نشریه علمی پژوهشی مهندسی فناوری اطلاعات مکانی

سال هفتم، شماره دوم، تابستان ۱۳۹۸

Vol.7, No.2, Summer 2019

۱۷۷-۱۹٤

مقاله پژوهشی

رویکردی هندسی برای بهبود کیفیت داده‌های مکانی داوطلبانه با استفاده از تاریخچه داده‌ها (مطالعه موردی: داده‌های خطی *OSM*)

افسانه نصیری دهج[۱]، رحیم علی عباسپور[۲*]، علیرضا چهرقان

۱- کارشناس ارشد *GIS* – دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی – پردیس دانشکده‌های فنی– دانشگاه تهران

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

۳- استادیار دانشکده مهندسی معدن – دانشگاه صنعتی سهند

تاریخ دریافت مقاله: ۱۳۹۶/۱۱/۱۷ تاریخ پذیرش مقاله: ۱۳۹۷/۰۴/۲۵

چکیده

پروژه نقشه خیابانی باز (*OSM*)، یکی از محبوب‌ترین و شناخته‌شده‌ترین پروژه‌های اطلاعات مکانی داوطلبانه است که تمامی داده‌های آن مانند هندسه عوارض، روابط، اطلاعات توصیفی و تمامی نسخه‌های قبلی آن‌ها در فایل تاریخچه ذخیره می‌شوند. با توجه به سهولت در دسترسی به داده‌ها و همچنین چالش های موجود در بحث کیفیت آن‌ها، مسئله کیفیت اطلاعات مکانی تولید شده توسط *OSM* موضوع جذابی برای محققین است. ازجمله رویکردهایی که در تحقیقات پیشین به آن توجه نشده است، استفاده از فایل تاریخچه داده جهت بهبود کیفیت اطلاعات مکانی داوطلبانه می‌باشد. از این‌رو هدف از این پژوهش، ارائه راهکاری برای بهبود دقت موقعیتی عوارض خطی از طریق تولید داده جدید با استفاده از تاریخچه داده‌ها می‌باشد. جهت رسیدن به این هدف، رویکردی هندسی مبتنی بر تناظریابی عوارض و نمودار ورونوی ارائه می‌شود. برای بررسی کارآیی راهکار پیشنهادی، منطقه شش شهر تهران به عنوان منطقه مورد مطالعه انتخاب شد. در نهایت برای برآورد دقت، کیفیت مجموعه داده استخراج شده با مجموعه داده تولیدی شهرداری تهران به‌عنوان مجموعه داده مرجع مورد مقایسه قرار گرفت. باتوجه به نتایج به‌دست آمده از این مقایسه مشخص شد که دقت کامل‌بودن ۹٫۰۳ درصد و دقت موقعیتی عوارض محیط *OSM*، ۸٫۸۸ درصد بهبود یافت.

کلیدواژه‌ها: داده‌های خطی *OSM*، تاریخچه داده‌ها، کیفیت مکانی، تناظریابی عوارض، نمودار ورونوی

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

تلفن: ۸۸۰۰۸۸۴۱

Email : abaspour@ut.ac.ir

[Downloaded from jgit.kntu.ac.ir on 2024-04-17]

[DOI: 10.29252/jgit.7.2.177]

۱- مقدمه

اطلاعات مکانی داوطلبانه روشی بصری و مستقیم برای جمع‌آوری اطلاعات مکانی به‌روز، از طریق مشارکت داوطلبانه افراد با هزینه بسیار پایین ارائه می‌دهند [۱]. در این زمینه، پروژه *OSM*[۱] را می‌توان به‌عنوان یکی از موفق‌ترین و بزرگ‌ترین نمونه‌های پروژه‌های اطلاعات مکانی داوطلبانه نام برد. این پروژه با هدف ایجاد و نگهداری یک نقشه برخط، رایگان و قابل ویرایش توسط کاربران، در سال ۲۰۰۴ از دانشگاه لندن آغاز شده و اکنون به سراسر جهان گسترش یافته است [۲]. در سال ۲۰۱۷، *OSM* اعلام کرد که شمار مشترکین ثبت شده در این پروژه بیش از چهار میلیون نفر بوده است [۳]. نیروهای محرک اصلی ایجاد و رشد *OSM*، محدودیت استفاده و عدم دسترسی به نقشه‌های سراسر جهان و همچنین ظهور سیستم‌های ردیابی ماهواره‌ای ارزان قیمت است [۴]. پروژه *OSM* از طریق مشارکت کاربران، مقدار زیادی از اطلاعات مکانی برچسب‌گذاری شده را تولید می‌کند. با این‌حال، این داده‌ها از محدودیت‌های قابل‌توجهی که به طورطبیعی برای داده‌های مکانی داوطلبانه رخ می‌دهد، رنج می‌برند. یکی از دلایل اصلی آن، عدم اطمینان در مورد نحوه جمع‌آوری این داده‌ها است [۵]. علاوه بر این، فقدان استانداردهای یکپارچه و عدم وجود مکانیزم‌های کنترل کیفیت در محیط *OSM*، کیفیت ناهمگن داده‌های این محیط را افزایش می‌دهد[۶].

تا به امروز حجم قابل توجهی از تحقیقات صورت گرفته در زمینه اطلاعات مکانی محیط *OSM* به ارزیابی کیفیت این داده‌ها پرداخته‌اند و بهبود کیفیت کمتر مورد توجه واقع شده است. در این تحقیق سعی بر آن شده است تا بااستفاده از اطلاعات موجود در فایل تاریخچه، به بهبود کیفیت آخرین نسخه اطلاعات عوارض موجود در سایت *OSM* پرداخته شود. از این رو هدف از این پژوهش، ارائه راهکاری برای بهبود کیفیت دقت موقعیتی عوارض خطی از طریق تولید داده جدید با استفاده از تاریخچه داده‌ها[۲] و کاهش عدم قطعیت در نسخه ارائه شده به کاربر است. تمرکز اصلی این تحقیق بر بهبود مولفه مکانی کیفیت بوده و سعی بر آن دارد تا با بررسی تغییرات در نسخه‌های قبلی یک عارضه در فایل تاریخچه، اثر تغییرات ناگهانی در یک عارضه که ممکن است ناشی از عدم آگاهی مشارکت‌کننده باشد، را در کیفیت نسخه نهایی کاهش دهد.

در این مقاله پس از مقدمه در بخش دوم پیشینه تحقیقات انجام شده بر روی کیفیت داده‌های *OSM* بیان می‌شود. بخش سوم به تشریح پایگاه داده *OSM* و فایل تاریخچه داده *OSM* می‌پردازد. در بخش چهارم شرح مفصلی از روش پیشنهادی ارائه می‌شود. سپس در بخش پنجم ضمن تشریح منطقه مورد مطالعه، چارچوب پیشنهادی بر روی این منطقه اعمال می‌شود و نتایج حاصل از تحقیق، مورد ارزیابی قرار می‌گیرند. در بخش ششم جمع‌بندی و پیشنهادات ارائه می‌شود.

۲- پیشینه تحقیق

از بین مولفه‌های متعدد کیفیت در تحقیقات پیشین، دو مولفه شامل اولی دقت موقعیتی و دوم کامل‌بودن داده‌ها به مراتب بیشتر از جنبه‌های دیگر مورد بررسی قرار گرفته‌اند. از آنجائیکه در این تحقیق با ارائه راهکاری پیشنهادی، به بهبود این دو مولفه پرداخته می‌شود، از این‌رو در ادامه، تحقیقات انجام شده با تمرکز بر دو مولفه دقت موقعیتی و کامل‌بودن داده‌ها بیان می‌شود.

۲-۱-دقت موقعیتی

تاکنون تحقیقات زیادی برای ارزیابی دقت مکانی داده‌های مکانی داوطلبانه بر مبنای روش‌های مقایسه اطلاعات مکانی داوطلبانه با یک مجموعه اطلاعات مکانی مرجع صورت گرفته است. این روش‌ها، برای چنین مقایسه‌ای، به وجود داده‌های مرجع با ویژگی‌های مشابه و یک چارچوب زمانی معتبر نیاز دارند. محققان در این روش‌ها برای ارزیابی دقت موقعیتی، عوارض *OSM* را به

[۱] *OpenStreetMap*

[2] *Data history*

[Downloaded from jgit.kntu.ac.ir on 2024-04-17]

[DOI: 10.29252/jgit.7.2.177]

یک ساختار مشابه مجموعه داده رسمی می‌برند و عوارض متناظر در هر دو مجموعه را به صورت دستی شناسایی می‌کنند. برای جلوگیری از هرگونه خطای پردازشی، تناظریابی دستی بیش از رویکرد تناظریابی خودکار ترجیح داده می‌شود [۴]. رایج‌ترین روشی که در این گروه از تحقیقات برای اندازه‌گیری دقت عوارض خطی استفاده می‌شود، آنالیز بافر است. در این روش بافری با شعاع معینی اطراف یک عارضه خطی از یک مجموعه معتبر زده می‌شود، سپس درصدی از عارضه مکانی داوطلبانه که درون بافر قرار گرفته، محاسبه می‌شود [۷]. تویا[1] و همکارش برای بررسی دقت موقعیتی عوارض خطی از دو فاصله هاسدورف و فاصله میانگین استفاده نمودند [۸]. در یک رویکرد نسبتاً متفاوت، محققان یک رویکرد مبتنی بر فتوگرامتری برای ارزیابی دقت موقعیتی عوارض جاده‌ای *OSM*، با استفاده از تصاویر استریو[2] و یک مدل تنظیم بردار[3] ارائه دادند. آن‌ها رویکرد پیشنهادی را در چند خیابان در شهر وست لافایت از پایگاه داده *OSM* مورد آزمایش قرار دادند و نتایج نشان داد که نقاط شکلی پس از تنظیم[4]، دقت موقعیتی را تا ۸۶ درصد بهبود دادند [۹]. محمدی و ملک در سال ۲۰۱۵ نیز برای ارزیابی کیفیت اطلاعات مکانی داوطلبانه، این داده‌ها را با داده‌های سازمان نقشه‌برداری ایران مقایسه کردند. سپس با استخراج الگو از داده‌های مرجع متناظر، دقت مکانی داده‌های *OSM* بدون متناظر را برآورد کردند [۱۰]. هایانگ[5] و همکاران در سال ۲۰۱۶ یک رویکرد ارزیابی بر اساس شباهت قوس متقارن[6] ارائه دادند و کیفیت هندسی شبکه جاده‌ای در اطلاعات مکانی داوطلبانه را بر اساس اختلاف با شبکه جاده‌ای زمینی مرجع متناظر ارزیابی کردند [۱۱]. بروولی[7] و همکاران در سال ۲۰۱۷ رویکردی برای ارزیابی دو مولفه کامل‌بودن و دقت موقعیتی شبکه جاده‌ای بر اساس تفاوت با داده‌های مرجع ارائه دادند که امکان تنظیم پارامترهای ارزیابی کیفیت به تناسب نیاز کاربر را فراهم می‌آورد. نویسندگان یک برنامه تحت وب معرفی کردند که در این برنامه با تعریف یک شبکه توسط کاربر، امکان تجزیه و تحلیل مولفه کیفیت را در هر قسمت از شبکه به طور جداگانه فراهم می‌آورد [۱۲].

برخی دیگر از محققان سعی کردند بدون استفاده از داده‌های مرجع و با استفاه از روش‌های اعتباردهی به کاربران و صرفا با توجه به منشا و منبع عوارض، کیفیت اطلاعات مکانی داوطلبانه را محاسبه کنند. برای نمونه در سال ۲۰۱۰، هاکلی[8] و همکاران قانون لینوس را اعمال کردند و متوجه شدند که مشارکت تعداد بیشتری از کاربران در یک واحد مکانی داده شده بر روی *OSM*، داده‌هایی با کیفیت بالاتری تولید می‌کند. این مطالعه نشان داد که مقایسه با مجموعه داده‌های مرجع، تنها روش برای ارزیابی کیفیت داده‌های *OSM* نیست و تعداد کاربران به عنوان معیاری برای دقت موقعیتی پیشنهاد شد [۱۳].

در گروه دیگری از تحقیقات، به پژوهش و تحقیق بر روی تاریخچه داده‌ها می‌پردازند و از تحلیل داده‌های تاریخی برای ارزیابی کیفیت ذاتی داده‌های مکانی داوطلبانه استفاده می‌کنند. برای نمونه آنتونیو[9] و همکاران در سال ۲۰۱۴ یک مدل برای ارزیابی اعتبار کاربر و اعتماد عوارض داده‌های مکانی داوطلبانه بر اساس تاریخچه خود اطلاعات معرفی کردند. این رویکرد شامل اختصاص یک ارزش قابل اعتماد به هر نسخه از عارضه است. برای این‌منظور با استفاده از خوشه‌بندی نسخه‌های یک عارضه با توجه به سه ویژگی معنایی، هندسی و کیفی و مقایسه نسخه‌های مشابه در یک خوشه، سطح تناسب برای هدف[10] هر عارضه مشخص شد [۱۴]. در سال ۲۰۱۲ محققان در مقاله‌ای پایگاه داده‌های تاریخچه‌ای *OSM*

[1] *Touya*
[2] *Stereo*
[3] *Vector adjustment model*
[4] *Post adjusted shape points*
[5] *Haiyang*
[6] *Symmetric arc similarity*
[7] *Brovelli*
[8] *Haklay*
[9] *Antonio*
[10] *Fitness for use*

[Downloaded from jgit.kntu.ac.ir on 2024-04-17]
[DOI: 10.29252/jgit.7.2.177]

انگلستان و ایرلند را مورد بررسی قرار دادند[۱۵]. نویسندگان تعداد و اطلاعات مشارکت‌کنندگان، ویرایش‌ها و تعداد تغییرات هندسی هر عارضه، برچسب‌های نسبت داده شده به هر عارضه و همچنین تغییرات هندسی عوارض را بررسی کردند. نتایج آن‌ها نشان داد که ۱۱٪ از همکاران، ۸۷٪ از داده‌های مکانی را بیش از ۱۵ بار ایجاد و یا ویرایش کرده‌اند [۱۵].

۲-۲- کامل‌بودن

کامل‌بودن داده‌های مکانی یکی از مهم‌ترین مولفه‌های کیفیت داده در کنار دقت موقعیتی است. کامل‌بودن داده‌های مکانی، وجود و عدم وجود عوارض مکانی داوطلبانه، اطلاعات توصیفی و روابط آن‌ها را در مقایسه با یک مجموعه داده مرجع توصیف می‌کند [۸]. برای تعیین میزان کامل‌بودن عوارض خطی، از مقایسه بین طول عوارض خطی موجود در *OSM* با طول عوارض خطی در مجموعه داده‌های مرجع استفاده می‌شود. برای نمونه هاکلی در سال ۲۰۱۰، برای ارزیابی مولفه کامل‌بودن داده‌های *OSM* کشور انگلستان، میزان طول کل راه‌ها در هر کیلومتر مربع از مجموعه *OSM* را با مجموعه داده سازمان نقشه‌برداری آن کشور مقایسه کرد [۱۶]. فرقانی و دلاور، براساس معیارهایی مانند طول جاده، مساحت حداقل محدوده محصور[1]، بیضی خطای استاندارد[2] و مرکز میانه[3] و با استفاده از منطق فازی میزان کامل‌بودن داده‌های *OSM* را در شهر تهران ارزیابی کردند [۱۷]. این روش‌ها نیاز به یک مجموعه داده مرجع و معتبر دارند. هزینه و محدودیت‌های دسترسی به داده‌های با دقت بالا، در بسیاری از موارد استفاده از این روش‌ها را با مشکل مواجه می‌سازد. از این‌رو، برای ارزیابی کامل‌بودن، از شاخص‌های درونی مانند برچسب‌های *OSM* و تاریخچه داده‌های آن‌ها استفاده شد [۱۸]. برای نمونه تحقیقی در هند انجام شد که از اطلاعات فایل تاریخچه برای ارزیابی کامل‌بودن داده‌های مکانی با استفاده از شاخص‌های ذاتی استفاده کرد [۱]. محققان فایل تاریخچه داده را در دو تاریخ، ژانویه سال ۲۰۱۶ و فوریه سال ۲۰۱۷ دریافت کردند و با مقایسه دو مجموعه داده به تعیین میزان کامل‌بودن طول جاده‌ها، کامل‌بودن مشخصات نام، حداکثر سرعت و مولفه دقت معنایی پرداختند. مدل توسعه یافته بر روی داده‌های *OSM* هند اعمال شد و نتایج بیانگر ضرورت انگیزه همکاران برای افزایش سطح تناسب برای هدف بود.

۳- پایگاه داده *OSM*

OSM در حال‌حاضر یکی از محبوب‌ترین پروژه‌های اطلاعات مکانی داوطلبانه بر روی اینترنت است [۱۳]. *OSM* از زمان آغاز آن در سال ۲۰۰۴ موفق بوده و در زمینه مسیریابی و ناوبری، بهبود کارتوگرافی، خدمات مبتنی بر مکان و مدل‌های سه بعدی شهری مورد استفاده قرار گرفته است. محیط *OSM* به طور مداوم در حال تغییر و به‌روزرسانی است. این محیط دارای سیستم بازبینی برای ویرایش‌های ارسال شده نیست و تغییرات را فقط چند دقیقه پس از ثبت و ذخیره توسط کاربر، منتشر می‌کند. *OSM* از داده‌های تولید شده از سیستم تعیین موقعیت‌یابی جهانی، تلفن‌های هوشمند و سایر سخت‌افزارهای نقشه‌برداری پشتیبانی می‌کند. هم‌چنین در بعضی از کشورها مانند ایالات متحده و کانادا داوطلبانی هستند که داده‌های معتبر را به پایگاه داده *OSM* وارد می‌کنند.

ساختار داده *OSM* یک ساختار داده توپولوژیکی شامل یک بلوکی از گره‌ها، راه‌ها و روابط برای نشان دادن عوارض مکانی است. علاوه بر این، *OSM* یک طرح برچسب‌گذاری (جفت کلید-ارزش[4]) برای ذخیره‌سازی اطلاعات توصیفی درباره عوارض فراهم می‌آورد. همه داده‌های *OSM* مانند نقاط، راه‌ها، روابط و نسخه‌های قبلی آن‌ها در فایل تاریخچه داده *OSM*، نگهداری می‌شوند. فایل تاریخچه داده *OSM*، شامل همه اطلاعات موقعیتی و توصیفی یک عارضه و همچنین اطلاعات

[1] *Minimum bounding geometry area*
[2] *Standard Deviational Ellipse*
[3] *Median Center*
[4] *Key-value*

[Downloaded from jgit.kntu.ac.ir on 2024-04-17]
[DOI: 10.29252/jgit.7.2.177]

مربوط به کاربران عارضه می‌باشد [۱۹]. محیط *OSM*، بر خلاف سایر محیط‌های جمع‌آوری کننده اطلاعات مکانی داوطلبانه، نه تنها آخرین نسخه اطلاعات عوارض را در دسترس کاربر قرار می‌دهد، بلکه امکان دانلود نسخه‌های قبلی و فایل تاریخچه تغییرات را نیز به کاربر می‌دهد. تمام داده‌های *OSM* که پس از معرفی *رابط کاربردی برنامه‌نویسی نسخه ۰٫۵*[1] به وجود آمده‌اند، در محیط *http://planet.openstreetmap.org* در قالب زبان نشانه‌گذاری گسترش‌پذیر *(XML*[2]*)* و قالب دودویی پروتکل بافر *(pbf*[3]*)* در دسترس عموم می‌باشند. در صورتی که کاربران به داده‌ی محدوده‌ای خاص نیاز داشته باشند، می‌توانند از سایر منابعی[4] که داده‌های فایل کامل تاریخچه را در قسمت‌های کوچک‌تر ارائه می‌دهند، استفاده کنند.

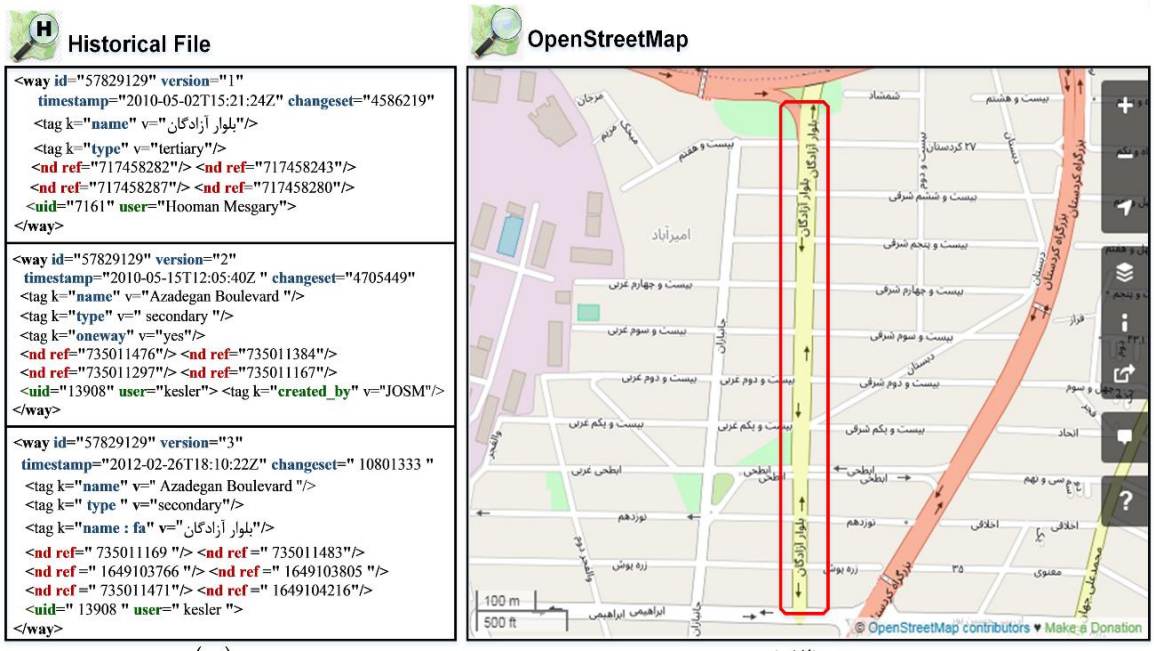
اطلاعات موجود در فایل تاریخچه با توجه به نوع عارضه، شناسه[5](*ID*) و نسخه[6] عوارض مرتب شده است. این فایل امکان دریافت اطلاعات یک عارضه‌ی خاص روی نقشه را به کاربر می‌دهد. برای نمونه شکل (۱) اطلاعات مربوط به سه ویرایش انجام شده برای بلوار آزادگان (*ID=57829129*) در فایل تاریخچه داده *OSM* را نشان می‌دهد. با توجه به شکل(۱)، اطلاعات مربوط به این عارضه را می‌توان به سه دسته اطلاعات هندسی عارضه، اطلاعات توصیفی عارضه و اطلاعات مربوط به کاربر تولیدکننده این عارضه تقسیم کرد. اطلاعات هندسی این عارضه، شامل شناسه مجموعه نقاط تشکیل‌دهنده این عارضه خطی بوده و اطلاعات توصیفی شامل شناسه عارضه، نسخه عارضه و زمان اعمال تغییرات، نوع راه و نام جاده می‌باشد. نام کاربر، شناسه کاربر و روش ثبت داده بر روی محیط *OSM*، به عنوان اطلاعات مربوط به کاربر عارضه خطی در فایل تاریخچه ثبت شده است.

۴- رویکرد پیشنهادی

هدف این تحقیق استفاده از فایل تاریخچه داده‌های محیط *OSM* جهت بهبود دقت موقعیتی و کامل‌بودن داده‌ها می‌باشد. جهت رسیدن به این هدف رویکردی هندسی مبتنی بر تناظریابی عوارض و نمودار ورونوی ارائه می‌شود. روش پیشنهادی مورد استفاده در این تحقیق با توجه به شکل(۲) شامل سه مرحله اصلی می‌باشد که در ادامه تشریح می‌شود. در این پژوهش فرض می‌شود که تمامی نسخه‌های قبلی یک عارضه موجود است. در این راستا، ابتدا باید فایل تاریخچه داده‌های *OSM* که شامل تمامی اطلاعات مربوط به همه عوارض *OSM* است، برای منطقه مورد مطالعه تهیه شود. بعد از تهیه فایل تاریخچه، اطلاعات آخرین نسخه موجود در *OSM* نیز استخراج می‌شود. تمرکز این تحقیق بر بهبود کیفیت داده‌های خطی بوده و داده‌های خطی از فایل تاریخچه جداسازی می‌شوند. سپس در مرحله اول برای هر عارضه در آخرین نسخه اطلاعات *OSM*، عوارض متناظر در تمامی نسخه‌های موجود در فایل تاریخچه شناسایی می‌شوند. برای این منظور از اطلاعات توصیفی، هندسی و توپولوژی مربوط به عوارض استفاده می‌شود. در مرحله دوم، بعد از دسته‌بندی عوارض متناظر، پیش‌پردازشی جهت شناسایی و حذف داده‌های به اشتباه ثبت‌شده برای هر عارضه، در بین تمامی نسخه‌های آن صورت می‌گیرد. در مرحله سوم، هدف استخراج نماینده‌ای برای یک عارضه از بین تمامی نسخه‌های موجود در فایل تاریخچه‌ای می‌باشد، که در این تحقیق از روش استخراج محور مرکزی از طریق اعمال نمودار ورونوی استفاده شده است.

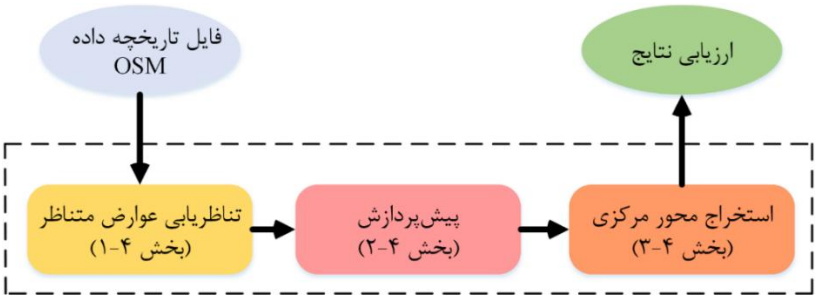
[1] *API 0.5*
[2] *EXtensible Markup Language*
[3] *Protocolbuffer Binary Format*
[4] *(Geofabrik.de) OR (http://osm.personalwerk.de/full-history-extracts)*
[5] *ID*
[6] *Version*

[Downloaded from jgit.kntu.ac.ir on 2024-04-17]
[DOI: 10.29252/jgit.7.2.177]



(ب) (الف)

شکل ۱: الف) قسمتی از شبکه جاده‌ای تهران در محیط *OSM*. ب) اطلاعات عارضه خطی *(ID=57829129)* در فایل تاریخچه.



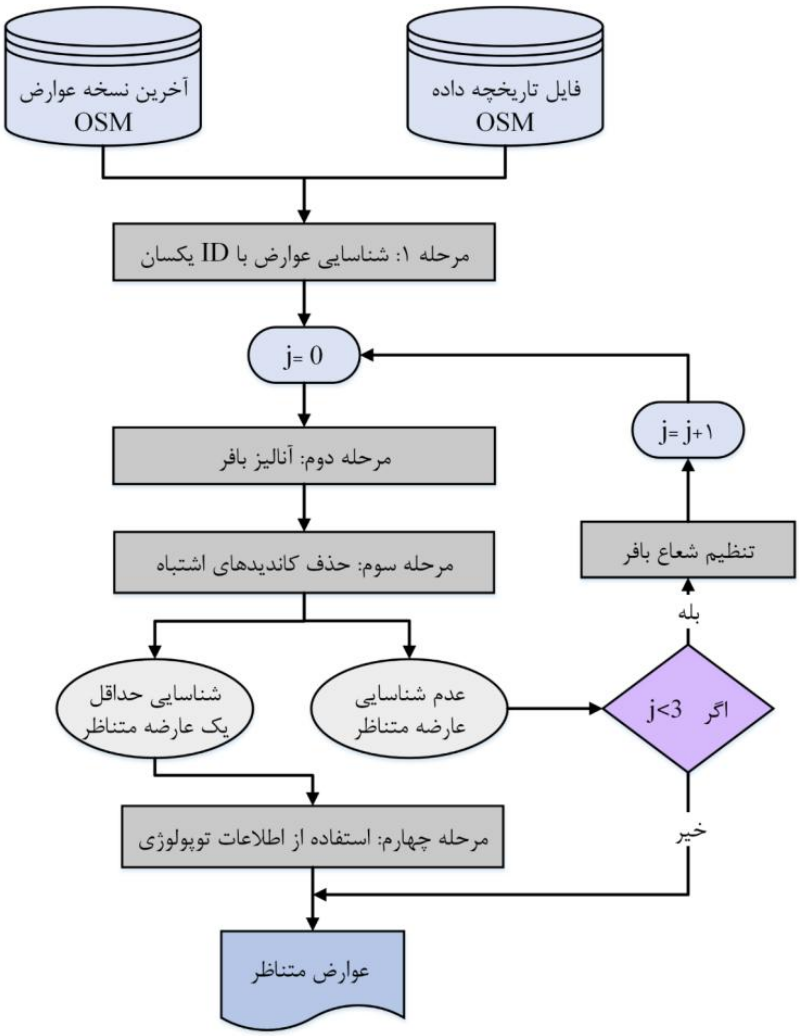
شکل ۲: مراحل کلی روش پیشنهادی

۴-۱- شناسایی عوارض متناظر در فایل تاریخچه

هدف اصلی از تناظریابی در این تحقیق شناسایی عوارض با ماهیت یکسان، در نسخه‌های متفاوت از آن در فایل تاریخچه *OSM* است. این تحقیق با استفاده از میزان مشابهت اطلاعات توصیفی، هندسی و توپولوژی عوارض به یکدیگر، در چهار مرحله به شناسایی عوارض متناظر در مجموعه داده‌های خطی می‌پردازد. شکل(۳) چهار مرحله اصلی مربوط به الگوریتم تناظریابی مورد استفاده در این تحقیق را نشان می‌دهد.

مرحله اول: تمامی عوارض موجود در پایگاه داده *OSM* دارای یک شناسه *OSM* هستند که به عنوان شناسه منحصر به فرد برای شناسایی عارضه، مورد استفاده قرار می‌گیرد. محیط *OSM* برای عوارض جدید از مقادیر منفی برای شناسه عوارض استفاده می‌کند ولی هرگز این مقادیر در پایگاه داده ذخیره نمی‌شود و تنها در هنگام ارسال داده‌ها به پایگاه داده *OSM* برای شناسایی عارضه جدید استفاده می‌شود.

[Downloaded from jgit.kntu.ac.ir on 2024-04-17]
[DOI: 10.29252/jgit.7.2.177]



شکل ۳: طرح کلی فرایند تناظریابی عوارض خطی

سپس سرور این شناسه موقت ارسال شده توسط برنامه ویرایشی را با اختصاص یک شناسه مثبت واقعی برای هر عارضه جدید، جایگزین می‌کند. مقدار شناسه عوارض (>0) در هنگام ویرایش و یا حذف عارضه، تغییر نمی‌کند. با بررسی اطلاعات توصیفی در فایل تاریخچه مشخص شد که برچسب شناسه عارضه تقریبا در تمامی نسخه‌های تغییرات یک عارضه، یکسان بوده است. به همین خاطر در این تحقیق، فرایند تناظریابی با مقایسه شناسه هر عارضه خطی در تمامی نسخه‌های آن آغاز می‌شود.

مرحله دوم: در این مرحله، برای شناسایی عوارض متناظر از آنالیز بافر معرفی شده توسط لانگاس[1] و همکارش استفاده شد [۲۰]. آنالیز بافر یک الگوریتم کارآمد برای تناظریابی عمومی خط است. همانطور که والتر[2] نشان داد، عارضه‌های متناظر با رابطه یک به یک[3] (۱:۱) و یک به چند[4] (۱:N) را می‌توان بوسیله بافر شناسایی کرد [۲۱]. با استفاده از یک آنالیز بافر، ابتدا لیستی از کاندیدهای احتمالی متناظر انتخاب می‌شود و سپس براساس پارامترهای هندسی، در مرحله سوم کاندیدهای اشتباه حذف می‌شوند. بزرگترین مزیت این روش

[1] *Langaas*
[2] *Walter*
[3] *One-to-One*
[4] *One-to-Many*

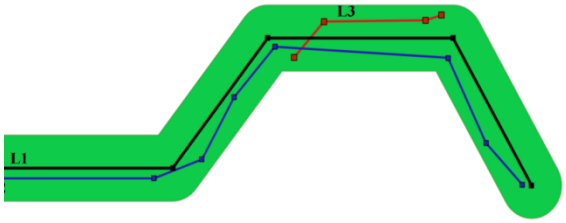
[Downloaded from jgit.kntu.ac.ir on 2024-04-17]
[DOI: 10.29252/jgit.7.2.177]

ازآن‌جهت است که محدوده جستجو و شناسایی عوارض کاندید در فایل تاریخچه کوچکتر شده و سرعت فرایند تناظریابی افزایش می‌یابد.

در این روش، در اطراف هر دو عارضه مرجع و عارضه کاندید یک بافر ایجاد می‌شود. سپس مساحت موجود بین منطقه مشترک به وجود آمده از حریم ایجاد شده برای دو عارضه خطی به عنوان مساحت حریم مشترک استفاده می‌شود [۲۲]. در این تحقیق، از رابطه (۱) برای محاسبه میزان اشتراک بین حریم عوارض استفاده شد. در این رابطه، A_{PL_1} مساحت حریم ایجاد شده برای عارضه خطی اول، A_{PL_2} مساحت حریم ایجاد شده برای عارضه خطی دوم و A_i مساحت منطقه مشترک بین دو حریم ایجاد شده است [۲۳].

$$M_i = \frac{2A_i}{A_{PL_1} + A_{PL_2}}$$ رابطه (۱)

مرحله سوم: از آنجا که بافر در مرحله دوم، در اطراف هر بخش از عارضه مورد نظر در آخرین نسخه *OSM* زده می‌شود، ممکن است مجموعه عوارض خطی که به عنوان کاندید متناظر انتخاب شده‌اند، شامل کاندیدهای اشتباه نیز باشند. برای نمونه در شکل (۴) عارضه L_3 یک کاندید اشتباه می‌باشد که برای دوعارضه دیگر شناسایی شده است.



شکل ۴: آنالیز بافر برای عارضه خطی L_1

در مرحله سوم از فرایند تناظریابی، برای شناسایی کاندیدهای اشتباه از دو شاخص فاصله و جهت[1] به منظور اندازه‌گیری شباهت بین عارضه مورد نظر و عوارض کاندید استفاده می‌شود.

فاصله: یک عارضه‌ی دنیای واقعی در نقشه‌های مختلف باید مکان مشابه‌ای داشته باشد. از این رو، مقایسه فاصله بین دو عارضه در دو مجموعه داده مختلف، ساده‌ترین راه برای شناسایی عوارض متناظر است [۲۴]. یکی از پرکاربردترین روش‌های محاسبه فاصله در علوم مکانی فاصله هاسدورف است [۲۵]. فاصله هاسدورف بیشتر در تناظریابی خودکار مجموعه داده‌های خطی به‌منظور تعیین حداکثر انحراف بین دو عارضه استفاده می‌شود [۲۴, ۲۶]. با این وجود فاصله هاسدورف نسبت به شکل دو عارضه، به خصوص نسبت به قسمت‌های دورافتاده از مرکز حساس است. از این رو تان[2] و همکارانش در سال ۲۰۱۴، فاصله هاسدورف میانه بر مبنای طول[3] را برای برآورد میزان نزدیکی بین عوارض خطی معرفی کردند [۲۷]. محققان نشان دادند که این روش در مقایسه با روش‌های هاسدورف و هاسدورف میانه دارای واریانس کمتر و عملکرد کارامدتری در مواجه با داده‌های پیچیده، جهت اندازه‌گیری فاصله بین عوارض خطی می‌باشد. فاصله هاسدورف میانه برمبنای طول بین دو عارضه خطی با استفاده از رابطه (۲) محاسبه می‌شود که در این رابطه $L(PL_1)$ و $L(PL_2)$ طول دو عارضه خطی بوده و $m(PL_1.PL_2)$ و $m(PL_2.PL_1)$ که بیانگر میانه کوتاه‌ترین فاصله بین هر نقطه از عارضه اول با نقاط عارضه دیگر است، با استفاده از روابط (۳) و (۴) بدست می‌آیند.

$$M_i = \begin{cases} m(PL_1.PL_2).\, if\ L(PL_1) < L(PL_2) \\ m(PL_2.PL_1).\, if\ L(PL_1) \geq L(PL_2) \end{cases}$$ رابطه (۲)

$$m(PL_1.PL_2) = median_{P_a \in PL_1}\{min_{P_b \in PL_2}\|P_a - L_b\|\}$$ رابطه (۳)

$$m(PL_2.PL_1) = median_{P_b \in PL_2}\{min_{P_a \in PL_1}\|P_b - L_a\|\}$$ رابطه (۴)

[1] *Orientation*
[2] *Tong*
[3] *Short-line Median Housdorff Distance (SMHD)*

[Downloaded from jgit.kntu.ac.ir on 2024-04-17]
[DOI: 10.29252/jgit.7.2.177]

جهت: معیار جهت برای مقایسه جهت‌گیری محلی دو عارضه خطی استفاده می‌شود. زاویه تشکیل شده بین خط تشکیل شده از نقاط ابتدایی و انتهایی عارضه و محور افقی، جهت یک عارضه خطی را مشخص می‌نماید [۲۸]. به طور مثال اگر جهت برای دو عارضه خطی PL_1 و PL_2 به ترتیب برابر α و β باشد و تفاوت زاویه α و β نزدیک به صفر باشد، نشان‌دهنده موازی بودن این دو عارضه خطی و داشتن جهت یکسان است. اگر مقدار تفاوت زاویه دو عارضه نزدیک به π باشد یعنی این دو عارضه خطی موازی اما در جهت‌های متفاوت هستند. میزان تفاوت جهت در دو عارضه خطی با استفاده از رابطه (۵) محاسبه می‌شود. در این رابطه $\vec{V}_{PL_1}$ و $\vec{V}_{PL_2}$ به ترتیب بردار تشکیل شده از گره‌های ابتدایی و انتهایی عارضه اول و عارضه دوم می‌باشد.

$$M_i = \cos^{-1}\left(\frac{\vec{V}_{PL_1}.\vec{V}_{PL_2}}{\|\vec{V}_{PL_1}\|.\|\vec{V}_{PL_2}\|}\right) \quad \text{رابطه (۵)}$$

گاهی اوقات ممکن است برای یک عارضه مرجع، اگرچه در واقعیت حداقل یک عارضه متناظر درست وجود دارد، اما در فرایند تناظریابی هیچ عارضه متناظری پیدا نشود. یکی از دلایل این اتفاق، انتخاب شعاع نامناسب (کوچک) برای آنالیز بافر است. ازسوی‌دیگر اگر در ابتدا شعاع بزرگی برای بافر تنظیم شود، ممکن است تعدادی کاندید اشتباه شناسایی شود و در نتیجه آن زمان محاسبه افزایش می‌یابد. برای غلبه بر این مشکل، همان‌طور که در شکل (۳) دیده می‌شود در بخش تنظیم شعاع بافر در الگوریتم تناظریابی، محدوده بافر به صورت تکراری در سه مرحله تنظیم می‌شود.

مرحله چهارم: پس از انجام مراحل فوق، بهترین کاندیدها برای هر عارضه از نظر هندسی و توصیفی شناسایی می‌شوند. با این وجود، در بسیاری از موارد بهترین کاندیدها، لزوما دقیق‌ترین نیستند. از آنجا که بافر در اطراف هر بخش از عارضه مرجع زده می‌شود، همیشه بیشتر از خود عارضه است و بهترین نتیجه تناظریابی هندسی احتمالا چند خطی است که طولانی‌تر از عارضه مرجع است. از این رو برای بهبود دقت تناظریابی، اطلاعات بیشتری مورد نیاز است. برای نمونه در شکل(۵)، چند خطی *1→N-2* به عنوان بهترین کاندید از نظر شکل و موقعیت عارضه برای چندخطی *OO'* شناسایی شده است، در صورتی که اگر روابط توپولوژی برای این تناظریابی در نظر گرفته شود چندخطی *3→N-1*، چند خطی صحیحی است که باید انتخاب شود.

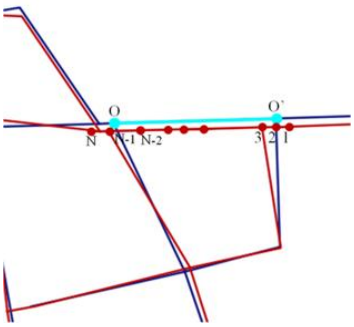
۴-۲- شناسایی و حذف داده‌های پرت

فرآیند شناسایی و حذف خطا از نسخه‌های یک عارضه در فایل تاریخچه، با تمرکز بر روی اطلاعات هندسی عارضه انجام می‌شود. از آنجا که تمامی نسخه‌ها و تغییرات یک عارضه در دسترس است، مبنای عارضه اشتباه بر اساس تفاوت معیار شکل یک نسخه از عارضه با سایر نسخه‌های آن در نظر گرفته شد. در این پژوهش، یک چارچوب ساده‌سازی - تشخیص برای شناسایی عوارض خطی اشتباه پیشنهاد می‌شود که هر عارضه را با مجموعه از پاره‌خط‌ها ساده می‌سازد و سپس با استفاده از تفاوت شکل، عارضه اشتباه را شناسایی می‌کند. برای نمونه در شکل (۶) ابتدا نسخه‌های یک عارضه خطی در مرحله ساده‌سازی با یک مجموعه‌ای از پاره‌خط‌ها تقریب زده می‌شوند. سپس، نسخه اشتباه (L_{V_3}) براساس تفاوت شکل شناسایی شده است.

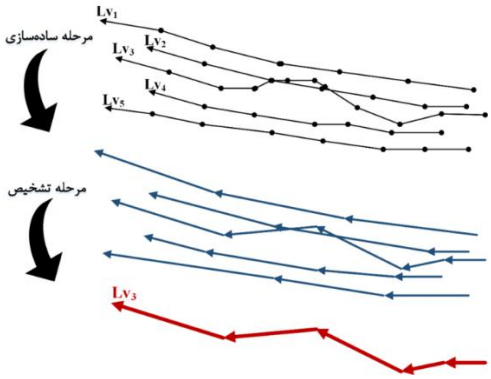
فرایند شناسایی و حذف داده‌های پرت بر اساس چارچوب ساده‌سازی-تشخیص، یک فرایند دو مرحله‌ای است که در گام نخست، برای ساده‌سازی خط از الگوریتم فشرده‌سازی داگلاس- پوکر[۲۹] استفاده می‌کند. هدف ساده‌سازی در این چارچوب، کاهش تعداد نقاط عارضه خطی بوده است و این چارچوب با کاهش تعداد نقاط، سرعت فرایند تشخیص و شناسایی مسیر پرت را افزایش می‌دهد. در مرحله ساده‌سازی عوارض خطی با یک چندخطی ساده جایگزین می‌شوند که این عوارض در شکل با یکدیگر متفاوت هستند. در گام دوم از معیار شکل به عنوان یک معیار هندسی برای تعیین میزان تفاوت عوارض متناظر شناسایی شده برای یک عارضه استفاده شد. یکی از توابع مرتبط با شکل عوارض، تابع تجمعی زاویه است [۳۰]. با توجه به شکل(۷)، این تابع

[Downloaded from jgit.kntu.ac.ir on 2024-04-17]
[DOI: 10.29252/jgit.7.2.177]

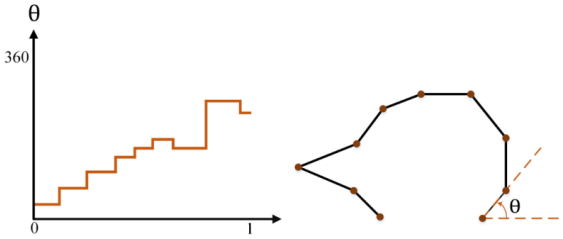
زاویه بین محور افقی و هر بخش از چندخطی را در خلاف جهت عقربه‌های ساعت برآورد می‌کند. تابع تجمعی زاویه یک تابع پیوسته است که در گره‌ها افزایش و یا کاهش می‌یابد و مقدار آن بین دو گره متوالی ثابت است.



شکل ۵: تعیین عارضه متناظر با استفاده از روابط توپولوژیکی (خطوط آبی: عوارض در آخرین نسخه *OSM* و خطوط قرمز: عوارض در فایل تاریخچه)



شکل ۶: یک نمونه از شناسایی نسخه اشتباه در چارچوب ساده‌سازی- تشخیص



شکل ۷: یک عارضه خطی و تابع تجمعی زاویه آن عارضه

برای محاسبه اختلاف شکل بین دو عارضه خطی از رابطه(۶) استفاده می‌شود. در این رابطه θ_{PL_1} تابع تجمعی زاویه عارضه خطی PL_1 می‌باشد [۲۸].

رابطه(۶)

$$M_i = \left\|\theta_{PL_1}(s) - \theta_{PL_2}(s)\right\|_p = \int_0^1 f\left(\theta_{PL_1}{}'\theta_{PL_2}\right)ds'\ f\left(\theta_{PL_1}{}'\theta_{PL_2}\right) = \begin{cases} \left|\theta_{PL_1} - \theta_{PL_2}\right|, & \text{if}\left|\theta_{PL_1} - \theta_{PL_2}\right| \le \pi \\ 2\pi - \left|\theta_{PL_1} - \theta_{PL_2}\right|, & \text{if}\left|\theta_{PL_1} - \theta_{PL_2}\right| > \pi \end{cases}$$

[Downloaded from jgit.kntu.ac.ir on 2024-04-17]
[DOI: 10.29252/jgit.7.2.177]

۴-۳- استخراج محور مرکزی

پس از اینکه تمامی عوارض متناظر با عارضه در آخرین نسخه *OSM* شناسایی شد، برای استخراج نسخه نماینده برای هر عارضه، از نمودار ورونوی جهت استخراج محور مرکزی استفاده می‌شود. محور مرکزی که برای هر جسم، منحصر به فرد و از لحاظ توپولوژیکی هم ارز با جسم اولیه می‌باشد عبارت است از مجموعه‌ای نقاط، که هرنقطه روی آن، از حداقل دو نقطه روی مرز جسم دارای فاصله‌ای برابر باشد [۳۱]. از آنجا که رئوس نمودار ورونوی تشکیل شده از مجموعه نقاط مرزی یک شکل می‌تواند به عنوان اسکلت شکل باشد، نمودار ورونوی و مثلث‌بندی دلونی به طور گسترده برای استخراج محور مرکزی استفاد می‌شوند. این دسته از روش‌ها برای استخراج محور مرکزی قادر به حفظ توپولوژی شکل هستند. روش‌های استخراج محور مرکزی از طریق ساختار مثلث‌بندی دلونی و نمودار ورونوی، از مجموعه نقاط نمونه‌برداری شده از مرز جسم استفاده می‌کنند. در این راستا ساختار مثلث‌بندی دلونی نقاط نمونه از مرز جسم تشکیل می‌شود. یال‌های مثلث‌بندی دلونی در سه دسته تقسیم می‌شوند: دسته اول یال‌هایی که مرز شکل را بازسازی کرده‌اند، دسته دوم یال‌هایی که به طور کامل در داخل شکل قرار دارند و دسته سوم یال‌هایی که به طور کامل خارج از شکل قرار دارند. یال‌های مثلث‌بندی که به طور کامل در خارج از مرز شکل قرار می‌گیرند، حذف می‌شوند. بعد از حذف یال‌های اضافه، ساختار نمودار ورونوی از طریق تشکیل دایره‌های محیطی مثلث‌های داخل ساختار مثلث‌بندی دلونی، تشکیل شده و رئوس نمودار ورونوی استخراج می‌شوند. نقاط رئوس نمودار ورونوی دارای ترتیب خاصی نبوده و صرفا مجموعه‌ای از نقاط هستند. در این تحقیق برای اتصال این نقاط از الگوریتم (۱) (که در شکل (۸) نشان داده شده است) استفاده می‌شود. در این الگوریتم ابتدا فواصل نقاط از هم محاسبه می‌شود و دو نقطه که بیشترین فاصله را دارند به عنوان نقطه ابتدا و انتهای عارضه در نظر گرفته می‌شوند و سپس سایر نقاط بر اساس کوتاه‌ترین فاصله از نقطه ماقبل خود مرتب می‌شوند. این روند تا رسیدن به نقطه انتهایی عارضه ادامه پیدا می‌کند.

۵- پیاده‌سازی و ارزیابی

در این بخش به پیاده سازی روش پیشنهادی بروری منطقه مطالعاتی پرداخته و نتایج حاصل مورد ارزیابی قرار خواهند گرفت.

۵-۱- منطقه مورد مطالعه

در این تحقیق به منظور پیاده‌سازی رویکرد پیشنهادی، منطقه ۶ شهر تهران از میان مناطق ۲۲ گانه شهری، با مساحت تقریبی ۲۱ کیلومترمربع به عنوان منطقه مورد مطالعه انتخاب شد. لازمه انجام این تحقیق داشتن تمامی نسخه‌های قبلی یک عارضه است. از این رو فایل جامع تاریخچه *OSM* از محیطی[1] در قالب پی‌دی‌اف استخراج شد. در این فایل به ترتیب نسخه‌های مختلف نقاط، راه‌ها و روابط وجود دارد. تمرکز این تحقیق بر بهبود کیفیت داده‌های خطی بوده و داده‌های خطی برای منطقه مورد مطالعه با استفاده از نرم‌افزار *osmconvert* از فایل جامع تاریخچه *OSM* جداسازی شده‌اند. این نرم افزار قادر است بر اساس مرز محدوده مورد نظر، داده‌های مکانی را از فایل جامع استخراج کند. همچنین با استفاده از این برنامه قالب فایل تاریخچه از قالب پی‌دی‌اف به قالب *osm* که قالب اصلی داده‌های دریافتی از *OSM* است، تبدیل شد. برای برآورد کیفیت و ارزیابی عملکرد روش پیشنهادی، داده‌های شبکه راه‌ها و معابر منطقه مطالعاتی تولیدشده توسط شهرداری تهران در مقیاس ۱:۲۰۰۰ به عنوان مجموعه‌داده مرجع مورد استفاده این تحقیق قرار گرفتند. شکل (۹- الف) نمایی از فایل تاریخچه، شکل (۹- ب) آخرین نسخه اطلاعات عوارض موجود در سایت *OSM* و شکل (۹- ج) مجموعه‌داده مرجع برای منطقه شش شهر تهران را نشان می‌دهد.

[1] *http://planet.openstreetmap.org*

[Downloaded from jgit.kntu.ac.ir on 2024-04-17]
[DOI: 10.29252/jgit.7.2.177]

Algorithm 1: Best fitting (P)

Input : not sort points P
Output : sorted point P'

```
For each point P_i ∈ P do
  For each point P_j ∈ P do
    If i≠j then
      d_i×j ← length(P_i, P_j );//computing distance matrix d_i×j
(r,c) ← max(d_i×j);// finding the maximum distance of d_i×j
Start ←r;
End ←c;
Index (1) ← start;
For each point P_i ∈ P do
    If i ≠ Start then
      Index (i) ← min(d(Index (i),:))
For each c_i ∈ Index do
    If c_i == End then
      Break;
        P' ← P(c_i)
Return sorted points P'
```

شکل ۸: شبه کد برازش خط به رئوس نمودار ورونوی

شکل ۹: الف) مجموعه داده فایل تاریخچه، ب) آخرین نسخه داده‌ها در محیط *OSM* و ج) مجموعه داده مرجع

۵-۲- پیاده‌سازی

در این پژوهش سعی شد تا با ارائه راهکاری با استفاده از اطلاعات موجود در فایل تاریخچه، مولفه‌های دقت موقعیتی و کامل‌بودن برای آخرین نسخه اطلاعات عوارض موجود در سایت *OSM* بهبود یابد. به‌منظور پیاده‌سازی رویکرد پیشنهادی، فایل تاریخچه منطقه مطالعاتی تهیه شد. این فایل، شامل تمامی تغییرات مربوط به یک عارضه از سال ۲۰۰۷ تا ۲۰۱۷ است. در مرحله نخست از رویکرد پیشنهادی سعی شد با استفاده از روش ارائه شده در بخش (۴-۱)، تمامی نسخه‌های یک عارضه شناسایی شوند. در این راستا از اطلاعات توصیفی، هندسی و توپولوژی مربوط به عوارض استفاده شد. پس از انجام مرحله نخست از رویکرد پیشنهادی، بیشترین تعداد ویرایش انجام شده برای یک عارضه برابر بیست‌ودو

[Downloaded from jgit.kntu.ac.ir on 2024-04-17]
[DOI: 10.29252/jgit.7.2.177]

مرتبه و کمترین تعداد ویرایش انجام شده برای عارضه دیگری سه مرتبه بود.

داده‌ها در محیط *OSM*، به دلیل محدودیت در تکنیک‌های جمع‌آوری داده و یا جنبه‌های مربوط به عدم آشنایی و تخصص کاربر در زمینه اطلاعات مکانی دارای خطا می‌باشند. لذا پالایش داده‌ها به منظور کاهش داده‌های پرت ضروری به نظر می‌رسد. ازاین‌رو برای حذف نسخه اشتباه از عوارض متناظر یک عارضه، در مرحله دوم از رویکرد پیشنهادی از چارچوب ساده‌سازی-تشخیص ارائه شده در بخش (۴-۲) استفاده شد.

پس از شناسایی و حذف داده‌های پرت، مجموعه عوارض متناظر برای هر عارضه به روش نمونه‌برداری به نقطه تبدیل شدند و سپس حریمی با فاصله ثابت ۱۰ متر حول مکان نقاط نمونه‌برداری شده از هر عارضه زده شد. انتخاب این شعاع به صورت تجربی و بر اساس تراکم نقاط نمونه‌برداری بوده است. در مرحله آخر از رویکرد پیشنهادی، جهت استخراج نماینده‌ای برای یک عارضه از بین تمامی نسخه‌های موجود در فایل تاریخچه‌ای، فرآیند استخراج محور مرکزی بر روی حریم حاصل از عوارض متناظر انجام شد. لازم به ذکر است که یک مرحله پیش‌پردازشی جهت ساده‌سازی و هموارسازی مرز شکل پیش از محاسبه‌ی محور مرکزی و همچنین یک مرحله پس‌پردازش جهت حذف شاخه‌های اضافی محور مرکزی صورت گرفت. بدین منظور از دو الگوریتم ساده‌سازی فاصله شعاعی ۲۸ و فاصله عمودی ۲۸ استفاده شد. الگوریتم فاصله شعاعی نقاط موجود در همسایگی با شعاع خاص را حذف می‌کند و الگوریتم فاصله عمودی یک روش تکراری است که از بین سه نقطه در صورتی که فاصله نقطه وسط از خط عبوری از دو نقطه دیگر، از مقدار حدآستانه کمتر باشد حذف می‌شود. هدف روش پیشنهادی، تولید یک نسخه جدید از منطقه مطالعاتی با کیفیت بالاتر نسبت به آخرین استخراج شده از *OSM* است. در این راستا روش پیشنهادی برای بهبود دقت مکانی داده‌ها از ویژگی‌های ذاتی داده‌های مکانی داوطلبانه در فایل تاریخچه استفاده کرد. در شکل(۱۰) چند نمونه از عوارضی که با استفاده از روش پیشنهادی در دو المان کامل‌بودن و دقت موقعیتی از کیفیت داده‌های مکانی بهبود پیدا کرده‌اند، آمده است. عوارض خطی در نسخه ایجاد شده، به داده‌های مرجع نزدیکتر شده‌اند و نسبت به آخرین نسخه بر روی محیط *OSM* از نظر کیفیت بهبود پیدا کرده‌اند.

۵-۳-ارزیابی

در این تحقیق برای ارزیابی روش پیشنهادی، دو مولفه کامل‌بودن و دقت موقعیتی کیفیت برآورد شد. بدین منظور کیفیت مجموعه‌داده آخرین نسخه بر روی محیط *OSM* و مجموعه‌داده استخراج شده توسط راهکار پیشنهادی، بر اساس میزان اختلاف با مجموعه‌داده رسمی مشخص شد و سپس کیفیت نسخه استخراج شده با آخرین نسخه اطلاعاتی مقایسه شدند. پیش از انجام ارزیابی مجموعه‌ای از پیش‌پردازش‌ها در دو مجموعه‌داده صورت گرفت. این پیش‌پردازش‌ها شامل حذف خطاهای سیستماتیک، یک قالب نمودن دو مجموعه‌داده و یکی کردن سیستم مختصات دو مجموعه‌داده می‌باشد.

برآورد دقت مکانی مجموعه‌داده استخراج شده و همچنین آخرین نسخه اطلاعات عوارض موجود در سایت *OSM* در مقایسه با مجموعه‌داده مرجع، از طریق محاسبه اختلافات هندسی صورت گرفت. در این تحقیق چهار معیار فاصله، شکل، جهت و مساحت حریم مشترک به عنوان ملاک ارزیابی عددی کیفیت هندسی عوارض، به کار گرفته شد و به هریک از این پارامترها به ترتیب وزن ۰/۶۲، ۰/۷، ۰/۳۱ و ۰/۶۵اختصاص داده شد [۲۳]. پس از محاسبه میزان تاثیر پارامترهای مختلف در مقدار اختلاف فاصله هندسی، از رابطه (۷) برای برآورد دقت موقعیتی استفاده شد. نتایج حاصل از این ارزیابی در شکل(۱۱) نشان داده شده است. نتایج این بررسی حاکی از بهبود دقت موقعیتی مجموعه‌داده استخراج شده نسبت به آخرین نسخه ارائه شده در محیط *OSM* است.

$$Q_p = \frac{\sum w_i M_i}{\sum w_i} \times 100$$ رابطه(۷)

[Downloaded from jgit.kntu.ac.ir on 2024-04-17]
[DOI: 10.29252/jgit.7.2.177]

در این رابطه، M_i مقدار معیار مورد استفاده و w_i وزن اختصاص داده شده به معیار می‌باشد.

کامل‌بودن مجموعه‌داده‌ها با محاسبه طول هر عارضه در آخرین نسخه اطلاعات عوارض موجود در سایت *OSM* و یا مجموعه‌داده استخراج شده و سپس مقایسه آن با نقشه مرجع در همان منطقه با استفاده از رابطه (۸) تعیین شد [۳۲].

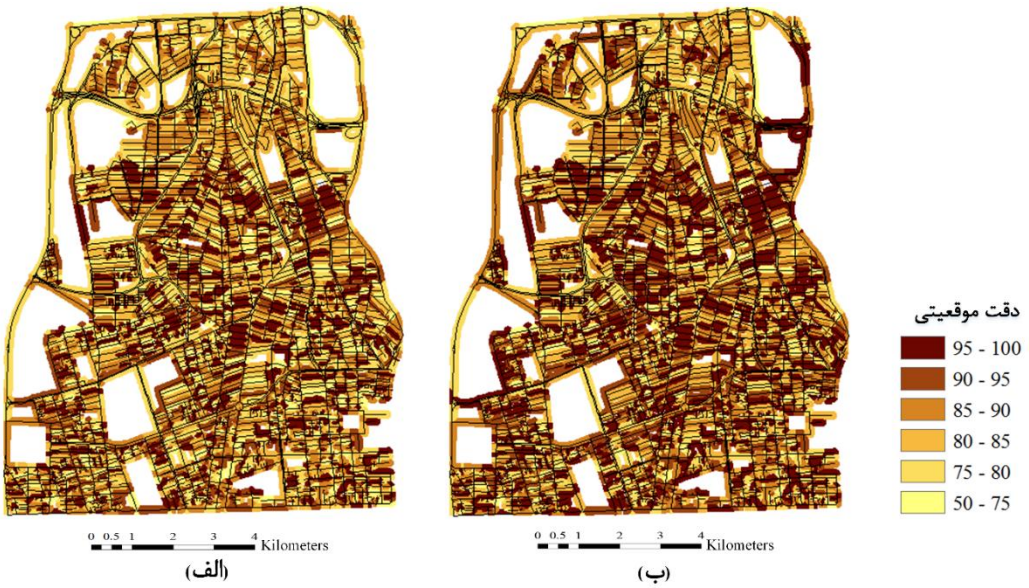
$$Q_C = \frac{L_{OSM}}{L_{Ref}} \times 100 \qquad \text{رابطه(۸)}$$

این معیار به صورت نسبت طول مجموعه‌داده *OSM* (L_{OSM}) به طول مجموعه‌داده مرجع (L_{Ref}) در یک منطقه یکسان محاسبه می‌شود. هر چه مقدار این رابطه به صد نزدیک‌تر باشد، حاکی از پوشش بیشتر داده‌های *OSM* است. نتایج حاصل از این مقایسه‌ها در شکل (۱۲) نشان داده شده است. با توجه به نتایج ارزیابی در دو شکل (۱۱) و (۱۲)، روش پیشنهادی موفق به بهبود دو پارامتر دقت مکانی و کامل‌بودن از کیفیت داده‌های مکانی *OSM* شده است. در این ارزیابی، مقدار میانگین دقت مکانی برای آخرین نسخه عوارض خطی *OSM*، در کل منطقه مطالعاتی برابر ۸۲٫۷۷٪ بوده و برای روش پیشنهادی مقدار این پارامتر به ۹۱٫۶۵٪ رسیده است. همچنین دقت کلی کامل‌بودن برای مجموعه‌داده آخرین نسخه *OSM* برابر ۸۱٫۹۴٪ بوده و در مجموعه‌داده استخراج شده برابر ۹۰٫۴۳٪ می‌باشد.

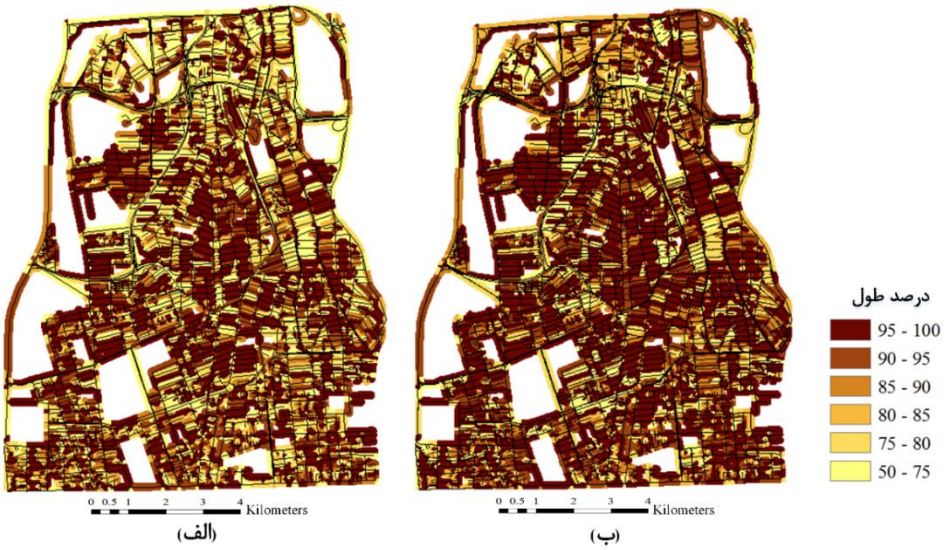
شکل ۱۰: چند نمونه از عوارض بهبود یافته توسط روش پیشنهادی در منطقه مطالعاتی

[Downloaded from jgit.kntu.ac.ir on 2024-04-17]

[DOI: 10.29252/jgit.7.2.177]



شکل ۱۱: الف) تفاوت دقت موقعیتی بین مجموعه داده استخراج شده و مجموعه داده مرجع. ب) تفاوت دقت موقعیتی بین آخرین نسخه *OSM* و مجموعه داده مرجع



شکل ۱۲: الف) تفاوت طول کلی بین راه‌های استخراج شده و مجموعه داده مرجع. ب) تفاوت طول کلی بین آخرین نسخه *OSM* و مجموعه داده مرجع

۶- نتیجه‌گیری

اطلاعات مکانی *OSM*، به عنوان یک نمونه موفق از پروژه‌های اطلاعات مکانی داوطلبانه، دارای مزایای متعدد می‌باشد؛ اما مشکلاتی در مورد استفاده از آن‌ها وجود دارد. یکی از این مشکلات کیفیت داده‌های مکانی است. با توجه به تحقیقات انجام گرفته توسط محققان مشخص شد که اغلب پژوهش‌های صورت گرفته در زمینه کیفیت اطلاعات مکانی داوطلبانه به جنبه‌های ارزیابی کیفیت این داده‌ها پرداخته‌اند. این پژوهش بر آن شد تا با استفاده از تاریخچه اطلاعات مکانی *OSM*، روشی برای

[Downloaded from jgit.kntu.ac.ir on 2024-04-17]
[DOI: 10.29252/jgit.7.2.177]

بهبود کیفیت اطلاعات مکانی این محیط ارائه نماید. پس از تهیه فایل تاریخچه داده‌های *OSM* شهر تهران، تمامی نسخه‌های مربوط به یک عارضه با استفاده از اطلاعات توصیفی، هندسی و توپولوژی شناسایی شدند. بعد از دسته‌بندی نسخه‌های یک عارضه، پیش‌پردازشی جهت حذف داده‌های پرت در هر دسته صورت گرفت و سپس نماینده‌ای برای یک عارضه از بین تمامی نسخه‌های موجود در فایل تاریخچه‌ای، از طریق اعمال نمودار ورونوی تولید شد. در این تحقیق از همه تغییرات داده‌های محیط *OSM* از سال ۲۰۰۷ تاکنون جهت استخراج عارضه جدید با کیفیت بالاتر از آخرین نسخه *OSM* استفاده شد و در نهایت نتایج حاصل از تحقیق مورد ارزیابی قرار گرفت. برای این منظور کیفیت دو مجموعه‌داده استخراج شده و آخرین نسخه *OSM*، با یک مجموعه‌داده با کیفیت بالا مقایسه شدند. نتایج این تحقیق حاکی از بهبود دقت موقعیتی و کامل‌بودن داده‌ها بر روی محیط *OSM* بوده و نشان داد با استفاده از اطلاعات موجود در فایل تاریخچه، می‌توان نسخه با کیفیت‌تری از داده‌ها تولید کرد.

تحقیقات زیادی با استفاده از تاریخچه اطلاعات به بررسی میزان تاثیر تعداد مشارکت‌کنندگان بر روی کیفیت داده‌ها پرداختند. نتایج حاصل از این تحقیقات نشان داد که در حالت کلی مشارکت تعداد بیشتری از کاربران در یک ناحیه منجر به رسیدن به دقت موقعیتی بیشتری خواهد شد. از این رو استفاده از اطلاعات مربوط به کاربران موجب بهبود عملکرد روش پیشنهادی می‌شود. استفاده از تاریخچه رفتار کاربران در کنار تاریخچه مکانی عوارض از جمله پیشنهادات برای مطالعات آتی در این زمینه می‌باشد.

مراجع

[1] *S. S. Sehra, J. Singh, and H. S. Rai, "Assessing OpenStreetMap Data Using Intrinsic Quality Indicators: An Extension to the QGIS Processing Toolbox," Future Internet, vol. 9, p. 15, 2017.*

[2] *P. Hashemi and R. A. Abbaspour, "Assessment of logical consistency in OpenStreetMap based on the spatial similarity concept," in OpenStreetMap in GIScience, ed: Springer, 2015, pp. 19-36.*

[3] *O. W. c. OSM statistics "http://wiki.openstreetmap.org/wiki/Stats," Accessed 24 June 2017.*

[4] *H. Senaratne, A. Mobasheri, A. L. Ali, C. Capineri, and M. Haklay, "A review of volunteered geographic information quality assessment methods," International Journal of Geographical Information Science, vol. 31, pp. 139-167, 2017.*

[5] *H. Chen, W. Zhang, C. Deng, N. Nie, and L. Yi, "Volunteered Geographic Information for Disaster Management with Application to Earthquake Disaster Databank & Sharing Platform," in IOP Conference Series: Earth and Environmental Science, 2017, p. 012015.*

[6] *R. Feick and S. Roche, "Understanding the Value of VGI," in Crowdsourcing geographic knowledge, ed: Springer, 2013, pp. 15-29.*

[7] *M. F. Goodchild and G. J. Hunter, "A simple positional accuracy measure for linear features," International journal of geographical information science, vol. 11, pp. 299-306, 1997.*

[8] *J. F. Girres and G. Touya, "Quality assessment of the French OpenStreetMap dataset," Transactions in GIS, vol. 14, pp. 435-459, 2010.*

[9] *R. Canavosio-Zuzelski, P. Agouris, and P. Doucette, "A photogrammetric approach for assessing positional accuracy of OpenStreetMap© roads," ISPRS International Journal of Geo-Information, vol. 2, pp. 276-301, 2013.*

[10] *N. Mohammadi and M. Malek, "VGI and Reference Data Correspondence Based on Location-Orientation Rotary Descriptor and Segment Matching," Transactions in GIS, vol. 19, pp. 619-639, 2015.*

[Downloaded from jgit.kntu.ac.ir on 2024-04-17]

[DOI: 10.29252/jgit.7.2.177]

[11] H. Lyu, Y. Sheng, N. Guo, B. Huang, and S. Zhang, "Geometric quality assessment of trajectory-generated VGI road networks based on the symmetric arc similarity," Transactions in GIS, 2017.

[12] M. A. Brovelli, M. Minghini, M. Molinari, and P. Mooney, "Towards an automated comparison of OpenStreetMap with authoritative road datasets," Transactions in GIS, vol. 21, pp. 191-206, 2017.

[13] M. Haklay, S. Basiouka, V. Antoniou, and A. Ather, "How many volunteers does it take to map an area well? The validity of Linus' law to volunteered geographic information," The Cartographic Journal, vol. 47, pp. 315-322, 2010.

[14] F. D'Antonio, P. Fogliaroni, and T. Kauppinen, "VGI edit history reveals data trustworthiness and user reputation," 2014.

[15] P. Mooney and P. Corcoran, "Characteristics of heavily edited objects in OpenStreetMap," Future Internet, vol. 4, pp. 285-305, 2012.

[16] M. Haklay, "How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets," Environment and planning B: Planning and design, vol. 37, pp. 682-703, 2010.

[17] M. Forghani and M. R. Delavar, "A quality study of the OpenStreetMap dataset for Tehran," ISPRS International Journal of Geo-Information, vol. 3, pp. 750-763, 2014.

[18] R. Devillers, M. Gervais, Y. Bédard, and R. Jeansoulin, "Spatial data quality: from metadata to quality indicators and contextual end-user manual," in OEEPE/ISPRS Joint Workshop on Spatial Data Quality Management, 2002, pp. 21-22.

[19] O. full history dump, "http://wiki.openstreetmap.org/wiki/Planet.osm/full," accessed 23 June 2017.

[20] H. Tveite, "An accuracy assessment method for geographical line data sets based on buffering," International journal of geographical information science, vol. 13, pp. 27-47, 1999.

[21] V. Walter, Zuordnung von raumbezogenen Daten: am Beispiel der Datenmodelle ATKIS und GDF: Beck, 1997.

[22] H. Fan, B. Yang, A. Zipf, and A. Rousell, "A polygon-based approach for matching OpenStreetMap road networks with regional transit authority data," International Journal of Geographical Information Science, vol. 30, pp. 748-764, 2016.

[23] A. Chehreghan and R. Ali Abbaspour, "A new descriptor for improving geometric-based matching of linear objects on multi-scale datasets," GIScience & Remote Sensing, vol. 54, pp. 836-861, 2017.

[24] S. Yuan and C. Tao, "Development of conflation components," Proceedings of Geoinformatics, Ann Arbor, pp. 1-13, 1999.

[25] I. Abbas, "Base de données vectorielles et erreur cartographique: Problèmes posés par le contrôle ponctuel, une méthode alternative fondée sur la distance de Hausdorff: le contrôle linéaire," 1994.

[26] S. Mustière and T. Devogele, "Matching networks with different levels of detail," GeoInformatica, vol. 12, pp. 435-453, 2008.

[27] X. Tong, D. Liang, and Y. Jin, "A linear road object matching method for conflation based on optimization and logistic regression," International Journal of Geographical Information Science, vol. 28, pp. 824-846, 2014.

[28] M. Zhang, "Methods and implementations of road-network matching," Unpublished PhD Dissertation, Technical University of Munich, 2009.

[29] D. H. Douglas and T. K. Peucker, "Algorithms for the reduction of the number of points required to represent a digitized line or its caricature," Cartographica: The International Journal for Geographic Information and Geovisualization, vol. 10, pp. 112-122, 1973.

[Downloaded from jgit.kntu.ac.ir on 2024-04-17]

[DOI: 10.29252/jgit.7.2.177]

[30] R. C. Veltkamp, "Shape matching: Similarity measures and algorithms," in Shape Modeling and Applications, SMI 2001 International Conference on., 2001, pp. 188-197.

[31] D. Attali, J.-D. Boissonnat, and H. Edelsbrunner, "Stability and computation of medial axes-a state-of-the-art report," in Mathematical foundations of scientific visualization, computer graphics, and massive data exploration, ed: Springer, 2009, pp. 109-125.

[32] M. Wang, Q. Li, Q. Hu, and M. Zhou, "Quality analysis of open street map data," International archives of the photogrammery, remote sensing and spatial information sciences, vol. 2, p. W1, 2013.

[Downloaded from jgit.kntu.ac.ir on 2024-04-17]

[DOI: 10.29252/jgit.7.2.177]

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

Journal of Geospatial Information Technology
Vol.7 No.2, Summer 2019

Research Paper

A Geometric Approach to Improve the Quality of voluntary spatial information Using Data History (Case Study: OSM Linear Data)

Afsaneh Nasiri [1], Rahim Ali Abaspou[2]*, Alireza Chehreghan[3]

1- MSc, School of Surveying and Geospatial Information Engineering, College of Engineering, University of Tehran.
2- Assistant professor, School of Surveying and Geospatial Information Engineering, College of Engineering, University of Tehran.
3- Assistant professor, School of Mining Engineering, College of Engineering, University of Sahand

Abstract

OSM is one of the most desired and recognized VGI projects. All information related to OSM data such as objects' geometry, relations, and descriptive information including all previous versions is stored in the history file. Given the ease in access and the challenges in their quality debate, the issue of the quality of spatial information produced by OSM is an attractive topic for researchers. In previous researches, the use of the data history file to improve the quality of voluntary spatial information has not been considered. Hence, the goal of this research is to present a solution for improving the quality of location precision of the linear objects through the generation of new data using the data records. In this article, to achieve this goal, a geometric approach is presented based on the Voronoi diagram and object matching methods. To evaluate the effectiveness of the proposed method, the District 6 of Tehran was selected as the study area. In order to estimate the quality, the quality of extracted dataset was compared to Dataset produced by the municipality of Tehran as a reference dataset. According to the results obtained from this comparison, it was found that the completeness and positional accuracy of OSM features is improved by about 9.03% and 8.88%, respectively.

Key words: *OSM Linear Data, Data History, Spatial Quality, Object Matching, Voronoi Diagram.*

[Downloaded from jgit.kntu.ac.ir on 2024-04-17]

[DOI: 10.29252/jgit.7.2.177]

Correspondence Address. *GIS Group, school of Surveying and Geospatial Information Engineering, College of Engineering, University of Tehran, Tehran.*
Tel: +۲۱۸۸۰۰۸۸۴۱
Email: *abaspour@ut.ac.ir*