

# توسعه چارچوبی مبتنی بر رایانش ابری برای تحلیل انطباق نقشه داده‌های خودروهای شناور

محمد مسعود رحیمی<sup>۱</sup>، فرشاد حکیم پور<sup>۲\*</sup>

<sup>۱</sup> دانشجوی کارشناسی ارشد سیستم‌های اطلاعات مکانی - دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی - پردیس دانشکده‌های فنی - دانشگاه تهران  
rahimi.masoud@ut.ac.ir

<sup>۲</sup> استادیار دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی - پردیس دانشکده‌های فنی - دانشگاه تهران  
fhakimpour@ut.ac.ir

(تاریخ دریافت خرداد ۱۳۹۶، تاریخ تصویب آبان ۱۳۹۶)

## چکیده

امروزه کاربردهای نوین الگوریتم انطباق نقشه در کنار تلاش برای حفظ نرخ بالای موفقیت، با دو چالش نرخ‌های نمونه‌برداری متغیر و سرعت پردازش داده‌ها روبرو هستند. در این تحقیق یک چارچوب یکپارچه و توزیع‌یافته برای تحلیل انطباق نقشه داده‌های خودروهای شناور ارائه شده است. در چارچوب پیشنهادی، الگوریتمی بهبودیافته و یک معماری کلی و مفهومی برای سکوی رایانشی تحلیل انطباق نقشه ارائه شده است. الگوریتم مذکور برای مواجهه با نرخ‌های نمونه‌برداری متفاوت طراحی شده و می‌تواند فارغ از وابستگی به حسگرهای اضافی عمل کند. همچنین استفاده از روش وزن‌دهی پویا باعث استقلال الگوریتم از پارامترهای محلی شده است. استفاده از مفاهیم سطوح اطمینان صحت عملکرد الگوریتم را نیز نسبت به یکی از روش‌های شناخته شده افزایش داده است. در الگوریتم پیشنهادی برای پاسخ به چالش نرخ نمونه‌برداری پایین از روش کوتاه‌ترین مسیر \*A در کنار روش وزن‌دهی پویا استفاده می‌شود. وزن‌دهی قید کوتاه‌ترین مسیر با کمک اطلاعات تعدیل دقت افقی (HDOP) انجام می‌شود. یکی از چالش‌های اصلی در به‌کارگیری داده‌های بزرگ FCD، ذخیره‌سازی، مدیریت، تحلیل و پردازش جمعی آن‌ها است. بدین منظور در این چارچوب از اصول رایانش ابری و الگوی نگاشت کاهش (MapReduce) بر روی سکوی هدوپ (Hadoop) استفاده شده است. بدین ترتیب چارچوب می‌تواند فوق تحلیلی توزیع‌یافته و مقیاس‌پذیر را بر روی سخت‌افزارهای معمولی ممکن می‌سازد. جهت ارزیابی عملکرد چارچوب ناوگان حمل‌ونقل شهر نیویورک به‌عنوان یکی از بزرگ‌ترین ناوگان‌های حمل‌ونقلی جهان مورد مطالعه قرار گرفته است. نتایج ارزیابی‌ها بیانگر نرخ موفقیت ۹۵٫۲ درصدی در نرخ نمونه‌برداری بالا (۱۰ ثانیه) و ۸۹٫۵ درصدی در نرخ نمونه‌برداری پایین (۱۲۰ ثانیه) است. همچنین مقایسه نتایج روش پیشنهادی با یکی از روش‌های شناخته شده نشان می‌دهد در نرخ نمونه‌برداری ۱۲۰ ثانیه روش پیشنهادی صحت انطباق نقشه را تا ۹٫۷ درصد بهبود داده است. ارزیابی‌ها نشان می‌دهد با حضور ۵ گره پردازشی چارچوب پیشنهادی توانسته بیش از ۷۰۰۰ نقطه در هر ثانیه را بر روی نقشه منطبق کند.

**واژگان کلیدی:** داده‌های خودروهای شناور (FCD)، تحلیل انطباق نقشه (MM)، نرخ نمونه‌برداری متغیر، رایانش ابری

\* نویسنده رابط

## ۱- مقدمه

تراکم ترافیک یکی از چالش‌های کلیدی در سامانه‌های حمل‌ونقلی است که با افزایش تعداد خودروها در سطح راه اتفاق می‌افتد. در سال ۲۰۱۴ تراکم ترافیک ۶٫۹ میلیارد ساعت زمان شهروندان و ۳٫۱ میلیارد گالن سوخت را به هدر داده که این امر ضرری برابر ۱۶۰ میلیارد دلار به اقتصاد ایالات متحده وارد کرده است [۱]. همچنین افزایش آلودگی‌های صوتی، بصری و محیطی، کاهش بهره‌وری و ناکارآمدی سامانه‌های امدادی نیز از دیگر عواقب منفی تراکم ترافیک است [۱]. برای پاسخگویی به این مشکلات و بهبود کیفیت زندگی انسان، حفظ محیط‌زیست، پاسخ به نیازهای جامعه، افزایش بهره‌وری، امنیت و رشد اقتصادی استراتژی‌های متعددی ارائه شده که از جمله آن‌ها می‌توان به توسعه سامانه‌های حمل‌ونقل هوشمند<sup>۱</sup> (ITS) اشاره نمود. همگام با توسعه فناوری‌های موقعیت‌یابی و شبکه‌های ارتباطی، داده‌های خودروهای شناور<sup>۲</sup> (FCD) به‌عنوان یکی از منابع داده اصلی در سامانه‌های حمل‌ونقل هوشمند پدیدار شده‌اند. FCD جریانی از نقاط زمانی مکانی است که خط سیر یک وسیله نقلیه متحرک را به تصویر می‌کشد [۲]. FCD معمولاً از یک ناوگان از وسایل نقلیه جمع‌آوری می‌شوند. هرکدام از این خودروها به یک گیرنده سیستم موقعیت‌یاب جهانی<sup>۳</sup> (GPS) و یک دستگاه ارتباط بی‌سیم مجهز شده‌اند. برتری FCD نسبت به سایر حسگرهای ترافیکی، ارائه آنی<sup>۴</sup> اطلاعات باکیفیت در کنار پوشش مکانی گسترده و هزینه راه‌اندازی پایین است [۳]. از این مجموعه داده‌های دقیق و آنی می‌توان در بسیاری از کاربردهای آنی ITS مانند مدیریت و نظارت آنی ترافیک استفاده نمود. همچنین با استفاده از خط سیرهای زمانی-مکانی FCD و با کمک روش‌های پس پردازشی و داده‌کاوی، کاربردهایی مانند مدل‌سازی انتخاب مسیر و پیشنهاد مسیر [۴]، تخمین ترافیک و زمان سفر [۵]، پیش‌بینی جریان ترافیک [۶] و غیره نیز برای این داده‌ها در نظر گرفته می‌شود. یکی از چالش‌های اساسی پیش روی کاربردهای داده‌های خودروهای شناور، عدم انطباق مکانی است. داده‌های خام خودروهای شناور دارای دقت محدود و نویز فراوان هستند.

علت این خطا عمدتاً خطای ایجادشده توسط ماهواره‌ها و گیرنده‌های GPS و خطای موجود در تولید نقشه و رقومی سازی آن است. عدم انطباق مکانی در شبکه‌های متراکم شهری مقدار صحت کلی را در سطح اطمینان ۹۵ درصد بین ۰ تا ۴۰ متر کاهش می‌دهد [۷] و مانع انطباق صحیح موقعیت حقیقی خودرو بر شبکه مکانی راه‌ها می‌شود. به‌کارگیری صحیح FCD مستلزم فرآیند پیش‌پردازشی به نام انطباق نقشه<sup>۵</sup> (MM) است. هدف از MM یکپارچه‌سازی موقعیت مکانی خودرو با شبکه راه‌ها، شناسایی یال حقیقی خودرو و تعیین مکان خودرو بر روی آن یال است. در سال‌های اخیر تحقیقات متعددی بر روی توسعه و بهبود الگوریتم‌های انطباق نقشه صورت گرفته است. بیشتر این تحقیقات بر روی انطباق نقشه یک خودرو تنها با نرخ نمونه‌برداری بالا متمرکز شده‌اند [۸-۱۱]. تعدادی دیگر از تحقیقات نیز به مسئله انطباق نقشه با نرخ نمونه‌برداری پایین پرداخته‌اند [۱۲-۱۵]. دسته دیگری از تحقیقات بر روی بهبود صحت انطباق نقشه با استفاده از حسگرهای موقعیت‌یابی اضافی متمرکز شده‌اند [۱۶، ۱۷]. با مروری بر تحقیقات انجام شده می‌توان دریافت که امروزه کاربردهای نوین الگوریتم انطباق نقشه در کنار تلاش برای حفظ نرخ بالای موفقیت، با دو چالش نرخ‌های نمونه‌برداری متغیر و سرعت پردازش داده‌ها روبرو هستند.

در این پژوهش با بررسی شهر نیویورک به‌عنوان یک مطالعه موردی، چالش‌ها و الزامات یک چارچوب یکپارچه تحلیل انطباق نقشه تبیین شده است. نیویورک دارای بزرگ‌ترین ناوگان‌های حمل‌ونقل شهری در ایالات متحده آمریکا، بیش از ۱۰۰۰۰۰ تاکسی و سرویس حمل‌ونقلی عمومی و خصوصی، بیش از ۵۷۰۰ اتوبوس و بیش از ۲٫۵ میلیون سفر روزانه درون‌شهری است [۱۸، ۱۹]. اگر تمامی این ناوگان دارای تجهیزات ردیاب خودرو باشند و داده‌های آن‌ها به‌صورت پیوسته و ۲۴ ساعته به سیستم مدیریت اطلاعات ترافیک دپارتمان حمل‌ونقل نیویورک [۲۰] ارسال شود حجم بسیار بزرگ و باارزشی از اطلاعات مکانی جمع‌آوری می‌شود که از آن می‌توان در بسیاری از کاربردهای سامانه‌های حمل‌ونقل هوشمند استفاده نمود. در این سناریو چندین چالش اساسی پیش روی چارچوب یکپارچه تحلیل انطباق نقشه است:

<sup>۱</sup> Intelligent Transportation Systems (ITS)

<sup>۲</sup> Floating Car Data (FCD)

<sup>۳</sup> Global Positioning System (GPS)

<sup>۴</sup> Real-Time

<sup>۵</sup> Map-Matching (MM)

- نرخ نمونه برداری پایین و قیود وزن دهی: اگر یک خودرو با سرعت ثابت ۵۴ کیلومتر بر ساعت در حال حرکت باشد و نرخ نمونه برداری آن بر روی دو دقیقه تنظیم شده باشد، فاصله میان دو نقطه GPS متوالی قرائت شده ۱٫۸ کیلومتر است. بدین ترتیب خودرو در این بازه زمانی ممکن است چندین یال از شبکه راهها را طی کرده باشد بدون آنکه قرائتی از آنها در دسترس باشد. در این شرایط به کارگیری قیود توپولوژیک مانند اتصال و گردش ممنوع (مورد استفاده در بسیاری از الگوریتم‌های معمول انطباق نقشه) در این شرایط باعث بروز خطا در عملکرد الگوریتم می‌شود.
  - نرخ نمونه برداری بالا و تحلیل آنی: اگر نرخ نمونه برداری تمامی این ۱۰۰۰۰۰ خودرو به صورت میانگین بر روی ۲۰ ثانیه تنظیم شده باشد نرخ بارگذاری داده در کارساز<sup>۱</sup> مرکزی به ۵۰۰۰ قرائت GPS در ثانیه نیز می‌رسد. لذا الگوریتم باید بتواند به صورت آنی این حجم از داده را پردازش و تحلیل کند.
  - تحلیل و پردازش جمعی داده‌های بزرگ FCD: اگر نرخ نمونه برداری تاکسی‌ها بر روی یک ثانیه تنظیم شده باشد نرخ بارگذاری داده در کارساز مرکزی به بیش از ۱۰۰۰۰۰ رکورد در هر ثانیه نیز می‌رسد. میدانیم در بیشتر تجهیزات مکان‌یابی خودکار وسایل نقلیه<sup>۲</sup> (AVL) حجم داده خودرو شناور حدوداً ۱۴۰ بایت در هر پیام است. بدین ترتیب در هر ثانیه حدود ۱۲ گیگابایت داده برای به کارگیری در کاربردهای آنی بالادستی ITS باید به صورت آنی و بی‌درنگ ذخیره‌سازی و پردازش شود. همچنین این داده‌ها را می‌توان برای کاربردهای متعددی از ITS بر روی کارساز مرکزی جمع کرده و سپس به صورت جمعی مورد پردازش قرارداد. لذا مدیریت، تحلیل و پردازش چنین حجم بالایی از این داده‌ها یک چالش اساسی است.
  - با شناخت این چالش‌ها الزامات یک چارچوب یکپارچه تحلیل انطباق نقشه عبارت‌اند از:
- الگوریتم پیاده‌سازی شده در چارچوب باید سراسری<sup>۳</sup> بوده و مستقل از پارامترهای محلی باشد.
  - چارچوب باید فارغ از وابستگی به حسگرهای اضافی بتواند نرخ موفقیت مناسبی را ارائه کند.
  - چارچوب باید بتواند به صورت آنی و مؤثر پاسخگوی نیاز تحلیل‌های بالادستی ITS باشد.
  - چارچوب باید داده‌هایی با نرخ نمونه برداری بالا یا پایین را دریافت و آنها را با دقت و سرعت مناسبی تحلیل کند.
  - چارچوب باید بتواند مجموعه داده‌های بزرگ FCD را ذخیره‌سازی، مدیریت، تحلیل و پردازش جمعی کند.
  - در این مقاله ما الزامات و چالش‌های نوین پیش روی تحلیل انطباق نقشه را در نظر گرفته و یک چارچوب یکپارچه برای پاسخ به نیازهای مسئله انطباق نقشه ارائه کرده‌ایم.
  - در چارچوب پیشنهادی برای پاسخ به چالش ذخیره‌سازی، مدیریت، تحلیل و پردازش داده‌های بزرگ FCD یک الگوریتم نوین MM توسعه یافته و بر روی سکویی مبتنی بر فناوری‌های رایانش ابری موازی‌سازی و پیاده‌سازی شده است. پیدایش فناوری رایانش ابری راه‌حلی سودمند، مقیاس‌پذیر و به صرفه در پاسخگویی به چالش‌های داده‌های بزرگ در کاربردهای حمل‌ونقلی و مکانی است. راه‌حل پیشنهادی این مقاله برای پاسخ به چالش مذکور تقسیم‌بندی مجموعه داده بزرگ و معرفی هر قسمت از پردازش به یک رایانه برای افزایش توان رایانشی است. عمده مشارکت‌های مقاله عبارت‌اند از: (۱) ارائه یک چارچوب یکپارچه و توزیع یافته تحلیل انطباق نقشه مبتنی بر فناوری‌های رایانش ابری برای پاسخ آنی به کاربردهای نوین تحلیل‌های بالادستی (۲) ارائه یک الگوریتم انطباق نقشه بهبود یافته و سازگار با انواع داده‌های FCD با نرخ‌های نمونه برداری متفاوت (۳) به کارگیری روش وزن‌دهی پویا برای قیود اختلاف میان زاویه‌های حامل راستای دو نقطه قرائت شده متوالی و راستای نقطه منطبق شده قبلی با نقطه کاندیدا و اختلاف کوتاه‌ترین مسیر باهدف بهبود صحت الگوریتم در مواجهه با داده‌های با نرخ نمونه برداری پایین. در ادامه مقاله ابتدا مروری بر تحقیقات

<sup>۳</sup> Global

<sup>۱</sup> Server

<sup>۲</sup> Automatic Vehicle Location (AVL)

اخیر انجام شده بر روی MM خواهیم داشت. سپس الگوریتم بهبودیافته و معماری چارچوب پیشنهادی را بررسی خواهیم نمود و در انتها صحت و عملکرد چارچوب پیشنهادی را مورد بررسی قرار خواهیم داد.

## ۲- پیشینه تحقیق

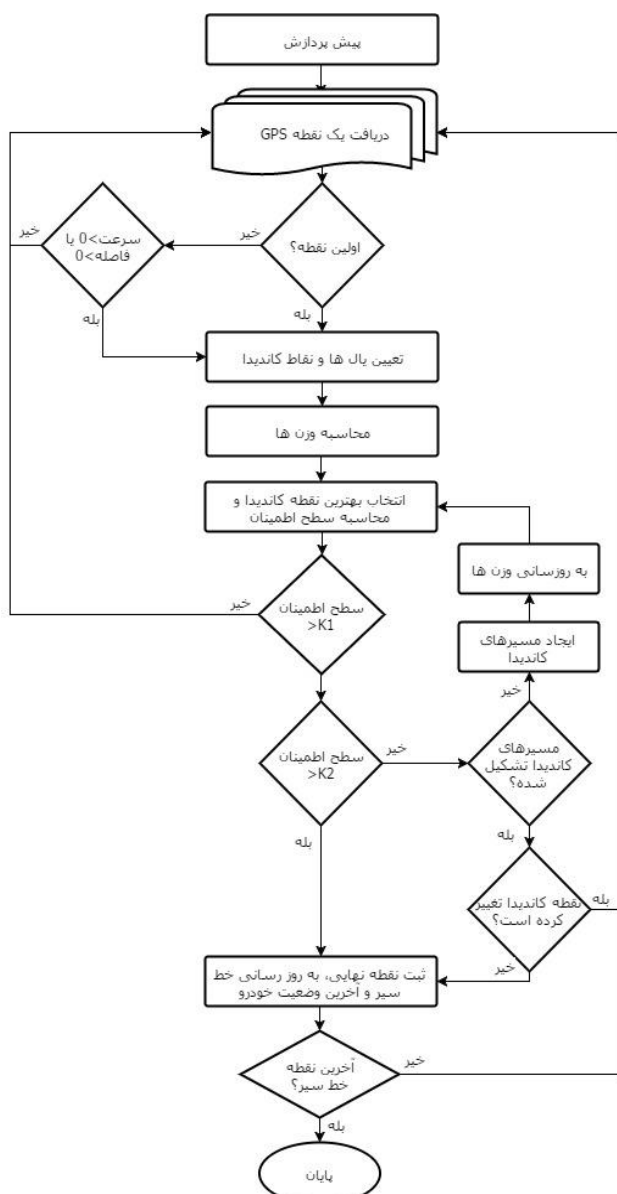
در سال‌های اخیر تحقیقات متعددی برای بهبود الگوریتم‌های انطباق نقشه انجام شده است. در این بخش تعدادی از تحقیقات اخیر انجام شده بر روی MM را بررسی و نقاط ضعف و قوت آن‌ها را بررسی خواهیم نمود. یکی از مهم‌ترین روش‌های ارائه شده اخیر برای مسئله MM روش ولاگا و همکاران [۱۶] است. در این تحقیق از همسایگی مکانی و اختلاف راستا به همراه روش امتیازدهی مثبت و منفی به یال‌های کاندیدا استفاده شده است. از جمله نواقص روش ولاگا و همکاران می‌توان به ضرایب وزن‌دهی ثابت محلی، وابستگی الگوریتم به اتصال یال‌های کاندیدا با یال منطبق شده برای قرائت پیشین و لذا ناتوانی در مواجهه با داده‌های با نرخ نمونه‌برداری پایین اشاره نمود که این نواقص در روش ارائه شده در این پژوهش به کمک وزن‌دهی پویا، بهره‌گیری از معیارهای هندسی و استفاده از معیار کوتاه‌ترین مسیر پاسخ داده شده‌اند.

هاشمی و کریمی [۹] یک الگوریتم وزن‌مبنا برای موقعیت‌یابی خودرو در شبکه‌های شهری پیچیده ارائه داده‌اند. هدف اصلی در این تحقیق ارائه یک الگوریتم کارآمد با حفظ صحت و نرخ موفقیت الگوریتم است. آن‌ها از سه قید فاصله میان نقطه GPS و قطعه‌راه، اختلاف میان راستای نقطه GPS و مسیر و اختلاف میان راستای نقاط متوالی GPS و راستای قطعات راه برای شناسایی بهترین قطعه‌راه در میان سایر کاندیداها استفاده کرده‌اند. مهم‌ترین خاصیت این الگوریتم نرخ بالای تشخیص صحیح قطعه‌راه در مناطق شهری است. از جمله معایب این روش همبستگی میان پارامترهای جهتی و به‌کارگیری پارامترهای توپولوژیک است که موجب ناکارآمدی الگوریتم در مواجهه با داده‌های با نرخ نمونه‌برداری پایین می‌گردد. بدین منظور در این مقاله از شباهت میان زاویه حامل میان راستای دو نقطه قرائت شده متوالی و راستای نقطه منطبق شده قبلی با نقطه کاندیدا استفاده شده است. همچنین در این پژوهش از معیار قید اختلاف کوتاه‌ترین

مسیر برای مواجهه با نرخ‌های نمونه‌برداری متفاوت استفاده شده است. بهره‌گیری از فناوری‌های رایانش ابری پردازش توزیع‌یافته MM را نیز ممکن کرده است. لئو و همکاران [۲۱] روشی برای انطباق نقشه وزن‌مبنا در کنار یک روش درون‌یابی بر پایه الگوریتم انطباق توسعه داده‌اند. آن‌ها از اطلاعات هندسی و توپولوژیک شبکه راه‌ها و بهره‌گیری از اطلاعات تاریخی خط سیر خودروها برای بهبود دقت روش خود بهره برده‌اند. از جمله نقایص عمده این روش می‌توان به استفاده از ضرایب وزن مساوی و استفاده از پارامترهای توپولوژیک اشاره نمود که در پژوهش حاضر نواقص مذکور به کمک مفاهیم سطوح اطمینان، روش وزن‌دهی پویا و استفاده از معیارهای هندسی پاسخ داده شده است.

برای پاسخ به چالش نرخ نمونه‌برداری پایین داده‌های FCD تحقیقات معدودی انجام شده است. میوا و همکاران [۱۲] روشی بر پایه انطباق نقاط بر اساس نقطه قبلی و نقطه بعدی ارائه کرده است. روش مذکور دارای سرعت محاسبات مناسبی است. در عین حال در این روش احتمال انتشار خطا در صورت انطباق نادرست نقطه قبلی به نقاط بعدی وجود دارد. چن و همکاران [۱۳] نیز روشی برای پاسخ به چالش نرخ نمونه‌برداری پایین ارائه کرده‌اند. در این مقاله الگوریتمی با برنامه‌نویسی چند معیاره پویا (MDP) برای کاهش تعداد یال‌های کاندیدا و افزایش دقت انتخاب بهترین یال استفاده شده است. همچنین در این مقاله یک الگوریتم دیکسترا بهبودیافته جهت تعیین مسیرهای کاندیدا ارائه شده است. قدوس و واشنگتن [۱۴] از الگوریتم کوتاه‌ترین مسیر  $A^*$  و داده خط سیر برای افزایش دقت فرآیند انطباق نقشه در داده‌های GPS با نرخ نمونه‌برداری پایین استفاده کرده‌اند.  $A^*$  برای تشخیص کوتاه‌ترین مسیر میان دو نقطه متوالی GPS استفاده شده است. از جمله نقایص این روش استفاده از الگوریتم ژنتیک برای تعیین ضرایب وزن قیود است. اما استفاده از الگوریتم‌های تکاملی حالت سراسری الگوریتم MM را از بین می‌برد چراکه الگوریتم باید مرتباً بر اساس پارامترهای محلی آموزش یابد. به علاوه این امر باعث کاهش کارآمدی و سرعت تحلیل MM نیز می‌گردد [۲۲]. این نقیصه در روش نیکولیچ و همکاران [۱۵] نیز مشاهده می‌شود.

۱ Multi-Criteria Dynamic Programming (MDP)



شکل ۱- نمودار جریان کار الگوریتم انطباق نقشه بهبودیافته

### ۳- مبانی نظری

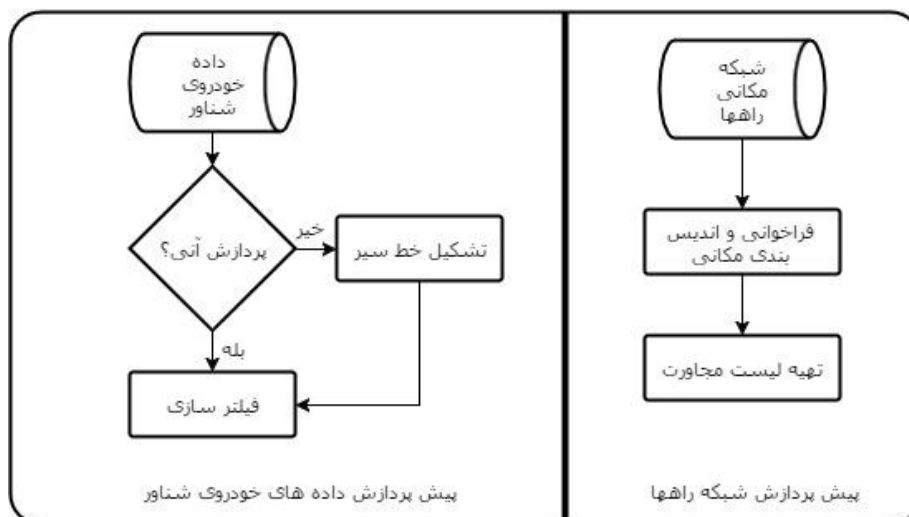
#### ۳-۱- الگوریتم انطباق نقشه بهبودیافته

در این پژوهش باهدف پاسخ به چالش‌های کاربردهای نوین ITS، الگوریتمی بهبودیافته برای تحلیل انطباق نقشه ارائه شده است. الگوریتم حاضر، در مواجهه با نرخ‌های نمونه برداری گوناگون مؤثر است و فارغ از وابستگی به حسگرهای اضافی دارای نرخ موفقیت مناسبی است. در این روش از شبکه مکانی راه‌ها در کنار داده‌های زمانی مکانی خودروهای شناور استفاده شده است.

عمده این نقایص در پژوهش حاضر به کمک روش وزن‌دهی پویا و استفاده از سطوح اطمینان پاسخ داده شده است.

باهدف بهبود عملکرد و کارایی الگوریتم‌های انطباق نقشه در مواجهه با داده‌های با نرخ نمونه برداری بالا، لی و همکاران [۲۳] الگوریتمی به نام \*Passby را ارائه کرده‌اند. این الگوریتم به‌جای انطباق تمامی نقاط GPS تنها آن‌هایی را که به تقاطعات نزدیک هستند شناسایی و منطبق می‌کند. بدین ترتیب حجم پردازشی کاهش یافته و از پردازش نقاط اضافی میانی جلوگیری به عمل می‌آید. نقص عمده این روش ناکارآمدی آن در مواجهه با داده‌های با نرخ نمونه برداری پایین است. هوانگ و همکاران [۸] نیز بر پایه روش لئو و همکاران [۲۱] و مشابه پژوهش لی و همکاران [۲۳] روشی بهبودیافته برای الگوریتم انطباق نقشه در داده‌های با نرخ نمونه برداری بالا ارائه کرده‌اند. روش مذکور مشابه با پرش از روی نقاط میانی تنها نقاط نزدیک به تقاطعات را بر روی نقشه منطبق می‌کند. برخلاف تمامی مقاله‌های بررسی شده نقطه قوت روش هوانگ و همکاران استفاده از سکوی هدوپ برای مواجهه با چالش ذخیره‌سازی، مدیریت، تحلیل و پردازش داده‌های GPS است. ازجمله نواقص مهم این روش نیز امکان انتشار خطا و ناکارآمدی در مواجهه با داده‌های با نرخ نمونه برداری پایین است که در پژوهش حاضر نواقص فوق به کمک به‌کارگیری معیارهای هندسی، معیار کوتاه‌ترین مسیر و وزن‌دهی پاسخ داده شده است.

شایان‌ذکر است پیش‌ازاین، نسخه ابتدایی چارچوب پیشنهادی، در مقاله رحیمی و حکیم پور [۲۴] ارائه شده که در پژوهش حاضر بهبودهایی مانند استفاده از معیارهای هندسی در کنار بهره‌گیری از معیار اختلاف کوتاه‌ترین مسیر جهت بهبود صحت عملکرد الگوریتم، مفاهیم سطوح اطمینان، روش‌های وزن‌دهی پویا و استفاده از اطلاعات HDOP برای بهبود عملکرد الگوریتم در مواجهه با داده‌های با نرخ نمونه برداری متفاوت به آن اضافه گردیده است.



شکل ۱- جریان کار مرحله پیش‌پردازش

حذف داده خودروهایی است که دارای قرائت‌های خارج از حدود شبکه راه‌ها هستند و یا نرخ نمونه‌برداری آن‌ها آن‌چنان قلیل بوده که فاصله زمانی میان دو گزارش متوالی از حد آستانه مشخصی بیشتر باشد. این حد آستانه به‌صورت تجربی و مشابه روش شناخته‌شده لئو و همکاران [۲۱] ۲۰ دقیقه در نظر گرفته شده است.

اطلاعات پیش‌پردازش شبکه راه‌ها پس از محاسبه در HDFS ذخیره‌سازی شده و تا زمان محاسبه مجدد بدون تغییر باقی می‌مانند. بازه محاسبه مجدد اطلاعات پیش‌پردازش شده می‌تواند وابسته به قیود زمانی و حساسیت داده تعیین شود. همچنین اطلاعات پیش‌پردازش خودرو شناور نیز می‌توانند بسته به نوع کاربرد (آنی یا پردازش جمعی) نگهداری شوند. نمودار جریان کار مرحله پیش‌پردازش در شکل ۲ نمایش داده شده است.

### ۳-۱-۲- تعیین یال‌های کاندیدا

می‌دانیم جستجوی تمامی شبکه برای یافتن یال‌های کاندیدای هر نقطه GPS امری زمان‌بر و پرهزینه است. لذا نیاز است بخش کوچکی از شبکه راه‌ها برای شناسایی یال‌های کاندیدا انتخاب شوند. یکی از رویکردهای انتخاب مناطق کاندیدا رویکرد ایجاد حوزه<sup>۱</sup> اطراف نقطه و انتخاب یال‌های متقاطع با محدوده حوزه است که این روش به علت هزینه‌های پردازشی روش مناسبی نیست. برای این منظور از شاخص‌بندی مکانی ایجادشده در مرحله

در روش پیشنهادی ابتدا گامی به نام پیش‌پردازش شبکه راه‌ها و داده‌های FCD را برای پردازش اصلی آماده‌سازی می‌کند. پس از پیش‌پردازش داده‌ها، الگوریتم پیشنهادی موقعیت نقاط بر روی یال‌ها را یک‌به‌یک تعیین می‌کند. همچنین در این الگوریتم از اطلاعات مسیر خودرو برای بهبود نرخ موفقیت و افزایش سطح اطمینان استفاده شده است. شکل ۱ نمایانگر جریان کار پیشنهادی برای الگوریتم انطباق نقشه بهبودیافته است که در ادامه به تشریح آن پرداخته می‌شود.

### ۳-۱-۱- گام پیش‌پردازش

مرحله پیش‌پردازش به دو بخش پیش‌پردازش شبکه راه‌ها و پیش‌پردازش داده‌های خودرو شناور تقسیم‌بندی می‌شود. در مرحله پیش‌پردازش شبکه راه‌ها، ابتدا شبکه راه‌ها در حافظه مرکزی فراخوانی می‌شود. سپس باهدف بهبود عملکرد پرس‌وجوهای مکانی روی شبکه راه‌ها یک شاخص مکانی دوبعدی برای تمامی شبکه ایجاد می‌شود. همچنین باهدف یافتن کوتاه‌ترین مسیر بر روی شبکه راه‌ها در شرایط نرخ نمونه‌برداری پایین، یک لیست مجاورت نیز در کنار شاخص مکانی ذخیره می‌شود. در این لیست برای هر گره مجموعه گره‌های مجاز جهت حرکت و طول یال‌های مربوطه ذخیره شده است. پیش‌پردازش داده‌های خودروی شناور نیز شامل دو گام است. در صورتی که الگوریتم به‌صورت پردازش جمعی اجرا شود گامی برای دسته‌بندی داده دریافتی بر اساس شناسه خودرو و استخراج خط سیر آن‌ها اجرا می‌شود. گام بعدی

<sup>۱</sup> Buffer

مشابه روش هاشمی و کریمی [۹] استفاده شده است. در این روش یک لیست از مقادیر سطوح اطمینان ۲۰ نقطه اخیر در قالب فایل متنی و بر روی HDFS ذخیره می‌شوند. سپس هر زمان که این لیست به‌روزرسانی گردید دو حد آستانه  $k_1$  (حد آستانه پایین) و  $k_2$  (حد آستانه بالا) مشابه روش هاشمی و کریمی [۹] و به کمک میانگین ( $\mu$ ) و انحراف معیار ( $\sigma$ ) سطوح اطمینان و در قالب روابط ۱ و ۲ تعیین می‌شوند.

$$k_1 = \mu - \sigma \quad (1)$$

$$k_2 = \mu + \sigma \quad (2)$$

در روش پیشنهادی ما در صورتی که سطح اطمینان هر نقطه مقداری بیشتر از  $k_2$  داشته باشد نقطه به‌عنوان نقطه نهایی انتخاب می‌شود. اما اگر سطح اطمینان مقداری بین  $k_1$  و  $k_2$  اتخاذ کرده باشد مسیرهای کاندیدای حرکت خودرو بازسازی شده و قید اختلاف کوتاه‌ترین مسیر برای تصمیم‌گیری نهایی مورد استفاده قرار می‌گیرند. بدین ترتیب قید اختلاف کوتاه‌ترین مسیر تنها در مواقع عدم اطمینان کافی به نتایج به‌کارگیری می‌شود. بدین ترتیب در کنار بهبود صحت تحلیل انطباق نقشه در نرخ‌های نمونه‌برداری متفاوت، با کاهش گام‌های محاسبات وزن، سرعت عملکرد روش پیشنهادی نیز نسبت به روش اولیه بهبود می‌یابد. در صورتی که پس از به‌کارگیری اختلاف کوتاه‌ترین مسیر مقدار سطح اطمینان همچنان کمتر از  $k_2$  باشد الگوریتم تنها در صورتی نقطه کاندیدا را به‌عنوان نقطه حقیقی انتخاب می‌کند که بهترین نقطه کاندیدا در فرآیند محاسبات مجدد وزن تغییری نکرده باشد. در غیر این صورت الگوریتم موقعیت نقطه را ثبت نکرده و آخرین وضعیت آن را به‌روزرسانی نمی‌کند. علت این امر عدم اطمینان کافی از صحت انتخاب بهترین نقطه کاندیدا است.

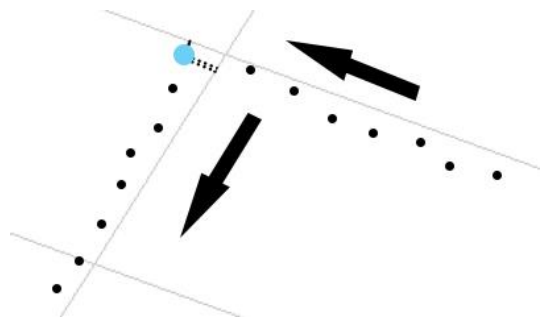
### ۳-۱-۴ - قیود وزن‌دهی

پیچیده‌ترین و زمان‌برترین گام در تمامی الگوریتم‌های انطباق نقشه گام وزن‌دهی است. در پژوهش حاضر، ما مشابه روش‌های گرینفیلد [۱۱]، قدوس و همکاران [۱۰] و ولاگا و همکاران [۱۶]، دو قید هندسی همسایگی مکانی و شباهت میان زاویه جهت‌گیری و راستای یال کاندیدا را به‌عنوان دو قید نخست وزن‌دهی انتخاب نموده‌ایم.

پیش‌پردازش استفاده می‌کنیم، لذا در عملیات جستجوی منطقه کاندیدا هیچ‌گونه پردازش اضافی به سیستم تحمیل نمی‌شود. در این روش برای یافتن مناطق کاندیدا نیاز است تنها سلول‌هایی از شاخص مکانی که فاصله نقطه GPS از مرکز آن‌ها از یک حد آستانه کمتر است را بیابیم. این حد آستانه مشابه روش قدوس [۲] و ولاگا [۱۶] با توجه به حداکثر خطای موقعیت‌یابی ۴۰ متری در مناطق متراکم شهری (صحت موقعیت‌یابی دوعبدهی در ۹۵٪ اوقات)  $40\sqrt{2}$  (به‌صورت تقریبی ۶۰ متر) انتخاب شده است. بعد از یافتن سلول‌های کاندیدا یال‌هایی که درون این سلول‌ها واقع شده‌اند و یا از این سلول‌ها گذشته‌اند به‌عنوان یال‌های کاندیدا فراخوانی شده و در لیست یال‌های کاندیدا قرار می‌گیرند. در گام بعدی یال‌های تکراری از این لیست حذف می‌شوند.

### ۳-۱-۳ - وزن‌دهی یال‌ها و نقاط کاندیدا و محاسبه سطح اطمینان

در تحقیقات پیشین انجام‌شده بر روی تحلیل انطباق نقشه وزن‌مبنا، قیود گوناگونی برای وزن‌دهی استفاده شده است. این قیود به دودسته قیود هندسی و توپولوژیک دسته‌بندی می‌شوند. استفاده از قیود توپولوژیک مانند بررسی اتصال یا گردش ممنوع در محاسبات وزن دارای یک نقص عمده است. اگر نرخ نمونه‌برداری داده پایین باشد، در فاصله میان دو نقطه، خودرو ممکن است از چندین یال عبور کرده و بر روی یالی قرار داشته باشد که دیگر به یال قبلی متصل نیست. بدین ترتیب وزن نقطه حقیقی کاهش‌یافته و الگوریتم دچار اشتباه می‌شود. برای مثال در شکل ۱ لحاظ نمودن قیود توپولوژیک ممکن است باعث بروز اشتباه در تعیین نقطه حقیقی شود. از طرفی اطلاعات حاصل از قیود هندسی همیشه برای تعیین موقعیت صحیح خودرو بر روی شبکه راه‌ها کافی نیست. لذا در این الگوریتم و باهدف پاسخ به چالش نرخ نمونه‌برداری متغیر در تحلیل انطباق نقشه ما روشی نوین ارائه کرده‌ایم. در این روش ابتدا وزن نقاط کاندیدا به کمک قیود هندسی تعیین‌شده و نقطه دارای بیشترین وزن به‌عنوان نقطه کاندیدای اصلی انتخاب می‌شود. سپس سطح اطمینان برای نقطه کاندیدای اصلی تعیین می‌شود. برای محاسبه سطح اطمینان و روبرو شدن با چالش عدم قطعیت موجود در موقعیت تعیین‌شده برای نقطه از روشی



شکل ۲- نمونه استفاده از اختلاف راستا میان زاویه جهت‌گیری خودرو و راستای یال در انطباق نقشه

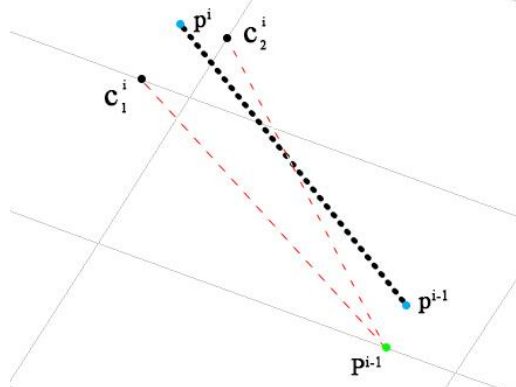
در روش پیشنهادی وزن اختلاف راستا به کمک رابطه ۴ بیان می‌شود:

$$f_H = \frac{\Delta\alpha_i}{\sum_{j=1}^N \Delta\alpha_j} \quad (4)$$

در رابطه  $\Delta\alpha_{i,4}$  اختلاف میان راستای خودرو و یال کاندید  $i$  و  $N$  تعداد یال‌های کاندید است.  $f_H$  تابعی با مقادیر مثبت (بین ۰ تا ۱) است. هرچه مقدار  $f_H$  کمتر باشد، احتمال آنکه وسیله نقلیه بر روی آن یال در حال حرکت بوده است افزایش می‌یابد. قید بعدی اختلاف میان زاویه‌های حامل راستای دو نقطه قرائت‌شده متوالی و راستای نقطه منطبق شده قبلی با نقطه کاندید است (شکل ۴) که در این مقاله آن را به اختصار اختلاف زوایای حامل می‌نامیم. در روش پیشنهادی وزن شباهت میان راستای دو نقطه GPS متوالی و راستای نقطه منطبق شده قبلی با نقطه کاندید به کمک رابطه ۵ بیان می‌شود:

$$f_B = \frac{\Delta\beta_i}{\sum_{j=1}^N \Delta\beta_j} \quad (5)$$

در رابطه  $\Delta\beta_{i,5}$  اختلاف میان راستای خودرو و یال کاندید  $i$  و  $N$  تعداد یال‌های کاندید است. هرچه مقدار  $f_B$  کمتر باشد، احتمال آنکه وسیله نقلیه بر روی آن یال در حال حرکت بوده است افزایش می‌یابد.



شکل ۳- نمونه به‌کارگیری قید اختلاف زوایای حامل

همان‌طور که پیش‌ازین نیز بیان شد یکی از اهداف الگوریتم ارائه‌شده سازگاری با نرخ‌های نمونه‌برداری متفاوت است. در بسیاری از تحقیقات پیشین استفاده از قیود توپولوژیک مانند تحلیل اتصال یا بررسی گردش ممنوع برای وزن‌دهی MM استفاده شده است. این در حالی است که در شرایط نرخ نمونه‌برداری پایین استفاده از قیود توپولوژیک منطقی نیست [۱۳]. گرینفیلد [۱۱] و قدوس [۲] ثابت نموده‌اند که شباهت در جهت‌گیری قید مهم‌تری نسبت به همسایگی مکانی است. بدین منظور می‌توان برای بهبود دقت انطباق نقشه یک قید جهت‌ی دیگر برای وزن‌دهی به محاسبات وارد نمود. هاشمی و کریمی [۹] از شباهت میان خط واصل میان دو نقطه GPS متوالی با راستای یال استفاده نموده است. نقص عمده این قید همبستگی با قید شباهت میان زاویه جهت‌گیری و راستای یال کاندید است. بدین منظور ما در این مقاله از شباهت میان زاویه حامل میان راستای دو نقطه قرائت‌شده متوالی و راستای نقطه منطبق شده قبلی با نقطه کاندید استفاده نموده‌ایم. در ادامه این قیود به تفصیل بررسی خواهند شد.

قید همسایگی مکانی بیانگر فاصله قائم میان نقطه GPS و یال کاندید است. این قید تابعی با مقادیر مثبت (بین ۰ تا ۱) است. هرچه این مقدار کمتر باشد، احتمال آنکه وسیله نقلیه بر روی آن یال در حال حرکت بوده است افزایش می‌یابد. در این روش تابع وزن برای همسایگی مکانی در قالب رابطه ۳ بیان می‌شود:

$$f_D = \frac{d_i}{\sum_{j=1}^N d_j} \quad (3)$$

در رابطه  $d_{i,3}$  فاصله میان نقطه قرائت‌شده تا یال  $i$  و  $N$  تعداد یال‌های کاندید است. قید بعدی اختلاف راستا میان زاویه جهت‌گیری خودرو و راستای یال است که در این مقاله آن را به اختصار اختلاف راستا می‌نامیم. زاویه جهت‌گیری توسط دستگاه GPS و به کمک جایجایی دوپلر امواج GPS محاسبه می‌شود [۲۵]. همچنین در هنگام وجود مدول مغناطیس‌سنج در دستگاه از اطلاعات به‌دست‌آمده از این مدول برای محاسبه اختلاف راستا استفاده می‌شود. راستای حرکت خودرو به‌صورت مستقیم به راستای یال عبوری وابسته است. همان‌طور که در شکل ۳ نمایان است در صورت عدم به‌کارگیری پارامتر جهت، نقطه آبی به‌صورت غلط بر روی یال دیگری منطبق می‌شود.



### ۳-۱-۵- ایجاد مسیرهای کاندیدا و محاسبه قید اختلاف کوتاه‌ترین مسیر

اگر یک خودرو با سرعت ۱۱۰ کیلومتر بر ساعت در حال حرکت باشد و نرخ نمونه‌برداری آن بر روی ۱ دقیقه تنظیم شده باشد حداقل دو داده دریافتی مسافتی در حدود ۱،۸ کیلومتر را طی کرده است. از آنجا هیچ اطلاعات دیگری در این حداقل در دسترس نیست، می‌توان فرض نمود که خودرو در این میان کوتاه‌ترین مسیر را طی نموده است. کوتاه‌ترین مسیر از  $u$  به  $v$  مسیری مانند  $p$  و با وزن  $w(p) = \delta(u, v)$  است به صورتی که وزن آن حداقل مقدار ممکن باشد. این کوتاه‌ترین مسیر را می‌توان با کمک نقطه منطبق شده قبلی و نقطه کاندیدای فعلی یافت. تحقیقات متعددی بر روی انواع روش‌های یافتن کوتاه‌ترین مسیر انجام شده است. با فرض آنکه تمامی وزن‌ها در گراف غیر منفی هستند، مسئله یافتن کوتاه‌ترین مسیر بین دو نقطه  $u$  و  $v$  (مسئله کوتاه‌ترین مسیر تک زوجه) را می‌توان با اعمال الگوریتم‌های ناآگاهانه مانند الگوریتم دیکسترا<sup>۱</sup> حل نمود. اگرچه الگوریتم دیکسترا روشی کامل و بهینه برای یافتن کوتاه‌ترین مسیر از گره آغازین تا سایر گره‌های گراف است، اما این الگوریتم یک روش جستجوی ناآگاهانه است. باهدف بهبود سرعت جستجو، در این الگوریتم از یک روش جستجوی کوتاه‌ترین مسیر آگاهانه به نام  $A^*$  بهره برده‌ایم.

مشابه رابطه ۶،  $A^*$  برای تخمین کوتاه‌ترین مسیر، هزینه رسیدن به گره  $n$  ( $g(n)$ ) را با تابع هیوریستیک هزینه رسیدن از گره  $n$  تا مقصد ( $h(n)$ ) جمع کرده تا مقدار تابع هزینه نهایی ( $f(n)$ ) حاصل شود. هارت و همکاران [۲۶] اثبات کرده‌اند که با انتخاب یک هیوریستیک شایسته بهینگی روش  $A^*$  تضمین شده است. همچنین با توجه به بازه نرخ نمونه‌برداری FCD، فاصله اقلیدسی میان دو نقطه قیدی مناسب برای توابع هزینه و هیوریستیک است. لذا در این تحقیق از فاصله اقلیدسی به‌عنوان قید اصلی توابع هزینه و هیوریستیک استفاده شده است.

$$f(n) = g(n) + h(n) \quad (۶)$$

پس از محاسبه کوتاه‌ترین مسیر میان نقطه منطبق شده قبلی با یک نقطه کاندیدا با کمک لیست مجاورت،

<sup>۱</sup> Dijkstra's Algorithm

می‌توان مقایسه‌ای میان این مقدار و فاصله اقلیدسی میان دو نقطه متوالی قرائت‌شده انجام داد. بدیهی است هرچه اختلاف این دو مقدار کمتر باشد احتمال آنکه نقطه کاندیدا همان موقعیت حقیقی خودرو بوده است بیشتر است. در روش پیشنهادی قید اختلاف کوتاه‌ترین مسیر به کمک رابطه ۷ محاسبه می‌شود:

$$f_{SP} = \frac{\Delta l_i}{\sum_{j=1}^N \Delta l_j} \quad (۷)$$

در رابطه ۷،  $\Delta l_i = |l_{SP} - l_e|$  اختلاف میان کوتاه‌ترین مسیر میان نقطه منطبق شده قبلی با یک نقطه کاندیدا  $i$  ام ( $l_{SP}$ ) و فاصله اقلیدسی میان دو نقطه آخر و متوالی قرائت‌شده ( $l_e$ ) است.  $N$  نیز تعداد نقاط کاندیدا برای نقطه آخر قرائت‌شده است.

### ۳-۱-۶- جمع وزن دار قیود

پس از محاسبه مقادیر حاصل از توابع قیود هندسی لازم است این مقادیر را جمع کنیم. لذا امتیاز نهایی برای هر نقطه کاندیدا در قالب رابطه ۸ تعیین می‌شود:

$$Score = W_D * f_D + W_H * f_H + W_B * f_B + W_{SP} * f_{SP} \quad (۸)$$

که در رابطه ۸،  $W_D$  و  $W_H$  و  $W_B$  و  $W_{SP}$  به ترتیب ضرایب تأثیر وزن قیود همسایگی مکانی، اختلاف راستا، اختلاف زوایای حامل و اختلاف کوتاه‌ترین مسیر است. یکی از چالش‌ها و منابع خطای اصلی در فرآیند انطباق نقشه مسئله تعیین ضریب تأثیر وزن قیود است. در تحقیقات انجام‌شده رویکردهای متفاوتی برای جمع وزن دار قیود انطباق نقشه اتخاذ شده است. گرینفیلد [۱۱] و لئو و همکاران [۲۱] ضریب تأثیر وزن‌ها را مساوی فرض نموده‌اند. قدوس و همکاران [۱۰] به‌صورت تجربی ضرایب ثابتی را برای الگوریتم خود تعیین نموده‌اند. ولاگا و همکاران [۱۶] با بررسی و بهینه‌سازی خطاهای حاصله، ضرایب از پیش تعیین‌شده و محلی را استفاده نموده است. همچنین قدوس و واشنگتن [۱۴] از الگوریتم ژنتیک برای تعیین وزن‌ها استفاده نموده که باعث کاهش سرعت عملکرد تحلیل انطباق نقشه می‌شود. در این مقاله و باهدف ارائه یک چارچوب یکپارچه و مستقل از پارامترهای محلی، ما از مفهوم وزن‌دهی پویا برای جمع وزن‌های محاسبه‌شده استفاده کرده‌ایم. در این روش وزن هر قید

$$W_{SP} = \begin{cases} 0 & H > H_2 \\ \frac{H_2 - H}{H_2 - H_1} & H_1 \leq H \leq H_2 \\ 1 & H < H_1 \end{cases} \quad (10)$$

در حالت پیش‌فرض مقدار  $W_{SP}$  برابر صفر است. در صورتی که سطح اطمینان محاسبه‌شده برای بهترین نقطه کاندیدا کمتر از  $k_2$  باشد، الگوریتم خط سیر طی شده بین نقطه منطبق شده قبلی و نقاط کاندیدای فعلی را بازسازی کرده و  $W_{SP}$  و  $f_{SP}$  را برای هر نقطه کاندیدا محاسبه می‌نماید. سپس این مقادیر به کمک سایر قیود آمده و این بار مرحله تجمیع وزن‌دار قیود به همراه مقادیر اختلاف کوتاه‌ترین مسیر تکرار می‌شود.

### ۳-۲- چارچوب تحلیل انطباق نقشه مبتنی بر فناوری‌های رایانش ابری

#### ۳-۲-۱- معماری هدوپ و الگوی نگاشت کاهش

هدوپ یک پروژه متن‌باز از آپاچی است که راه‌حلهایی برای پردازشی توزیع‌یافته، مقیاس‌پذیر و مطمئن ارائه می‌کند. دو جزء اصلی در هدوپ، فایل سیستم توزیع‌یافته هدوپ<sup>۲</sup> (HDFS) و روش برنامه‌نویسی نگاشت کاهش<sup>۳</sup> است که در این پژوهش هر دو این اجزا به‌کارگیری شده‌اند. همچنین در چارچوب پیشنهادی از عامل فلووم<sup>۴</sup> و اوزی<sup>۵</sup> به‌عنوان دو عضو دیگر از اکوسیستم هدوپ استفاده شده است. نگاشت-کاهش [۲۷] یک فناوری انقلابی در پردازش و تحلیل داده‌ها و یکی از محبوب‌ترین فناوری‌ها در زمینه داده‌های بزرگ است.

به‌صورت مستقل برای هر نقطه GPS و بر اساس شرایط خاص همان نقطه محاسبه‌شده است که این امر می‌تواند الگوریتم را از استفاده از پارامترهای محلی و یا استفاده از الگوریتم‌های فرا ابتکاری مستقل کند. بدین منظور مشابه روش هاشمی و کریمی [۹] وزن دو قید همسایگی مکانی و اختلاف راستا ( $W_H$  و  $W_D$ ) تعیین می‌شود. هر چه مسافت پیموده شده میان دو نقطه GPS طولانی‌تر باشد راستای GPS کمتر تحت تأثیر خطای GPS قرار می‌گیرد [۹]. لذا اختلاف میان زاویه‌های حامل راستای دو نقطه قرائت‌شده متوالی و راستای نقطه منطبق شده قبلی با نقطه کاندیدا نیز به‌صورت مستقیم به مسافت پیموده شده میان دو نقطه GPS وابسته است. لذا در روش پیشنهادی  $W_B$  به کمک رابطه ۹ تعیین می‌شود. هرچه این مسافت طولانی‌تر باشد مقدار  $W_B$  نیز افزایش می‌یابد.

$$W_B = \begin{cases} 1 & l \geq D \\ \frac{l}{D} & l < D \end{cases} \quad (9)$$

که در این رابطه  $l$  مسافت پیموده شده بین دو نقطه قرائت‌شده و  $D$  حد آستانه فاصله پیموده شده است. طبق ارزیابی‌های هاشمی و کریمی [۹] زمانی که فاصله طی شده میان دو نقطه GPS از ۲۵ متر بیشتر شود خطا در راستای میان دو نقطه از ۱۱ درجه بیشتر نخواهد شد. لذا مقدار  $D$  برابر 25 متر در نظر گرفته می‌شود. از آنجاکه اطمینان‌پذیری فاصله میان دو نقطه متوالی و کوتاه‌ترین مسیر محاسبه‌شده به صحت افقی نقاط GPS وابسته است، از  $W_{SP}$  از طریق رابطه ۴ و بر پایه تعدیل دقت افقی<sup>۱</sup> (HDOP) محاسبه می‌شود. لذا بدیهی است که هرچه مقدار HDOP کوچک‌تر باشد مقدار  $W_{SP}$  بیشتر خواهد بود. هاشمی و کریمی [۹] با بررسی ۲۱۰۰۰ نقطه GPS و ترسیم هیستوگرام آن‌ها دریافتند که هیچ HDOP ای کمتر از ۰,۷ در این قرائت‌ها وجود نداشته است و تنها ۲,۱٪ از نقاط HDOP بالای ۴,۵ داشته‌اند. لذا در این تحقیق ما دو حد آستانه برای HDOP در نظر گرفته و برای حد آستانه پایین ( $H_1$ ) مقدار ۰,۷ و برای حد آستانه بالا ( $H_2$ ) مقدار ۴,۵ را در نظر می‌گیریم. در روش پیشنهادی ضریب تأثیر وزن قید کوتاه‌ترین مسیر به کمک رابطه ۱۰ تعیین می‌شود.

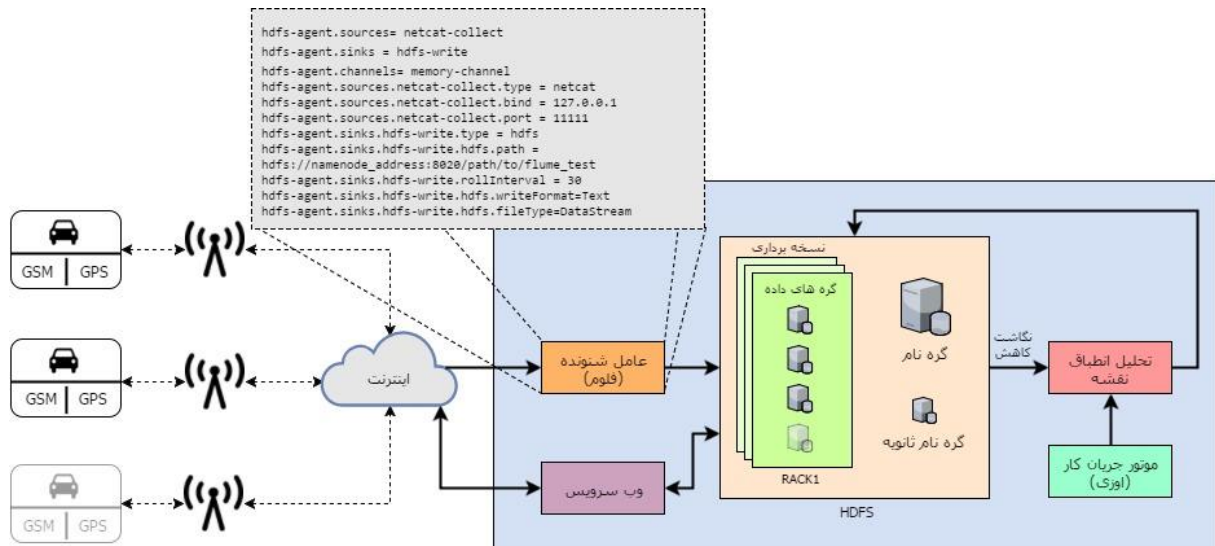
<sup>۲</sup> Hadoop Distributed File System (HDFS)

<sup>۳</sup> MapReduce

<sup>۴</sup> Flume

<sup>۵</sup> Oozie

<sup>۱</sup> Horizontal Dilution Of Precision (HDOP)



شکل ۴- معماری چارچوب تحلیل انطباق نقشه پیشنهادی

مشکل در یک گره است که HDFS با نسخه برداری از هر بلوک در گره‌های متعدد این مشکل را حل نموده است. بر روی کارساز اصلی HDFS (گره نام<sup>۱</sup>) برای هر داده ابر داده‌ای ذخیره شده است که نشان می‌دهد داده شامل چه بلوک‌هایی است و هر کدام از این بلوک‌ها در کدام بخش از خوشه ذخیره‌سازی شده‌اند.

### ۳-۲-۳- معماری عمومی چارچوب تحلیل انطباق نقشه مبتنی بر فناوری‌های رایانش ابری

همان‌طور که پیش‌ازین نیز بیان شد داده‌های FCD به صورت پیوسته در حال ارسال به سوی یک کارساز مرکزی هستند. با در نظر گرفتن تعداد خودروها و نرخ‌های نمونه برداری می‌توان گفت در هر ثانیه حجم بالایی از داده‌های FCD در حال مخابره به کارساز مرکزی است. با توجه به کاربردهای متعدد FCD، کارساز مرکزی باید بتواند این حجم داده‌های بزرگ را به صورت آنی دریافت، مدیریت، ذخیره‌سازی و پردازش کند. همچنین این داده‌ها را می‌توان پس از جمع بر روی کارساز مرکزی، برای کاربردهای بالادستی ITS مورد پردازش قرار داد. اما مدیریت، تحلیل و پردازش چنین حجم بالایی از داده‌ها یک چالش اساسی است. زمانی که حجم داده از حد مشخصی فراتر رود ذخیره‌سازی و مدیریت این داده‌ها امری پرهزینه و زمان‌بر است. به علاوه پردازش حجم بالایی از داده‌ها نیازمند منابع قدرتمند پردازشی است که این امر

نگاشت-کاهش محاسبات را به دو عملیات کلیدی دسته‌بندی می‌کند: تابع نگاشت<sup>۱</sup> که مسئول تقسیم مجموعه داده بزرگ به زیرمجموعه‌ها و پردازش هر کدام از آن‌ها به صورت مستقل است و تابع کاهش<sup>۲</sup> که نتایج حاصل از نگاشت را تجمیع و ادغام می‌کند. یک کار<sup>۳</sup> نگاشت-کاهش از چندین فرآیند مانند جداسازی و توزیع داده‌ها، نگاشت و کاهش کدها و نوشتن نتایج بر روی فایل سیستم توزیع یافته تشکیل شده است. تابع نگاشت داده ورودی را به زوج‌های کلید-مقدار تقسیم کرده و خروجی هر وظیفه نگاشت توسط کلید آن مرتبط سازی می‌شود. تابع کاهش مقادیر را تجمیع کرده و خروجی نهایی حاصل می‌شود.

### ۳-۲-۳- فایل سیستم توزیع یافته هدوپ (HDFS)

HDFS یک فایل سیستم ذخیره‌سازی توزیع یافته برای ذخیره‌سازی و انتقال مطمئن حجم بسیار بالایی از داده‌های ساختاریافته یا غیر ساختاریافته بر روی خوشه‌های رایانشی است [۲۸]. در HDFS فایل‌ها به بلوک‌هایی با ابعاد ثابت تقسیم شده و این بلوک‌ها در سطح گره‌های پردازشی پخش می‌شوند. چنین معماری به سیستم اجازه می‌دهد که بلوک‌های یک فایل را بر روی گره‌های متعددی ذخیره‌سازی کند. یکی از نقایص اصلی در فایل سیستم‌های سنتی از دست رفتن داده به علت بروز

۱ Map  
۲ Reduce  
۳ Job

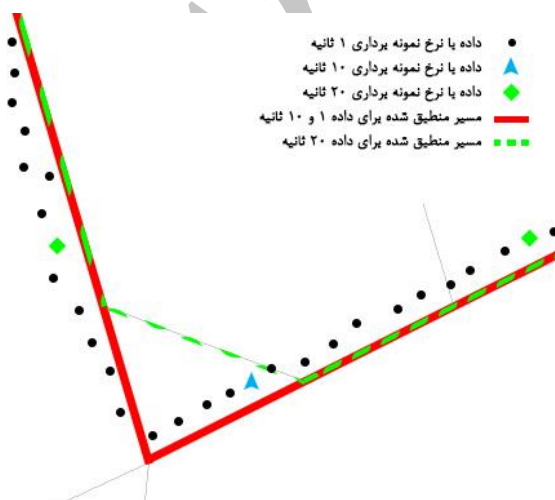
۴ NameNode

قرار گرفته است. شهر نیویورک دارای یکی از بزرگ‌ترین ناوگان‌های حمل‌ونقل در جهان و بیش از ۱۰۰۰۰۰ تاکسی و سرویس حمل‌ونقلی عمومی و خصوصی و بیش از ۲٫۵ میلیون سفر روزانه درون‌شهری است [۱۸]. مطالعه موردی بر روی سیستم حمل‌ونقل ابر شهر نیویورک می‌تواند محک مناسبی برای چارچوب پیشنهادی در پاسخ به الزامات تعیین‌شده برای یک چارچوب یکپارچه و نوین تحلیل انطباق نقشه باشد. بدین منظور در این تحقیق از مجموعه داده‌های واقعی (OSM) OpenStreetMap شهر نیویورک در ایالات متحده آمریکا استفاده شده است. در این تحقیق بیش از ۷۰ خط سیر متفاوت با نرخ نمونه‌برداری ۱ ثانیه و با ویژگی‌های زیر به‌کارگیری شده است:

۱. خط سیر GPS متعلق به حرکت یک وسیله نقلیه متحرک باشد.

۲. تمامی نقاط خط سیر GPS دارای اطلاعات سرعت، جهت‌گیری و HDOP باشند.

سپس جهت ارزیابی عملکرد الگوریتم در نرخ‌های نمونه‌برداری متفاوت، خط سیرهای انتخاب‌شده در بازه‌های زمانی ۱۰ ثانیه‌ای مورد نمونه‌برداری مجدد قرار گرفتند. در نهایت مجموعاً بیش از چهار میلیون رکورد داده با نرخ‌های نمونه‌برداری متفاوت (۱ ثانیه الی ۲ دقیقه) جهت ارزیابی نحوه عملکرد چارچوب به‌کارگیری شده است. همچنین برای شبکه راه‌ها نیز از شبکه رقومی راه‌های OSM استفاده شده است. شبکه راه‌های شهر نیویورک به‌خودی‌خود دارای بیش از ۵۳۰۰۰ گره و ۷۲۰۰۰ یال است.



شکل ۵- نمونه داده خط سیر GPS در نرخ‌های نمونه‌برداری متفاوت

نیز نیازمند هزینه‌های گزاف جهت خرید و نگهداری زیرساخت‌های حجیم رایانشی است. لذا ذخیره‌سازی، پردازش و تحلیل داده‌های FCD در سکوی سنتی چالشی بزرگ محسوب می‌شود. برای پاسخ به این چالش داده‌های بزرگ FCD در چارچوب ارائه‌شده از فناوری‌های رایانش ابری استفاده شده است. پیدایش فناوری رایانش ابری راه‌حلی سودمند، مقیاس‌پذیر و به‌صرفه در پاسخگویی به چالش‌های داده‌های بزرگ در کاربردهای حمل‌ونقلی و مکانی است. راه‌حل پیشنهادی این مقاله برای پاسخ به چالش مذکور تقسیم‌بندی مجموعه داده بزرگ و معرفی هر قسمت از پردازش به یک رایانه برای افزایش توان رایانشی است. بدین منظور ما از الگوی برنامه‌نویسی نگاشت کاهش بر روی سکوی هدوپ استفاده کردیم. شکل ۵ بیانگر معماری چارچوب تحلیل انطباق نقشه پیشنهادی است. داده‌های خودروهای شناور توسط AVL اندازه‌گیری شده و توسط فناوری‌های ارتباطی مانند سامانه‌های ارتباطی سلولی (3G/4G) به سمت کارساز مرکزی تعیین‌شده ارسال می‌شوند. کارساز به کمک عامل فلوم آپاچی داده‌ها را دریافت کرده و در HDFS ذخیره می‌کند. فلوم یک خدمت مطمئن و توزیع‌یافته برای جمع‌آوری، تجمیع و انتقال مؤثر مجموعه داده‌های بزرگ است. معماری فلوم ساده و منعطف است و اساس آن جاری‌سازی داده<sup>۱</sup> است. پایداری<sup>۲</sup> و مقاومت در برابر خطا با مکانیسم اطمینان‌پذیری قابل تنظیم و مکانیسم‌های بازیابی از خطا از دیگر قابلیت‌های فلوم است. در چارچوب پیشنهادی برای تنظیم جریان کار نیز از عامل اوزی استفاده شده است. اوزی یک نرم‌افزار تحت وب جاوا است که برای زمان‌بندی کارهای آپاچی هدوپ استفاده می‌شود. اوزی چندین کار مختلف را باهم ترکیب کرده و یک واحد منطقی از کارها را ایجاد می‌کند.

#### ۴- پیاده‌سازی و ارزیابی

##### ۴-۱- داده‌های مورد مطالعه و محیط پیاده‌سازی

در این بخش، نتایج حاصل از پیاده‌سازی الگوریتم توسعه‌یافته بر روی چارچوب پیشنهادی مورد ارزیابی

<sup>۱</sup> Data Streaming  
<sup>۲</sup> Robustness

## ۴-۲- موزی سازی الگوریتم انطباق نقشه بهبود یافته

در این بخش و با استفاده از الگوی نگاشت کاهش، الگوریتم پیشنهادی انطباق نقشه برای استفاده بر روی سکوهاى رایانش ابری موزی سازی شده و بر روی چارچوب پیشنهادی پیاده سازی شده است. الگوریتم موزی انطباق نقشه پیشنهادی از دو گام کلی تشکیل شده است. این گامها به صورت خطی اجرا شده و خروجی گام نخست ورودی گام بعدی است.

گام نخست (Job1) شامل یافتن سلولها، یالها و در نهایت نقاط کاندیدا برای داده ورودی است (شکل ۷). ورودی Job1 داده خام متنی دریافتی از دادههای خودروهایی شناور است که بر روی HDFS ذخیره شده است. در Job1 ابتدا نیاز است مناطق کاندیدا شناسایی شود. این کار توسط تابع نگاشت و در قالب رابطه ۱۱ انجام می پذیرد. به تابع نگاشت کلیدهای ورودی معرفی شده و برای هر کلید پردازنده ای اختصاص داده شده تا تمام مقادیر مورد پردازش قرار گیرند.

$$Map(PID, Pinfo) \rightarrow (PID, list(GridID)) \quad (11)$$

که در رابطه ۱۱، GridID شناسه یکتای یک سلول شاخص مکانی، PID شناسه یکتای هر نقطه قرائت شده (حاصل ترکیب شناسه خودرو و زمان مخابره اطلاعات) و list(GridID) آرایه ای از GridID های متناظر با هر PID است. بدین ترتیب از PID می توان به عنوان یک کلید یکتا در محاسبات نگاشت کاهش استفاده نمود. Pinfo نیز اطلاعاتی مانند طول و عرض جغرافیایی، سرعت وسیله نقلیه، HDOP و سایر اطلاعات مورد نیاز برای محاسبات نگاشت کاهش است. هدوپ می تواند به صورت خودکار تمامی اطلاعات مربوط به یک PID خاص را جمع کند. در این پژوهش برای یافتن فاصله میان دو نقطه GPS از فرمول فاصله کمانی هاورسین [۲۹] استفاده شده است.

تابع کاهش، خروجی تابع نگاشت که توسط تابع برزننده<sup>۴</sup> ادغام شده است را در قالب زوجهای کلید-مقدار به عنوان ورودی خود می پذیرد که در این گام کلید شناسه یک نقطه GPS و مقدار فهرستی از شناسه های سلولهای کاندیدا است.

شکل ۶ بیانگر نمونه ای از داده خط سیر GPS در نرخهای نمونه برداری متفاوت و مسیرهای منطبق شده است. قالب داده خط سیر GPS ورودی به سیستم پیاده سازی شده در جدول ۱ تشریح شده است.

جدول ۱- قالب داده های خط سیر ورودی به سیستم

|                  |                |                |
|------------------|----------------|----------------|
| Lat              | طول جغرافیایی  |                |
| Lon              | عرض جغرافیایی  |                |
| Ele              | ارتفاع         |                |
| Time             | زمان           |                |
| HDOP             | تعدیل دقت افقی |                |
| Speed            | سرعت           | اطلاعات الحاقی |
| Compass          | جهت گیری       |                |
| Compass Accuracy | صحت جهت گیری   |                |

برای ارزیابی نحوه عملکرد چارچوب پیشنهادی، برنامه نویسی الگوریتم به زبان جاوا انجام شده است. همچنین برای ارزیابی حقیقی نحوه عملکرد چارچوب، ما از خوشه پردازشی نانسی در میز آزمایش Grid5000<sup>۱</sup> استفاده کرده ایم. Grid5000 یک بستر آزمایشی بزرگ مقیاس و متنوع برای تحقیقات آزمایش مبنای در تمامی زمینه های علوم رایانه با تمرکز بر رایانش موزی و توزیع یافته مانند رایانش ابری و داده های بزرگ است. در پژوهش حاضر ما از خوشه دانشگاه نانسی<sup>۲</sup> با ۵ گره پردازشی استفاده نموده ایم. گره پردازشی اصلی پاسخگوی پردازش داده ها و مدیریت خوشه است. سایر گره های پردازشی پیرو<sup>۳</sup> تنها مسئول پردازش داده می باشند. سخت افزارها و نرم افزارهای مورد استفاده در این پژوهش در جدول ۲ نمایش داده شده اند.

جدول ۲- محیط مورد استفاده در پیاده سازی

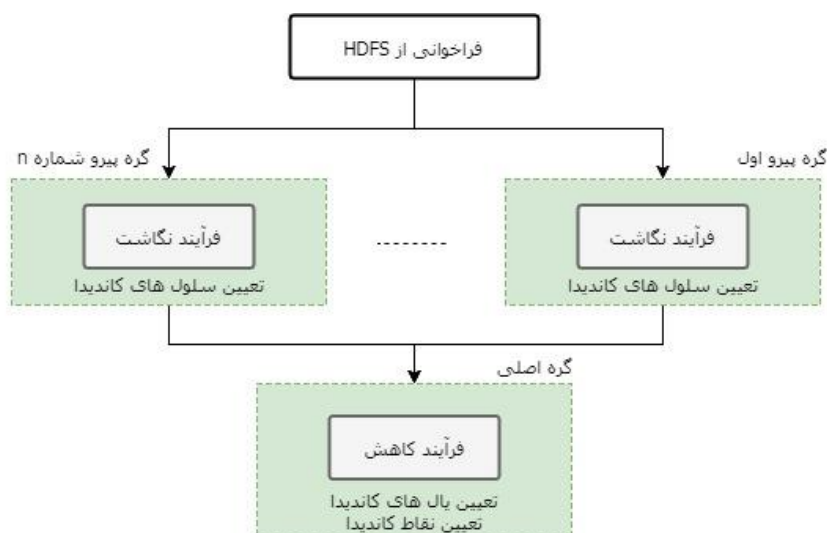
| نام فناوری      | توضیحات                                   |
|-----------------|---|
| نسخه هدوپ       | 2.7.1                                     |
| نسخه HBase      | 1.1.2                                     |
| نسخه Flume      | 1.5.2                                     |
| مدل خوشه        | Dell PowerEdge R730                       |
| پردازنده گره ها | Intel(R) Xeon(R) CPU E5-2620 v3 @ 2.40GHz |
| حافظه           | 64GB RAM                                  |
| شبکه            | 10 Gigabit Ethernet DA/SFP+               |

<sup>۱</sup> <https://www.grid5000.fr>

<sup>۲</sup> Nancy

<sup>۳</sup> Slave

۴ Shuffle



شکل ۶- گام نخست الگوریتم موازی انطباق نقشه بهبودیافته

پس از محاسبه نقاط کاندیدا برای هر نقطه GPS، در گام دوم (Job2) امتیازات کلی برای نقاط کاندیدا محاسبه می‌شود (شکل ۸). در الگوریتم پیشنهادی Job2 دارای دو فاز است. در فاز اولیه، به کمک مرحله نگاشت وزن‌ها برای هر نقطه کاندیدا محاسبه می‌شود. بدین منظور ابتدا در تابع نگاشت Job2 و بر اساس قیود و روابط بیان‌شده، وزن هر نقطه کاندیدا را می‌یابیم. ورودی تابع نگاشت نقطه کاندیدی  $i$  ( $CP_i$ ) حاصل از فرآیند Job1 و نقطه GPS است و خروجی این تابع مقدار امتیاز  $CP_i$  است. این فرآیند در قالب رابطه ۱۳ انجام می‌پذیرد.

$$Map(PID, CP_i) \rightarrow (PID, W_{CP_i}) \quad (13)$$

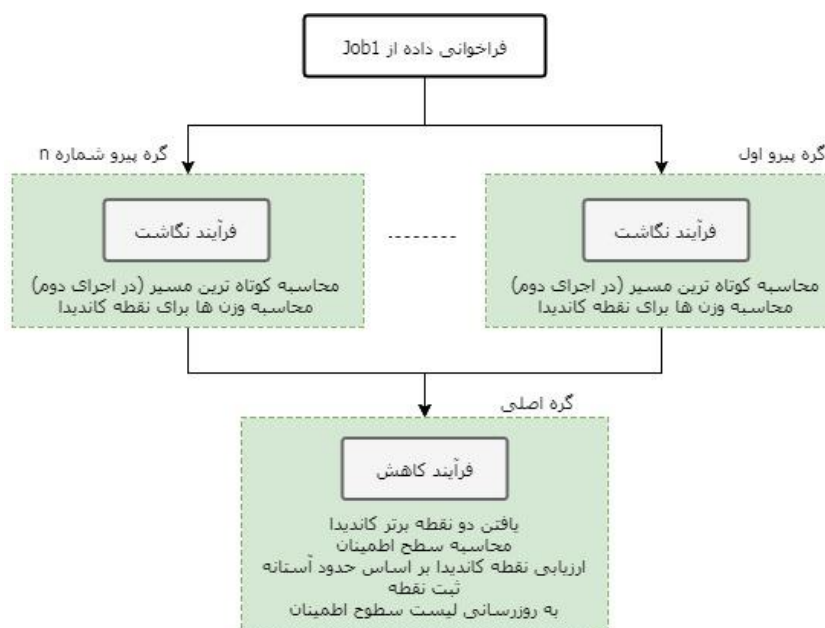
در مرحله کاهش، وزن‌ها برای هر PID جمع شده و بهترین نقطه کاندیدا و سطح اطمینان مربوط به آن مشخص می‌شود. بنا بر روش پیشنهادی در صورتی که سطح اطمینان مقداری بین  $k_1$  و  $k_2$  را اتخاذ نماید مسیرهای کاندیدا بازسازی شده و گام Job2 مجدداً و این بار در فاز دوم با مقادیر وزن جدید اجرا می‌شود. در فاز دوم برای هر نقطه کاندیدا مسیرهای کاندیدا محاسبه شده و طول کوتاه‌ترین مسیر در محاسبات وزن نقطه کاندیدا وارد می‌شود. بدین ترتیب چارچوب در عین حفظ کارآمدی، صحت خود را نیز بهبود می‌دهد. سپس در مرحله کاهش وزن جدید برای محاسبه نقطه کاندیدی جدید لحاظ می‌شود. در نهایت نیز با تأیید نقطه به‌عنوان نقطه حقیقی، خط سیر خودرو و آخرین وضعیت خودرو به‌روزرسانی می‌شود.

سپس یال‌های درون مناطق مرتبط با هر کلید به‌عنوان یال‌های کاندیدا شناسایی شده و نزدیک‌ترین نقطه یال به نقطه GPS (پای عمود) به‌عنوان نقطه کاندیدا ثبت می‌شود. هذوپ به‌صورت خودکار مدیریت اطلاعات میانی و حذف آن‌ها را بر عهده می‌گیرد. در روش پیشنهادی رابطه ۱۲ بیانگر نحوه عملکرد تابع کاهش در Job1 است. در رابطه ۱۲، CP یک نقطه کاندیدا برای نقطه GPS است.

$$Reduce(PID, list(GridID)) \rightarrow list(PID, list(CP)) \quad (12)$$

شبه کد Job1 به شرح زیر است:

1. **procedure:** Job1
2. **input parameters**
3. PID: point id
4. Pinfo: vehicle\_tag, lat, lng, timestamp, speed, hdrop
5. **begin**
6. //Map
7. index:=retrieveIndexfromHDFS();
8. **for** i:=1 to |index| **do**
9. //finding center of a grid cell
10. (cx,cy):=findingCellCenter(index(i));
11. d:=haversineDistance(PID,Pinfo,cx,cy);
12. **if** (d<threshold) **then**
13. gridID.collect(pid, index(i));
14. **end if**
15. **end for**
16. { wait until hadoop nodes have achieved work }
17. //Reduce
18. edges := findEdges(gridID);
19. candidatePoints :=projection(PID,Pinfo,edges);
20. **return** candidatePoints



شکل ۷- گام دوم الگوریتم موازی انطباق نقشه بهبودیافته

```

28. pointRegistration(PID, bestCp);
29. updatethresholdList(wBest);
30. else
31.   if (secondRun is false) then
32.     //saves info, sets secondRun to true
33.     saveInfo(PID,bestCp,true);
34.     JobClient.runJob(job2);
35.   else
36.     bestCpLast:=loadInfo(PID);
37.     if(bestCp== bestCpLast) then
38.       pointRegistration(PID, bestCp);
39.       updatethresholdList(wBest);
40.     else
41.       break;
42.     end if
43.   end if
44. end if

```

#### ۳-۴- ارزیابی عملکرد چارچوب پیشنهادی

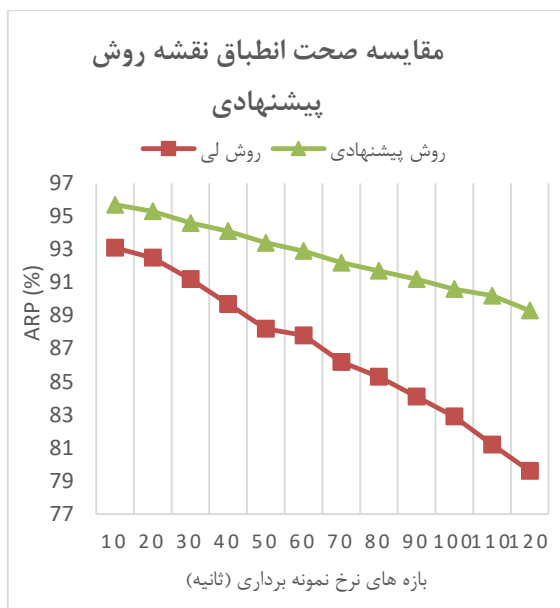
برای ارزیابی عملکرد چارچوب پیشنهادی باید تحقق الزامات تعیین شده در بخش ۱ توسط چارچوب یکپارچه تحلیل انطباق نقشه پیشنهادی نسبت به داده‌های در دسترس برای ارزیابی مورد بررسی قرار گیرد. در این تحقیق حدود آستانه در تحلیل انطباق نقشه به کمک میانگین و انحراف معیار سطوح اطمینان محاسبه می‌شود تا از هرگونه به‌کارگیری مقادیر محلی اجتناب شود. تعیین حدود آستانه سلول و حدود آستانه زمانی در بخش پیش‌پردازش نیز با کمک مراجع معتبر قبلی مانند قدوس [۲]، ولاگا [۱۶] و لئو

این فرآیند در قالب رابطه ۱۴ انجام می‌شود.

$$Reduce(PID, list(W_{CP})) \rightarrow list(PID, P_{Matched}) \quad (14)$$

شبکه کد Job2 به شرح زیر است:

1. **procedure:** Job2
2. **input parameters**
3. PID: point id
4. Pinfo: vehicle\_tag, lat, lng, timestamp, speed, hdp
5. candidatePoints: retrieved from Job1 for PID
6. cp: a point from candidatePoints list
7. mapResult: PID + weights for each cp
8. **begin**
9. //Map
10. secondRun:=isSecondRun(PID);
11. **if** (secondRun) **then**
12. lp:=retrieveLastMatchedPoint();
13. pathList:=shortestPath(cp,lp);
14. **end if**
15. //weights calculation wd,wh,wb,wsp
16. w:=calculateWeights(PID, Pinfo, cp, candidatePoints, pathList);
17. mapResult.collect(PID, cp\_w);
18. { wait until hadoop nodes have achieved work }
19. //Reduce
20. (bestCp, wBest):= getBestCp(mapResult);
21. (secondBestCp,wSBest):= getSecondBestCp(mapResult);
22. cl:=confidenceLevel(bestCp, wBest, secondBestCp, wSBest);
23. (k1,k2):=getThresholds();
24. secondRun:=isSecondRun(PID);
25. **if** (cl<k1) **then**
26. break;
27. **else if** (cl>k2) **then**



شکل ۸ مقایسه صحت روش انطباق نقشه پیشنهادی

همچنین در شکل ۹ مشاهده می‌کنیم که با افزایش بازه نمونه‌برداری صحت هر دو روش کاهش می‌یابد که این نتیجه با توجه به وجود فرضیات کوتاه‌ترین مسیر قابل‌انتظار است. در فرآیند وزن‌دهی قیود ما فرض نمودیم که خودرو در فاصله میان دو نقطه کوتاه‌ترین مسیر را طی نموده است. بدین ترتیب با افزایش بازه نمونه‌برداری فاصله میان دو نقطه GPS افزایش یافته و شانس بیشتری برای عبور خودرو از مسیرهای جایگزین وجود دارد. شکل ۱۰ بیانگر اثر قید کوتاه‌ترین مسیر بر روی صحت نهایی است. همان‌طور که مشاهده می‌شود با افزایش بازه نمونه‌برداری و افزایش عدم قطعیت در فرآیند انطباق نقشه، وجود قید کوتاه‌ترین مسیر باعث بهبود صحت کلی عملکرد الگوریتم شده است. تحقیقات نشان می‌دهد میزان اختلاف بین الگوریتم با و بدون قید کوتاه‌ترین مسیر در  $si = 120$  به ۳،۱ درصد می‌رسد.

#### ۴-۳-۲- ارزیابی سرعت عملکرد چارچوب پیشنهادی

در این تحقیق یک چارچوب مقیاس‌پذیر و توزیع‌یافته برای تحلیل انطباق نقشه ارائه شده است.

[۲۱] تعیین شده است. بدین ترتیب الگوریتم توسعه‌یافته در این پژوهش سراسری<sup>۱</sup> و مستقل از پارامترهای محلی است. چارچوب پیشنهادی فارغ از به‌کارگیری هرگونه حسگرهای اضافی، باید بتواند در روبرو شدن با نرخ‌های نمونه‌برداری متفاوت صحت مناسبی داشته باشد. همچنین این چارچوب باید بتواند در مواجهه با داده‌های بزرگ FCD سرعت پردازش مناسبی را از خود به نمایش بگذارد. لذا در ادامه این مقاله سیستم پیاده‌سازی شده بر اساس این دو عامل یعنی صحت و سرعت عملکرد موردبررسی قرار گرفته‌اند.

#### ۴-۳-۱- ارزیابی صحت عملکرد چارچوب پیشنهادی

برای اثبات صحت عملکرد الگوریتم پیشنهادی، ما پس از پیاده‌سازی روش شناخته‌شده لئو و همکاران [۲۱] نتایج دو روش را مورد مقایسه قرار داده‌ایم. همچنین در این تحقیق از شاخص شناخته‌شده نسبت صحت نقاط منطبق شده<sup>۲</sup> (ARP) به‌عنوان معیار اندازه‌گیری نرخ موفقیت الگوریتم استفاده شده است (رابطه ۱۵).

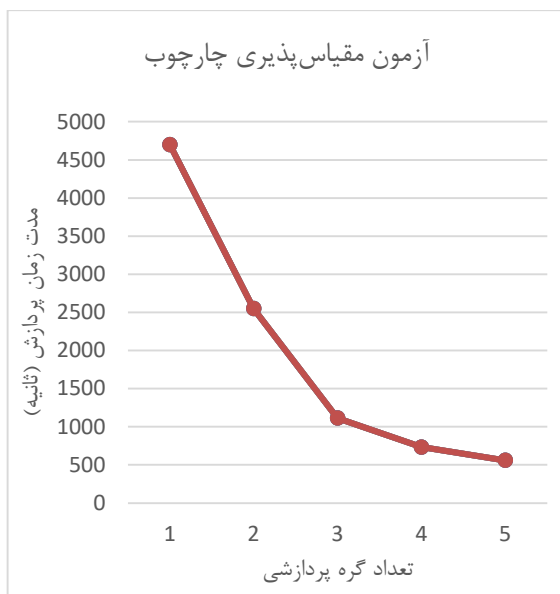
$$ARP = \frac{N_{correct}}{N_{original}} * 100\% \quad (15)$$

که در رابطه ۱۵،  $N_{correct}$  تعداد نقاط منطبق شده صحیح و  $N_{original}$  تعداد کل نقاط در بازه نمونه‌برداری است. برای تشخیص صحت انطباق نقطه  $P$  در نرخ نمونه‌برداری  $t$ ، نقطه منطبق شده  $P_m^t$  با نقطه منطبق شده در نرخ نمونه‌برداری یک ثانیه  $P_m^1$  مقایسه می‌شود. شکل ۹ صحت انطباق نقشه الگوریتم پیشنهادی را با روش لئو و همکاران [۲۱] در نرخ‌های نمونه‌برداری متفاوت نمایش می‌دهد. همان‌طور که شکل ۹ نشان می‌دهد در نرخ نمونه‌برداری ۱۰ ثانیه ( $si = 10$ )، فاصله عملکردی الگوریتم پیشنهادی با الگوریتم لئو و همکاران [۲۱] ۲،۶٪ است. این در حالی است که در  $si = 120$  این اختلاف به ۹،۷ درصد می‌رسد که این نشان از برتری روش پیشنهادی این مقاله نسبت به روش لئو و همکاران [۲۱] دارد.

<sup>۱</sup> Global

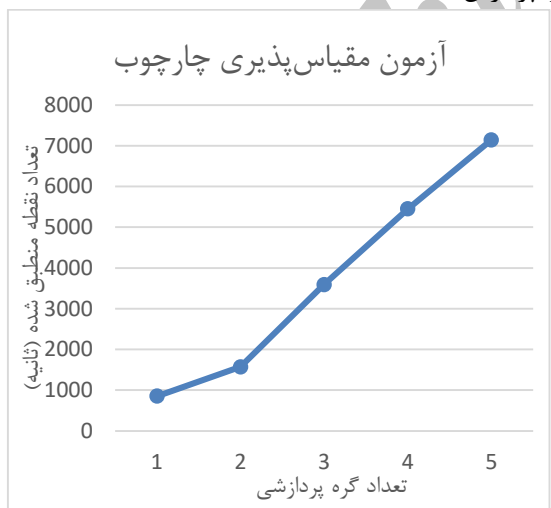
<sup>۲</sup> Accuracy Ratio of Points Matched





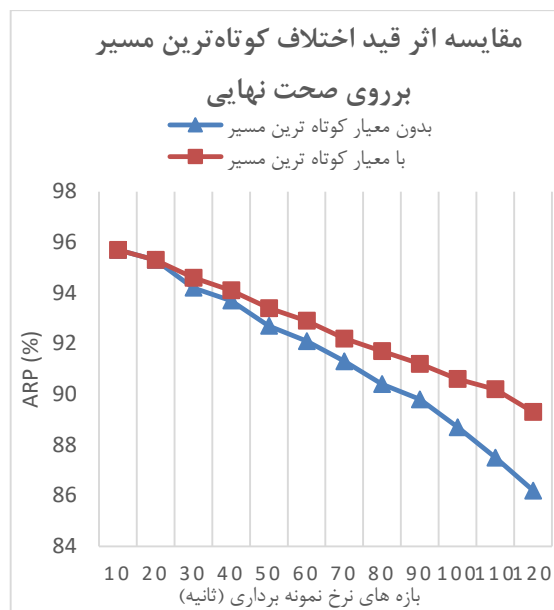
شکل ۱۰- آزمون مقیاس پذیری چارچوب (ارزیابی مدت زمان پردازش برحسب تعداد گره پردازشی)

همان طور که شکل ۱۲ نشان می دهد با افزایش تعداد گره های پردازشی تعداد نقاط منطبق شده در هر ثانیه به سرعت افزایش یافته است. بدین صورت که با حضور ۵ گره پردازشی چارچوب پیشنهادی توانسته بیش از ۷۰۰۰ نقطه در هر ثانیه را بر روی نقشه منطبق کند. این در حالی است که روش لئو و همکاران [۲۱] در بهترین حالت توانست بر روی یک گره تنها، حدود ۱۰۰۰ نقطه در ثانیه را پردازش نماید.



شکل ۱۱- آزمون مقیاس پذیری چارچوب (ارزیابی تعداد نقاط منطبق شده در ثانیه برحسب تعداد گره پردازشی)

ما نرخ کارآمدی چارچوب را به کمک ۵ گره پردازشی و برای حجم های متفاوتی از داده مورد ارزیابی قرار دادیم که حاصل آن در شکل ۱۳ نمایش داده شده است.



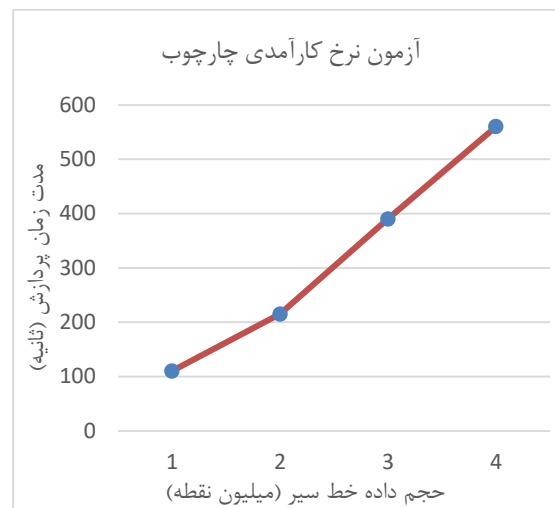
شکل ۹- مقایسه اثر قید کوتاه ترین مسیر بر روی صحت نهایی

این چارچوب بر اساس فناوری های رایانش ابری و اکوسیستم هدوپ توسعه یافته است. برای ارزیابی عملکرد چارچوب از یک سیستم موازی با ۵ گره پردازشی استفاده شده است. همچنین برای ارزیابی عملکرد چارچوب با یک گره از حالت شبه توزیع یافته<sup>۱</sup> هدوپ استفاده شده است. برای ارزیابی مقیاس پذیری چارچوب پیشنهادی، در ابتدا مدت زمان پردازش MM برای تعداد متغیری از گره های پردازشی مورد بررسی قرار گرفت. همان طور که شکل ۱۱ مشاهده می کنید با به کارگیری ۵ گره پردازشی مدت زمان پردازش تحلیل انطباق نقشه نسبت به تک گره حدود ۷۰٪ کاهش یافته است.

همچنین در شکل ۱۰ مشاهده می شود که با افزایش تعداد گره های پردازشی، نرخ بهبود مدت زمان پردازش  $(\Delta t_p)$  کاهش می یابد. بدین صورت که نرخ کاهش مدت زمان پردازش بین حالت تک گره و دو گره حدوداً ۸۴٪ است که این نرخ میان حالت سه گره و چهار گره حدوداً ۳۱٪ است. ما همچنین سرعت تحلیل انطباق نقشه در چارچوب پیشنهادی را با تعداد گره های پردازشی مختلف مورد ارزیابی قرار داده ایم.

<sup>۱</sup> Pseudo-Distributed Mode

همان‌طور که مشاهده می‌کنید با افزایش حجم داده، مدت‌زمان پردازش نیز افزایش یافته است. همچنین این نمودار نشان می‌دهد افزایش حجم تأثیر مشخصی بر روی نرخ تغییر مدت‌زمان پردازش نداشته است.



شکل ۱۲- آزمون نرخ کارآمدی چارچوب

#### ۴- نتیجه‌گیری و تحقیقات آتی

مدیریت هوشمند حمل‌ونقل نیازمند تحلیل‌های آتی و لحظه‌ای داده‌های حجیم ترافیکی است. یکی از مهم‌ترین داده‌های حجیم ترافیکی، داده‌های خودروهای شناور است. از آنجاکه در کاربردهای سامانه‌های حمل‌ونقل هوشمند معمولاً باید مسیر سفر وسیله نقلیه مشخص باشد، داده موقعیت خودرو شناور باید با شبکه راه‌ها منطبق شود که این امر به کمک تحلیل انطباق نقشه صورت می‌پذیرد. امروزه کاربردهای نوین الگوریتم انطباق نقشه در کنار تلاش برای حفظ نرخ بالای موفقیت، با دو چالش نرخ‌های متغیر نمونه‌برداری و سرعت پردازش داده‌ها روبرو هستند. در این مقاله ما الزامات و چالش‌های نوین پیش روی تحلیل انطباق نقشه را در نظر گرفته و یک چارچوب

#### مراجع

یکپارچه و مقیاس‌پذیر را برای پاسخ به نیازهای مسئله انطباق نقشه ارائه کرده‌ایم. نتایج ارزیابی‌ها بر اساس داده‌های موجود نشان می‌دهد روش پیشنهادی نسبت به روش لئو و همکاران [۲۱] دارای صحت و سرعت بهتری است. عمده مشارکت‌های این مقاله عبارت‌اند از:

۱. در این مقاله یک الگوریتم انطباق نقشه نوین، کلی و مستقل از حسگرهای اضافی ارائه گردیده است. الگوریتم مذکور می‌تواند نسبت به روش لئو و همکاران [۲۱] داده‌های با نرخ‌های نمونه‌برداری متفاوت را با سرعت و صحت بهتری تحلیل کند. نتایج حاصله نشان می‌دهد در صورت استفاده از تمامی اطلاعات حاصل از قرائت GPS در فرآیند وزن‌دهی، یک الگوریتم وزن‌مبنای انطباق نقشه نیز می‌تواند دقتی مشابه روش‌های پیشرفته MM به دست آورد.

۲. در این مقاله چارچوبی یکپارچه و مقیاس‌پذیر برای تحلیل انطباق نقشه ارائه گردیده است. این چارچوب می‌تواند پاسخ به کاربردهای نوین تحلیل‌های بالادستی ITS ارائه گردید. استفاده از فناوری‌های رایانش ابری در چارچوب مذکور امکاناتی مانند دسترسی‌پذیری، تحمل خطا و تعادل بخشی خودکار بار را هدیه می‌کند.

با توجه به نتایج به‌دست‌آمده از این تحقیق در تحقیقات آتی ما چارچوب پیشنهادی را با کاربردهای متعدد سامانه‌های حمل‌ونقل هوشمند مانند مدل‌سازی انتخاب مسیر و پیشنهاد مسیر، تخمین ترافیک و زمان سفر و پیش‌بینی جریان ترافیک یکپارچه‌سازی می‌کنیم. همچنین بر روی استانداردهای ارسال، دریافت و ذخیره‌سازی داده مکانی در HDFS تمرکز خواهیم داشت.

- [1] B. O. T. STATISTICS, "Transportation Statistics Annual Report ", U. S. D. o. Transportation, Ed., ed, 2015.
- [2] M. A. Quddus, "High integrity map matching algorithms for advanced transport telematics applications," Citeseer, 2006.
- [3] X. Chang, B. Y. Chen, Q. Li, X. Cui, L. Tang, and C. Liu, "Estimating real-time traffic carbon dioxide emissions based on intelligent transportation system technologies," IEEE Transactions on Intelligent Transportation Systems, vol. 14, pp. 469-479, 2013.
- [4] J. Yuan, Y. Zheng, C. Zhang, W. Xie, X. Xie, G. Sun, et al., "T-drive: driving directions based on taxi trajectories," in Proceedings of the 18th SIGSPATIAL International conference on advances in geographic information systems, 2010, pp. 99-108.

- [5] R. Ebdndt, A. Sohr, T. Tcheumadjeu, L. Calvin, and P. Wagner, "Utilizing historical and current travel times based on floating car data for management of an express truck fleet," 2010.
- [6] C. De Fabritiis, R. Ragona, and G. Valenti, "Traffic estimation and prediction based on real time floating car data," in Intelligent Transportation Systems, 2008. ITSC 2008. 11th International IEEE Conference on, 2008, pp. 197-203.
- [7] M. A. Quddus, W. Y. Ochieng, and R. B. Noland, "Current map-matching algorithms for transport applications: State-of-the art and future research directions," Transportation research part c: Emerging technologies, vol. 15, pp. 312-328, 2007.
- [8] J. Huang, J. Qie, C. Liu, S. Li, J. Weng, and W. Lv, "Cloud computing-based map-matching for transportation data center," Electronic Commerce Research and Applications, vol. 14, pp. 431-443, 2015.
- [9] M. Hashemi and H. A. Karimi, "A weight-based map-matching algorithm for vehicle navigation in complex urban networks," Journal of Intelligent Transportation Systems ,pp. 00-00, 2016.
- [10] M. A. Quddus, W. Y. Ochieng, L. Zhao, and R. B. Noland, "A general map matching algorithm for transport telematics applications," GPS solutions, vol. 7, pp. 157-167, 2003.
- [11] J. S. Greenfeld, "Matching GPS observations to locations on a digital map," in Transportation Research Board 81st Annual Meeting, 2002.
- [12] T. Miwa, D. Kiuchi, T. Yamamoto, and T. Morikawa, "Development of map matching algorithm for low frequency probe data," Transportation Research Part C: Emerging Technologies, vol. 22, pp. 132-145, 2012.
- [13] B. Y. Chen, H. Yuan, Q. Li, W. H. Lam, S.-L. Shaw, and K. Yan, "Map-matching algorithm for large-scale low-frequency floating car data," International Journal of Geographical Information Science, vol. 28, pp. 22-38, 2014.
- [14] M. Quddus and S. Washington, "Shortest path and vehicle trajectory aided map-matching for low frequency GPS data," Transportation Research Part C: Emerging Technologies, vol. 55, pp. 328-339, 2015.
- [15] M. Nikolić and J. Jović, "Implementation of generic algorithm in map-matching model," Expert Systems with Applications, vol. 72, pp. 283-292, 2017.
- [16] N. R. Velaga, M. A. Quddus, and A. L. Bristow, "Developing an enhanced weight-based topological map-matching algorithm for intelligent transport systems," Transportation Research Part C: Emerging Technologies, vol. 17, pp. 672-683, 2009.
- [17] L. Li, M. Quddus, and L. Zhao, "High accuracy tightly-coupled integrity monitoring algorithm for map-matching," Transportation Research Part C: Emerging Technologies, vol. 36, pp. 13-26, 2013.
- [18] N. Y. C. T. a. L. Commission, "Annual Report to the New York City Council," 2016.
- [19] M. T. Authority, "Facts and Figures," 2015.
- [20] A. Rahman, M. Marsico, and K. Howard, "New York City Department of Transportation Traffic Information Management System (TIMS)," presented at the Transportation Research Board 93rd Annual Meeting, 2014.
- [21] S. Liu, C. Liu, Q. Luo, L. M. Ni, and R. Krishnan, "Calibrating large scale vehicle trajectory data," in 2012 IEEE 13th International Conference on Mobile Data Management, 2012, pp. 222-231.
- [22] M. Hashemi and H. A. Karimi, "A critical review of real-time map-matching algorithms: Current issues and future directions," Computers, Environment and Urban Systems, vol. 48, pp. ۱۶۰-۱۷۴, ۲۰۱۴.
- [23] Y. Li, C. Liu, K. Liu, J. Xu, F. He, and Z. Ding, "On efficient map-matching according to intersections you pass by," in International Conference on Database and Expert Systems Applications, 2013, pp. 42-56.
- [24] M. M. Rahimi and F. Hakimpour, "A Framework For Weighted-based Map-Matching Based on Map-Reduce and Cloud Computing," presented at the The 2nd National Conference on Geospatial Information Technology (NCGIT), Faculty of Geomatics Engineering at K.N. Toosi University of Technology, 2017.
- [25] K. Ohno, T. Tsubouchi, B. Shigematsu, S. Maeyama, and S. i. Yuta, "Outdoor navigation of a mobile robot between buildings based on DGPS and odometry data fusion," in Robotics and Automation, 2003. Proceedings. ICRA'03. IEEE International Conference on, 2003, pp. 1978-1984.
- [26] P. E. Hart, N. J. Nilsson, and B. Raphael, "A formal basis for the heuristic determination of minimum cost paths," IEEE transactions on Systems Science and Cybernetics, vol. 4, pp. 100-107, 1968.
- [27] J. Dean and S. Ghemawat, "MapReduce: simplified data processing on large clusters," Communications of the ACM, vol. 51, pp. 107-113, 2008.
- [28] T. White, Hadoop: The definitive guide: " O'Reilly Media, Inc.", 2012.
- [29] C. C. Robusto, "The cosine-haversine formula," The American Mathematical Monthly, vol. 64, pp. 38-40, 1957.