

یکی از روش هایی که می توان با آن تعداد محدودی از خطوط سیر موجود در پایگاه داده را انتخاب و در بازسازی از آن ها استفاده نمود، روش حد آستانه می باشد. منظور از حد آستانه در اینجا در نظر گرفتن یک مقدار مرزی به گونه ای که میزان شباهت خطوط سیر از این مقدار مرزی کمتر باشد. با اعمال شرط حد آستانه، تنها خطوط سیری که مقادیر شباهت آن ها از مقدار حد آستانه کمتر باشد، انتخاب و در فرآیند بازسازی شرکت داده می شوند. لازم به ذکر است که مقادیر حد آستانه در این پژوهش با توجه به تراکم خطوط سیر دو منطقه مورد مطالعه و در نظر گرفتن این مسئله که خطوط سیر دارای گپ، حداقل یک خط سیر مشابه برای بازسازی داشته باشند، بصورت تجربی ۰/۲ و ۰/۰۲ تعریف می شوند. با این حال این اعداد می توانند برای مناطق مختلف با تراکم و طول خطوط سیر متفاوت، متغیر باشند.

### ۲-۳-۳-۲- انتخاب خطوط سیر مشابه منتخب با روش تعداد ثابت

در روشی دیگر می توان بجای تعریف حد آستانه از تعداد مشخصی خط سیر مشابه استفاده نمود. روش کار بدین شکل می باشد که مقادیر شباهت خطوط سیر پایگاه داده که در مرحله قبلی بدست آمده اند، مرتب می شوند و تنها تعداد محدودی خط سیر با شباهت بالا انتخاب می شوند. در این پژوهش پنج خط سیری که بیشترین میزان شباهت را به خط سیر هدف دارند، در فرآیند بازسازی شرکت داده شده اند. انتخاب خطوط سیر زیاد، بدلیل کاهش شباهت خطوط سیر می تواند منجر به کاهش دقت بازسازی شود. از طرفی استفاده از شبیه ترین خط سیر با بالاترین میزان تشابه شاید در حالت های کلی دقت بازسازی را افزایش دهد. اما بدلیل وابستگی بیش از اندازه ی خط سیر هدف به یک خط سیر، در حالتی که خطوط سیر مشابه، شباهت زیادی به خط سیر هدف نداشته باشند، دقت بازسازی را کاهش می دهد [۷]. با این حال انتخاب تعداد خطوط سیر مشابه، بسته به تراکم منطقه مورد مطالعه و میزان شباهت خطوط سیر مجاور به خط سیر هدف، می تواند متفاوت باشد.

### ۲-۳-۳-۲- ایجاد گپ مصنوعی

به منظور مقایسه و ارزیابی هرچه بهتر فرآیند بازسازی، در این پژوهش از گپ مصنوعی بجای گپ طبیعی موجود در داده ها استفاده شده است. بدین صورت که تعدادی از نقاط که موقعیت آن ها مشخص می باشند از درون خط سیر هدف استخراج و به عنوان گپ فرضی و مصنوعی در نظر گرفته می شوند. تعداد و موقعیت نقاط گپ مصنوعی ممکن است بر روی نتایج تاثیر متفاوتی بگذارند. در نظر گرفتن گپ مصنوعی با تعداد نقاط خیلی بالا، امکان ارزیابی خطوط سیر با طول کوتاه را از بین می برد. از سوی دیگر کم بودن تعداد نقاط گپ نیز بعضا تعریف گپ را زیر سوال می برد. به عنوان مثال در نظر گرفتن پنج نقطه گپ که مسافت زمانی و مکانی قابل توجهی را شامل نمی شود، با تعریف گپ که از دست دادن موقعیت مکانی در یک بازه زمانی و مکانی قابل توجه است، در تناقض است. به همین دلیل در این پژوهش به منظور تعمیم و ارزیابی دقت بازسازی در خطوط سیر با طول های متفاوت، ابتدا از ۱۰ نقطه گپ مصنوعی و در ادامه



از ۲۰ نقطه گپ مصنوعی استفاده شده است. همچنین، به منظور ارزیابی تاثیر موقعیت گپ در دقت بازسازی، ابتدا گپ مصنوعی در میانه خطوطسیر تعریف شده اند سپس به انتهای خطوطسیر منتقل می شوند تا تاثیر موقعیت نقاط گپ در بازسازی مشخص شود.

## ۲-۳-۳-۴- بازسازی با میانگین گیری از مقادیر خطوطسیر مشابه منتخب

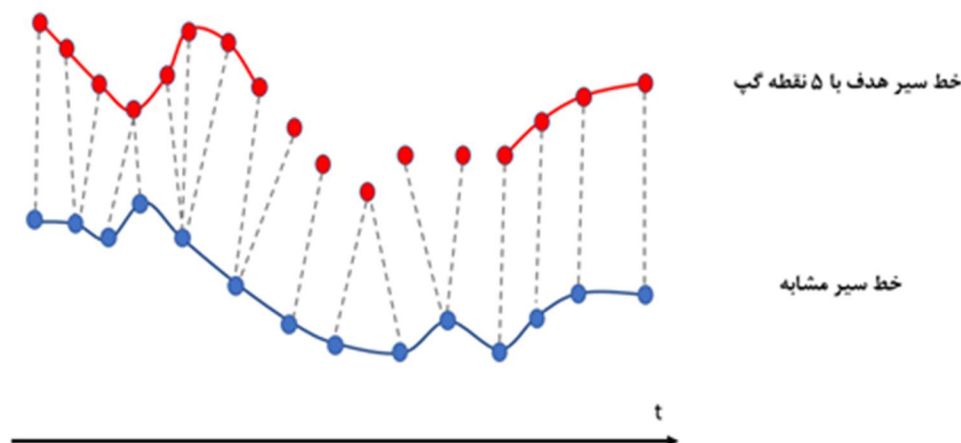
پس از تعریف و ایجاد گپ مصنوعی در خطسیر هدف، مراحل بازسازی به شرح ذیل صورت می گیرد: (۱) استخراج خطوطسیر مشابه منتخب از پایگاه داده با استفاده از روش های ذکر شده: تعداد محدودی از خطوطسیر مشابه که شباهت بیشتری به خطسیر هدف دارند، توسط روش حد آستانه یا تعداد ثابت از پایگاه داده استخراج می شوند. (۲) استخراج نقاط متناظر گپ از خطوطسیر مشابه منتخب: در این مرحله مجدداً از روش CDTW برای بازسازی استفاده می شود. بدین صورت که نقاط متناظر گپ مصنوعی از خطوطسیر مشابه منتخب توسط روش CDTW استخراج می شوند. با توجه به تعریف روش CDTW، هر نقطه از خطسیر هدف که دارای گپ مصنوعی باشد با یک یا چند نقطه از خطسیر مشابه تناظرایی می شوند. بدین ترتیب با توجه به اینکه موقعیت نقاط گپ مصنوعی مشخص می باشد، می توان نقاط متناظر آن ها را از درون خطوطسیر مشابه استخراج نمود. در مواردی که نقاط گپ با چندین نقطه متناظر می شوند، نقطه ای که فاصله کمتری با نقطه گپ داشته باشد به عنوان نقطه متناظر آن انتخاب می شود. (۳) میانگین گیری از مقادیر متناظر گپ و بازسازی گپ: در نهایت با میانگین گیری از موقعیت نقاط متناظر، نقاط

بازسازی شده بدست می آیند. در ادامه مراحل ۲ و ۳ برای تمامی نقاط گپ تکرار می شود تا تمامی آن ها بازسازی شوند. با فرض آنکه  $P_i$  ها موقعیت نقاط متناظر گپ از خطوطسیر مشابه می باشند که از مرحله ۲ بدست آمده اند، موقعیت بازسازی شده هر نقطه گپ با میانگین گیری و با استفاده از رابطه (۸) بدست می آید. در این پژوهش وزن تمامی خطوطسیر مشابه منتخب، یکسان در نظر گرفته شده اند.

$$G = \frac{P_1 + P_2 + \dots + P_n}{n} \quad (8)$$

جائیکه  $n$  تعداد خطوطسیر منتخب مورد استفاده برای بازسازی می باشد. در حالت حد آستانه این مقدار متغیر و در حالت خطوطسیر با تعداد ثابت، این مقدار پنج می باشد.  $P_1$  موقعیت نقطه متناظر گپ از خطسیر مشابه اول،  $P_2$  موقعیت نقطه متناظر گپ از خطسیر مشابه دوم و  $P_n$  موقعیت نقطه متناظر گپ از خطسیر مشابه  $n$ ام می باشند. در نهایت مقدار بازسازی شده برای نقطه گپ اول  $G_1(x, y)$  بدست می آید. با فرض داشتن  $m$  نقطه گپ، مراحل بالا برای تمامی نقاط گپ تکرار می شود و  $G_1$  تا  $G_m$  بدست می آیند. در نهایت به منظور ارزیابی دقت بازسازی، مقادیر بازسازی شده برای هر کدام از نقاط گپ با مقادیر واقعی آن ها مقایسه می شوند که در ادامه نحوه انجام آن شرح داده شده است.

در شکل ۴ سازوکار بازسازی با استفاده از تنها یک خطسیر مشابه و برای پنج نقطه گپ نشان داده شده است. تناظرایی میان نقاط خطسیر دارای گپ با خطسیر مشابه، توسط CDTW انجام و با خطوط نقطه چین نشان داده شده است.



شکل ۴- بازسازی خطسیر دارای گپ مصنوعی با استفاده از یک خطسیر مشابه

## ۲-۳-۴- ارزیابی

پس از بدست آوردن نقاط بازسازی شده از مرحله قبل ( $G_i$  ها) و با در اختیار داشتن موقعیت نقاط گپ مصنوعی و مقایسه این مقادیر با یکدیگر، دقت روش مورد ارزیابی قرار می‌گیرد. در این پژوهش از دو معیار جذر میانگین کمترین مربعات<sup>۱</sup> (RMSE) و خطای مطلق میانگین<sup>۲</sup> (MAE) استفاده شده‌است. دو معیار ذکر شده به ترتیب از روابط (۹) و (۱۰) محاسبه شده‌اند.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (9)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (10)$$

جائیکه  $y_i$  و  $\hat{y}_i$  به ترتیب نشان‌دهنده مقدار واقعی و مقدار بازسازی شده ( $G_i$  ها) می‌باشند. منظور از مقدار واقعی، موقعیت نقاط گپ مصنوعی می‌باشد که در دسترس هستند و مقدار بازسازی شده همان مقادیر میانگین‌گیری شده از نقاط متناظر می‌باشند.

## ۳- نتایج و بحث

پس از پیش‌پردازش داده‌های خام AIS، خطوط مسیر کلی هر کشتی بصورت مجزا تشکیل و پس از آن برای هر کشتی سفرهای مجزا بدست آمدند. با توجه به پراکندگی خطوط مسیر با طول‌های متفاوت در دو منطقه، در منطقه اول تعداد ۱۰۴۴ خط‌سیر و در منطقه دوم ۲۸۸ خط‌سیر بدست آمدند. با در نظر گرفتن این نکته که اغلب خطوط مسیر با طول‌های کوتاه در منطقه اول و خطوط مسیر با طول‌های بلندتر اغلب در منطقه دوم قرار دارند، خطوط مسیر هدف با طول‌های کوتاه و تا ۵۰ نقطه از منطقه اول و خطوط مسیر با طول‌های بین ۱۵۰ تا ۲۰۰ نقطه از منطقه دوم انتخاب شده‌اند (شکل ۵).

در این پژوهش از دو روش حد آستانه و تعداد خطوط مسیر ثابت برای بازسازی نقاط گپ درون خطوط مسیر هدف استفاده شده‌است. به منظور مقایسه نتایج در شرایط

مختلف، در هر دو روش طول گپ ۱۰ و ۲۰ نقطه در میانه و انتهای خطوط مسیر در نظر گرفته شده‌اند که میانگین مقادیر RMSE و MAE در تمامی این حالات در جدول ۲ آورده شده‌است.

در جدول ۲، در تمامی حالات وزن پارامتر مکانی بیشتر از پارامترهای سرعت و جهت کشتی در نظر گرفته شده‌است. با این حال به منظور ارزیابی تاثیر پارامترها بر روی نتایج بازسازی، حالتی که در آن وزن تمامی پارامترها یکسان در نظر گرفته شده‌است نیز مورد بررسی قرار گرفت. نتایج نشان‌دهنده آن است که وزن‌دهی بیشتر به پارامتر مکانی در مقایسه با پارامتر جهت و سرعت، باعث بهبود دقت بازسازی می‌شود.

همچنین از خطوط مسیر کشتی‌ها در دو منطقه با طول‌های متفاوت استفاده شده‌است. به دلیل ماهیت و تفاوت تراکم خطوط مسیر در هر منطقه و مقایسه تاثیر طول خطوط، در منطقه اول از خطوط مسیر با طول‌های کوتاه‌تر (۲۰ تا ۵۰ نقطه) و در منطقه دوم خطوط مسیر با طول‌های بلندتر (۱۵۰ تا ۲۰۰ نقطه) به عنوان خطوط مسیر هدف و دارای گپ استفاده شده‌اند. نتایج نشان‌دهنده آن است که با افزایش طول خطوط مسیر در هر منطقه، میانگین مقادیر RMSE و MAE کمتر و دقت بازسازی بهتر شده‌است. دلیل این امر می‌تواند در فرآیند شباهت‌یابی باشد. هر چه طول خطوط مسیر بلندتر باشد، تابع CDTW نقاط و داده ورودی بیشتری برای محاسبه میزان شباهت دو خط‌سیر و تناظریابی نقاط دارد. از طرفی خطوط مسیر بلندتر را می‌توان به عنوان مسیرهای عبور و مرور متداول کشتی‌ها در مناطق مورد مطالعه در نظر گرفت که توسط بسیاری از کشتی‌های دیگر یا خود کشتی در بازه‌های زمانی مختلف تکرار شده‌است. در حالی که خطوط مسیر با طول‌های کوتاه می‌توانند تنها بیانگر حرکات تصادفی و مانور کشتی‌ها در شرایط خاص باشند که از الگوی خاصی پیروی نمی‌کنند و شباهت‌یابی مناسبی میان آن‌ها صورت نمی‌گیرد.

با مقایسه میانگین مقادیر RMSE و MAE می‌توان به این نتیجه رسید که انتخاب خطوط مسیر مشابه با تعداد مشخص، عملکرد بهتری نسبت به تعریف حد آستانه در دو منطقه مورد مطالعه دارد. با این حال از آنجائیکه مقادیر حدود آستانه بصورت تجربی در نظر گرفته شده‌اند و تراکم خطوط مسیر در مناطق مورد مطالعه متفاوت است، می‌توان با تغییر مقادیر حد آستانه و

<sup>۱</sup> Root Mean Square Error

<sup>۲</sup> Mean Absolute Error

پیاپی سازی این روش در مناطق دیگر عملکرد این روش را بهتر مورد ارزیابی قرار داد. مقادیر نشان داده شده در جدول ۲ بیانگر آن است که افزایش طول گپ از ۱۰ نقطه به ۲۰ نقطه تاثیر چندانی در دقت بازسازی خطوط سیر در دو منطقه نداشته است. با این حال به منظور ارزیابی هرچه بهتر تاثیر طول گپ در دقت بازایی، می توان از مناطق متعدد با تراکم خطوط سیر بیشتر در کنار خطوط سیر با هندسه های مختلف استفاده نمود.

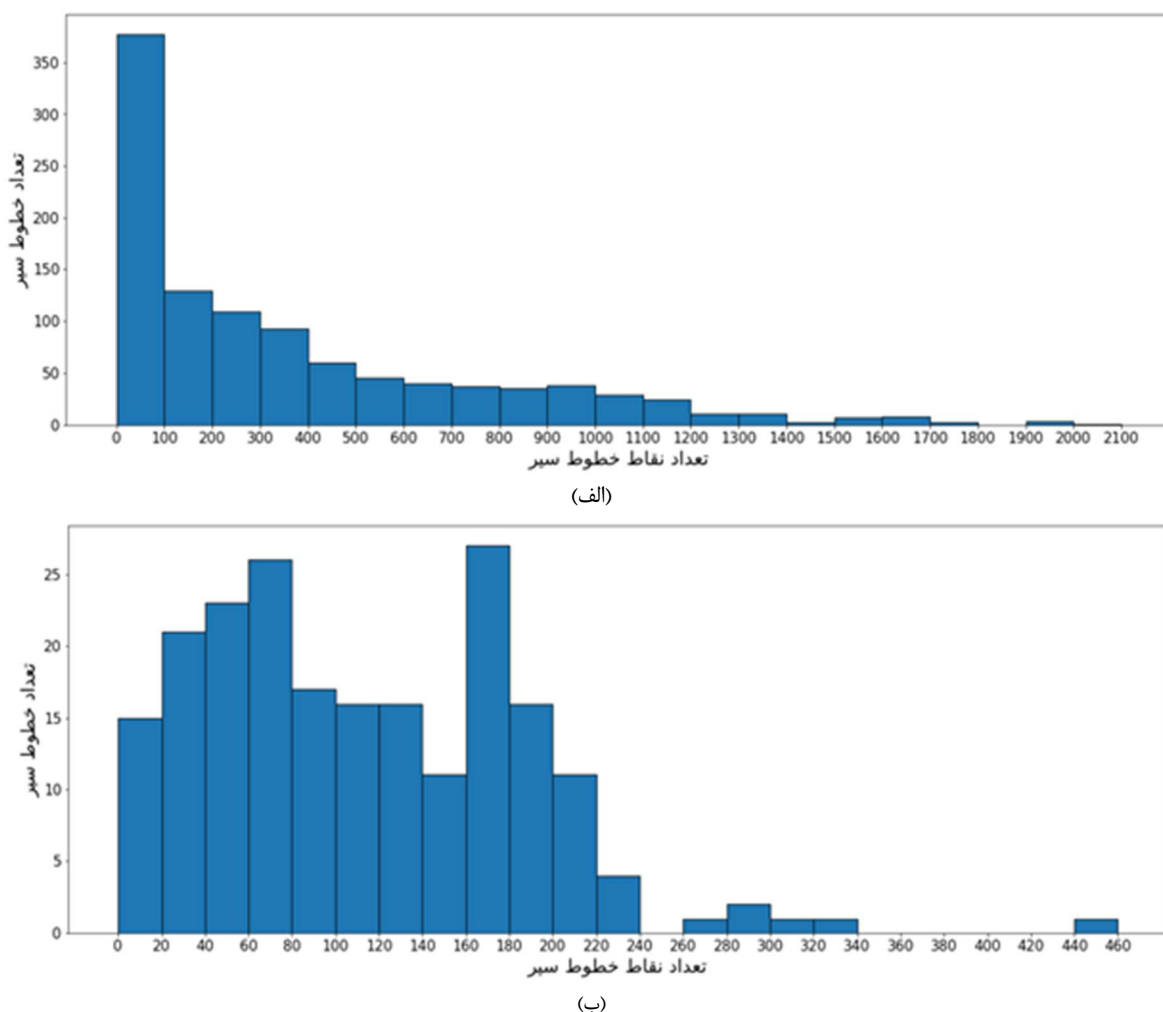
پارامتر دیگری که می تواند بر دقت بازسازی نقاط گپ تاثیرگذار باشد، مکان گپ است. در این پژوهش از نقاط گپ در میانه و انتهای خطوط سیر استفاده شده است. بطوری که صرف نظر از طول گپ و طول خطوط سیر، ابتدا نقاط گپ در میانه خطوط تعریف و در ادامه به انتهای خطوط سیر

منتقل شده اند. با این حال با مقایسه مقادیر جدول ۲ می توان به این نتیجه رسید که تغییر مکان گپ از میانه به انتها تاثیری در دقت بازسازی گپ نداشته است.

استفاده از الگوهای تکراری و شباهت یابی میان خطوط سیر و داده های گذشته در کنار پارامترهای حرکتی، نقش مهمی در بازسازی خطوط سیر دارد و این موارد را می توان از مزایای این روش در مقایسه با روش های معمول بازسازی همچون روش های درون یابی دانست. به عنوان نمونه در روش ارائه شده توسط [۷] با وجود استفاده از شباهت یابی در فرآیند بازسازی، تنها خطوطی که در یک کیلومتری خط سیر هدف و در مجاورت نقاط گپ بودند، به عنوان خطوط سیر مشابه در فرآیند بازسازی استفاده شدند.

جدول ۲- مقادیر بازسازی شده در منطقه اول (اقیانوس اطلس) و منطقه دوم (اقیانوس آرام)

بازسازی با تعداد خطوطسیر ثابت						
میانگین مقادیر MAE	میانگین مقادیر RMSE	محل گپ	طول گپ (نقطه)	تعداد خطوطسیر مشابه	طول خطسیر هدف (نقطه)	
۰/۴۶	۰/۵۱	وسط	۱۰	۵	۵۰ تا ۲۰	منطقه اول
۰/۴۷	۰/۵۳	انتها	۱۰			
۰/۴۷	۰/۵۲	وسط	۲۰			
۰/۴۷	۰/۵۲	انتها	۲۰			
۰/۰۹	۰/۱۳	وسط	۱۰	۵	۲۰۰ تا ۱۵۰	منطقه دوم
۰/۱۴	۰/۱۶	انتها	۱۰			
۰/۰۹	۰/۱۱	وسط	۲۰			
۰/۱۳	۰/۱۵	انتها	۲۰			
بازسازی با حد آستانه						
میانگین مقادیر MAE	میانگین مقادیر RMSE	محل گپ	طول گپ (نقطه)	حد آستانه	طول خطسیر هدف (نقطه)	
۱/۴	۱/۵	وسط	۱۰	کمتر از ۰/۰۲	۵۰ تا ۲۰	منطقه اول
۱/۴	۱/۵	انتها	۱۰			
۱/۴	۱/۵	وسط	۲۰			
۱/۴	۱/۵	انتها	۲۰			
۰/۲۱	۰/۲۶	وسط	۱۰	کمتر از ۰/۲	۲۰۰ تا ۱۵۰	منطقه دوم
۰/۲۸	۰/۳۱	انتها	۱۰			
۰/۲۱	۰/۲۷	وسط	۲۰			
۰/۳۳	۰/۳۱	انتها	۲۰			



شکل ۵- توزیع خطوط سیر با تعداد نقاط و طول‌های متفاوت واقع در (الف) منطقه اول (اقیانوس اطلس) و (ب) منطقه دوم (اقیانوس آرام)

خطوط سیر را همزمان برای اشیاء متحرک دارد. انتخاب مقدار مناسب پارامتر وزن برای هر کدام از ابعاد را می‌توان از محدودیت CDTW نام برد. چرا که عملکرد این روش رابطه مستقیمی با مقادیر وزن انتخاب شده برای هر پارامتر دارد. به همین دلیل تعیین مقادیر وزنی مناسب چالش برانگیز و بسته به کاربرد متفاوت می‌باشد. از دیگر نقاط ضعف، هزینه محاسباتی بالای روش CDTW است. از محدودیت‌های این پژوهش می‌توان به عدم امکان تعریف حد آستانه یکسان در مناطق مختلف، به منظور انتخاب خطوط سیر مشابه منتخب اشاره کرد.

#### ۴- نتیجه‌گیری

در این پژوهش به بازسازی خطوط سیر کشتی‌ها از طریق فرآیند شباهت‌یابی و با استفاده از روش CDTW پرداخته شد. در این راستا شباهت میان خطوط سیر نه تنها

همچنین، انتخاب خطوط سیر مشابه تنها از بعد مکانی و بدون در نظر گرفتن پارامترهای حرکتی سرعت و جهت انجام شده‌است. در حالی که در این پژوهش میزان شباهت تمامی خطوط سیر موجود در منطقه صرف‌نظر از میزان مجاورت آن‌ها به خط سیر هدف محاسبه شده و علاوه بر بعد مکانی از سرعت و جهت کشتی‌ها نیز در فرآیند شباهت‌یابی استفاده شده‌است.

روش CDTW که برای محاسبه میزان شباهت خطوط سیر و در ادامه برای تناظریابی میان نقاط خطوط سیر مورد استفاده قرار گرفت، صرف‌نظر از طول خطوط سیر، شباهت‌یابی و تناظریابی را برای تمامی نقاط انجام می‌دهد. همچنین، با توجه به تنوع خطوط سیر اشیاء نقطه‌ای و عدم دسترسی دائم به حجم بالایی از داده‌های حرکتی برای آموزش و استفاده از روش‌های یادگیری ماشین، روش CDTW تنها با دسترسی به حجم اندکی از داده‌های حرکتی، توانایی شباهت‌یابی و بازسازی

از بُعد مکانی بلکه با در نظر گرفتن پارامترهای حرکتی سرعت و جهت کشتی بررسی گردید. همچنین، تاثیر طول گپ، مکان گپ و طول خطوطسیر در فرآیند بازسازی مورد ارزیابی قرار گرفت. نتایج نشان‌دهنده آن است که افزایش وزن پارامتر مکانی در مقایسه با پارامترهای سرعت و جهت، دقت بازسازی را افزایش می‌دهد. همچنین، استفاده از تعداد خطوطسیر مشخص و ثابت در مقایسه با روش حد آستانه

دقت بازسازی را به طرز قابل توجهی بهبود می‌بخشد. در پژوهش‌های آتی می‌توان از سایر روش‌های اندازه‌گیری شباهت هم در فرآیند شباهت‌یابی و هم در فرآیند بازسازی استفاده نمود و نقاط ضعف و قوت آن روش‌ها را در مقایسه با روش CDTW مورد ارزیابی قرار داد. همچنین، می‌توان صرف‌نظر از مناطق مورد مطالعه و تراکم خطوطسیر، مقدار مشخصی برای حد آستانه تعریف نمود.

## مراجع

- [۱] Y. Zheng, "Trajectory data mining: an overview," *ACM Trans. Intell. Syst. Technol.*, vol. 6, no. 3, pp. 1–41, 2015.
- [۲] Z. Fan, A. Arai, X. Song, A. Witayangkurn, H. Kanasugi, and R. Shibasaki, "A collaborative filtering approach to citywide human mobility completion from sparse call records," *IJCAI Int. Jt. Conf. Artif. Intell.*, vol. 2016-Janua, pp. 2500–2506, 2016.
- [۳] D. Hecker, H. Stange, C. Körner, and M. May, "Sample Bias due to Missing Data in Mobility Surveys," in *2010 IEEE International Conference on Data Mining Workshops*, 2010, pp. 241–248. doi: 10.1109/ICDMW.2010.162.
- [۴] J. A. Long, "Kinematic interpolation of movement data," *Int. J. Geogr. Inf. Sci.*, vol. 30, no. 5, pp. 854–868, 2016, doi: 10.1080/13658816.2015.1081909.
- [۵] M. Li, S. Gao, F. Lu, and H. Zhang, "Reconstruction of human movement trajectories from large-scale low-frequency mobile phone data," *Comput. Environ. Urban Syst.*, vol. 77, p. 101346, 2019, doi: <https://doi.org/10.1016/j.compenvurbsys.2019.101346>.
- [۶] M. Liang, R. W. Liu, Q. Zhong, J. Liu, and J. Zhang, "Neural Network-Based Automatic Reconstruction of Missing Vessel Trajectory Data," in *2019 IEEE 4th International Conference on Big Data Analytics (ICBDA)*, 2019, pp. 426–430. doi: 10.1109/ICBDA.2019.8713215.
- [۷] S. Pathak, M. He, S. Malinchik, and S. Sobolevsky, "Pattern Ensembling for Spatial Trajectory Reconstruction," Jan. 2021, [Online]. Available: <http://arxiv.org/abs/2101.09844>
- [۸] K. Corder, U. Ekelund, R. M. Steele, N. J. Wareham, and S. Brage, "Assessment of physical activity in youth," *J. Appl. Physiol.*, vol. 105, no. 3, pp. 977–987, 2008, doi: 10.1152/jappphysiol.00094.2008.
- [۹] E. A. Codling, M. J. Plank, and S. Benhamou, "Random walk models in biology," *J. R. Soc. Interface*, vol. 5, no. 25, pp. 813–834, 2008, doi: 10.1098/rsif.2008.0014.
- [۱۰] Y. Tremblay *et al.*, "Interpolation of animal tracking data in a fluid environment," *J. Exp. Biol.*, vol. 209, no. 1, pp. 128–140, 2006, doi: 10.1242/jeb.01970.
- [۱۱] D. Seng, Q. Zhang, X. Zhang, G. Chen, and X. Chen, "Spatiotemporal prediction of air quality based on LSTM neural network," *Alexandria Eng. J.*, vol. 60, no. 2, pp. 2021–2032, 2021, doi: <https://doi.org/10.1016/j.aej.2020.12.009>.
- [۱۲] Y. Tao *et al.*, "A comparative analysis of trajectory similarity measures," *GIScience Remote Sens.*, vol. 58, no. 5, pp. 643–669, 2021, doi: 10.1080/15481603.2021.1908927.
- [۱۳] M. Sharif and A. A. Alesheikh, "A Review on the Process of Point Objects' Movement and Their Trajectory Similarity Measurement Approaches," *issge-gej*, vol. 7, no. 1, pp. 41–54, Mar. 2016, [Online]. Available: <http://gej.issge.ir/article-1-151-en.html>
- [۱۴] M. Sharif and A. A. Alesheikh, "Context-awareness in similarity measures and pattern discoveries of trajectories: a context-based dynamic time warping method," *GIScience Remote Sens.*, vol. 54, no. 3, pp. 426–452, May 2017, doi: 10.1080/15481603.2017.1278644.
- [۱۵] D. Alizadeh, A. A. Alesheikh, and M. Sharif, "Vessel Trajectory Prediction Using Historical Automatic Identification System Data," *J. Navig.*, vol. 74, no. 1, pp. 156–174, 2021, doi: 10.1017/S0373463320000442.