نشریه علمی پژوهشی مهندسی فناوری اطلاعات مکانی

سال ششم، شماره دوم، تابستان ۱۳۹۷

Vol.6, No.2, Summer 2018

۱۰۵-۱۲۴

ارزیابی الگوریتم‌های بهینه‌سازی در تناظریابی داده‌های مکانی چندمقیاسی مبتنی بر ویژگی‌های هندسی

علیرضا چهرقان[۱]، رحیم علی عباسپور[۲*]

۱- دکترای سیستم‌های اطلاعات مکانی، دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی، پردیس دانشکده‌های فنی، دانشگاه تهران

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

تاریخ دریافت مقاله: ۱۳۹۵/۱۲/۰۸ تاریخ پذیرش مقاله: ۱۳۹۶/۰۷/۰۲

چکیده

شناسایی عوارض با ماهیّت یکسان در مجموعه داده‌های مختلف تحت عنوان تناظریابی عوارض شناخته می‌شود. تناظریابی کاربردهای مستقیم و غیر مستقیم بسیاری نظیر تلفیق، ارزیابی کیفیّت، به روز رسانی داده‌ها و انجام آنالیزهای چندمقیاسی دارد. از این رو در این تحقیق راهکاری نوین جهت تناظریابی عوارض ارائه می‌گردد که ضمن در نظر گرفتن تنها معیارهای هندسی (خصوصیات هندسی و توپولوژیکی) استخراج شده از عوارض، هرگونه وابستگی اولیه به پارامترهای تجربی مرسوم نظیر حد آستانه درجه شباهت مکانی، فاصله بافر و وزن معیارها حذف و تناظریابی در مجموعه داده‌های مختلف انجام می‌گیرد. در رویکرد پیشنهادی تمامی روابط یک به هیچ، هیچ به یک، یک به یک، یک به چند، چند به یک و چند به چند در نظر گرفته می‌شود. همچنین در این تحقیق کارایی الگوریتم‌های ژنتیک، توده ذرات و جستجوی غذای زنبور عسل برای تناظریابی عوارض خطی در مجموعه داده‌های مختلف با استفاده از بهینه‌سازی معیارهای هندسی مورد بررسی قرار می‌گیرد. برای ارزیابی کارایی رویکرد پیشنهادی از سه مجموعه داده در مقیاس‌ها و منابع مختلف استفاده می‌گردد. نتایج نشان داد که چارچوب پیشنهادی به خوبی توانایی شناسایی عوارض متناظر در مجموعه داده‌های مختلف را دارا می‌باشد، همچنین نتایج نشان داد که الگوریتم ژنتیک در مقایسه با دو روش دیگر دارای کارایی بالاتری برای بهینه‌سازی پارامترهای موجود در تناظریابی عوارض خطی می‌باشد.

کلیدواژه‌ها: تناظریابی عوارض خطی، معیارهای هندسی، روش‌های بهینه‌سازی، آنالیز حساسیّت، الگوریتم ژنتیک..

*نویسنده مکاتبه کننده: خیابان کارگرشمالی، دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی، پردیس دانشکده‌های فنی، دانشگاه تهران

تلفن: ۶۱۱۱۴۵۲۵-۰۲۱

Email: abaspour@ut.ac.ir

[Downloaded from jgit.kntu.ac.ir on 2024-04-25]

[DOI: 10.29252/jgit.6.2.105]

۱- مقدمه

امروزه مجموعه داده‌های مختلفی در دسترس کاربران قرار دارد که هریک نمایش‌های مختلفی از دنیای واقعی را ارائه می‌دهند. این نمایش‌های مختلف می‌تواند برای تولیدکنندگان داده و یا استفاده کنندگان آن در پردازش‌هایی نظیر تلفیق، ارزیابی کیفیت، به‌روز رسانی داده‌ها و آنالیزهای چندمقیاسی ایجاد مشکل نماید. از این رو نیاز است تا عوارض با ماهیّت یکسان در مجموعه داده‌های مختلف به یکدیگر متصل شوند. این فرایند در تحقیقات گذشته تحت عنوان تناظریابی داده یا تناظریابی عوارض شناخته می‌شود [۱]. چالش‌های پیش‌رو در فرایند تناظریابی شامل مقیاس متفاوت، دقت و صحت مکانی و توصیفی متفاوت، سطح جزئیات متفاوت، نحوه نمایش متفاوت و سیستم مختصات متفاوت در هریک از مجموعه داده‌ها می‌باشد. از این رو تناظریابی عوارض به عنوان یک موضوع چالش برانگیز در علوم مکانی مطرح می‌باشد [۲]. ساختار روش‌های تناظریابی برای انواع مختلف داده‌های برداری (نقطه، خط و چندضلعی) با یکدیگر متفاوت می‌باشد [۳ و ۴] که موضوع مورد بحث این تحقیق تناظریابی عوارض خطی می‌باشد.

از جمله اولین تلاش‌ها در این زمینه به‌پروژه‌ای در اواسط دهه ۱۹۸۰ بر می‌گردد که تناظریابی بین دو مجموعه داده تهیه شده توسط سازمان زمین شناسی و اداره سرشماری آمریکا صورت پذیرفت [۵]. پس از آن تلاش‌های بسیاری برای توسعه روش‌های تناظریابی عوارض خطی انجام شد [۲-۴ و ۶-۱۶]. برای درک بهتر اقدامات صورت گرفته، تحقیقات گذشته را می‌توان در سه دسته ارتباط بین عوارض، معیارهای مورد استفاده و رویکردهای حل مسئله مورد بررسی قرار داد.

در دسته اول در حالت کلی شش نوع رابطه بین عوارض وجود دارد که شامل یک به هیچ، هیچ به یک، یک به یک، یک به چند، چند به یک و چند به چند می‌باشد [۱۶ و ۱۷]. روش‌های معرفی شده توسط لی و گودچایلد (۲۰۱۰)، سالفلد (۱۹۹۳) و سالفلد (۱۹۸۸) تنها می‌توانند رابطه یک به یک بین عوارض را شناسایی کند [۷، ۱۴ و ۱۸]. همچنین در روش تناظریابی معرفی شده توسط لی و گودچایلد (۲۰۱۱) و سستر و همکاران (۲۰۰۷) علاوه بر رابطه یک به یک توانایی شناسایی رابطه یک به چند نیز وجود دارد [۲ و ۱۹]. در تحقیقات تونگ و همکاران (۲۰۱۴)، سافرا و همکاران (۲۰۱۳) و والتر و فریتچ (۱۹۹۹) ضمن شناسایی روابط یک به یک و یک به چند، توانایی شناسایی رابطه چند به چند نیز وجود دارد [۴، ۱۰ و ۱۶]. از جمله تحقیقاتی که تمامی روابط بین عوارض را در نظر گرفتند می توان به وانگ و همکاران (۲۰۱۵)، وانگ و همکاران (۲۰۱۴)، یانگ و همکاران (۲۰۱۳)، ژانگ و منگ (۲۰۰۸) و ژانگ و منگ (۲۰۰۷) اشاره کرد [۱۵، ۲۰، ۲۱، ۲۲ و ۲۳].

در دسته دوم، در دو دهه گذشته تحقیقات مختلف معیارهای متفاوتی را جهت شناسایی عوارض استفاده نمودند. بصورت کلی معیارهای مورد استفاده را می‌توان در دو نوع هندسی (شامل خصوصیات هندسی و توپولوژیکی) و معنایی در نظر گرفت و با توجه به مجموعه داده‌های در دسترس از هریک و یا ترکیبی از این معیارها جهت تناظریابی استفاده نمود [۱۷]. پرکاربردترین معیارهای هندسی را می‌توان طول، فاصله، جهت و مساحت نام برد که مهمترین و پرکاربردترین آنها فاصله بین دو عارضه می‌باشد. از جمله پرکاربردترین روش‌های محاسبه فاصله بین دو عارضه می‌توان فاصله اقلیدسی بین نقاط [۸ و ۲۲] ، فاصله هاسدورف و مشتقات آن [۲، ۴، ۲۴، ۲۵ و ۲۶]، فاصله کمترین مربعات [۲۷] و فاصله فرشت[1] [۱۲، ۲۸ و ۲۹] نام برد. تحقیقات دیگری نظیر سافرا و همکاران (۲۰۱۳)، یانگ و همکاران (۲۰۱۳)، سونگ و

[1] Frechet

[Downloaded from jgit.kntu.ac.ir on 2024-04-25]

[DOI: 10.29252/jgit.6.2.105]

همکـــاران (۲۰۱۱) و تونـــگ و همکـــاران (۲۰۰۹) و موســوی و آل شــیخ (۲۰۰۸) ضــمن در نظــر گــرفتن خصوصیات هندســی از خصوصــیات توپولــوژیکی نظیــر درجه گره‌ها و روابط مکـانی عـوارض بـا یکـدیگر بـرای شناسایی عوارض متناظر استفاده نمودند [۱۰، ۲۱، ۳۰، ۳۱ و ۳۲]. تحقیقات دو و همکاران (۲۰۱۶)، دیمونـد و همکاران (۲۰۱۵)، رودریگز و اگنهافر (۲۰۰۴)، کوهن و همکاران (۲۰۰۳) و والتـر و فـریتچ (۱۹۹۹) عـلاوه بـر در نظر گرفتن معیارهای هندسی از معیارهـای معنـایی مانند نوع و نام عارضه برای شناسـایی عـوارض متنـاظر استفاده نمودند [۱۶، ۳۳، ۳۴، ۳۵ و ۳۶].

در دسـته سـوم در حالـت کلـی دو نـوع رویکـرد بـرای تناظریابی عوارض می توان در نظـر گرفـت. در رویکـرد اول بسیاری از تحقیقات از طریق مقایسه درجه شباهت به‌دست آمده از طریق معیارهـای هندسـی و معنـایی و همچنـین در نظـر گـرفتن فاصـله بـافر و حـد آسـتانه شباهت، عوارض متناظر را تعیین می‌کنند [۳، ۱۳، ۱۵، ۱۶، ۲۰، ۳۱، ۳۷ و ۳۸]. در این نوع رویکرد، سه نکتـه قابل ملاحظه وجود دارد:

۱. در بسیاری از ایـن تحقیقـات بـا تغییـر مقیـاس و افزایش یا کاهش جزئیات از کـارایی راهکـار ارائـه شده کاسته می‌شود. از جمله دلایل ایـن امـر وزن هریک از معیارهای مورد بررسی مـی‌باشـد کـه از طریق کارشناسان تعیـین مـی‌گـردد، زیـرا تـاثیر این معیارها در اخـتلاف مقیـاس‌هـای مختلـف بـا یکدیگر تفاوت دارد [۱۰، ۳۹، ۴۰ و ۴۱].

۲. در تحقیقات مبتنی بـر مقـدار حـد آسـتانه بـرای مقدار درجـه شـباهت عـوارض و معیارهـای مـورد استفاده، با انتخاب نامناسب این مقادیر نتایج مورد تاثیر قرار می‌گیرد [۱۵، ۲۰، ۳۷، ۳۸، ۳۹ و ۴۲].

۳. در تحقیقات مبتنی بر فاصله بافر، مقدار این فاصله به‌صورت تجربـی [۱۰، ۱۵، ۱۶، ۴۳ و ۴۴] و یـا از طریـق دقـت مکـانی دو مجموعـه داده [۱۰] به‌دست می‌آیـد، در حـالی کـه ایـن مقـدار تـاثیر قابل توجهی بر روی دقت نهایی تناظریـابی دارد و باید به‌صورت مقدار بهینه محاسبه گردد.

در رویکـرد دوم از دسته سوم که اخیـراً مـورد اسـتفاده قرار گرفته اسـت، در نظـر گـرفتن مسـئله تنـاظریابی به‌عنـوان یـک مسئـلـه بهینـه‌سـازی مـی‌باشـد. لـی و گودچایلد (۲۰۱۰) با ارائه یک روش بهینـه‌سـازی و بـا در نظر گرفتن معیار فاصـله، رابطـه یـک بـه یـک بـین عوارض را شناسایی نمود [۱۸]. در ادامه لی و گودچایلد (۲۰۱۱) با توسعه روش فوق قادر بـه شناسـایی روابـط یک به چند نیز گردید [۲]. همچنین تونگ و همکـاران (۲۰۱۴) با استفاده از راهکار ارائـه شـده توسـط لـی و گودچایلد (۲۰۱۱) و در نظر گرفتن رگرسیون لجستیک راهکاری مبتنی بر تکرار برای شناسایی روابـط یـک بـه یک، یک به چنـد و چنـد بـه چنـد ارائـه نمودنـد [۴]. در این نوع رویکرد که مسئله تناظریابی به‌عنوان مسئله بهینه‌سازی فاصله در نظـر گرفتـه مـی‌شـود، بـه دلیـل در نظـر گـرفتن تنهـا معیـار فاصـله از کـارآیی آن کاسته می‌شود. زیرا در بسیاری از مجموعه داده‌ها صرفاً معیار طول نمی‌تواند جهت پیدا کردن عـوارض متنـاظر کافی باشد.

همانطور که بیان گردید در تحقیقات انجام شـده تا بـه امـروز چنـدین کاسـتی وجـود دارد کـه از اهـداف مقـاله حاضـر می‌باشد. مقاله حاضر صـرفا از معیارهـای هندسی استفاده می‌کند تا در صورت فقـدان مجموعـه داده‌ها از اطلاعات معنایی، کـارایی رویکـرد پیشـنهادی کاسـته نشـود. در مقالـه حاضـر بـا اسـتفاده از روش بهینـه‌سـازی کـارایی معیارهـای هندسـی در مجموعـه داده‌های مختلف مورد بررسی قرار می‌گیرد و معیارهای بهینه با توجه به مقیاس مجموعـه داده مـورد اسـتفاده تعیـین مـی‌گـردد. عـلاوه بـر بهینـه‌سـازی معیارهـا، مقادیر بهینه پارامترهایی نظیر فاصله بافر (برای تعیـین عوارض کاندید)، حد آسـتانه درجـه شبـاهت مکـانی و وزن معیـارهـا نیز با توجه به مجموعه داده‌هـای ورودی تعیـــین مـــی‌گـــردد. در ایـــن تحقیـــق کـــارایی

[Downloaded from jgit.kntu.ac.ir on 2024-04-25]

[DOI: 10.29252/jgit.6.2.105]

روش‌های بهینه‌سازی مرسوم نظیر الگوریتم ژنتیک، توده ذرات و جستجوی غذای زنبور عسل مورد ارزیابی قرار می‌گیرد. برای رسیدن به نتایج واقعی پیاده‌سازی از طریق در نظر گرفتن سه مجموعه داده مختلف صورت می‌گیرد.

ادامه این مقاله بدین شرح است: پس از مقدمه در بخش ۲ چارچوب پیشنهادی ارائه و جزئیات آن تشریح می‌گردد. در بخش ۳ ضمن تشریح منطقه مورد مطالعه، چارچوب پیشنهادی بر روی این منطقه اعمال و نتایج مورد بررسی قرار می‌گیرد. در انتها نیز در بخش ۴ نتیجه گیری و پیشنهادات ارائه می‌گردد.

۲- رویکرد پیشنهادی

با توجّه به کاستی‌های مطرح در تحقیقات پیشین، نیاز است تا مقادیر بهینه فاصله حریم، حد آستانه شباهت و وزن معیارها برای مجموعه داده‌های ورودی تعیین گردد. از این رو در این تحقیق راهکاری مبتنی بر بهینه‌سازی پارامترها ارائه می‌گردد. در ادامه ضمن معرفی پارامترهای مورد بررسی، جزئیات راهکار ارائه شده بیان می‌گردد.

با توجّه به کاستی‌های مطرح شده در بخش اول، نیاز است تا پارامترهایی نظیر وجود و یا عدم وجود معیارهای هندسی (M)، وزن معیارها ($W = \{w_1, w_2, \dots, w_n\}$)، مقدار حدّ آستانه درجه شباهت مکانی (τ) و مقدار فاصله حریم (β) با توجّه به مجموعه داده‌های ورودی بهینه گردند.

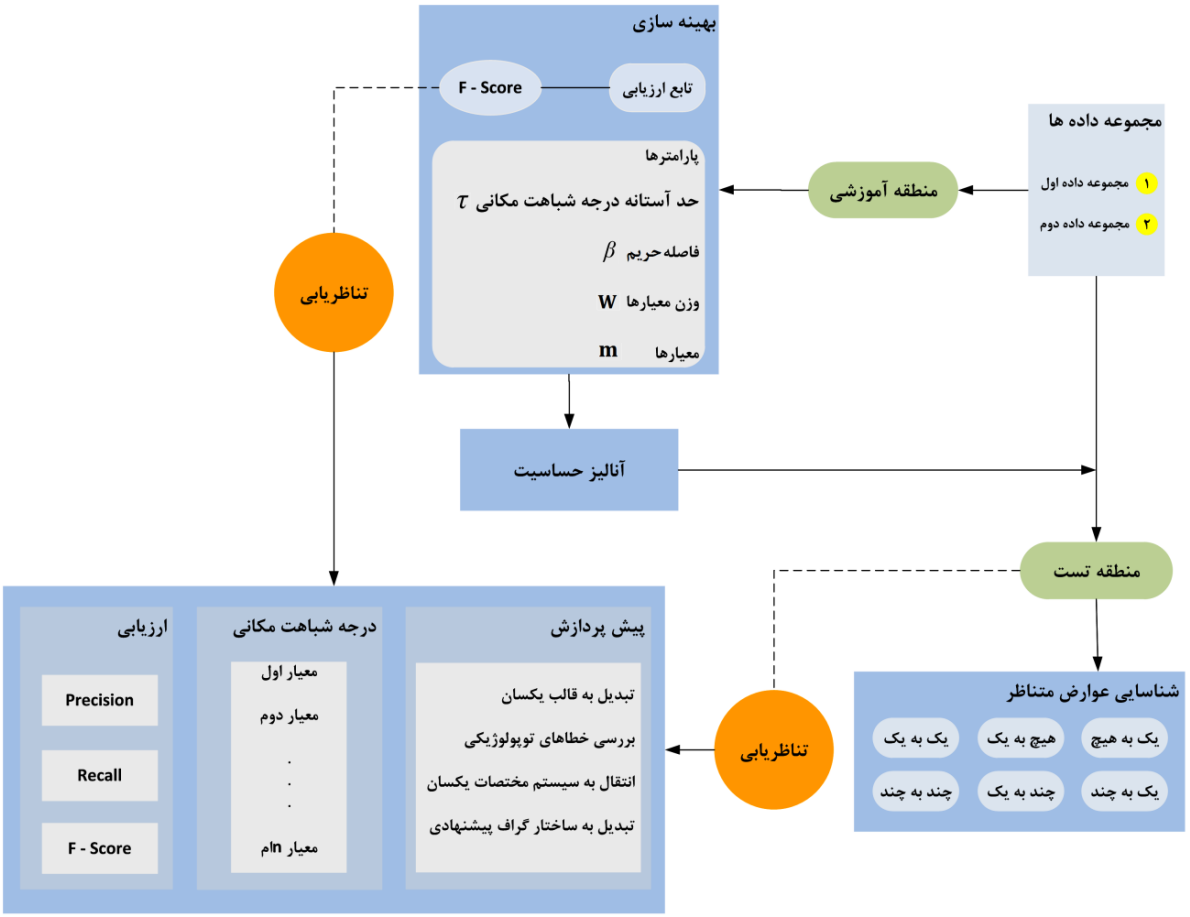
- پارامتر M به این اشاره دارد که بر اساس مجموعه داده‌های ورودی، کدامیک از معیارهای مورد استفاده در فرایند تناظریابی حضور داشته باشند.
- پارامتر W به میزان اهمیّت هریک از معیارهایی اشاره دارد که در تناظریابی حضور دارند.
- پارامتر τ به مقدار حدّ آستانه شباهت جهت اینکه دو عارضه متناظر یکدیگر باشند، بر می‌گردد. مقدار بالای حدّ آستانه شباهت باعث شناسایی تعدادی کمی از جفت متناظرها شده و مقدار پایین آن باعث شناسایی تعداد زیادی از عوارض نامتناظر به عنوان جفت متناظر می‌گردد. از این رو تعیین مقادیر بهینه این پارامتر نیز بر اساس مجموعه داده‌های ورودی ضروری به‌نظر می‌رسد.
- پارامتر β به فاصله حریم مورد استفاده در فرایند تناظریابی اشاره دارد. مقدار فاصله حریم کم باعث عدم شناسایی عوارض متناظر شده و مقدار فاصله حریم زیاد علاوه بر کاهش کارایی روش تناظریابی، باعث افزایش زمان محاسبات می‌گردد.

با توجه به شکل (۱) در این تحقیق راهکاری مبتنی بر روش بهینه‌سازی ارائه می‌شود که مجموعه دادهای ورودی را به دو قسمت مناطق آموزشی و تست تقسیم‌بندی می‌کند. در مناطق آموزشی با توجّه به مجموعه داده‌های ورودی مقادیر بهینه پارامترهای M، $W = \{w_1, w_2, \dots, w_n\}$، τ و β محاسبه می‌گردد. سپس در منطقه تست، تناظریابی از با استفاده از مقدار بهینه پارامترها صورت گرفته و نتایج مورد ارزیابی قرار می‌گیرد. در فرایند بهینه‌سازی τ مقداری بین ۰ تا ۱۰۰ (درصد شباهت مکانی) و β با عنوان فاصله حریم شامل هر مقدار مثبتی می‌باشد. W به‌عنوان وزن معیارها بصورت ماتریسی است که به اندازه تعداد معیارها (n) درایه دارد. همچنین M به‌صورت ماتریسی با تعداد n درایه است که وجود و یا عدم وجود معیارها در فرایند بهینه‌سازی را شامل می‌شود. این پارامتر دارای عناصر صفر و یک می‌باشد، به‌طوری‌که هر درایه‌ای مقدار صفر داشته باشد، معیار متعلق به آن درایه از فرایند تناظریابی حذف می‌شود.

همانطور که در شکل (۱) نمایان است، رویکرد تناظریابی پیشنهادی شامل سه مرحله پیش‌پردازش، درجه شباهت مکانی و ارزیابی می‌باشد. همچنین این رویکرد در دو مرحله مورد استفاده قرار می‌گیرد. مرحله اول برای محاسبه مقدار تابع برازش (مقدار F-

[Downloaded from jgit.kntu.ac.ir on 2024-04-25]
[DOI: 10.29252/jgit.6.2.105]

score) و مرحله دوم برای تناظریابی در منطقه تست با استفاده از مقادیر بهینه پارامترها می‌باشد.



شکل ۱: چارچوب کلی برای تناظریابی در مجموعه داده‌های مختلف

۲-۱- بهینه‌سازی پارامترهای τ، β، W و M

در روش پیشنهادی برای تعیین مقدار بهینه پارامترهای τ، β، W و M برای منطقه مورد مطالعه، از روش‌های بهینه‌سازی استفاده می‌شود. از آنجائی‌که برای حل مسأله مورد بحث در این مقاله هیچ الگوریتم شناخته شده قابل اجرا در زمان چندجمله‌ای وجود ندارد، مسأله حاضر به عنوان یک مسأله NP[1] کامل مطرح می‌باشد. به بیان دیگر علاوه بر اینکه مسأله مذکور در زمان معینی (زمان در نظر گرفته شده بوسیله یک چندخطی) قابل حل نمی‌باشد، حل آن به‌صورت بهینه کامل نیز ضروری به نظر نمی‌رسد. در مسائل NP کامل امکان این وجود دارد که پس از اجرای برنامه در زمان بسیار طولانی مقدار جواب حاصل به اندازه ناچیزی افزایش یابد که مورد نظر این تحقیق نمی‌باشد. از آنجائی‌که برای حل این مسئله نیاز به الگوریتم‌هایی است که ضمن حل مسئله توانایی خارج شده از مقادیر بهینه محلی و رسیدن به مقدار بهینه سراسری را داشته باشد، الگوریتم‌های فرا ابتکاری مناسب حل مسئله می‌باشند. الگوریتم‌های فرا ابتکاری شامل دو دسته روش جستجوی مبتنی بر جمعیّت و جستجوی تک نقطه‌ای می‌باشد. امّا به دلیل تعداد بالای پارامترها نیاز به روشی است که بر روی جمعیّتی از جواب‌ها فرایند بهینه‌سازی انجام گیرد، زیرا روش‌های

[1] Non-Deterministic Polynomial

[Downloaded from jgit.kntu.ac.ir on 2024-04-25]
[DOI: 10.29252/jgit.6.2.105]

جستجوی تک نقطه‌ای همزمان روی راه‌حل‌های واحدی کار می‌کنند که منجر به طولانی‌شدن بیش از حدّ فرایند بهینه‌سازی و نرسیدن به جواب نزدیک به جواب بهینه سراسری (در محدوده شرط تعیین شده برای توقف الگوریتم) می‌گردد.

الگوریتم ژنتیک یکی از پرکاربردترین روش‌های بهینه‌سازی فرا ابتکاری جمعیّت-مبنا جهت حل مسائل NP کامل است که در تحقیقات مختلف کارایی خود را برای حل مسائل بهینه‌سازی مرتبط با علوم اطّلاعات مکانی نشان داده است [۴۵-۴۷]، از این رو جهت تعیین مقادیر بهینه پارامترهای معرفی شده برای مجموعه داده‌های ورودی از الگوریتم ژنتیک استفاده می‌گردد. به‌صورت کلی الگوریتم ژنتیک دارای مراحلی نظیر ارزیابی[1]، انتخاب[2]، تقاطع[3] و جهش[4] می‌باشد. از این رو در ادامه هریک از این مراحل در راهکار مورد استفاده تشریح می‌گردد.

شایان ذکر است جهت ارزیابی الگوریتم ژنتیک، نتایج حاصل از الگوریتم‌های دیگر نظیر توده ذرات (PSO) و زنبور عسل مصنوعی (ABC) با نتایج به‌دست آمده از الگوریتم ژنتیک مورد مقایسه قرار می‌گیرد.

۲-۱-۱- ساختار کروموزم پیشنهادی

با توجّه به ساختار الگوریتم ژنتیک، ابتدا پارامترهای موردنظر به صورت ساختار کروموزم در نظر گرفته می‌شوند. امّا با توجّه به اینکه تعداد پارامترهایی که بایستی بهینه شوند بالاست، امکان در نظر گرفتن ساختار کروموزم بصورت دودویی[5] وجود ندارد. از این رو از الگوریتم ژنتیک با کدگذاری واقعی (RCGA[6]) استفاده می‌شود. در الگوریتم RCGA مقادیر پارامترها بصورت مقدار واقعی در کروموزم در نظر گرفته می‌شوند [۴۸-۵۰]. شکل (۲) ساختار کروموزم در نظر گرفته شده در الگوریتم RCGA را نشان می‌دهد. در ساختار کروموزم پیشنهادی ابتدا پارامترهای τ (ژن اوّل) و β (ژن دوّم)، سپس وزن معیارها (ژن سوّم تا دوازدهم) (W) و در نهایت تعداد معیارهای هندسی (ژن سیزدهم تا بیست و دوم) (M) وجود دارد. هر کروموزم شامل تعداد $2n + 2$ ژن می‌باشد، n تعداد معیارهایی هستند که در فرایند تناظریابی در نظر گرفته می‌شوند (در این تحقیق ۱۰ معیار). شایان ذکر است درصورتیکه در مراحل بهینه‌سازی، خانه i ام از ماتریس M در ساختار کروموزم برابر مقدار صفر گردد، خانه معادل آن در ماتریس W نیز معادل صفر شده و این معیار از فرایند تناظریابی حذف می‌گردد.

۲-۱-۲- مرحله ارزیابی

در این مرحله، تابعی تحت عنوان تابع برازش[7] در نظرگرفته می‌شود و برای هریک از کروموزم‌ها این مقدار محاسبه می‌گردد. در حقیقت هدف کلی بهینه سازی، کمینه (بیشینه) کردن تابع هدف می‌باشد. هدف در مسأله تناظریابی عوارض خطی، شناسایی تعداد بالاتری عوارض متناظر در مجموعه داده‌های مورد بررسی می‌باشد. برای این منظور از F-score استفاده می‌شود. پارامتر F-score همزمان با در نظر گرفتن پارامترهای Precision و Recall، عدم تعادل در هریک از این پارامترها را در مورد توجه قرار می‌دهد. ممکن است مقادیر Precision و Recall مخالف یکدیگر باشند، ممکن است Precision مقدار بالا و Recall مقدار پائینی داشته باشد و یا Precision مقدار پائین و Recall مقدار بالائی

[1] Evaluation
[2] Selection
[3] Crossover
[4] Mutation
[5] Binary

[6] Real Coded Genetic Algorithm
[7] Fitness Function

[Downloaded from jgit.kntu.ac.ir on 2024-04-25]
[DOI: 10.29252/jgit.6.2.105]

به‌دست آید. از این رو می‌توان از F-score به‌عنوان یک پارامتر نهایی جهت ارزیابی نتایج استفاده نمود. در این تحقیق تابع هدف در فرایند بهینه‌سازی، بیشینه کردن مقدار F-score در فرایند بهینه‌سازی می‌باشد. رابطه (۱) تابع هدف را نشان می‌دهد.

رابطه (۱)

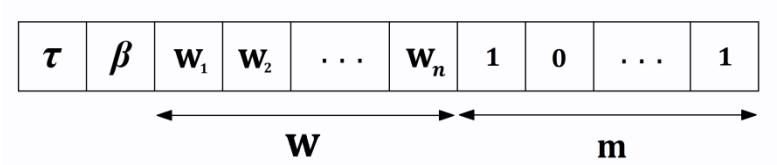
$$\text{Max}\ (\ \text{F}-\text{score})\quad ,\quad \text{F}-\text{score}=2\times\frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}$$

همچنین رابطه (۲) محاسبه Precision و Recall را نشان می‌دهد.

رابطه (۲)

$$\text{Precision}=\frac{\text{TP}}{\text{TP}+\text{FP}}\times 100\ ,\ \text{Recall}=\frac{\text{TP}}{\text{TP}+\text{FN}}\times 100\%$$

در این روابط TP تعداد روابطی هستند که به درستی کشف شده‌اند، FP تعداد روابطی هستند که به اشتباه شناسایی شده‌اند و FN تعداد روابطی هستند که کشف نشده‌اند.



شکل ۲: ساختار کروموزم پیشنهادی

۲-۱-۳- مرحله انتخاب

پس از اینکه برای تمامی کروموزم‌های جمعیّت مقدار تابع برازش محاسبه گردید، نیاز است تا بهترین (بالاترین مقادیر $\text{F}-\text{score}$) مقادیر برای نسل آینده انتخاب گردند. کروموزم‌های انتخاب شده تحت عنوان کروموزم‌های والد شناخته می‌شوند. در نظر گرفتن درصد بالای کروموزم‌ها از بین کروموزم‌های جمعیّت به عنوان کروموزم والد، باعث کاهش سرعت رسیدن به جواب بهینه می‌گردد. همچنین در نظر گرفتن درصد پائین کروموزم‌ها از بین کروموزم‌های جمعیّت به‌عنوان کروموزم والد، باعث یکسان شدن تعداد زیادی از کروموزم‌های فرزند در مرحله بعد و کاهش سرعت رسیدن به جواب بهینه می‌گردد. از این رو به طور معمول ۴۰ تا ۵۰ درصد از جمعیّت هر نسل به‌عنوان کروموزم والد در نظر گرفته می‌شود.

۲-۱-۴- مرحله تقاطع

پس از اینکه کروموزم‌های والد در مرحله انتخاب مشخص گردیدند، نیاز است تا به تعداد کروموزم‌های حذف شده کروموزم‌های فرزند ایجاد گردد. به فرایند ترکیب کروموزم‌های والد با یکدیگر و تشکیل کروموزم‌های فرزند تقاطع گفته می‌شود. برای این منظور دو کروموزم به‌صورت تصادفی از بین تعداد کروموزم‌های والد انتخاب می‌گردند و از طریق ترکیب آنها فرزندان ایجاد می‌شوند.

با توجّه به شکل (۳) ساختار کروموزم پیشنهادی شامل مقادیر دودویی در M و مقادیر واقعی در τ، β و W می‌باشد. از این رو روش ترکیب دو کروموزم والد برای هریک از این دو مقادیر با یکدیگر متفاوت می‌باشد. برای قسمت دودویی کروموزم روش‌های مختلفی نظیر تقاطع یک نقطه‌ای[1]، دو نقطه‌ای[2]، یکنواخت[3] و میانه[4] ارائه شده است که انتخاب هریک از این روش‌ها تأثیری بر جواب نهایی حاصل از بهینه‌سازی نخواهد داشت. از این رو در این تحقیق از روش یکنواخت به‌عنوان یکی از ساده‌ترین و پرکاربردترین آن‌ها جهت تقاطع دو کروموزم والد استفاده شده است. در روش یکنواخت ژن‌های کروموزم فرزندان از طریق انتخاب تصادفی با

[1] Single Point Crossover
[2] Two Point Crossover
[3] Uniform Crossover
[4] Intermediate Crossover

[Downloaded from jgit.kntu.ac.ir on 2024-04-25]
[DOI: 10.29252/jgit.6.2.105]

احتمال برابر از بین ژن‌های دو کروموزم والد انتخاب می‌گردد [۵۱]. از آنجائیکه در ساختار کروموزم پارامترهای τ، β و W دارای مقادیر واقعی هستند، از روش تقاطع متفاوتی استفاده می‌شود. در این تحقیق، فرزندان از طریق روش تقاطع BLX-alpha ایجاد می‌گردند. روابط (۳ و ۴) مقادیر واقعی به‌دست آمده برای قسمت τ، β و W ژن کروموزم فرزندان با استفاده از این روش را نشان می‌دهد [۵۲].

رابطه (۳) $$Ch^i = Random[Min\{P_1^i, P_2^i\} - D\,\alpha^i, Max\{P_1^i, P_2^i\} + D\,\alpha^i]$$

رابطه (۴) $$D = Max\{P_1^i, P_2^i\} - Min\{P_1^i, P_2^i\}$$

در این رابطه Ch^i ژن iام ایجاد شده برای فرزندان، α^i مقدار تصادفی انتخاب شده در بازه $[0,1]$ برای ژن iام، P_1^i و P_2^i به ترتیب مقدار واقعی ژن iام در والدین اوّل و دوّم، $Max\{.\}$ و $Min\{.\}$ عملگرهایی هستند که به ترتیب بیشترین و کمترین مقدار را محاسبه می‌کنند. $Random[a, b]$ عملگری است که مقداری به‌صورت تصادفی بین بازه a تا b انتخاب می‌کند.

۲-۱-۵- مرحله جهش

فرایند جهش به هنگام حرکت از جمعیّت حاضر به جمعیّت جدید باعث می‌شود که میزان تنوّع در جمعیّت جدید بالا برود و این تنوّع اساس تکامل و پیشرفت در رسیدن به جواب نهایی می‌باشد. تعداد جهش طبق رابطه (۵) به‌دست می‌آید.

رابطه (۵) $$N = \mu\,(N_P - 1)N_G$$

که در این رابطه N تعداد جهش، μ نرخ جهش، N_P تعداد کروموزم و N_G تعداد ژن هر کروموزم می‌باشد.

۲-۱-۶- شرط همگرایی

در این تحقیق از «تغییر نکردن بیشترین مقدار F-score طی ۲۰۰ تکرار» به عنوان شرط همگرایی و توقف بهینه‌سازی در نظر گرفته می‌شود. همچنین حدّاکثر تعداد تکرار نیز ۱۰۰۰ در نظر گرفته شد تا درصورتی‌که مسئله همگرا نشد، بهینه‌سازی متوقف گردد.

۲-۱-۷- آنالیز حساسیّت

پس از اینکه مقادیر پارامترها محاسبه و معیارهای بهینه تعیین شدند، وزن هریک از معیارها از طریق آنالیز حساسیّت سنجیده می‌شوند تا در صورتیکه امکان حذف معیاری در این مرحله وجود داشته باشد، آن معیار از فرایند تناظریابی حذف گردد. برای مثال فرض کنید در فرایند بهینه‌سازی در منطقه آموزش $w_3 = 0.004$ محاسبه شده است، در این‌صورت در مرحله آنالیز حساسیّت بررسی می‌شود که آیا امکان قرار دادن $w_3 = 0$ وجود دارد و یا اینکه حذف این معیار باعث کاهش کارایی تناظریابی می‌گردد. در این مرحله وزن هریک از معیارها با فرض ثابت بودن سایر معیارها، از مقدار صفر تا یک به‌صورت تصاعدی (محور X) افزایش پیدا می‌کند و میزان تغییرات F-score (محور Y) مورد بررسی قرار می‌گیرد.

۲-۲- رویکرد تناظریابی

با توجه به ساختار پیشنهادی در شکل (۱)، تناظریابی در دو مرحله مورد استفاده قرار می‌گیرد. در مرحله اول برای محاسبه مقدار تابع ارزیابی هر کروموزم، تناظریابی از طریق معیارهای انتخابی(M) و مقدار پارامترهای τ، β و W انجام گرفته و مقدار F-score به عنوان مقدار تابع ارزیابی تعیین می‌گردد. در مرحله دوم زمانیکه معیارهای بهینه انتخاب و مقدار بهینه پارامترهای τ ، β و W برای مناطق آموزشی محاسبه گردید، از این معیارها و پارامترها برای تناظریابی در داده‌های تست استفاده شده و تناظریابی با استفاده از این مقادیر انجام گرفته و نتایج مورد بررسی قرار می‌گیرد. با تواجه به شکل (۱) مراحل تناظریابی عوارض خطی با استفاده از معیارهای هندسی شامل سه مرحله پیش پردازش، درجه شباهت مکانی و ارزیابی می‌باشد.

[Downloaded from jgit.kntu.ac.ir on 2024-04-25]

[DOI: 10.29252/jgit.6.2.105]

۲-۲-۱- پیش پردازش

برای تناظریابی مجموعه داده شبکه راه‌ها با مقیاس‌ها و منابع متفاوت، در ابتدا نیاز است که یک مرحله پیش پردازش بر روی مجموعه داده‌ها صورت گیرد. این پیش پردازش شامل تبدیل به قالب یکسان، انتقال به سیستم مختصات یکسان، حذف خطاهای توپولوژیکی و تبدیل به ساختار گراف در نظر گرفته شده می‌باشد.

در مرحله پیش پردازش در ابتدا دو مجموعه داده تبدیل به قالب یکسان می‌گردند، سپس در صورتی‌که دو مجموعه داده سیستم مختصات متفاوتی از یکدیگر داشته باشند، در هر دو مجموعه داده سیستم مختصات یکسانی تعیین می‌گردد، همچنین برای هریک از دو مجموعه داده خطاهای توپولوژیکی حذف می‌شوند. در نهایت برای اینکه در تعریف عوارض در هریک از دو مجموعه داده ابهام ایجاد نگردد از تئوری گراف‌ها برای توصیف شبکه راه‌ها به‌عنوان یکسری نقاط و خطوط متصل به هم استفاده می‌گردد.

در ریاضیات گراف مرتبط با خطوط شبکه می‌تواند بصورت زوج مرتب $G = (V, E)$ نمایش داده شود که V شامل مجموعه رئوس شبکه و E شامل مجموعه یال‌های شبکه می‌باشد. هر یال توسط یک جفت از رئوس قابل شناسایی می‌باشد که درجه هریک از این رئوس، تعداد یال‌های متصل به آن می‌باشد [۵۳]. در شبکه راه‌های شهری هر چندخطی شامل چندین رأس و یال می باشد. چندخطی مفروض PL_i شامل نقاط $P_{i,1}$، $P_{i,2}$، . . . ، $P_{i,n}$ می‌باشد که هر دو نقطه $P_{i,j}$ و $P_{i,j+1}$ یک یال از چندخطی را تشکیل می‌دهند که نقاط $P_{i,1}$ و $P_{i,n}$ نقاط ابتدایی و انتهایی چندخطی PL_i می‌باشند. اما در این تحقیق عوارض بصورتی تعریف می‌گردد که در ساختار گراف این چندخطی‌ها درجه ابتدایی و یا انتهایی رئوس آن مخالف با دو باشد. در نتیجه در فرایند تناظریابی هریک از این چندخطی‌ها به‌صورت یک عارضه در نظر گرفته می‌شود. بنابراین هر چندخطی ممکن است با یک تقاطع شروع و یا پایان یابد ولی هرگز شامل یک تقاطع در نقاط میانی نخواهد بود [۱۰].

۲-۲-۲- درجه شباهت مکانی

روش‌های مختلفی تا به امروز برای شناسایی عوارض متناظر در مجموعه داده‌های مختلف ارائه شده است، اما یکی از معمولترین و پر کاربردترین این روش‌ها تعیین عوارض متناظر از طریق محاسبه درجه شباهت مکانی بین عوارض می باشد [۱۳, ۱۵, ۳۵, ۵۴-۵۶]. برای این منظور فرض کنید PL_1 و PL_2 دو چند خطی در دو مجموعه داده متفاوت می‌باشند که به یک ماهیت در دنیای واقعی اشاره می‌کنند، درجه شباهت مکانی (به درصد) دو عارضه خطی PL_1 و PL_2 از طریق رابطه (۶) محاسبه می‌گردد [۵۶].

$$S_{PL_1,PL_2} = \frac{\sum_{i=1}^{n} W_i S_{PL_1,PL_2}^{C_i}}{\sum_{i=1}^{n} W_i} \times 100 \qquad \text{رابطه (۶)}$$

در این رابطه $Sim_{PL_1,PL_2}^{C_i}$ مقدار نرمال شده معیار iام، W_i وزن معیار iام و S_{PL_1,PL_2} درجه شباهت مکانی بین چندخطی PL_1 و PL_2 می‌باشد که دارای مقادیری بین ۰ تا ۱۰۰ است. برای محاسبه درجه شباهت مکانی بین دو عارضه معیارهای هندسی موقعیت، طول، فاصله، اندازه، انحنا، پیچیدگی، جهت، مساحت، شکل، ناحیه مشترک بین حریم عوارض و درجه گره‌ها در نظر گرفته می‌شود.

- **موقعیت:** یکی از معیارهای هندسی که در تناظریابی عوارض مورد استفاده قرار می‌گیرد، اختلاف موقعیت گره‌های ابتدایی و انتهایی عوارض از یکدیگر می باشد که با C_1 نشان داده می‌شود.
- **طول:** معیار هندسی دیگری که در بسیاری از تحقیقات مورد استفاده قرار گرفته است در نظر گرفتن طول عوارض می باشد که با C_2 نشان داده می‌شود.

[Downloaded from jgit.kntu.ac.ir on 2024-04-25]

[DOI: 10.29252/jgit.6.2.105]

– **فاصله:** روش‌های مختلفی برای محاسبه فاصله وجود دارد، اما تونگ و همکاران (۲۰۱۴) با معرفی هاسدورف میانه بر مبنای طول نشان داد که این روش در مقایسه با سایر روشهای مبتنی بر فاصله هاسدورف، دارای واریانس کمرتر و عملکرد کارامدتری جهت اندازه‌گیری فاصله بین عوارض می‌باشد. رابطه (۷) فاصله هاسدورف میانه بر مبنای طول بین دو عارضه خطی PL_1 و PL_2 را نشان می‌دهد [۴].

رابطه (۷)

$$C_3(PL_1, PL_2) = \begin{cases} m(PL_1, PL_2), & \text{if} \quad L_{PL_1} < L_{PL_2} \\ m(PL_2, PL_1), & \text{if} \quad L_{PL_1} \geq L_{PL_2} \end{cases}$$

$$m(PL_1, PL_2) = \text{median}_{P_a \in PL_1}\{\min_{P_b \in PL_2} \|P_a - L_b\|\}$$

$$m(PL_2, PL_1) = \text{median}_{P_b \in PL_2}\{\min_{P_a \in PL_1} \|P_b - L_a\|\}$$

در این روابط L_a و L_b دو یال اختیاری از عوارض خطی PL_1 و PL_2، $\|P_a - L_b\|$ فاصله عمودی یکی از گره‌های عارضه PL_1 (P_a) از یکی از یال‌های عارضه PL_2 (L_b) و $\|P_b - L_a\|$ فاصله عمودی یکی از نودهای عارضه PL_2 (P_b) از یکی از یال‌های عارضه PL_1 (L_a) می‌باشد.

– **اندازه:** در یک عارضه خطی به فاصله اقلیدسی بین گره ابتدایی و انتهایی گفته می‌شود که که با C_4 نشان داده می‌شود [۱۷].

– **انحنا:** از تقسیم اندازه یک عارضه خطی بر طول آن معیاری به نام انحنا ایجاد می‌گردد که می‌توان از آن برای مقایسه دو عارضه خطی استفاده نمود [۵۷]. معیار انحنا با C_5 نشان داده می‌شود.

– **پیچیدگی:** پیچیدگی خطوط از جمله معیارهای دیگری است که می توان از آن جهت محاسبه درجه شباهت مکانی عوارض استفاده نمود. پیچیدگی خطوط را می توان از طریق در نظر گرفتن میانگین وزن دار فاصله رئوس عارضه از خط واصل ایجاد شده بین گره ابتدایی و انتهایی، محاسبه نمود[۵۷]. رابطه (۸) اختلاف پیچیدگی دو عارضه PL_1 و PL_2 را نشان می‌دهد.

رابطه (۸)

$$C_6(PL_1, PL_2) = \left|Com_{PL_1} - Com_{PL_2}\right| , \quad Com_{PL} = \sum_{i=1}^{t-1}\left(\left(\frac{h_i + h_{i+1}}{2}\right).\left(\frac{d_i}{D}\right)\right)$$

در این رابطه h_i فاصله عمودی رأس iام از خط واصل بین گره ابتدایی و انتهایی، d_i طول یال iام، D طول خط واصل بین گره ابتدایی و انتهایی و t تعداد گره عارضه PL می‌باشد.

– **جهت:** معیار هندسی دیگری که می‌توان برای مسئله تناظریابی عوارض خطی مور استفاده قرار داد، اختلاف جهت عوارض خطی از یکدیگر می‌باشد که می‌تواند نقش مهمی را در مسئله ایفا کند [۵۸]. اختلاف جهت برای دو عارضه خطی PL_1 و PL_2 با جهت های α و β از رابطه (۹) به‌دست می‌آید.

رابطه (۹)

$$C\_7\ (PL\_1, PL\_2) = |\alpha - \beta| = \cos^{-1}\left(\frac{\vec{V}_{PL_1}.\vec{V}_{PL_2}}{\|\vec{V}_{PL_1}\|.\|\vec{V}_{PL_2}\|}\right)$$

در این رابطه $\vec{V}_{PL_1}$ بردار تشکیل شده ازنودهای ابتدایی و انتهایی عارضه اول، $\vec{V}_{PL_2}$ بردار تشکیل شده از نودهای ابتدایی و انتهایی عارضه دوم و عملگر $\| \cdot \|$ فاصله اقلیدسی بین نودهای ابتدایی و انتهایی عارضه مورد نظر می‌باشد.

– **مساحت:** از طریق اتصال نود ابتدایی به انتهایی در عوارض خطی یک چندضلعی ایجاد می‌شود که می‌توان از طریق در نظر گرفتن مساحت

[Downloaded from jgit.kntu.ac.ir on 2024-04-25]

[DOI: 10.29252/jgit.6.2.105]

ایجاد شده توسط این چندضلعی و فاصله گره ابتدایی از انتهایی، عوارض مختلف را با یکدیگر مورد مقایسه قرار داد. رابطه (۱۰) نحوه محاسبه معیار مساحت بین این دو عارضه خطی را نشان می‌دهد [۱۷].

$$C_8(PL_1, PL_2) = \left|\frac{A_1}{D_1} - \frac{A_2}{D_2}\right| \quad \text{رابطه (۱۰)}$$

در این رابطه A_1 و A_2 مساحت‌های ایجاد شده برای دو عارضه PL_1 و PL_2 ، D_1 و D_2 فاصله اقلیدسی بین گره‌های ابتدایی و انتهایی در دو عارضه PL_1 و PL_2 می‌باشد.

- **شکل:** عوارض خطی در مقیاسهای مختلف می‌توانند از نظر شکل با یکدیگر متفاوت باشند. در نظر گرفتن این تفاوت می‌تواند به‌عنوان یک معیار هندسی در تعیین مقدار درجه شباهت مکانی عوارض به یکدیگر مورد استفاده قرار گیرد. یکی از شناخته شده ترین و پرکاربردترین توصیفگرهای مرتبط با شکل عوارض، تابع پیچش[1] می‌باشد [۵۹]. در این تابع برای هر کدام از گره‌ها میزان چرخش یال متصل به گره نسبت به محور افقی در نظر گرفته می‌شود. در انتها نیز مقدار مساحت محصور بین دو تابع به‌عنوان اختلاف شکل دو عارضه خطی محاسبه می‌گردد. رابطه (۱۱) نحوه محاسبه اختلاف شکل دو عارضه PL_1 و PL_2 را نشان می‌دهد [۱۷] .

رابطه (۱۱)

$$C_9(PL_1, PL_2) = \int_0^1 f(\theta_{PL_1}, \theta_{PL_2})\, ds \ , \quad f(\theta_{PL_1}, \theta_{PL_2}) = \begin{cases} |\theta_{PL_1} - \theta_{PL_2}| \ , & \text{if } |\theta_{PL_1} - \theta_{PL_2}| \le \pi \\ 2\pi - |\theta_{PL_1} - \theta_{PL_2}| \ , & \text{if } |\theta_{PL_1} - \theta_{PL_2}| > \pi \end{cases}$$

در این رابطه θ (s) تابع پیچش عارضه مورد نظر می‌باشد. در این تابع محور افقی طول و محور قائم زاویه چرخش یال متصل به گره برحسب رادیان می‌باشد.

- **ناحیه مشترک بین حریم عوارض:** از جمله معیارهای هندسی دیگر در نظر گرفتن مساحت موجود بین منطقه مشترک بوجود آمده از حریم ایجاد شده بین عوارض می‌باشد [۶۰]. رابطه (۱۲) نحوه محاسبه اختلاف مساحت بین حریم عوارض برای دو عارضه PL_1 و PL_2 را نشان می‌دهد که هرچه قدر مقدار این مساحت به یک نزدیکتر باشد دو عارضه از نظر هندسی به یکدیگر شبیه‌تر هستند.

$$C_{10}(PL_1, PL_2) = \frac{2A_i}{A_{PL_1} + A_{PL_2}} \quad \text{رابطه (۱۲)}$$

در این رابطه A_{PL_1} مساحت حریم ایجاد شده برای عارضه PL_1، A_{PL_2} مساحت حریم ایجاد شده برای عارضه PL_2 و A_i مساحت منطقه مشترک بین دو حریم ایجاد شده می‌باشد.

- **درجه گره‌ها:** در هریک از عوارض می توان از اختلاف درجه گره‌های ابتدایی و انتهایی برای محاسبه درجه شباهت مکانی بین عوارض استفاده نمود [۱۰].

۲-۲-۳- ارزیابی

در هنگام تناظریابی بین مجموعه داده‌های مکانی ممکن است نمایش عارضه‌ای در یکی از مجموعه‌ها نسبت به دیگری متفاوت باشد .از این رو، روش‌های تناظریابی بایستی امکان شناسایی روابط مختلف بین عوارض را دارا باشند [۱۷]. روابط در نظر گرفته شده بین عوارض در تحقیقات مختلف، در حالت کلی شامل شش رابطه می‌باشد [۲، ۱۰، ۱۷، ۶۱ و ۶۲].

— رابطه یک به هیچ (۱:۰): در این حالت برای عارضه‌ای در مجموعه داده اوّل هیچ عارضه متناظری در مجموعه داده دوّم وجود ندارد.

[1] Turning Function

[Downloaded from jgit.kntu.ac.ir on 2024-04-25]

[DOI: 10.29252/jgit.6.2.105]

- هیچ به یک (۰:۱): این حالت عکس رابطه یک به هیچ است. در این حالت برای عارضه‌ای در مجموعه داده دوّم هیچ عارضه متناظری در مجموعه داده اوّل وجود ندارد.
- یک به یک (۱:۱): برای هر عارضه در یک مجموعه داده یک عارضه مشابه با آن در مجموعه داده دیگر وجود دارد.
- رابطه یک به چند (۱:M): در این رابطه برای یک عارضه در مجموعه داده اوّل بیش از یک عارضه در مجموعه داده دوّم وجود دارد.
- رابطه چند به یک (M:۱): این حالت عکس حالت یک به چند می‌باشد. در این حالت برای چند عارضه در مجموعه داده اوّل یک عارضه در مجموعه داده دوّم وجود دارد.
- رابطه چند به چند (M:N): در این حالت گروهی از عوارض در مجموعه داده اوّل متناظر با گروهی دیگر از عوارض در مجموعه داده دوّم می‌باشند.

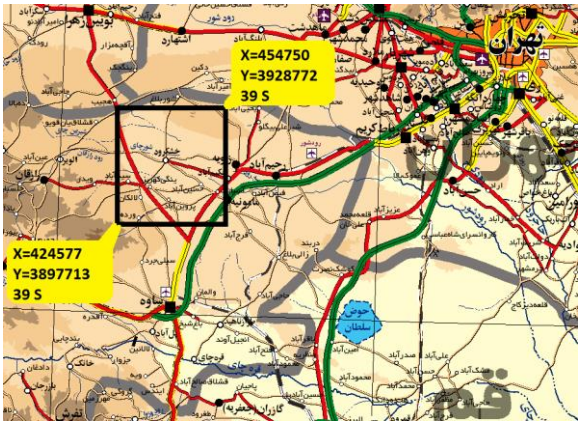
پس از اینکه نوع روابط هریک از عوارض از طریق درجه شباهت مکانی شناسایی گردید، با استفاده از روابط (۱ و ۲) مقدار F-score محاسبه می‌گردد.

۳- پیاده‌سازی

برای ارزیابی رویکرد پیشنهادی از داده‌های قسمتی از شهرستان ساوه با مساحت ۸۸۷/۰۸۸ کیلومترمربع و از نوع شبکه راه‌های برون شهری استفاده شده است (شکل (۳)). با توجه به شکل (۴) مجموعه داده اول دارای مقیاس ۱:۲۵۰۰۰، مجموعه داده دوم دارای مقیاس ۱:۵۰۰۰۰ و مجموعه داده سوم دارای مقیاس ۱:۱۰۰۰۰۰ می‌باشد که تمامی این مجموعه‌ها در سال ۱۳۹۱ به روش فتوگرامتری و توسط سازمان نقشه‌برداری کشور تولید شده است. با توجه به چارچوب پیشنهادی نیاز است تا در ابتدا بر روی مجموعه داده‌ها پیش پردازش صورت گیرد. پس از تبدیل مجموعه داده‌ها در هر منطقه به قالب و سیستم مختصات یکسان، خطاهای توپولوژیکی (نظیر خطاهای ردشدگی و نرسیدگی) هریک از مجموعه داده‌ها حذف می‌گردد. سپس برای اینکه در تعریف عوارض ابهام ایجاد نگردد مجموعه داده‌ها تبدیل به ساختار گراف تعریف شده می‌گردند. جدول (۱) تعداد عوارض هریک از مجموعه داده‌ها را قبل و بعد از پیش پردازش نشان می‌دهد.

جدول ۱: تعداد عوارض در مجموعه داده‌های مورد استفاده قبل و بعد از پیش پردازش

مجموعه داده‌ها	تعداد عوارض	
	قبل از پیش پردازش	بعد از پیش پردازش
مجموعه داده اول	۸۵۳	۶۴۳
مجموعه داده دوم	۵۷۶	۴۴۱
مجموعه داده سوم	۴۰۳	۳۴۸



شکل ۳: محدوده منطقه مورد مطالعه

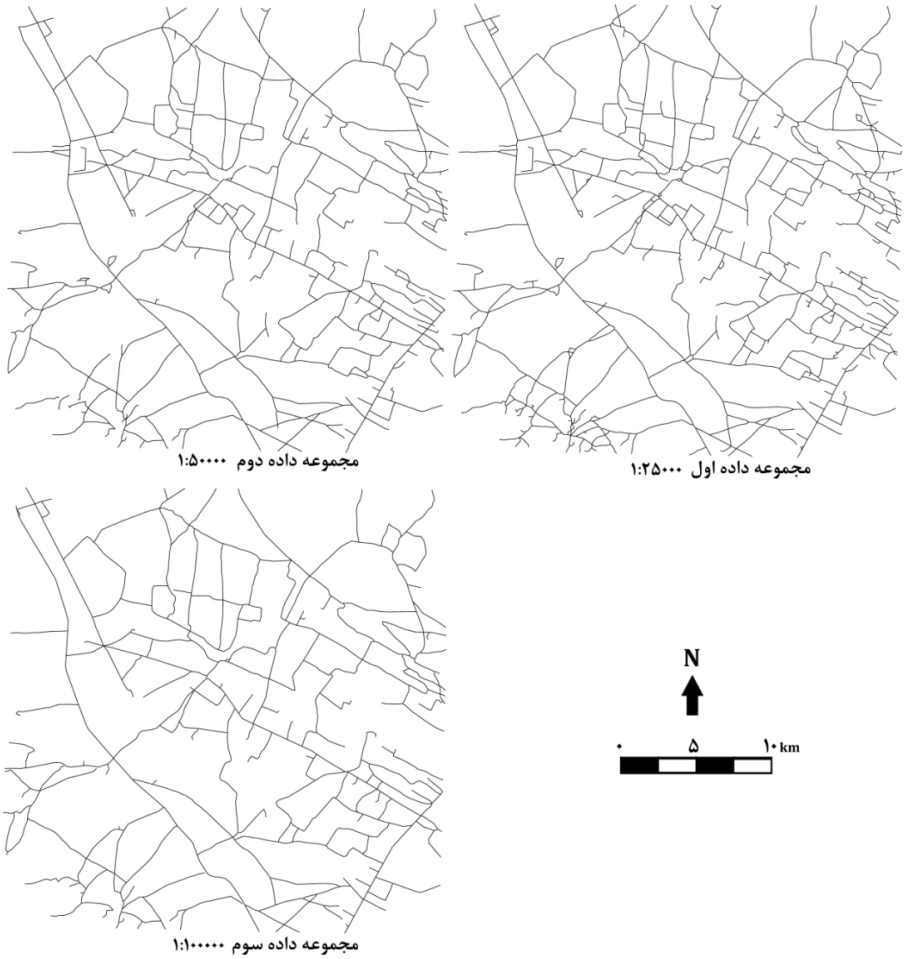
۳-۱- ارزیابی روش‌های بهینه‌سازی در منطقه آموزشی

در مرحله اول هریک از الگوریتم‌های ژنتیک، توده ذرات و جستجوی غذای زنبور عسل پارامترهای τ ، β ، $\mathbf{W}$ و $\mathbf{m}$ را در مناطق آموزشی هریک از مجموعه داده‌ها محاسبه می‌کند. تجربه نشان داده است تقریبا ۱۰٪ از مجموعه داده‌ها به‌عنوان منطقه آموزشی و باقی مانده عوارض به‌عنوان منطقه تست در نظر گرفته می‌شود. همچنین داده‌های آموزشی در منطقه‌ای

[Downloaded from jgit.kntu.ac.ir on 2024-04-25]
[DOI: 10.29252/jgit.6.2.105]

انتخاب می‌گردد که دارای هر شش حالت تناظریابی شامل یک به هیچ، هیچ به یک، یک به یک، یک به چند، چند به یک و چند به چند باشد. جدول (۲) گروه‌های مختلف مورد مطالعه برای تناظریابی و همچنین تعداد عوارض و روابط هریک را نشان می‌دهد.

برای مثال برای گروه اول، منطقه آموزشی در قسمتی از مجموعه داده‌ها انتخاب می‌گردد که ضمن وجود داشتن هر شش رابطه، حداقل ۵۸ رابطه در آن منطقه وجود داشته باشد



شکل ۴: مجموعه داد های مورد استفاده برای منطقه دوم مورد مطالعه

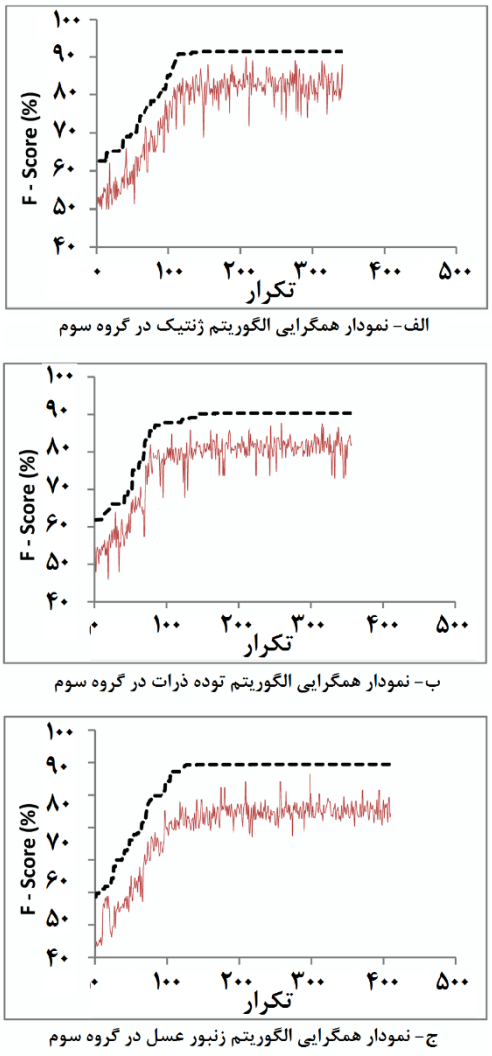
جدول ۲: تعداد روابط بین عوارض در گروههای مختلف مورد مطالعه

گروه‌ها	داده‌های تناظریابی	روابط مکانی						
		کل	یک به هیچ	هیچ به یک	یک به یک	یک به چند	چند به یک	چند به چند
۱	اول و دوم	۵۷۴	۵	۲	۳۸۲	۱۷۶	۲	۷
۲	اول و سوم	۴۲۱	۸	۳	۲۳۵	۱۵۶	۲	۱۷
۳	دوم و سوم	۵۲۳	۷	۵	۳۶۸	۱۳۰	۴	۹

[Downloaded from jgit.kntu.ac.ir on 2024-04-25]
[DOI: 10.29252/jgit.6.2.105]

با توجه به اینکه پارامترهای مورد نیاز برای الگوریتم بهینه‌سازی از طریق آزمون و خطا به دست می آیند، در نتیجه تعداد جمعیت اولیه برای الگوریتم ژنتیک ۳۰ کروموزم، نرخ جهش ۰/۲ (اعمال تنها بر روی ژن‌هایی با مقدار تصادفی بالای نرخ جهش) و باقی‌ماندن ۴۰ درصد از جمعیت هر نسل برای حل مسئله درنظر گرفته می‌شود. برای الگوریتم توده ذرات نیز ضرائب c_1 و c_2 برابر یک و w برابر ۰/۸ در نظر گرفته می‌شود. همچنین شرط همگرایی الگوریتم‌ها، تغییر نکردن بیشترین F-Score پس از ۲۰۰ تکرار می باشد. برای غلبه بر ماهیت احتمالی مسئله، کل فرایند الگوریتم ۳۰ بار تکرار می‌گردد.

به‌عنوان شکل (۵) نمونه نمودارهای همگرایی گروه سوم برای هر سه الگوریتم را نشان می‌دهد. الگوریتم ژنتیک بعد از ۳۳۲ تکرار به مقدار F-Score برابر ۹۱/۷۲ ٪، الگوریتم توده ذرات بعد از ۳۵۹ تکرار به مقدار F-Score برابر ۹۰/۹۳ ٪ و الگوریتم جستجوی غذای زنبور عسل نیز بعد از ۴۰۸ تکرار به مقدار F-Score برابر ۹۰/۷۵ ٪ در گروه سوم رسیده است. با بررسی نمودار همگرایی در گروه‌های سه‌گانه مورد مطالعه و از آنجائیکه هدف این تحقیق رسیدن به مقدار بالاتر F-Score صرف نظر از تعداد تکرارها می‌باشد، الگوریتم ژنتیک با داشتن پایداری در رسیدن به مقدار بالاتر F-Score به‌عنوان روش بهینه‌سازی مناسب جهت محاسبه مقدار بهینه پارامترهای τ ، β ، **W** و **m** در مسئله تناظریابی عوارض در مجموعه داده‌های مختلف انتخاب می‌گردد. شایان ذکر است که الگوریتم‌های توده ذرات و جستجوی غذای زنبور عسل علاوه بر رسیدن به مقدار F-Score پائین‌تر نسبت به الگوریتم ژنتیک، نسبت به یکدیگر نیز در تعدادی از گروه‌های مورد مطالعه الگوریتم توده ذرات عملکرد بهتری داشته است و در تعدادی دیگر از گروه‌ها الگوریتم جستجوی غذای زنبور عسل عملکرد مناسب‌تری از خود نشان داده است.



شکل ۵: نمودار همگرایی الگوریتم‌های ژنتیک، توده ذرات و زنبور عسل در گروه سوم

جدول (۳) مقادیر نهایی هریک از پارامترها را در گروه‌های مورد مطالعه نشان می‌دهد. همانطور که در جدول (۳) نمایان است، نمی‌توان از تمام معیارها در تمامی مجموعه گروه‌ها استفاده نمود و در صورت استفاده، از کارایی الگوریتم تناظریابی کاسته می‌شود. برای مثال معیار طول در هیچ یک از گروه‌های تناظریابی باعث بهبود فرایند تناظریابی نمی‌گردد. همچنین معیار موقعیت گره‌ها در گروه اول و سوم باعث بهبود فرایند تناظریابی می‌گردد در حالی که در

[Downloaded from jgit.kntu.ac.ir on 2024-04-25]
[DOI: 10.29252/jgit.6.2.105]

گروه دوم از فرایند تناظریابی حذف شده است. نتایج نشان می‌دهد مقادیر حد آستانه درجه شباهت مکانی و فاصله بافر برای هریک از گروه‌ها مقادیر متفاوتی است و انتخاب نادرست این مقادیر می‌تواند باعث کاهش کارایی الگوریتم‌های تناظریابی گردد. همانطور که در جدول (۳) نمایان است مقادیر فاصله بافر از ۷۰/۶۵ متر تا ۸۱/۲۱ متر و مقدار حد آستانه درجه شباهت مکانی از ۵۰/۵۲ ٪ تا ۵۹/۲۹ ٪ برای گروه‌های داده با مقیاس مختلف متفاوت است.

جدول ۳: مقدار بهینه محاسبه شده برای مناطق آموزشی در گروههای مورد مطالعه

گروه‌ها	F-score(%)	پارامترها												
		τ (%)	β (m)	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}
۱	۹۵/۱۱	۵۹/۲۹	۷۰/۶۵	۰/۱۲	۰	۰/۸۹	۰	۰/۷۶	۰	۰/۶۶	۰	۰/۲۸	۰/۵۹	۰/۱۸
۲	۸۹/۸۵	۵۰/۵۲	۷۸/۷۱	۰	۰	۰/۴۶	۰/۱۵	۰/۸۱	۰	۰/۲۵	۰	۰/۳۲	۰/۸۹	۰
۳	۹۱/۷۲	۵۷/۸۶	۸۱/۲۱	۰/۲۸	۰	۰/۶۸	۰/۰۸	۰/۳۸	۰	۰/۳۱	۰	۰/۱۱	۰/۸۳	۰/۲۱

۳-۲- تناظریابی در منطقه تست

پس از اینکه معیارهای موثر و مقدار بهینه پارامترها در مناطق آموزشی هر گروه از مجموعه داده‌ها محاسبه گردید، تناظریابی از طریق این مقادیر و معیارها در منطقه تست هریک از گروه‌ها صورت می‌گیرد. جدول (۴) نتایج به‌دست آمده از تناظریابی در منطقه تست برای گروه‌های مورد مطالعه را نشان می‌دهد.

جدول ۴: نتایج حاصل از تناظریابی در منطقه تست

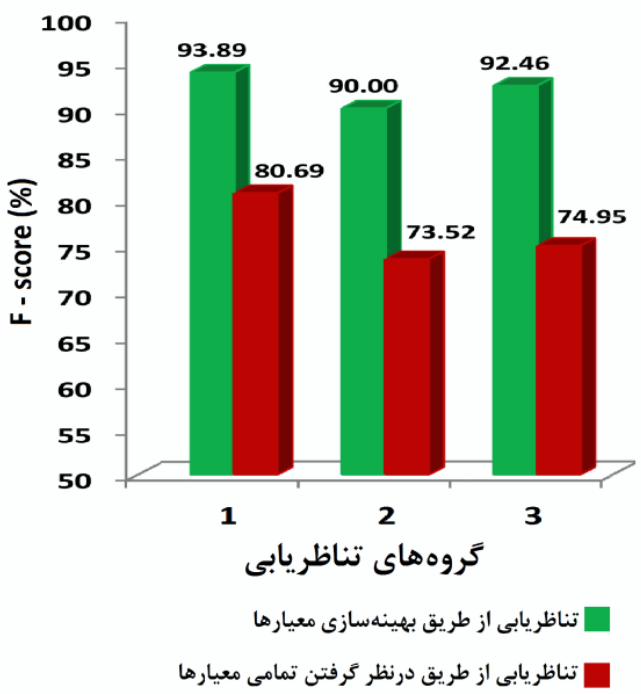
گروه‌ها	TP	FP	FN	P (%)	R (%)	F-score (%)
۱	۵۳۰	۲۵	۴۴	۹۵/۵۰	۹۲/۳۳	۹۳/۸۹
۲	۳۶۹	۳۰	۵۲	۹۲/۴۸	۸۷/۶۵	۹۰
۳	۴۷۲	۲۶	۵۱	۹۴/۷۸	۹۰/۲۵	۹۲/۴۶

۳-۳- ارزیابی نتایج

برای ارزیابی کارایی نتایج حاصل از بهینه‌سازی معیارهای هندسی، این نتایج با نتایج به‌دست آمده از تناظریابی از طریق در نظر گرفتن تمامی معیارهای هندسی (۱۱ معیار معرفی شده) مورد مقایسه قرار می‌گیرد. شکل (۶) نتایج نهایی به‌دست آمده از هر دو روش را نشان می‌دهد. همانطور که در شکل (۶) نمایان است چارچوب پیشنهادی در تمامی گروه‌های مورد مطالعه به مقدار بالاتری از F-Score در مقایسه با روش دیگر رسیده است و این نشان از کارایی بالای چارچوب پیشنهادی در تناظریابی عوارض در مجموعه داده‌های با مقیاس متفاوت می‌باشد. میانگین اختلاف مقدار F-score به‌دست آمده در چارچوب پیشنهادی نسبت به روش دیگر برابر ۱۴/۷۴ ٪ می‌باشد. چارچوب پیشنهادی برخلاف روش‌های دیگر برای هر گروه از مجموعه داده‌ها، آموزش دیده و پارامترهای مورد نیاز را با توجه به جزئیات و مقیاس آن گروه تعیین می‌کند. در نتیجه با فزایش یا کاهش سطح جزئیات و یا تغییر مقیاس در مجموعه داده‌ها از کارائی آن کاسته نخواهد شد.

[Downloaded from jgit.kntu.ac.ir on 2024-04-25]

[DOI: 10.29252/jgit.6.2.105]



شکل ۶: ارزیابی بهینه‌سازی معیارها در گروههای نه‌گانه مورد مطالعه

۴- نتیجه‌گیری و پیشنهادها

در این تحقیق کارایی الگوریتم‌های ژنتیک، توده ذرات و جستجوی غذای زنبور عسل برای تناظریابی عوارض خطی در مجموعه داده‌های مختلف با استفاده از بهینه‌سازی معیارهای هندسی مورد بررسی قرار گرفت. همچنین در این تحقیق علاوه بر تعیین بهینه‌ترین معیارهای هندسی، مقادیر بهینه حد آستانه درجه شباهت مکانی، مقدار بهینه فاصله بافر و همچنین وزن معیارها برای هر گروه از مجموعه دادها به دست آمد. نتایج نشان داد که الگوریتم ژنتیک در مقایسه با دو روش دیگر دارای کارایی بالاتری برای بهینه‌سازی پارامترهای τ، β، **W** و **M** در مسئله تناظریابی هندسی عوارض خطی می‌باشد. همچنین در این تحقیق در فرایند تناظریابی صرفاً از معیارهای هندسی (شامل خصوصیات هندسی و توپولوژیکی) استفاده گردید تا از کارایی تناظریابی در مجموعه داده‌های فاقد اطلاعات معنایی کاسته نشود. برخلاف بسیاری از تحقیقات گذشته چارچوب پیشنهادی توانایی شناسایی تمامی شش حالت یک به هیچ، هیچ به یک، یک به یک، یک به چند، چند به یک و چند به چند بین عوارض را داراست.

در این تحقیق صرفا از معیارهای هندسی (شامل خصوصیات هندسی و توپولوژیکی) استفاده گردید، در نتیجه در صورت وجود اطلاعات معنایی در مجموعه داده‌های ورودی می‌توان از آنها برای بهبود نتایج استفاده نمود. همچنین تحقیق حاضر صرفا برای شناسایی عوارض خطی (راه‌ها) متناظر ارائه شده است، از این رو با توسعه چارچوب فوق می‌توان برای سایر عوارض خطی و یا عوارض چندضلعی مورد استفاده قرار گیرد.

مراجع

[1] E. Xavier, F. J. Ariza-López, and M. A. Ureña-Cámara, "A Survey of Measures and Methods for Matching Geospatial Vector Datasets," ACM Computing Surveys (CSUR), vol. 49, pp. 1-34, 2016.

[Downloaded from jgit.kntu.ac.ir on 2024-04-25]

[DOI: 10.29252/jgit.6.2.105]

[2] L. Li and M. F. Goodchild, "An optimisation model for linear feature matching in geographical data conflation," International Journal of Image and Data Fusion, vol. 2, pp. 309-328, 2011/12/01 2011.

[3] A. Samal, S. Seth, and K. Cueto "A feature-based approach to conflation of geospatial sources," International Journal of Geographical Information Science, vol. 18, pp. 459-489, 2004.

[4] 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.

[5] B. Rosen and A. Saalfeld, "Match Criteria for Automatic Alignment," in Proceedings, Auto-Carto VII., 1985.

[6] A. Lupien and W. Moreland, "A general approach to map conflation," in Proceedings of 8th International Symposium on Computer Assisted Cartography (AutoCarto 8), 1987, pp. 630-639.

[7] A. Saalfeld, "Conflation Automated map compilation," International Journal of Geographical Information Systems, vol. 2, pp. 217-228, 1988/01/01 1988.

[8] R. M. Pendyala, Development of GIS-based conflation tools for data integration and matching: Florida Department of Transportation, 2002.

[9] D. Sheeren, S. Mustière, and J.-D. Zucker, "How to Integrate Heterogeneous Spatial Databases in a Consistent Way?," in Advances in Databases and Information Systems. vol. 3255, A. Benczúr, J. Demetrovics, and G. Gottlob, Eds., ed: Springer Berlin Heidelberg, 2004, pp. 364-378.

[10] E. Safra, Y. Kanza, Y. Sagiv, and Y. Doytsher, "Ad hoc matching of vectorial road networks," International Journal of Geographical Information Science, vol. 27, pp. 114-153, 2013.

[11] H. Stigmar, "Matching route data and topographic data in a real-time environment," in 10th Scandinavian Research Conference on Geographical Information Science, 2005, pp. 13-15.

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

[13] S. Mustière, "Results of experiments on automated matching of networks at different scales," International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 36, pp. 92-100, 2006.

[14] A. J. Saalfeld, "Automated map conflation. Washington, DC: University of Maryland.," 1993.

[15] M. Zhang and L. Meng, "An iterative road-matching approach for the integration of postal data," Computers, Environment and Urban Systems, vol. 31, pp. 597-615, 2007.

[16] V. Walter and D. Fritsch, "Matching spatial data sets: a statistical approach," International Journal of Geographical Information Science, vol. 13, pp. 445-473, 1999.

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

[18] L. Li and M. Goodchild, "Automatically and accurately matching objects in geospatial datasets," in Proceedings of joint international conference on theory, data handling and modelling in geospatial information science, Hong Kong, 2010, pp. 98-103.

[19] M. Sester, G. Gösseln, and B. Kieler, "Identification and adjustment of corresponding objects in data sets of different origin," in 10th AGILE International Conference on Geographic Information Science 2007, 2007.

[20] M. Zhang and L. Meng, "Delimited stroke oriented algorithm-working principle and implementation for the matching of road networks," Geographic Information

[Downloaded from jgit.kntu.ac.ir on 2024-04-25]

[DOI: 10.29252/jgit.6.2.105]

Sciences, vol. 14, pp. 44-53, 2008.

[21] B. Yang, Y. Zhang, and X. Luan, "A probabilistic relaxation approach for matching road networks," International Journal of Geographical Information Science, vol. 27, pp. 319-338, 2013.

[22] Y. Wang, Q. Du, F. Ren, and Z. Zhao, "A Propagating Update Method of Multi-Represented Vector Map Data Based on Spatial Objective Similarity and Unified Geographic Entity Code," in Cartography from Pole to Pole: Selected Contributions to the XXVIth International Conference of the ICA, Dresden 2013, M. Buchroithner, N. Prechtel, and D. Burghardt, Eds., ed Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, pp. 139-153.

[23] Y. Wang, H. Lv, X. Chen, and Q. Du, "A PSO-Neural Network-Based Feature Matching Approach in Data Integration," in Cartography-Maps Connecting the World, ed: Springer, 2015, pp. 189-219.

[24] D. Min, L. Zhilin, and C. Xiaoyong, "Extended Hausdorff distance for spatial objects in GIS," International Journal of Geographical Information Science, vol. 21, pp. 459-475, 2007/04/01 2007.

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

[26] J. Hangouet, "Computation of the Hausdorff distance between plane vector polylines," in AUTOCARTO-CONFERENCE-, 1995, pp. 1-10.

[27] G. Touya, A. Coupé, J. L. Jollec, O. Dorie, and F. Fuchs, "Conflation optimized by least squares to maintain geographic shapes," ISPRS International Journal of Geo-Information, vol. 2, pp. 621-644, 2013.

[28] A. Mascret, T. Devogele, I. Le Berre, and A. Hénaff, Coastline matching process based on the discrete Fréchet distance: Springer, 2006.

[29] T. Devogele, "A new Merging process for data integration based on the discrete Fréchet distance," in Advances in spatial data handling, ed: Springer, 2002, pp. 167-181.

[30] X. Tong, W. Shi, and S. Deng, "A probability-based multi-measure feature matching method in map conflation," International Journal of Remote Sensing, vol. 30, pp. 5453-5472, 2009.

[31] W. Song, J. M. Keller, T. L. Haithcoat, and C. H. Davis, "Relaxation - Based Point Feature Matching for Vector Map Conflation," Transactions in GIS, vol. 15, pp. 43-60, 2011.

[32] A. Moosavi and A. A. Alesheikh, "Developing of Vector Matching Algorithm Considering Topologic Relations," Proceedings of Map Middle East, 2008.

[33] W. Cohen, P. Ravikumar, and S. Fienberg, "A comparison of string metrics for matching names and records," in Kdd workshop on data cleaning and object consolidation, 2003, pp. 73-78.

[34] M. A. Rodriguez and M. J. Egenhofer, "Comparing geospatial entity classes: an asymmetric and context-dependent similarity measure," International Journal of Geographical Information Science, vol. 18, pp. 229-256, 2004.

[35] A.-M. Olteanu-Raimond, S. Mustière, and A. Ruas, "Knowledge formalisation for vector data matching using Belief Theory," Journal of Spatial Information Science, vol. 10, pp. 21-46, 2015.

[36] H. Du, N. Alechina, M. Jackson, and G. Hart, "A Method for Matching Crowd - sourced and Authoritative Geospatial Data," Transactions in GIS, 2016.

[37] C. Beeri, Y. Kanza, E. Safra, and Y. Sagiv, "Object fusion in geographic information systems," in Proceedings of the Thirtieth international conference on Very large data bases-Volume 30, 2004, pp. 816-827.

[38] C. Beeri, Y. Doytsher, Y. Kanza, E. Safra, and Y. Sagiv, "Finding corresponding objects when integrating several geo-spatial datasets," in Proceedings of the 13th annual ACM international workshop on Geographic

[Downloaded from jgit.kntu.ac.ir on 2024-04-25]

[DOI: 10.29252/jgit.6.2.105]

information systems, 2005, pp. 87-96.

[39] M. Zhang, W. Shi, and L. Meng, "A generic matching algorithm for line networks of different resolutions," in Workshop of ICA Commission on Generalization and Multiple Representation Computering Faculty of A Coruña University-Campus de Elviña, Spain, 2005.

[40] E. Safra, Y. Kanza, Y. Sagiv, and Y. Doytsher, "Efficient integration of road maps," in Proceedings of the 14th annual ACM international symposium on Advances in geographic information systems, 2006, pp. 59-66.

[41] J. O. Kim, K. Yu, J. Heo, and W. H. Lee, "A new method for matching objects in two different geospatial datasets based on the geographic context," Computers & Geosciences, vol. 36, pp. 1115-1122, 2010.

[42] Y. Gabay and Y. Doytsher, "Automatic adjustment of line maps," in proceedings of GIS/LIS, 1994, pp. 332-340.

[43] P. Lüscher, D. Burghardt, and R. Weibel, "Matching road data of scales with an order of magnitude difference," in Proceeding of XXIII International Cartographic Conference, 2007.

[44] D. Mantel and U. Lipeck, "Matching cartographic objects in spatial databases," Int. Archives of Photogrammetry, Remote Sensing and Spatial Inf. Sciences, vol. 35, pp. 172-176, 2004.

[45] D. Demetriou, L. See, and J. Stillwell, "A spatial genetic algorithm for automating land partitioning," International Journal of Geographical Information Science, vol. 27, pp. 2391-2409, 2013/12/01 2013.

[46] N. N. Vinh and B. Le, "Incremental Spatial Clustering in Data Mining Using Genetic Algorithm and R-Tree," in Simulated Evolution and Learning: 9th International Conference, SEAL 2012, Hanoi, Vietnam, December 16-19, 2012. Proceedings, L. T. Bui, Y. S. Ong, N. X. Hoai, H. Ishibuchi, and P. N. Suganthan, Eds., ed Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 270-279.

[47] G. N. Yücenur and N. Ç. Demirel, "A new geometric shape-based genetic clustering algorithm for the multi-depot vehicle routing problem," Expert Systems with Applications, vol. 38, pp. 11859-11865, 2011.

[48] L. M. Li, K. D. Lu, G. Q. Zeng, L. Wu, and M. R. Chen, "A novel real-coded population-based extremal optimization algorithm with polynomial mutation: A non-parametric statistical study on continuous optimization problems," Neurocomputing, vol. 174, pp. 577-587, 2016.

[49] S. V. Suggala and P. K. Bhattacharya, "Real coded genetic algorithm for optimization of pervaporation process parameters for removal of volatile organics from water," Industrial & engineering chemistry research, vol. 42, pp. 3118-3128, 2003.

[50] A. Blanco, M. Delgado, and M. Pegalajar, "A real-coded genetic algorithm for training recurrent neural networks," Neural networks, vol. 14, pp. 93-105, 2001.

[51] R. L. Haupt and S. E. Haupt, Practical genetic algorithms: John Wiley & Sons, 2004.

[52] J. Eshelman Larry and J. Schaffer David, "Real-coded Genetic Algorithms and Interval-Schemata," Foundations of Genetic Algorithms, vol. 2, pp. 187-202, 1992.

[53] W. A. Mackaness and G. A. Mackechnie, "Automating the detection and simplification of junctions in road networks," GeoInformatica, vol. 3, pp. 185-200, 1999.

[54] A. Chehreghan and R. Ali Abbaspour, "An assessment of spatial similarity degree between polylines on multi-scale, multi-source maps," Geocarto International, vol. 32, pp. 471-487, 2017.

[55] Y. Wang, D. Chen, Z. Zhao, F. Ren, and Q. Du, "A Back-Propagation Neural Network-Based Approach for Multi-Represented Feature Matching in Update Propagation," Transactions in GIS, vol. 19, pp. 964-993,

[Downloaded from jgit.kntu.ac.ir on 2024-04-25]

[DOI: 10.29252/jgit.6.2.105]

2015.

[56] H. Yan and J. Li, Spatial Similarity Relations in Multi-scale Map Spaces: Springer, 2014.

[57] D. L. Anderson, D. P. Ames, and P. Yang, "Quantitative Methods for Comparing Different Polyline Stream Network Models," Journal of Geographic Information System, vol. 6, pp. 88-98, 2014.

[58] A.-M. Olteanu Raimond and S. Mustière, "Data Matching – a Matter of Belief," in Headway in Spatial Data Handling, A. Ruas and C. Gold, Eds., ed: Springer Berlin Heidelberg, 2008, pp. 501-519.

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

[60] 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.

[61] A. Cecconi, "Integration of cartographic generalization and multi-scale databases for enhanced web mapping," Universität Zürich, 2003.

[62] C. Parent and S. Spaccapietra, "Database integration: the key to data interoperability," ed, 2000.

[Downloaded from jgit.kntu.ac.ir on 2024-04-25]

[DOI: 10.29252/jgit.6.2.105]

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

Journal of Geospatial Information Technology
Vol.6, No.2, Summer 2018

Assessment of Optimization Algorithms on Multi-scale Matching of Spatial Datasets Based on Geometric Properties

Alireza Chehreghan[1], Rahim Ali Abbaspour[*2]

1-PhD of GIS, School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran
2-Assisstant professor, School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran

Abstract

Identification of objects referring to the same entity in different datasets is known as objects matching, which is both directly and indirectly used in a wide range of applications including conflation, quality assessment, data updating, and multi-scale analysis. Hence, a novel object matching approach is presented in this article, in which, in addition to take only geometric property into account, i.e. geometric and topological criteria, extracted from objects, any initial dependency on empirical parameters such as threshold of spatial similarity degree, buffer distance, and metric weights is eliminated, through which matching procedure may then be conducted in different datasets. All the relations in the proposed approach are considered including: one-to-null, null-to-one, one-to-one, one-to-many, many-to-one, and many-to-many. Moreover, efficiency of linear object matching using Real Coded Genetic Algorithm (RCGA), Particle Swarm Optimization (PSO) algorithm, and Artificial Bee Colony (ABC) algorithm in different datasets were investigated through optimization of geometric criteria. In order to assess the efficiency of the proposed approach, three datasets of different scales from various sources were used. As indicated by the results, the proposed framework was able to appropriately identify corresponding objects in different datasets. Additionally, it was revealed that GA outperformed the other two algorithms in terms of optimizing the parameters present in linear object matching.

Key words:. Linear Object Matching; Multi-scale Datasets; Geometric Property; Optimization Algorithms; Real Coded Genetic Algorithm.

[Downloaded from jgit.kntu.ac.ir on 2024-04-25]

[DOI: 10.29252/jgit.6.2.105]

Correspondence Address: GIS Group, School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran
Tel: +98-2188008841
Email : abaspour@ut.ac.ir