نشریه علمی مهندسی فناوری اطلاعات مکانی

سال نهم، شماره سوم، پاییز ۱۴۰۰

Vol.9, No.3, Autemn 2021

۱ -۲٤

مقاله پژوهشی

DOR: 20.1001.1.20089635.1400.9.3.5.6

تناظریابی عوارض چندضلعی با استفاده از بهینه‌سازی معیارهای هندسی

علی معینی رودبالی[۱]، رحیم علی عباسپور[۲*]، علیرضا چهرقان[۳]

۱- دانشجوی کارشناسی ارشد سیستم‌های اطلاعات مکانی، دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی، دانشگاه تهران

۲- دانشیار گروه سیستم‌های اطلاعات مکانی، دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی، دانشگاه تهران

۳- استادیار دانشکده مهندسی معدن، دانشگاه صنعتی سهند

تاریخ دریافت مقاله: ۱۴۰۰/۰۳/۰۲ تاریخ پذیرش مقاله: ۱۴۰۰/۰۸/۲۵

چکیده

در تناظریابی هندسی عوارض چندضلعی در مجموعه داده‌های چندمقیاسی معیارهای متفاوتی مورد استفاده قرار می‌گیرند. این معیارها در تناظریابی عوارض چندضلعی در مجموعه داده‌های برداری با مقیاس‌های مختلف عملکرد متفاوتی دارند. به عنوان نمونه معیار مساحت هم‌پوشانی در تناظریابی مجموعه داده‌های ۱:۲۵۰۰۰ و ۱:۵۰۰۰۰ اثرگذاری متفاوتی نسبت به تناظریابی در مجموعه داده‌های ۱:۵۰۰۰۰ و ۱:۱۰۰۰۰۰ در نتایج تناظریابی دارد. از این رو جهت رسیدن به نتایج مطلوب‌تر، تعیین مقادیر بهینه وزن معیارهای هندسی برای هریک از مجموعه داده‌ها ضروری به نظر می‌رسد. در تحقیق حاضر رویکردی برمبنای الگوریتم ژنتیک جهت تعیین مقدار بهینه میزان اثرگذاری معیارهای هندسی در تناظریابی عوارض چندضلعی با بهینه‌سازی وزن معیارها پیشنهاد می‌گردد. این رویکرد به منظور یافتن عوارض متناظر در مجموعه داده‌های مکانی از پنج معیار مساحت هم‌پوشانی، فاصله اقلیدسی، راستای عوارض، فاصله هاسدورف و شباهت شکل عوارض بصورت همزمان استفاده نموده و با بهره‌گیری از الگوریتم ژنتیک، تناظریابی عوارض را براساس بهینه‌سازی معیارها انجام می‌دهد. جهت ارزیابی رویکرد پیشنهادی از مجموعه داده‌های مکانی متنوعی استفاده شده است. داده‌های مورد استفاده شامل بخشی از عوارض مسکونی شهر بندرعباس در مقیاس‌های ۱:۲۵۰۰۰، ۱:۵۰۰۰۰ و ۱:۱۰۰۰۰۰، عوارض مسکونی منطقه ۶ شهر تهران در مقیاس‌های ۱:۲۵۰۰۰ و ۱:۵۰۰۰۰ و بخشی از عوارض مسکونی شهر رشت در مقیاس‌های ۱:۲۵۰۰۰، ۱:۵۰۰۰۰ و ۱:۱۰۰۰۰۰ می‌باشد. نتایج نشان داد تناظریابی با رویکرد پیشنهادی نسبت به حالتی که تمام معیارها با وزن برابر وارد تناظریابی شوند، به مقدار ۲۸٫۶۱ درصد و نسبت به حالتی که وزن معیارها بر اساس نظر کارشناس وارد تناظریابی شوند به مقدار ۹٫۱۳ درصد بهبود می‌یابد.

کلید واژه‌ها : تناظریابی عوارض چندضلعی، معیارهای هندسی، بهینه‌سازی، الگوریتم ژنتیک.

* نویسنده مکاتبه کننده: گروه سیستم‌های اطلاعات مکانی، دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی، پردیس دانشکده‌های فنی، دانشگاه تهران، تهران، ایران.

تلفن: ۰۲۱۸۸۰۰۸۸۴۱

Email: abaspour@ut.ac.ir

۱- مقدمه

با توجه به اهداف مختلفی که برای پروژه‌های مکانی تعریف می‌شود، بسیار اتفاق می‌افتد که برای یک منطقه واحد داده‌های مکانی متفاوتی موجود باشد. یک عارضه با ماهیت واحد در مجموعه داده‌های مختلف ممکن است با روابط مکانی، معنایی و هندسه متفاوتی ذخیره و نمایش داده شود [۱]. بدین ترتیب ممکن است از یک منطقه جغرافیایی واحد، نقشه‌های متفاوتی تولید شود [۲]. در این‌صورت، مفهوم پایگاه داده‌های مکانی چندمقیاسی مطرح می‌شود. در بسیاری از کاربردها نیاز است عوارضی که در دنیای واقعی موجودیت واحد دارند، در مجموعه داده‌های مختلف به عنوان یک عارضه متناظر تشخیص داده شوند. این فرآیند، تناظریابی نامیده شده و کلید اساسی در تناظریابی، یکسان بودن این عوارض در دنیای واقعی است [۱]. تناظریابی کاربردهای بسیاری دارد که می‌توان آنها را به چهار دسته کلی ادغام داده‌ها [۳]، غنی‌سازی داده‌های مکانی [۴]، ارزیابی کیفیت داده‌ها [۵] و مدیریت داده‌ها [۵ و ۶] تقسیم نمود.

در دهه گذشته تمرکز بسیاری از تحقیقات بر ارائه راهکارهایی جهت تناظریابی عوارض نقطه‌ای و خطی بوده است [۷، ۸، ۹، ۱۰، ۱۱ و ۱۲] و تناظریابی عوارض چندضلعی کمتر مورد توجه قرار گرفته است. برخی از پژوهش‌های انجام شده در یک مقیاس مشخص عملکرد خوبی دارند و در مقیاس‌های دیگر کیفیت نتایج پایین خواهد آمد [۱۳ و ۱۴]. در سال‌های اخیر تناظریابی عوارض چندضلعی نیز مورد توجه محققین قرار گرفته است و هریک رویکردی را برای این مسئله پیشنهاد کرده‌اند [۱۵ و ۱۶]. تناظریابی در عوارض چندضلعی بر اساس معیارهای اندازه‌گیری شباهت صورت می‌گیرد. این معیارها به دو دسته کلی معیارهای هندسی و معیارهای توصیفی تقسیم می‌شوند [۱۷]. از آنجائیکه معیارهای هندسی امکان استخراج از هر مجموعه داده‌ای را دارد و نیاز به اطلاعات اضافی (نظیر اطلاعات توصیفی) در آن نیست، در این تحقیق از معیارهای هندسی جهت تناظریابی استفاده می‌شود. در پژوهش‌های پیشین معیارهای هندسی مختلفی برای محاسبه شباهت عوارض مبنا قرار داده شده‌است؛ از جمله موقعیت مکانی عوارض [۱۸ و ۱۹]، طول عوارض [۱۳ و ۲۰]، فاصله اقلیدسی دو عارضه [۸ و ۲۱]، راستای عوارض [۱۲ و ۲۲]، مساحت [۲۳] و شکل عوارض [۷]. با توجه به وجود معیارهای مختلف در تناظریابی لازم است مقدار کارایی نتایج تناظریابی با استفاده از معیارهای مختلف بررسی شود و مشخص گردد استفاده همزمان از معیارها به چه نحوی می‌تواند نتایج تناظریابی را بهبود بخشد.

در تناظریابی عوارض چندضلعی معیارهای شباهت تعیین شده و با اندازه‌گیری مقدار شباهت دو عارضه و مقایسه آن با حد آستانه، عوارض متناظر شناسایی می‌شوند. در صورت استفاده از یک معیار کافی است مقدار شباهت محاسبه شده با مقدار حد آستانه مقایسه شود و اگر کمتر از حد آستانه نباشد دو عارضه متناظر شناخته می‌شوند. اما چنانچه بیش از یک معیار برای اندازه‌گیری مقدار شباهت دو عارضه استفاده شود باید ترتیبی اتخاذ شود که تمامی معیارها در تناظریابی دخیل باشند. به عنوان نمونه می‌توان میانگین مقادیر شباهت اندازه‌گیری شده را به عنوان معیار شباهت در نظر گرفت. به منظور شرکت دادن تمامی معیارها در تناظریابی از ترکیبی استفاده می‌شود که حاصل جمع هر معیار در یک ضریب است. این ضریب با عنوان وزن معیار شناخته شده و تعیین کننده مقدار تاثیر هر معیار در اندازه‌گیری شباهت دو عارضه خواهد بود. به تعداد معیارهای اندازه‌گیری تعیین شده وزن وجود دارد و باید مقدار این وزن‌ها تعیین شود. با متفاوت بودن مقادیر وزن معیارها نتایج تناظریابی نیز متفاوت خواهد بود. لذا باید ترکیبی از وزن‌ها را یافت به گونه‌ای که نتایج تناظریابی بهترین حالت ممکن باشد. این عملیات با نام بهینه‌سازی وزن معیارها مشخص شده و هدف اصلی پژوهش حاضر می‌باشد.

برای این منظور در این تحقیق رویکردی برمبنای

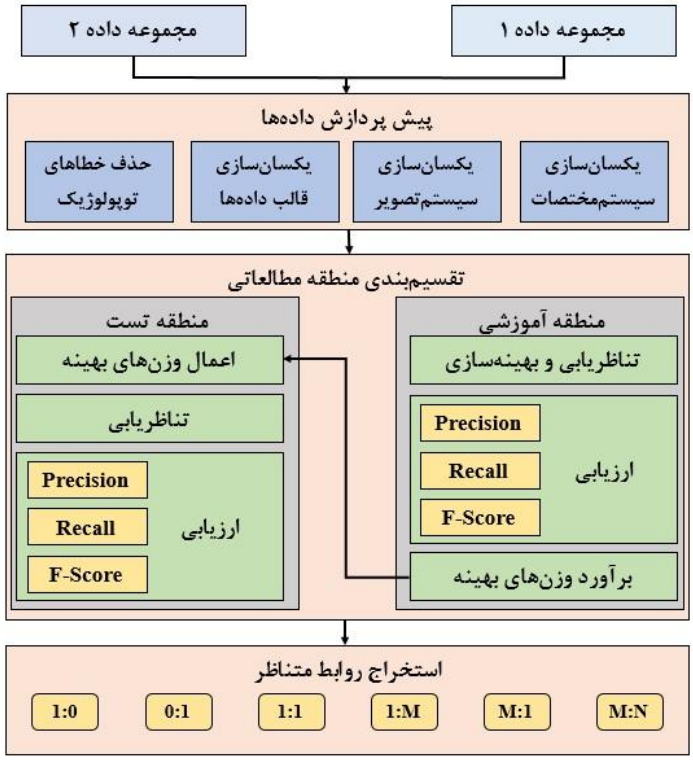
[DOR: 20.1001.1.20089635.1400.9.3.5.6]
[DOI: 10.52547/jgit.9.3.1]

الگوریتم ژنتیک جهت تعیین مقدار بهینه میزان اثرگذاری معیارهای هندسی در تناظریابی عوارض چندضلعی پیشنهاد می‌گردد. این رویکرد به منظور یافتن عوارض متناظر در مجموعه داده‌های مکانی از پنج معیار مساحت هم‌پوشانی، فاصله اقلیدسی، راستای عوارض، فاصله هاسدورف و شباهت شکل عوارض بصورت همزمان استفاده نموده و با بهره‌گیری از الگوریتم ژنتیک تناظریابی عوارض را براساس بهینه-سازی معیارها انجام می‌دهد. جهت ارزیابی جامع رویکرد پیشنهادی از مجموعه داده‌های مکانی با مقیاس‌های مختلف استفاده می‌گردد. همچنین نتایج بدست آمده با دو رویکرد معمول تناظریابی با در نظر گرفتن وزن برابر معیارها و تناظریابی براساس وزن اخذ شده از نظرات کارشناسان مورد مقایسه قرار می‌گیرد.

ادامه این مقاله بدین شرح خواهد بود: پس از مقدمه، در بخش ۲ رویکرد پیشنهادی ارائه شده و به تشریح جزئیات آن پرداخته خواهد شد. در بخش ۳ مناطق مورد مطالعه معرفی و نتایج اعمال روش پیشنهادی بر این مناطق بررسی و گزارش می‌شوند. در انتها، در بخش ۴ نتیجه‌گیری و پیشنهادها ارائه می‌گردد.

۲- رویکرد پیشنهادی

چارچوب کلی رویکرد پیشنهادی در شکل (۱) نشان داده شده‌است. در این پژوهش، به طور ویژه بر تناظریابی عوارض چندضلعی در مجموعه داده‌های برداری تمرکز شده‌است و برای تناظریابی تنها از معیارهای هندسی استفاده می‌شود. برای این منظور با حل مساله تناظریابی به عنوان یک مساله بهینه‌سازی و با بهره‌گیری از الگوریتم ژنتیک به عنوان یکی از پرکاربردترین و کاراترین الگوریتم‌های بهینه‌سازی سعی خواهد شد در هر مجموعه داده مکانی ترکیبی از وزن‌های متناسب با هر معیار به گونه‌ای یافت شود که نتایج تناظریابی بهترین حالت ممکن شود.



شکل ۱: چارچوب کلی رویکرد پیشنهادی

۲-۱- پیش پردازش

مجموعه داده‌های ورودی به تناظریابی از منابع مختلف و توسط سازمان‌های مختلف با استانداردها و ابزار، دقت، مقیاس، قالب و با اهداف گوناگون تولید شده‌اند؛ لذا بایستی قبل از اجرای تناظریابی، داده‌های مکانی در سیستم‌مختصات و سیستم‌تصویر یکسانی آماده شده و در قالب داده مکانی مشابهی ذخیره شوند. به علاوه در این مرحله همه مجموعه داده‌های مکانی از نظر خطاهای توپولوژیک بررسی شده و عوارض دارای خطای توپولوژیک از مجموعه داده‌ها حذف می‌گردند. عمده خطاهای توپولوژیکی که ممکن است در مجموعه داده‌های مکانی وجود داشته باشد و در تناظریابی ایجاد خطا کنند عبارتند از:

- وجود عوارض نقطه‌ای یا خطی در مجموعه داده مکانی: ممکن است عارضه نقطه‌ای و یا خطی به اشتباه به عنوان یک عارضه چندضلعی در مجموعه داده‌ها ذخیره شده باشند. از این رو شناسایی و حذف آنها در مرحله پیش پردازش ضروری است.
- وجود عوارض چندضلعی با مساحت مشترک در مجموعه داده مکانی: ممکن است در یک مجموعه داده مکانی عوارضی وجود داشته باشد که دارای مساحت مشترک باشند. به عنوان نمونه ذخیره همزمان یک واحد مسکونی و بلوک در بر گیرنده آن عارضه. وجود چنین عوارضی در نتایج تناظریابی موثر بوده و باعث نتیجه‌گیری اشتباه می‌گردد.
- وجود عوارض چندضلعی تکراری در مجموعه داده مکانی: در مراحل برداشت، ذخیره‌سازی و آماده‌سازی داده‌های مکانی ممکن است عوارض چندضلعی بیشتر از یک مرتبه در پایگاه داده ذخیره شوند. وجود چنین عوارضی باعث بوجود آمدن روابط متناظر نادرست می‌گردد.
- وجود عوارض نامرتبط در مجموعه داده مکانی: اگر هدف مساله، تناظریابی عوارض مسکونی در مجموعه داده مکانی باشد، بایستی در داده‌های مکانی فقط و فقط عوارض چندضلعی مربوط به ساختمان‌های مسکونی ذخیره شده‌باشد و وجود عوارض دیگر در این مجموعه داده‌ها (به عنوان نمونه فضاهای آموزشی یا بهداشتی) عامل ایجاد خطا در تناظریابی می‌گردد.

۲-۲- تفکیک مجموعه داده مکانی به مناطق آموزشی و تست

پس از اجرای پیش‌پردازش و حذف عوارض دارای خطای توپولوژیک، هر مجموعه داده به دو منطقه آموزشی و تست تقسیم می‌شود. در منطقه آموزشی تناظریابی بصورت همزمان با بهینه‌سازی انجام می‌شود. بهینه‌سازی بر روی وزن معیارها انجام شده و نتیجه آن برآورد وزن بهینه برای معیارهای پنج‌گانه خواهد بود. در گام بعدی وزن‌های بهینه در منطقه تست به معیارها اعمال شده و نتایج تناظریابی ارزیابی می‌شوند.

به منظور تفکیک مجموعه داده‌های مکانی به دو منطقه آموزشی و تست، ابتدا مرز جداکننده دو منطقه تعیین شده و عوارض موجود در مجموعه داده‌ها به دو دسته تقسیم می‌شوند. نمی‌توان عوارضی را بصورت تصادفی انتخاب نموده و به عنوان منطقه آموزشی در نظر گرفت؛ زیرا در اینصورت ممکن است عوارض از محدوده‌ای انتخاب شوند که تمامی شش نوع رابطه متناظر وجود نداشته باشد و بدین ترتیب هدف تحقیق که استخراج تمامی شش نوع رابطه متناظر می‌باشد، محقق نخواهد شد.

۲-۳- بهینه‌سازی وزن معیارها در منطقه آموزشی

در منطقه آموزشی همزمان با تناظریابی بهینه‌سازی نیز اجرا شده و مساله تناظریابی به عنوان یک مساله بهینه‌سازی حل می‌شود. برای این منظور از الگوریتم ژنتیک به عنوان یک الگوریتم کارآمد و پرکاربرد بهره برده می‌شود. در نهایت و پس از پایان یافتن این مرحله، پنج وزن برای پنج معیار هندسی استخراج می‌شود که در منطقه تست به عنوان وزن‌های بهینه به معیارها اعمال شده و نتایج تناظریابی حاصل شده

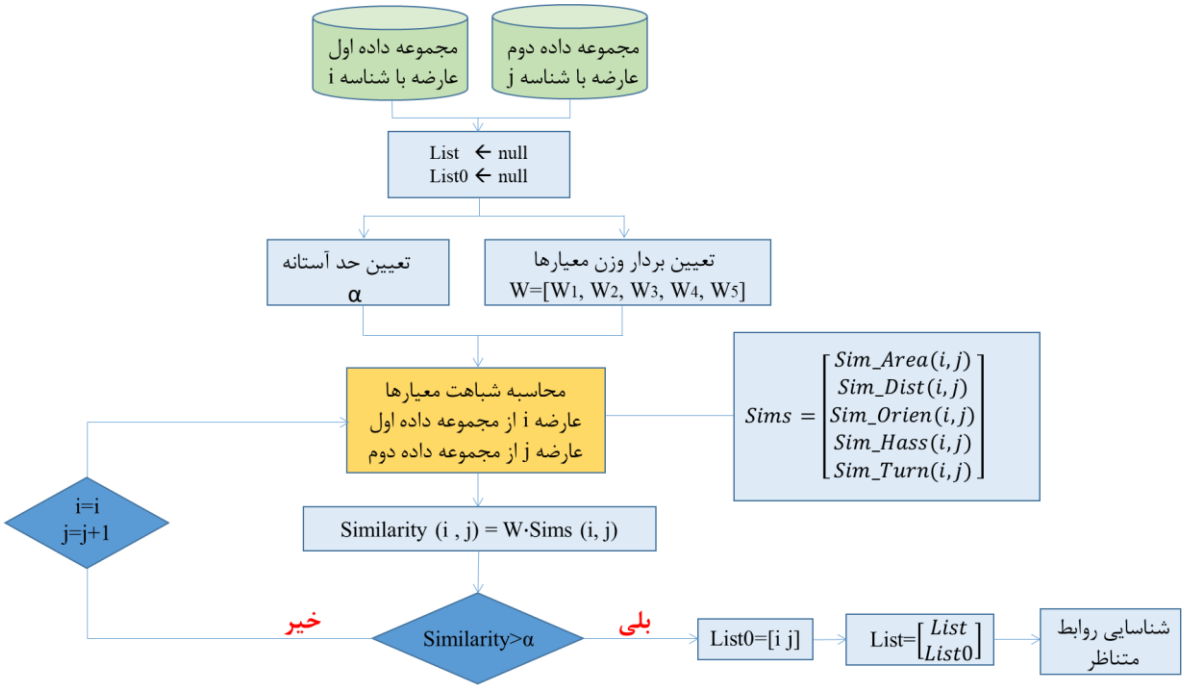
[DOR: 20.1001.1.20089635.1400.9.3.5.6]
[DOI: 10.52547/jgit.9.3.1]

ارزیابی می‌گردد. بدین ترتیب برآوردی از میزان موفقیت رویکرد پیشنهادی در مجموعه داده مکانی به دست خواهد آمد.

۲-۳-۱- تناظریابی عوارض چندضلعی

به منظور اجرای تناظریابی، برای هر عارضه در مجموعه داده اول عوارض کاندید در مجموعه داده دوم تعیین می‌شود. منظور از عوارض کاندید، عوارضی است که محاسبه شباهت‌ها تنها برای آنها صورت می‌گیرد. می‌توان در ساده‌ترین حالت شعاعی از عارضه چندضلعی مجموعه داده اول را در نظر گرفت و تنها شباهت عوارضی از مجموعه داده دوم را بررسی نمود که در این محدوده قرار دارند. در مرحله بعد درجه شباهت هندسی برای هر عارضه از مجموعه داده اول با عوارض کاندید از مجموعه داده دوم محاسبه می‌شود. جفت عوارضی که درجه شباهت آنها بیشتر از حد آستانه شباهت باشد به عنوان جفت عارضه متناظر تعیین می‌شوند. در شکل (۲) فلوچارت مراحل تناظریابی نشان داده شده‌است.

همچنین شبه‌کد الگوریتم استفاده‌شده برای تناظریابی دو عارضه در ادامه آمده‌است.



شکل ۲: فلوچارت الگوریتم تناظریابی مورد استفاده

Algorithm 1: Proposed Matching Method

```
'Data1[]: Polygons of dataset1
'Data2[]: Polygons of dataset2
function matching (Data1[], Data2[])
        for i ϵ Data1[] do
            list0 = null
            list = null
            for j ϵ Dataset2[] do
                calculate: Sim_Area (i, j),
                calculate: Sim_ Dist (i, j),
                calculate: Sim_ Orien (i, j),
                calculate: Sim_ Hass (i, j),
                calculate: Sim_ Turn (i, j),
                Sims = [Sim_Area (i, j), Sim_Dist (i, j), Sim_Orien (i, j), Sim_Hass (i, j), Sim_Turn (i, j)]
                define threshold: α
                define criteria weights: w (1*5)
                Similarity = inner product (w, Sims)
                if Similarity > α then
                    list0 = [i, j]
                end if
                list = [list; list0]
            end for
        end for
        return list
end function
```

۲-۳-۲- معیارهای هندسی

در پیشینه تحقیق معیارهای مورد استفاده به دو دسته تقسیم می‌شوند: معیارهای هندسی [۴، ۱۷ و ۲۴] و معیارهای معنایی [۲۵]. معیارهای هندسی متنوعی را می‌توان مبنای مقایسه شباهت دو عارضه مکانی قرار داد که در رویکرد پیشنهادی از پنج معیار هندسی پرکاربرد نظیر مساحت هم‌پوشانی، فاصله اقلیدسی، راستای عوارض، فاصله هاسدورف و شباهت شکل عوارض استفاده شده است.

مساحت هم‌پوشانی: مساحت هم‌پوشانی در پژوهش‌های فن و همکاران (۲۰۱۴)، هوآنگ و همکاران (۲۰۱۰)، کیم و یو (۲۰۱۵) و ژونگلیانگا و جیانهوا (۲۰۰۸) استفاده شده‌است [۵، ۲۶، ۲۷ و ۲۸]. مساحت هم‌پوشانی عارضه چندضلعی A از مجموعه داده نخست با عارضه چندضلعی B از مجموعه داده دوم با رابطه (۱) محاسبه می‌گردد [۴].

رابطه(۱)
$$Sim_{Area(A,B)} = \frac{Area(A \cap B)}{\min(Area(A), Area(B))}$$

در رابطه(۱)، $Area(A \cap B)$ مساحت مشترک دو عارضه، min(...) عملگری است که کمترین مقدار مساحت عوارض ورودی را برمی‌گرداند و $Sim_{Area(A,B)}$ عددی بین صفر و یک می‌باشد که بیانگر شباهت از منظر مساحت هم‌پوشانی دو عارضه است. شایان ذکر است هرچه مقدار این متغیر بیشتر باشد به معنای شباهت بیشتر دو عارضه تفسیر خواهد شد.

فاصله اقلیدسی: فاصله اقلیدسی مراکز هندسی دو عارضه نیز می‌تواند معیار مقایسه شباهت دو عارضه باشد [۶ و ۲۶]. برای تناظریابی و مقایسه شباهت دو عارضه بر مبنای فاصله اقلیدسی، با استفاده از تابع عضویت z شکل[۱] که در رابطه (۲) نحوه محاسبه آن برای مقادیر مختلف x با توجه به کران‌های بالا و پایین a و b نشان داده شده‌است، مقادیر فاصله اقلیدسی بین عوارض به عددی در بازه $[۰,۱]$ تبدیل می‌شود [۲۹]. شکل(۳) نمونه‌ای از تابع عضویت جهت نرمال‌سازی این

[۱] *Z-Shape Membership Function (ZMF)*

مقدار را نشان می‌دهد. مقادیر کران بالا و پائین پس از ارزیابی مجموعه داده‌های گوناگون برابر ۲۰ و ۸۰ تعیین می‌گردد. در رابطه (۲) مقادیر a و b کران‌های پایین و بالای بازه می‌باشند و x فاصله اقلیدسی بین مرکز ثقل دو عارضه چندضلعی می‌باشد.

رابطه(۲)

$$Sim_{Dist}=\begin{cases}۱ & x\le a\\ ۱-۲(\frac{x-a}{b-a})^{۲} & a\le x\le \frac{a+b}{۲}\\ ۲(\frac{x-b}{b-a})^{۲} & \frac{a+b}{۲}\le x\le b\\ ۰ & b\le x\end{cases}$$

شکل ۳: تبدیل مقادیر فاصله اقلیدسی به بازه $[۰,۱]$ از طریق تابع عضویت z شکل

راستای عارضه: یکی دیگر از معیارهای مقایسه شباهت، محاسبه اختلاف راستای دو عارضه می‌باشد [۳۰]. در حالت مقایسه دو عارضه چندضلعی، یکی از روش‌های محاسبه اختلاف راستا می‌تواند مقایسه راستای کوچکترین مستطیل محیطی عارضه چندضلعی[۱] باشد [۷]. در شکل (۴) کوچکترین مستطیل محیطی و زاویه معرف راستای آن برای دو عارضه A و B نشان داده شده است.

برای تبدیل مقدار اختلاف راستای عارضه A از مجموعه داده نخست با عارضه B از مجموعه داده دوم به عددی در بازه صفر و یک از رابطه (۳) استفاده می‌شود [۳۱].

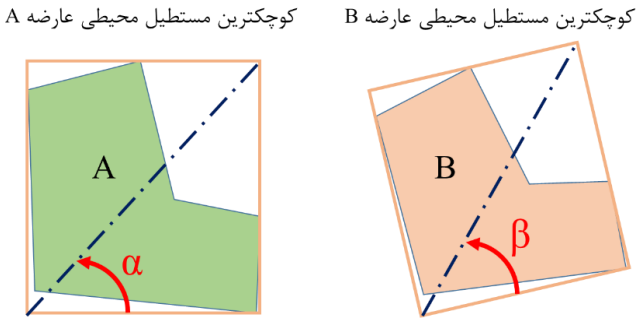
در رابطه(۳)، $Orien_A$ راستای عارضه A و $Orien_B$ راستای عارضه B می‌باشد.

رابطه(۳)

$$Sim_{Orien(A,B)}=۱-(\frac{Orien_{diff}}{۹۰})$$

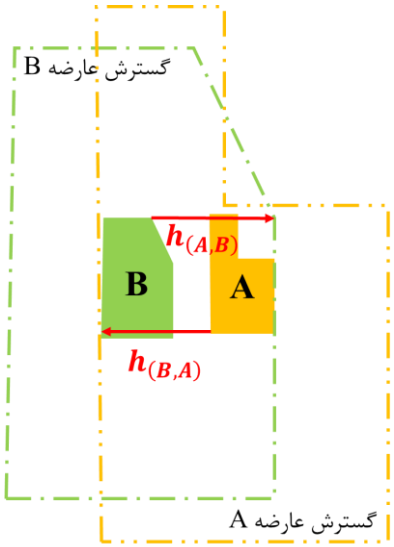
$$Orien_{diff}=|Orien_A-Orien_B|$$

[۱] *Minimum Bounding Box*



شکل ۴: راستای عوارض چندضلعی *A* و *B* با کوچکترین مستطیل محیطی

فاصله هاسدورف: علاوه بر فاصله اقلیدسی می‌توان از فاصله‌های دیگری نیز برای مقایسه شباهت دو عارضه بهره برد؛ ازجمله پرکاربردترین آنها فاصله هاسدورف می‌باشد [۳۲]. فاصله هاسدورف، بصورت حداکثر مقدار از بین حداقل فاصله یک مجموعه از مجموعه دیگر تعریف می‌شود. در محاسبه فاصله هاسدورف همه نقاط چندضلعی در محاسبه فاصله تاثیرگذار هستند و این موضوع می‌تواند از جنبه متفاوت‌تری نسبت به فاصله اقلیدسی دو عارضه را مورد مقایسه قرار دهد. همچنین شکل و جهت عارضه نیز در محاسبه مقدار فاصله هاسدورف موثر است. فاصله هاسدورف بین دو عارضه A و B در شکل (۵) نمایش داده شده‌است.



شکل ۵: نمایش فاصله هاسدورف دو عارضه چندضلعی

برای محاسبه فاصله هاسدورف بین دو عارضه، ابتدا مقادیر $h_{(A,B)}$ (شعاع کوچکترین دایره‌ای که اگر عارضه B به آن مقدار گسترش یابد، تمام عارضه A را پوشش می‌دهد) و $h_{(B,A)}$ (کوچکترین شعاع دایره‌ای که با گسترش عارضه A به آن اندازه، تمام عارضه B در درون عارضه A قرار گیرد) مطابق با رابطه (۴) محاسبه شده و مقدار فاصله هاسدورف بین دو عارضه طبق رابطه (۵) محاسبه می‌شود [۳۱].

$$h_{(A,B)} = \min\{\varepsilon : A \subset B \oplus S(\varepsilon)\}$$
$$h_{(B,A)} = \min\{\varepsilon : B \subset A \oplus S(\varepsilon)\} \quad \text{رابطه(۴)}$$

$$d_{Hausdorff}(A,B) = \max\{h_{(A,B)}, h_{(B,A)}\} \quad \text{رابطه(۵)}$$

در رابطه(۴)، علامت $\oplus$ نمایش‌دهنده عملگر گسترش (*Dilation*) در مورفولوژی ریاضی و $S_{(\varepsilon)}$ برابر با دایره‌ای به شعاع ε می‌باشند. پس از محاسبه مقدار فاصله هاسدورف، مشابه معیار فاصله اقلیدسی، برای نگاشت این مقدار به بازه صفر و یک نیز از تابع عضویت z شکل استفاده شده‌است.

تابع چرخش(*Turning Function*): این معیار برای هر عارضه چندضلعی، برآوردی از شکل آن عارضه را بصورت یک نمودار پله‌ای ارائه می‌دهد. با مقایسه این نمودارها برای دو عارضه چندضلعی می‌توان شکل هندسی دو عارضه را مقایسه نمود [۳۳، ۳۴ و ۳۵]. مقادیر تابع چرخش از رابطه(۶) محاسبه می‌شود [۳۶]. شکل(۶) نمودارهای ایجاد شده از اجرای تابع چرخش برای دو عارضه مفروض را نشان می‌دهد. به منظور نگاشت مقدار اختلاف تابع چرخش دو شکل به بازه $[۰,۱]$ از رابطه (۷) استفاده می‌شود.

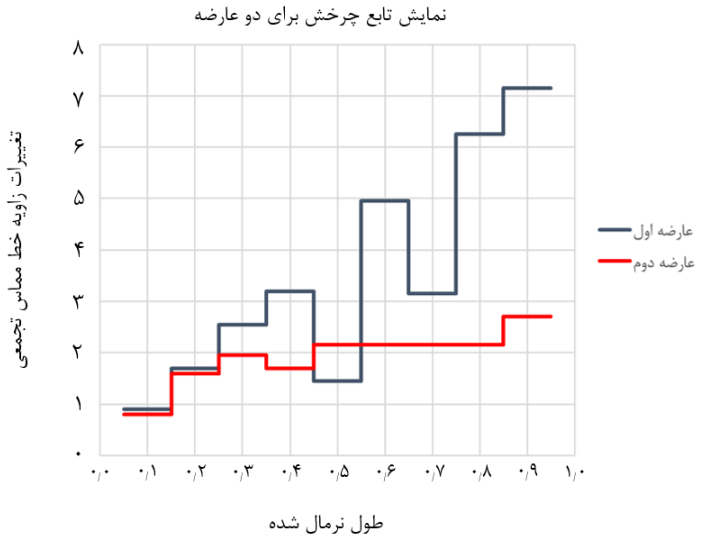
رابطه(۶)

$$d_{TF}(A,B)=\|T_A-T_B\|_۲=\left(\int_{۰}^{۱}|T_A(l)-T_B(l)|^۲\,dl\right)^{\frac{۱}{۲}}$$

رابطه(۷)

$$Sim_{Shape}(A,B)=۱-\frac{d_{TF}(A,B)}{U}$$

در روابط (۶) و (۷)، $d_{TF}(A,B)$ مقدار اختلاف تابع چرخش دو عارضه چندضلعی، T_A و T_B به ترتیب تابع چرخش چندضلعی‌های A و B و $\|...\|_۲$ نشانگر اپراتور نرم $l_۲$ می‌باشند و U حداکثر اختلاف در دو تابع چرخش هستند.



شکل ۶: نمایش تابع چرخش برای دو عارضه چندضلعی

۲-۳-۳- تابع ارزیابی

در فرایند بهینه‌سازی معیارهای هندسی در تناظریابی عوارض چندضلعی، محاسبه شباهت هندسی از طریق معیارهای معرفی شده از اولین مراحل می‌باشد. برای این منظور رابطه(۸) نحوه محاسبه مقدار شباهت هندسی دو عارضه چندضلعی A و B را نشان می‌دهد [۱۱]. در رابطه(۸)، W_i مقدار وزن هریک از معیارها و $Similarity_{(A,B)}$ مقدار شباهت محاسبه شده برای دو عارضه را نشان می‌دهد.

پس از محاسبه مقدار شباهت هندسی برای دو عارضه، این مقدار با حد آستانه شباهت مورد مقایسه قرار می‌گیرد. اگر مقدار شباهت هندسی محاسبه شده از حد

آستانه بیشتر باشد به معنای متناظر بودن دو عارضه تفسیر می‌شود. سپس جهت ارزیابی نتایج تناظریابی، تمام عوارضی که مقدار شباهت‌شان بیشتر از حد آستانه باشد در ماتریسی ذخیره می‌شوند. این ماتریس به ماتریس روابط خروجی نامگذاری می‌شود. در گام بعدی این ماتریس با فایل روابط صحیح که بصورت دستی و توسط کارشناس از طریق شناسایی عوارض متناظر تولید شده است مقایسه می‌شود.

برای مقایسه ماتریس روابط خروجی با فایل روابط صحیح روش‌های مختلفی وجود دارد، از جمله پرکاربرترین آنها مقایسه مقادیر *Precision*، *Recall* و *F-Score* می‌باشد. در این پژوهش از معیار *F-Score* به عنوان تابع ارزیابی استفاده شده است که از طریق رابطه(۹) محاسبه می‌شود. در این رابطه، مقادیر *Precision* و *Recall* از رابطه (۱۰) محاسبه می‌شوند. در رابطه(۱۰)، *TP* تعداد روابط تشخیص داده شده صحیح، *FP* تعداد روابط تشخیص داده شده اشتباه و *FN* تعداد روابط صحیح تشخیص داده نشده می‌باشند.

رابطه(۸)
$$Similarity_{(A,B)} = \frac{W_1 \times Sim_{Area} + W_2 \times Sim_{Dist} + W_3 \times Sim_{Orien} + W_4 \times Sim_{Has} + W_5 \times Sim_{Turn}}{\sum_{i=1}^{5} W_i}$$

رابطه(۹)
$$F - Score = 2 \times \frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}$$

رابطه(۱۰)
$$\mathrm{Precision} = \frac{TP}{TP + FP} \quad , \quad \mathrm{Recall} = \frac{TP}{TP + FN}$$

۲-۳-۴- ساختار کروموزوم

در الگوریتم ژنتیک هر کروموزوم به عنوان یکی از پاسخ‌های مساله می‌باشد که از طریق مقدار تابع ارزیابی، شایستگی آن سنجیده می‌شود. در این پژوهش به منظور بهینه‌سازی معیارها از الگوریتم ژنتیک با کدگذاری واقعی استفاده می‌شود، لذا ساختار کروموزوم در این تحقیق بصورت شکل (۷) و دارای پنج ژن خواهد بود که هر کدام از ژن‌ها نمایانگر وزن یکی از معیارهای هندسی می‌باشد. شایان ذکر است هر کدام از وزن‌ها دارای مقادیری بین ۰ تا ۱ می‌باشند.

شکل ۷: ساختار کروموزوم در رویکرد پیشنهادی

۲-۳-۵- تعداد اعضای جامعه و مقداردهی اولیه

با توجه به فرایند الگوریتم ژنتیک، پس از تعیین کروموزوم بایستی جمعیت اولیه را مقداردهی نمود. تعداد جمعیت در زمان رسیدن به جواب‌های بهینه بسیار موثر است. اگر تعداد اولیه جمعیت کم باشد برنامه نوشته شده، برای رسیدن به جواب‌های بهینه نسل‌های بیشتری را تولید می‌کند. تعداد بالای جمعیت اولیه نیز باعث افزایش قابل توجه زمان محاسبات می‌شود. لذا پس از آزمون و خطا در این پژوهش از ۱۰۰ عضو برای جمعیت اولیه استفاده شده است.

پس از تعیین تعداد اعضای جامعه، برای مقداردهی اولیه می‌توان از هر مقداری استفاده نمود اما با توجه به اینکه در این پژوهش، هدف یافتن وزن‌های بهینه است و این وزن‌ها بایستی در مجموع مقدار واحد داشته باشند، برای مقداردهی اولیه بایستی مقادیر هرکدام از وزن‌ها در بازه $[0,1]$ باشد. در ادامه از بین اعضای جامعه اعضای برتر انتخاب شده و در تولیدمثل شرکت داده می‌شوند. در ادامه در زیربخش بعدی چگونگی انتخاب اعضای برتر جامعه و نحوه تولید فرزندان مطالب مربوطه ارائه خواهد شد.

۲-۳-۶- انتخاب، تقاطع و جهش

پس از اینکه برای تمام اعضای جمعیت اولیه مقدار *F-Score* محاسبه شد، اعضا بر اساس مقدار *F-Score* مرتب می‌شوند. اعضایی که دارای *F-Score* بالاتری هستند، در تولید نسل‌های بعدی شرکت داده می‌شوند. در این مقاله از ۲۰ درصد بالایی این جامعه مرتب شده برای تولید مثل استفاده می‌شود و فرزندان تولید شده جایگزین اعضایی از جامعه خواهند شد که پایین‌ترین مقادیر *F-Score* را داشته‌اند. در الگوریتم ژنتیک با انتخاب دو عضو از جامعه به عنوان والدین، دو فرزند تولید می‌شوند. در مرحله تقاطع، چگونگی تولید فرزندان از والدین تعیین می‌شود. به منظور اجرای تقاطع می‌توان از روش‌های تقاطع یک‌نقطه‌ای، تقاطع دونقطه‌ای و تقاطع یکنواخت بهره برد. در این پژوهش برای تقاطع از روش ترکیب یکنواخت استفاده شده است. نحوه وراثت فرزندان در شکل (۸) نشان داده شده‌است. همانطور که در شکل (۸) نیز مشخص است، در تقاطع یکنواخت احتمال اینکه وزن معیار از والد اول یا دوم اخذ شود برابر است. هرکدام از فرزندان پس از نرمال شدن به عنوان یک عضو جدید به جامعه اضافه می‌شوند. همچنین با توجه به نبود دانش قبلی نسبت به وزن‌های بهینه و به منظور تکمیل الگوریتم ژنتیک عملگر جهش نیز بایستی اعمال شود. با اجرای این عملگر، به جامعه آزادی عمل داده می‌شود تا با تولید وزن‌هایی خارج از روند تقاطع، اعضایی جدید و دارای تنوع بیشتری تولید شوند. در تعیین نرخ جهش باید توجه داشت که کوچک بودن این عدد باعث کاهش تنوع شده و عملا اعضای متنوعی تولید نخواهند شد. برعکس با بزرگ بودن مقدار نرخ جهش تنوع در اعضا افزایش یافته و باعث واگرا شدن جواب مساله می‌گردد. در رویکرد پیشنهادی عملگر جهش با نرخ ۰/۰۲ اعمال می‌شود.

شکل ۸: روش ترکیب یکنواخت در تقاطع

۲-۳-۷- همگرایی و توقف برنامه

شرط همگرایی برابر با تغییر نکردن بهترین جواب بعد از ۵۰۰ تکرار در نظر گرفته می‌شود. همچنین برای جلوگیری از تصادفی بودن جواب کل فرایند به تعداد ۱۰ بار اجرا می‌شود.

۲-۴- تناظریابی در منطقه تست

در منطقه تست هیچگونه بهینه‌سازی بر روی وزن‌ها و نتایج صورت نخواهد گرفت و نتایج به‌دست‌آمده دقیقا حاصل اعمال وزن‌های بهینه حاصل از تناظریابی در منطقه آموزشی هستند. روند اجرای تناظریابی در منطقه تست بدین صورت است که در ابتدا مقادیر شباهت هر پنج معیار برای تمامی عوارض محاسبه و ذخیره می‌شود، سپس برای هر عارضه از مجموعه داده مکانی نخست با هر عضو کاندید از مجموعه داده مکانی دوم یک بردار شباهت، دارای ۵ درایه ذخیره می‌گردد. برای اعمال وزن‌های بهینه بر این بردار با ضرب داخلی نمودن بردار وزن‌های بهینه در بردار شباهت‌های محاسبه شده یک مقدار عددی به عنوان مقدار شباهت کلی حاصل می‌شود. در مرحله بعد این مقدار با حد آستانه مقایسه شده و اگر مقدار شباهت کلی بین دو عارضه کمتر از حد آستانه نباشد، به عنوان دو عارضه متناظر تشخیص داده شده و در ماتریس روابط کلی ذخیره می‌شود.

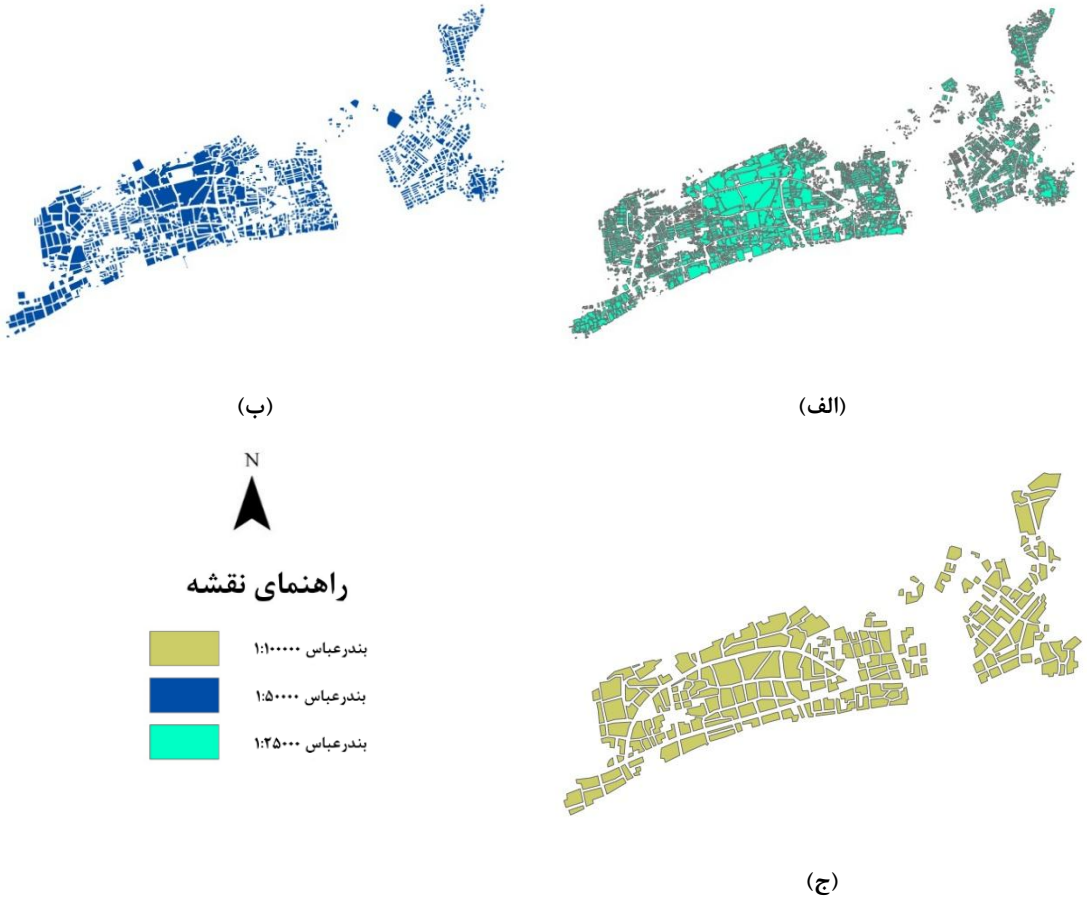
۲-۵- استخراج روابط متناظر

در آخرین گام از تناظریابی و برای اتمام آن بایستی روابط متناظر به دسته‌های شش‌گانه (یک به هیچ، هیچ

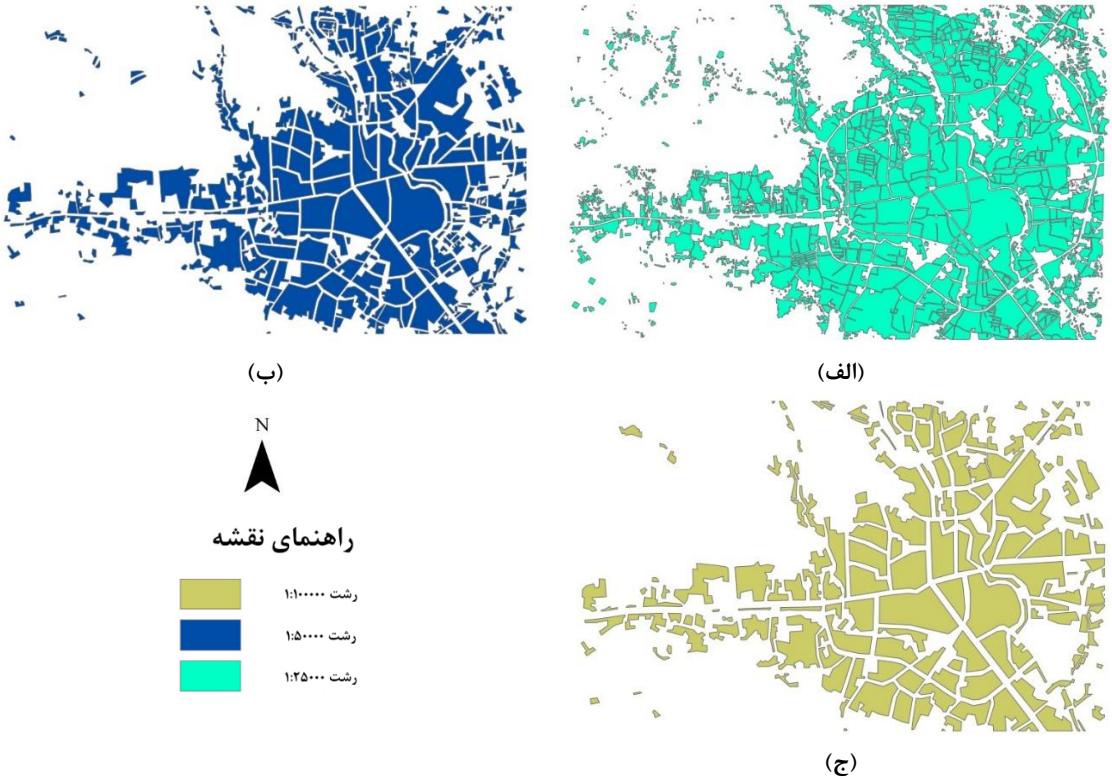
به یک، یک به یک، یک به چند، چند به یک و چند به چند) تفکیک شوند. با اجرای فرآیند تناظریابی، ماتریسی از شناسه‌های عوارض ذخیره می‌شود که به عنوان خروجی نهایی تناظریابی گزارش می‌شود. با استخراج و تفکیک روابط متناظر به دسته‌های شش‌گانه تناظریابی پایان یافته و می‌توان نتایج را مورد ارزیابی قرار داد.

۳- پیاده‌سازی

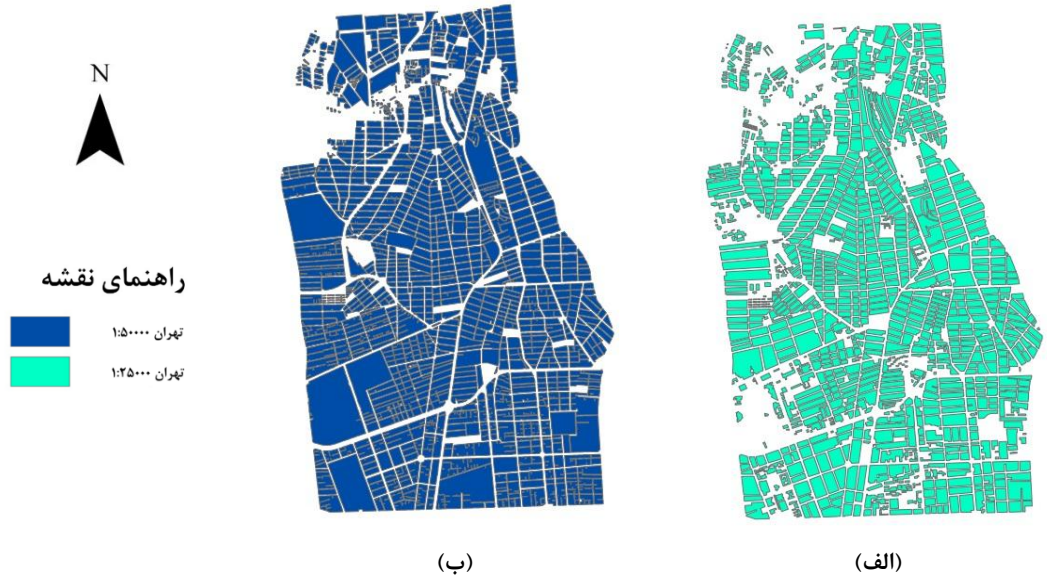
به منظور اجرای روش پیشنهادی، داده‌های مکانی مربوط به عوارض ساختمانی به عنوان عوارض چند ضلعی در سه مجموعه داده متفاوت آماده شده اند. مجموعه داده اول بخشی از شهر بندرعباس در مقیاس‌های ۱:۲۵۰۰۰، ۱:۵۰۰۰۰ و ۱:۱۰۰۰۰۰ (شکل (۹-الف)، شکل (۹-ب) و شکل (۹-ج))، مجموعه داده دوم بخشی از شهر رشت در سه مقیاس ۱:۲۵۰۰۰، ۱:۵۰۰۰۰ و ۱:۱۰۰۰۰۰ (شکل (۱۰-الف)، شکل (۱۰-ب) و شکل (۱۰-ج)) و مجموعه داده سوم بخشی از منطقه ۶ شهر تهران در مقیاس‌های ۱:۲۵۰۰۰ و ۱:۵۰۰۰۰ (شکل (۱۱-الف) و شکل (۱۱-ب)) می‌باشد.



شکل ۹: مجموعه داده‌های مکانی شهر بندرعباس: الف) بندرعباس در مقیاس ۱:۲۵۰۰۰، ب) بندرعباس در مقیاس ۱:۵۰۰۰۰، ج) بندرعباس در مقیاس ۱:۱۰۰۰۰۰



شکل ۱۰: مجموعه داده‌های مکانی شهر رشت: الف) رشت در مقیاس ۱:۲۵۰۰۰، ب) رشت در مقیاس ۱:۵۰۰۰۰، ج) رشت در مقیاس ۱:۱۰۰۰۰۰



شکل ۱۱: مجموعه داده‌های مکانی شهر تهران: الف) تهران در مقیاس ۱:۲۵۰۰۰، ب) تهران در مقیاس ۱:۵۰۰۰۰

۳-۱- پیش پردازش

مجموعه داده‌های مکانی در ابتدا از نظر قالب، سیستم‌مختصات و سیستم تصویر یکسان شده و عوارض دارای خطای توپولوژیک از مجموعه داده‌های مکانی حذف شدند. جدول (۱) تعداد عوارض هر مجموعه داده قبل و بعد از پیش‌پردازش را نشان می‌دهد. در ستون‌های بعدی این جدول، تعداد عوارض در دو منطقه آموزشی و تست آمده است. منطقه آموزشی بخشی از مجموعه داده‌هاست که برای عوارض موجود در این ناحیه تناظریابی به همراه بهینه‌سازی اجرا شده و وزن‌های بهینه برآورد می‌شوند. این وزن‌های برآورد شده در تناظریابی عوارض موجود در ناحیه تست اعمال شده و نتیجه تناظریابی بر اساس اعمال این وزن‌ها در منطقه تست، برآوردی از کیفیت تناظریابی ارائه می‌دهد.

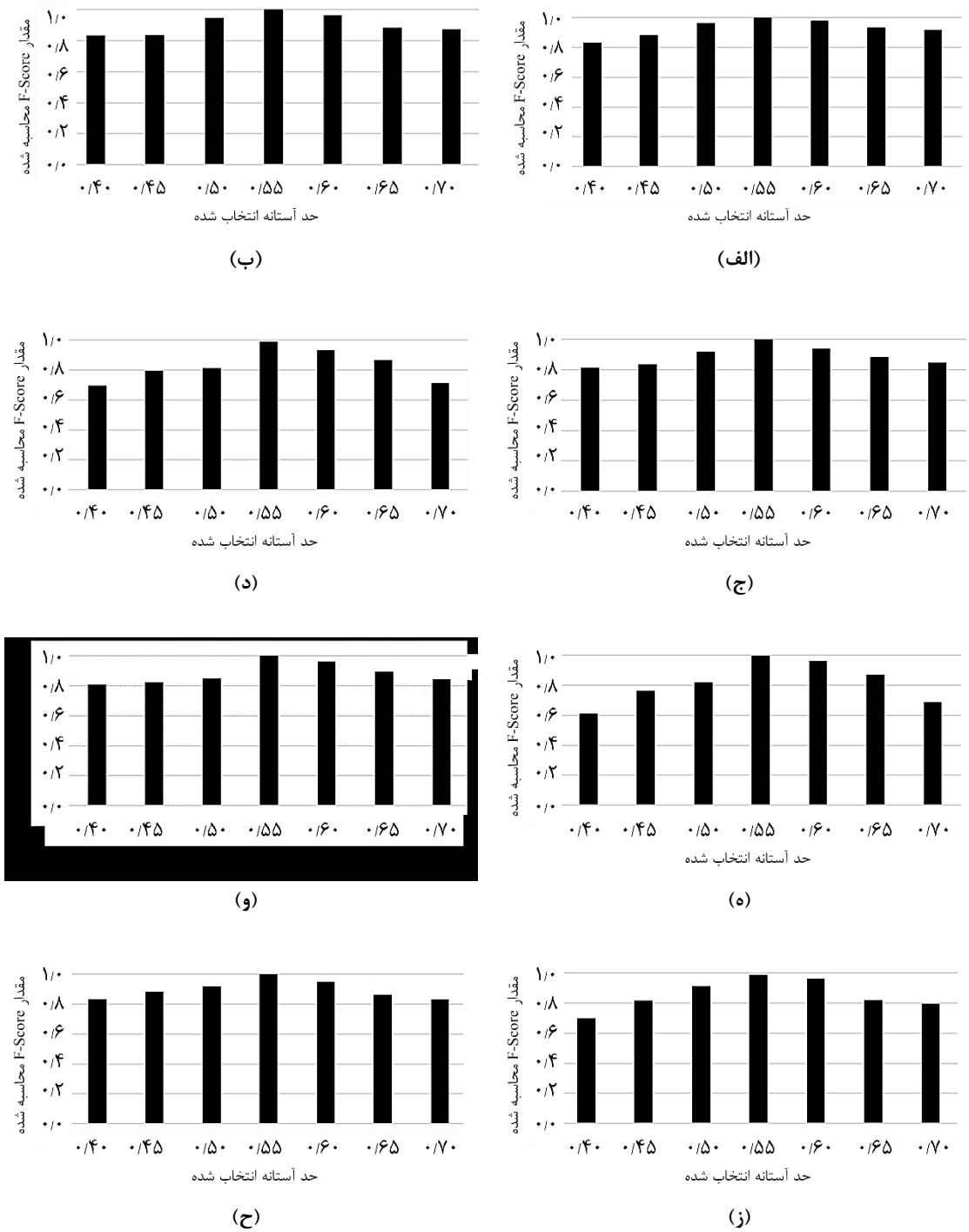
جدول ۱: تعداد عوارض در هر مجموعه داده، قبل و بعد از پیش‌پردازش و تفکیک تعداد عوارض در مناطق آموزشی و تست

مجموعه داده مکانی	مقیاس	تعداد عوارض		تعداد عوارض	
		قبل از پیش‌پردازش	بعد از پیش‌پردازش	در منطقه آموزشی	در منطقه تست
بندرعباس	۱:۲۵۰۰۰	۲۷۶۸	۲۷۶۳	۲۲۰۱	۵۶۲
	۱:۵۰۰۰۰	۱۱۶۰	۱۱۴۶	۹۱۶	۲۳۰
	۱:۱۰۰۰۰۰	۲۲۰	۲۲۰	۱۷۳	۴۷
رشت	۱:۲۵۰۰۰	۱۸۲۶	۱۸۲۰	۱۴۲۶	۳۹۴
	۱:۵۰۰۰۰	۳۶۱	۳۶۰	۲۸۳	۷۷
	۱:۱۰۰۰۰۰	۱۸۳	۱۸۳	۱۴۴	۳۹
تهران	۱:۲۵۰۰۰	۱۲۸۱	۱۲۷۵	۱۰۲۳	۲۵۲
	۱:۵۰۰۰۰	۹۴۶	۹۳۹	۷۶۸	۱۷۱

۳-۲- برآورد وزن معیارها در منطقه آموزشی

در نخستین گام مجموعه داده‌های مکانی به دو منطقه آموزشی و تست تقسیم شدند، در منطقه آموزشی که حدود ۷۰ درصد از عوارض کل مجموعه داده مکانی را داراست، تناظریابی در تکرارهای متعدد انجام می‌شود و در هر تکرار هر کروموزوم به عنوان وزن معیارها در مقدار شباهت هر معیار ضرب می‌شود .پس از محاسبه مقدار شباهت مکانی، این مقدار که درجه شباهت هندسی عوارض است، مبنای تشخیص متناظر بودن یا نبودن دو عارضه قرار می‌گیرد و با مقایسه آن با حد آستانه، می‌توان دو عارضه را متناظر یا نامتناظر دانست.

به منظور تعیین حد آستانه جهت تشخیص متناظر بودن یا نبودن دو عارضه، در بازه ۰/۴ تا ۰/۷ و با حد فاصل ۰/۰۵ حدود آستانه مختلف بررسی شدند و در هر مورد با محاسبه بیشترین مقدار *F-Score* به دست آمده، تاثیر مقدار حد آستانه بر تناظریابی مشخص گردید. در نهایت تناظریابی انجام‌شده با مقدار حد آستانه ۰/۵۵ دارای بیشترین مقدار *F-Score* گردید که این مقدار به عنوان حد آستانه نهایی انتخاب شد. نمودار تغییرات مقدار *F-Score* به ازای مقادیر مختلف حد آستانه در مجموعه داده‌های مکانی مختلف در شکل (۱۲) آمده است.



شکل ۱۲: تاثیر انتخاب حد آستانه بر مقدار *F-Score* نهایی در تناظریابی مجموعه داده‌ها، الف) از ۱:۲۵۰۰۰ به ۱:۵۰۰۰۰ منطقه ۶ شهر تهران ، ب) از ۱:۵۰۰۰۰ به ۱:۲۵۰۰۰ منطقه ۶ شهر تهران ، ج) از ۱:۲۵۰۰۰ به ۱:۵۰۰۰۰ شهر رشت ، د) از ۱:۲۵۰۰۰ به ۱:۱۰۰۰۰۰ شهر رشت ، هـ) از ۱:۵۰۰۰۰ به ۱:۱۰۰۰۰۰ شهر رشت ، و) از ۱:۵۰۰۰۰ به ۱:۲۵۰۰۰ شهر بندرعباس ، ز) از ۱:۱۰۰۰۰۰ به ۱:۲۵۰۰۰ شهر بندرعباس و ح) از ۱:۱۰۰۰۰۰ به ۱:۵۰۰۰۰ شهر بندرعباس

با اجرای الگوریتم پیشنهادی در ابتدا وزن‌های تصادفی در مقادیر شباهت محاسبه شده ضرب می‌شوند و در هر تکرار با توجه به مقدار *F-Score* محاسبه شده بهترین عضو جامعه انتخاب شده و نسل بعدی تولید می‌شود. به عنوان نمونه در شکل (۱۳) که بخشی از محدوده مورد مطالعه در این تناظریابی است، عارضه با شناسه ۰ از مجموعه داده شهر تهران با مقیاس ۱:۵۰۰۰۰ مبنا قرار گرفته شده و عوارض کاندید برای آن انتخاب می‌شوند. این عوارض که در این شکل با رنگ سبز قابل تشخیص هستند، در شعاع مشخصی از مرکز هندسی این عارضه قرار دارند. این شعاع با توجه به منطقه مطالعاتی تعیین می‌شود؛ در این مثال مقدار ۲۰۰ متر به عنوان محدوده بافر تعیین شده است. برای این عوارض معیارهای پنجگانه محاسبه شده و مقدار محاسبه شده برای این پارامترها به عنوان مقادیر شباهت وارد مرحله بعد می‌شوند. در ادامه تناظریابی و با بهینه‌سازی و برآورد وزن‌های بهینه، با توجه به بالاترین مقدار *F-Score* محاسب شده، در نهایت عوارض با شناسه‌های ۱۴ و ۱۵ از مجموعه داده ۱:۲۵۰۰۰ به عنوان عوارض متناظر با آن تشخیص داده شده و شناسه این عوارض در فایل رابطه یک به چند ذخیره می‌شوند. بخشی از مقادیر محاسبه شده شباهت بر مبنای هریک از معیارها برای این نمونه از تناظریابی در جدول (۲) آورده شده است. در ستون هشتم این جدول مقادیر شباهت هندسی با اعمال وزن‌های بهینه برآوردشده در آخرین تکرار نشان داده شده است.

شکل ۱۳: نمونه‌ای از تناظریابی و بهینه‌سازی وزن‌ها در منطقه آموزشی

جدول ۲: مقادیر شباهت معیارهای مختلف و مقدار شباهت کلی

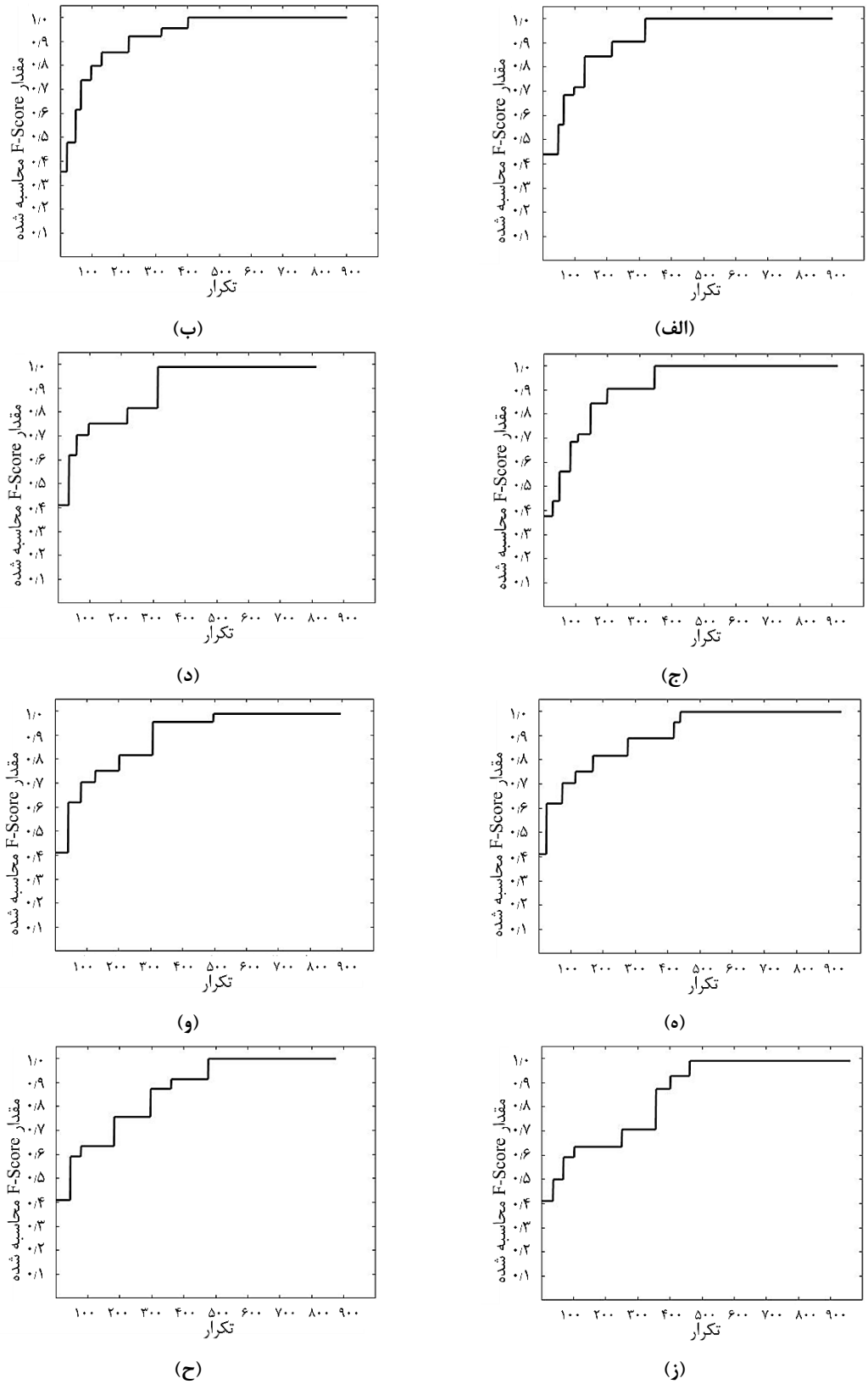
شناسه عارضه		مقادیر محاسبه شده شباهت برای معیارها					مقدار شباهت هندسی
در داده با مقیاس ۱:۵۰۰۰۰	در داده با مقیاس ۱:۲۵۰۰۰	مساحت هم‌پوشانی	فاصله اقلیدسی	راستای عارضه	فاصله هاسدورف	تابع چرخش	
۰	۱۲	۰	۰٫۴۱۷	۰٫۸۵۲	۰٫۳۱۶	۰٫۱۲۰	۰٫۱۸۹
۰	۱۳	۰	۰٫۹۳۷	۰٫۸۲۹	۰٫۷۷۳	۰٫۰۱۷	۰٫۳۶۷
۰	۱۴	۰٫۹۶۸	۰٫۹۹۸	۰٫۹۸۸	۱٫۰۰۰	۰٫۲۱۷	۰٫۶۹۲
۰	۱۵	۱٫۰۰۰	۰٫۹۹۶	۰٫۹۳۷	۰٫۹۴۹	۰٫۲۰۹	۰٫۹۷۵
۰	۱۶	۰	۰٫۹۵۹	۰٫۸۶۹	۰٫۹۳۲	۰٫۰۹۶	۰٫۳۸۲
۰	۱۷	۰	۰٫۷۹۷	۰٫۹۸۵	۰٫۲۶۸	۰٫۱۳۵	۰٫۳۱۸

در ستون آخر جدول(۲) مقادیر شباهت هندسی بین عارضه ۰ از مجموعه داده نخست با عوارض ۱۲ تا ۱۷ از مجموعه داده دوم نشان داده شده‌است. در مورد عوارض ۱۴ و ۱۵ از مجموعه داده دوم تناظریابی از حد آستانه (۰/۵۵) بالاتر بوده و این عوارض به عنوان عوارض متناظر با عارضه ۰ از مجموعه داده نخست تشخیص داده می‌شوند. پس از بررسی مقادیر شباهت بین عارضه ۰ و تمام عوارض کاندید از مجموعه داده دوم، عارضه ۱ مبنا قرار گرفته و مقادیر شباهت برای عوارض کاندید آن در مجموعه داده دوم محاسبه می‌شود. این مراحل به ازای تمام عوارض موجود در مجموعه داده نخست اجرا شده و مقادیر شباهت ناشی از انتخاب پنج معیار برای تمام عوارض محاسبه می‌شود. در هر مورد از تناظریابی با رویکرد پیشنهادی، بهینه‌سازی وزن معیارها تا جایی ادامه خواهد یافت که شرط ثابت ماندن بیشترین مقدار *F-Score* در ۵۰۰ تکرار برقرار شود. شکل (۱۴) نمودار همگرایی را در مناطق آموزشی مجموعه داده‌های مورد استفاده نشان می‌دهد.

در تناظریابی مجموعه داده‌ها پس از ثابت ماندن بیشترین مقدار *F-Score* در ۵۰۰ نسل، تولید نسل متوقف شده و وزن‌های بهینه برآورد شده آماده ارزیابی در منطقه تست خواهند بود. این مراحل در تمام مجموعه داده‌ها اعمال شده و نتایج آن که محاسبه وزن‌های بهینه در تناظریابی مجموعه داده‌های مکانی است که در جدول (۳) ارائه شده است. در ستون‌های چهارم تا هشتم جدول (۳) مقادیر نهایی وزن‌های بهینه نمایان است. در ستون نهم این جدول، مقادیر بیشینه بدست‌آمده برای *F-Score* در منطقه آموزشی نمایش داده شده است.

جدول ۳: وزن‌های بهینه برآورد شده و مقدار *F-Score* متناسب آنها در حالت‌های مختلف تناظریابی

منطقه	تناظریابی		وزن‌های بهینه برآورد شده برای معیارها					مقدار *F-Score*
	از	به	مساحت همپوشانی	فاصله اقلیدسی	راستای عارضه	فاصله هاسدورف	تابع چرخش	
تهران	۱:۵۰۰۰۰	۱:۲۵۰۰۰	۰٫۶۰۹	۰٫۲۳۹	۰٫۰۲۷	۰٫۰۳۹	۰٫۰۸۶	۰٫۹۹۹
	۱:۲۵۰۰۰	۱:۵۰۰۰۰	۰٫۵۸۳	۰٫۳۱۵	۰٫۰۰۷	۰٫۰۳۶	۰٫۰۵۹	۰٫۹۹۹
رشت	۱:۲۵۰۰۰	۱:۵۰۰۰۰	۰٫۵۷۳	۰٫۳۶۱	۰٫۰۰۶	۰٫۰۳۲	۰٫۰۲۸	۰٫۹۹۸
	۱:۲۵۰۰۰	۱:۱۰۰۰۰۰	۰٫۶۰۲	۰٫۳۱۲	۰٫۰۰۴	۰٫۰۰۲	۰٫۰۸۰	۰٫۹۸۸
	۱:۵۰۰۰۰	۱:۱۰۰۰۰۰	۰٫۵۶۱	۰٫۳۳۱	۰٫۰۱۵	۰٫۰۶۵	۰٫۰۲۸	۰٫۹۹۷
بندرعباس	۱:۵۰۰۰۰	۱:۲۵۰۰۰	۰٫۵۷۴	۰٫۳۲۶	۰٫۰۴۹	۰٫۰۲۷	۰٫۰۲۴	۰٫۹۸۸
	۱:۱۰۰۰۰۰	۱:۲۵۰۰۰	۰٫۵۵۷	۰٫۳۴۰	۰٫۰۴۶	۰٫۰۲۸	۰٫۰۲۹	۰٫۹۸۹
	۱:۱۰۰۰۰۰	۱:۵۰۰۰۰	۰٫۵۵۳	۰٫۳۱۴	۰٫۰۲۹	۰٫۰۳۸	۰٫۰۶۶	۰٫۹۹۸

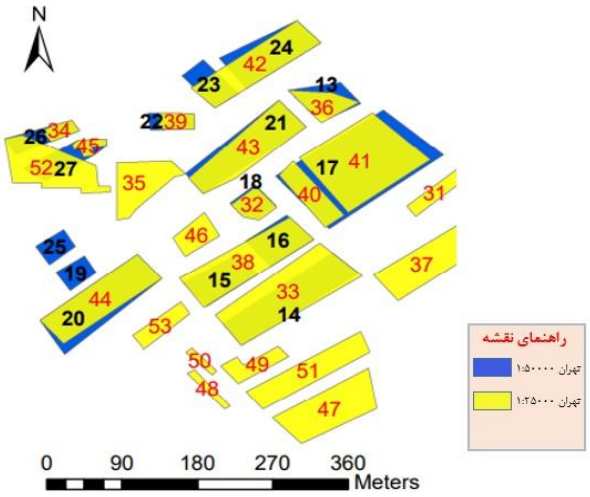


شکل ۱۴: نمودار همگرایی در منطقه آموزشی در تناظریابی مجموعه داده‌ها: الف) از ۱:۲۵۰۰۰ به ۱:۵۰۰۰۰ منطقه ۶ شهر تهران،

ب) از ۱:۵۰۰۰۰ به ۱:۲۵۰۰۰ منطقه ۶ شهر تهران ، ج) از ۱:۲۵۰۰۰ به ۱:۵۰۰۰۰ شهر رشت ، د) از ۱:۲۵۰۰۰ به ۱:۱۰۰۰۰۰ شهر رشت ، هـ) از ۱:۵۰۰۰۰ به ۱:۱۰۰۰۰۰ شهر رشت ، و) از ۱:۵۰۰۰۰ به ۱:۲۵۰۰۰ شهر بندرعباس ، ز) از ۱:۱۰۰۰۰۰ به ۱:۲۵۰۰۰ شهر بندرعباس و ح) از ۱:۱۰۰۰۰۰ به ۱:۵۰۰۰۰ شهر بندرعباس

۳-۳- اعمال وزن‌های بهینه در منطقه تست

به منظور ارزیابی نتایج وزن‌های بهینه برآورد شده از منطقه آموزشی در منطقه تست اعمال شده و بدین ترتیب می‌توان برآوردی از موفقیت رویکرد پیشنهادی در تناظریابی عوارض چندضلعی بدست آورد. در شکل(۱۵) نمونه‌ای از تناظریابی انجام شده در منطقه تست ارائه شده‌است. چنانکه در این شکل مشخص است، تناظریابی از مجموعه داده ۱:۲۵۰۰۰ تهران (عوارض به رنگ زرد و شناسه‌ها به رنگ قرمز) به مجموعه داده ۱:۵۰۰۰۰ تهران (عوارض به رنگ آبی و شناسه‌ها به رنگ مشکی) انجام شده و پس از اتمام تناظریابی، روابط شش‌گانه استخراج شده‌اند. در جدول(۴) نتیجه تفکیک روابط موجود در شکل (۱۳) گزارش شده است.



شکل ۱۵: بخشی از روابط تناظریابی استخراج شده در منطقه تست

جدول ۴: بخشی از روابط شش‌گانه استخراج شده بعد از تناظریابی از مجموعه داده ۱:۲۵۰۰۰ به مجموعه داده ۱:۵۰۰۰۰

شناسه در روابط استخراج شده											
یک به صفر		صفر به یک		یک به یک		یک به چند		چند به یک		چند به چند	
۳۱		۱۹		۳۲	۱۸	۴۲	۲۳	۴۰	۱۷	۳۴	۲۶
۳۵		۲۵		۳۳	۱۴		۲۴	۴۱		۴۵	۲۷
۴۶				۳۶	۱۳	۳۸	۱۵			۵۲	
۴۸				۳۹	۲۲		۱۶				
۴۹				۴۳	۲۱						
۵۰				۴۴	۲۰						
۵۱											

در جدول (۵) تعداد روابط متناظر تشخیص داده شده با اجـرای رویکـرد پیشـنهادی در حالـت‌هـای مختلـف تناظریابی در منطقه تسـت بـه دسـته‌هـای شـش‌گانـه تفکیک شده و همچنـین تعـداد روابـط تشـخیص داده نشده نیز آمده‌است. به عنوان نمونه در ردیـف اول ایـن جـدول تناظریـابی از مجموعـه داده ۱:۵۰۰۰۰ بـه مجموعـه داده ۱:۲۵۰۰۰ اجـرا شـده و نتـایج ایـن تناظریابی که تفکیک روابط بـه دسـته‌هـای شـش‌گانـه می‌باشد در ستون‌های چهارم تا هشـتم جـدول نمـایش داده شده است. به علاوه در سـتون دهـم ایـن جـدول، تعداد روابط متناظری که پـس از تناظریـابی تشـخیص داده نشده‌است (*FN*) و در ردیف یـازدهم تعـداد روابـط متناظری که به صورت اشتباه تشخیص داده شده (*FP*) آورده شده‌اند. در ستون‌های بعدی این جدول به ترتیب مقدار *Precision*، *Recall* و *F-Score* متناسـب بـا هـر مورد از تناظریابی ثبت شده است.

جدول ۵: تعداد روابط صحیح و اشتباه تشخیص داده شده در منطقه تست توسط روش پیشنهادی

مجموعه داده	تناظریابی		تعداد روابط صحیح						*FN*	*FP*	*Precision*	*Recall*	*F-Score*
	از	به	۱:۰	۰:۱	۱:۱	۱:*M*	*M*:۱	*M*:*N*					
تهران	۱:۵۰۰۰۰	۱:۲۵۰۰۰	۲۴	۶	۱۵۷	۱۲	۴۳	۱۵۸	۷	۱۲	۰٫۹۷۱	۰٫۹۸۳	۰٫۹۷۷
	۱:۲۵۰۰۰	۱:۵۰۰۰۰	۲۶	۸	۱۷۱	۷	۵۴	۱۵۷	۹	۱۵	۰٫۹۶۶	۰٫۹۷۹	۰٫۹۷۲
رشت	۱:۲۵۰۰۰	۱:۵۰۰۰۰	۲۳	۱۲	۶۱	۲۱	۱۴	۱۹۸	۱۲	۸	۰٫۹۷۶	۰٫۹۶۵	۰٫۹۷۱
	۱:۲۵۰۰۰	۱:۱۰۰۰۰۰	۱۶	۳	۳۲	۴	۳۳	۹۴	۶	۹	۰٫۹۵۳	۰٫۹۶۸	۰٫۹۶۰
	۱:۵۰۰۰۰	۱:۱۰۰۰۰۰	۹	۳	۴۴	۱۴	۱۷	۵۳	۳	۴	۰٫۹۷۲	۰٫۹۷۹	۰٫۹۷۶
بندرعباس	۱:۵۰۰۰۰	۱:۲۵۰۰۰	۱۶	۱۴	۶۱	۱۲	۲۱	۲۰۱	۱۰	۱۲	۰٫۹۶۴	۰٫۹۷۰	۰٫۹۶۷
	۱:۱۰۰۰۰۰	۱:۲۵۰۰۰	۴	۷	۲۷	۳۵	۷	۹۸	۴	۱۱	۰٫۹۴۲	۰٫۹۷۸	۰٫۹۵۹
	۱:۱۰۰۰۰۰	۱:۵۰۰۰۰	۴	۱۱	۴۸	۱۸	۱۵	۵۷	۷	۲	۰٫۹۸۷	۰٫۹۵۶	۰٫۹۷۱

۳-۴- ارزیابی نتایج

به منظور مقایسـه نتـایج حاصـل از پیـاده‌سـازی روش پیشـنهادی، دو ارزیـابی انجـام مـی‌پـذیرد. در ارزیـابی نخست، نتایج حاصـل از ایـن روش بـا حـالتی مقایسـه می‌شود که برای تمـامی معیارهـا وزن یکسـان در نظـر گرفته می‌شود. ارزیابی دیگـر مقایسـه نتـایج حاصـل از پیاده‌سازی روش پیشنهادی با رویکرد ارائه شده توسـط چمنی و همکاران می‌باشد [۳۷]. در پـژوهش چمنـی و همکـاران از چهـار معیـار فاصـله اقلیدسـی، مسـاحت هم‌پوشانی، راستای عوارض و شباهت شـکل دو عارضـه (تابع چرخش) استفاده شده و به ایـن چهـار معیـار بـه ترتیـب وزن‌هـای ۰٫۴۵، ۰٫۷۳، ۰٫۵۰ و ۰٫۶۲ اعمـال شده‌اسـت. جـدول (۶) نتـایج حاصـل از مقایسـه روش پیشنهادی با دو حالت دیگر نشان داده شده است. نتایج این جدول نشان داد تناظریـابی بـا رویکـرد پیشـنهادی نسبت به حـالتی کـه تمـام معیارهـا بـا وزن برابـر وارد تناظریابی شوند بـه مقـدار ۲۸٫۶۱ درصـد و نسـبت بـه رویکرد ارائه شده توسط چمنـی و همکـاران بـه مقـدار ۹٫۱۳ درصد بهبود می‌یابد [۳۷].

جدول ۶: مقایسه مقدار *F-Score* در دو حالت: با وزن برابر معیارها و با استفاده از وزن‌های بهینه

ردیف	تناظریابی		منطقه	مقدار *F-Score* محاسبه شده با		
	از	به		وزن برابر معیارها	چمنی و همکاران [۳۷]	رویکرد پیشنهادی
الف	۱:۲۵۰۰۰	۱:۵۰۰۰۰	تهران	۰٫۷۱۲	۰٫۸۹۲	۰٫۹۷۲
ب	۱:۲۵۰۰۰	۱:۵۰۰۰۰	رشت	۰٫۷۲۹	۰٫۸۸۸	۰٫۹۷۱
ج	۱:۲۵۰۰۰	۱:۱۰۰۰۰۰	رشت	۰٫۶۴۳	۰٫۸۴۹	۰٫۹۶۰
د	۱:۵۰۰۰۰	۱:۱۰۰۰۰۰	رشت	۰٫۶۹۸	۰٫۹۰۱	۰٫۹۷۶
هـ	۱:۵۰۰۰۰	۱:۲۵۰۰۰	بندرعباس	۰٫۶۷۱	۰٫۸۹۹	۰٫۹۶۷
و	۱:۵۰۰۰۰	۱:۲۵۰۰۰	تهران	۰٫۶۹۲	۰٫۹۱۹	۰٫۹۷۷
ز	۱:۱۰۰۰۰۰	۱:۲۵۰۰۰	بندرعباس	۰٫۶۱۷	۰٫۸۴۱	۰٫۹۵۹
ح	۱:۱۰۰۰۰۰	۱:۵۰۰۰۰	بندرعباس	۰٫۷۰۲	۰٫۸۶۱	۰٫۹۷۱

۴- نتیجه‌گیری و پیشنهادها

در این تحقیق رویکردی برای تناظریابی عوارض چندضلعی در مجموعه داده‌های چندمقیاسی با اندازه‌گیری معیارهای مختلف مقایسه شباهت بین عوارض ارائه شد. در این رویکرد برای استفاده همزمان از همه معیارها از الگوریتم ژنتیک به عنوان یک الگوریتم بهینه‌سازی کمک گرفته شده و مساله تناظریابی به عنوان یک مساله بهینه‌سازی در نظر گرفته شد. معیار ارزیابی تناظریابی در این تحقیق *F-Score* بود که نسبت به سایر معیارها از جمله *Precision* و *Recall* برتری دارد.

معیارهای شباهت هندسی به مقیاس مجموعه داده‌های مکانی وابسته‌اند؛ زیرا مستقیما با هندسه عوارض در ارتباط بوده و یک عارضه واحد ممکن است در مجموعه داده‌های با مقیاس‌های متفاوت، هندسه متفاوتی داشته باشد. در تحقیق حاضر مقدار تاثیر پنج معیار هندسی مساحت هم‌پوشانی دو عارضه، فاصله اقلیدسی بین دو عارضه، اختلاف راستای دو عارضه، فاصله هاسدورف بین دو عارضه و همچنین اختلاف تابع چرخش دو عارضه در تناظریابی مجموعه داده‌های مختلف بررسی شد. به منظور تاثیر همزمان همه این پنج معیار، ترکیب وزن‌داری از مقدار شباهت محاسبه شده از هرکدام از این پنج معیار تولید شد که مقدار وزن هر معیار به کمک الگوریتم ژنتیک و در نسل‌های مختلف بهبود یافته و نهایتاً وزن‌هایی که منجر به بالاترین مقدار *F-Score* شدند به عنوان بهترین وزن‌ها در تناظریابی هر مجموعه داده استخراج شدند.

نتایج نشان داد که استفاده از الگوریتم ژنتیک در شناسایی مقدار تاثیر هر معیار در بهترین تناظریابی عملکرد خوبی داشته است و در منطقه تست که فقط از وزن‌های بهینه شده برای تناظریابی استفاده شد نتایج قابل قبولی حاصل شد. از آنجا که در منطقه تست هیچگونه بهینه‌سازی اعمال نشده، هیچ دانش قبلی از عوارض و ارتباط بین عوارض در مجموعه داده‌های مختلف وجود نداشته و تنها از وزن‌های استخراج شده از منطقه آموزشی برای تعیین وزن هر معیار در مقدار شباهت کلی استفاده شده است. در این تحقیق با ارزیابی مجموعه داده‌های گوناگون به طور میانگین

مقدار *F-Score* به اندازه ۲۸٫۶۱ درصد نسبت به اعمال وزن برابر برای معیارها و همچنین ۹٫۱۳ درصد نسبت به رویکرد ارائه شده توسط چمنی و همکاران بهبود یافته است [۳۷].

نتایج تحقیق نشان می‌دهد بیشترین تاثیر را در تناظریابی بهینه معیارهای مساحت هم‌پوشانی و فاصله اقلیدسی دارند و سایر معیارها از وزن به مراتب کمتری در تناظریابی برخوردار هستند. همچنین در این تحقیق پس از تناظریابی و تشخیص عوارض متناظر، روابط متناظر بین عوارض که در شش دسته کلی تقسیم‌بندی می‌شوند استخراج شده و نهایتا عوارض متناظر در دسته‌های یک به هیچ، هیچ به یک، یک به یک، یک به چند، چند به یک و چند به چند قرار گرفتند. از جمله نقاط قوت رویکرد پیشنهادی توانایی تشخیص و تفکیک تمامی شش نوع روابط متناظر می‌باشد.

به منظور ارائه پیشنهاداتی برای تحقیقات آتی در حوزه‌های مشابه می‌توان استفاده از سایر روش‌های بهینه‌سازی را پیشنهاد نمود. در این صورت می‌توان نظیر آنچه در تحقیق حاضر و با بهره بردن از الگوریتم ژنتیک اجرا شد، مساله تناظریابی را به یک مساله بهینه‌سازی تبدیل نموده و با تعریف پارامترهای مورد نیاز (ازجمله تابع هزینه) نسبت به بهینه‌سازی تناظریابی اقدام نمود.

همچنین می‌توان با بهره‌گیری از آنالیز حساسیت نسبت به تعیین میزان تاثیر هریک از معیارهای هندسی در نتایج تناظریابی اقدام نمود و وابستگی تناظریابی به هریک از معیارها را برآورد نمود. بدین ترتیب می‌توان مهمترین معیار(های) اندازه‌گیری شباهت را تعیین نمود.

مراجع

[1] Wang, Y., et al., "A back - propagation neural network - based approach for multi - represented feature matching in update propagation" , Transactions in GIS, Vol.19, pp. 964-993, 2015.

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

[3] Lei, T. and Z. Lei, "Optimal spatial data matching for conflation: A network flow - based approach" , Transactions in GIS, Vol.23, pp. 1152-1176, 2019.

[4] Fan, H., et al., "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.

[5] Fan, H., et al., "Quality assessment for building footprints data on OpenStreetMap", International Journal of Geographical Information Science, Vol.28, pp. 700-719, 2014.

[6] Wang, Y., et al., "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: Springer, 2014, pp. 139-153.

[7] Tong, X., 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.

[8] Kim, I.-H., C.-C. Feng, and Y.-C. Wang, "A simplified linear feature matching method using decision tree analysis, weighted linear directional mean, and topological relationships", International Journal of Geographical Information Science, Vol.31, pp. 1042-1060, 2017.

[9] Tong, X., 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.

[10] Chehreghan, A. 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.

[11] Chehreghan, A. and R. Ali Abbaspour, "A geometric-based approach for road matching on multi-scale datasets using a genetic algorithm", Cartography and Geographic Information Science, Vol.45, pp. 255-269, 2018.

[12] Olteanu-Raimond, A.-M., S. Mustiere, and A. Ruas, "Knowledge formalization for vector data matching using belief theory", Journal of Spatial Information Science, pp. 21-46, 2015.

[13] Farahanipooya, A., et al., "Roads matching in a multi-scale spatial database using a least square line", Journal of Geomatics Science and Technology, Vol.3, pp. 87-104, 2013.

[14] Li, L. and M. Goodchild, "Automatically and accurately matching objects in geospatial datasets", in Adv. Geo-Spat. Inf. Science: 2012, pp. 71-79.

[15] Zhang, X., et al., "A multi-scale residential areas matching method using relevance vector machine and active learning", ISPRS International Journal of Geo-Information, Vol.6, p. 70, 2017.

[16] Fu, Z., et al., "A moment-based shape similarity measurement for areal entities in geographical vector data", ISPRS International Journal of Geo-Information, Vol.7, p. 208, 2018.

[17] Wang, Y., et al., "A PSO-neural network-based feature matching approach in data integration", in Cartography-Maps connecting the world: Springer, 2015, pp. 189-219.

[18] Du, H., et al., "A method for matching crowd - sourced and authoritative geospatial data", Transactions in GIS, Vol.21, pp. 406-427, 2017.

[19] Abdolmajidi, E., et al., "Matching authority and VGI road networks using an extended node-based matching algorithm", Geo-Spatial Information Science, Vol.18, pp. 65-80, 2015.

[20] Yang, B., 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.

[21] Zhang, M. and L. Meng, "Delimited stroke oriented algorithm-working principle and implementation for the matching of road networks", Geographic Information Sciences, Vol.14, pp. 44-53, 2008.

[22] Kieler, B., et al., "Matching river datasets of different scales", in Advances in GIScience: Springer, 2009, pp. 135-154.

[23] Hastings, J., "Automated conflation of digital gazetteer data", International Journal of Geographical Information Science, Vol.22, pp. 1109-1127, 2008.

[24] Harrie, L. and A. Hellstrom, "A case study of propagating updates between cartographic data sets", presented at 19th International Cartographic Conference, 11th General Assembly of ICA, Ottawa. 1999.

[25] Levenshtein, V.I., "Binary codes capable of correcting deletions, insertions, and reversals". Soviet physics doklady: Soviet Union, Vol.10, pp.707-710, 1966.

[26] Huang, L., et al., "Feature matching in cadastral map integration with a case study of Beijing", presented at 18th International Conference on Geoinformatics, 2010.

[27] Kim, J. and K. Yu, "Areal feature matching based on similarity using CRITIC method", The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol.40, p. 75, 2015.

[28] Zhonglianga, F. and W. Jianhuaa, "Entity matching in vector spatial data", presented at XXI ISPRS Congress. 2008.

[29] Mandal, S.N., J.P. Choudhury, and S.B.

and Technology, Vol.7, pp. 73-87, 2018.

Chaudhuri, "In search of suitable fuzzy membership function in prediction of time series data", International Journal of Computer Science Issues, Vol.9, pp. 293-302, 2012.

[30] Wenjing, T., et al., "Research on areal feature matching algorithm based on spatial similarity", presented at Chinese control and decision conference, 2008.

[31] Rucklidge, W., Efficient visual recognition using the Hausdorff distance. Springer-Verlag, 1996.

[32] Huh, Y., K. Yu, and J. Heo, "Detecting conjugate-point pairs for map alignment between two polygon datasets", Computers, Environment and Urban Systems, Vol.35, pp. 250-262, 2011.

[33] Ying, S., et al., "Probabilistic matching of map objects in multi-scale space", presented at 25th International Cartographic Conference, 2011.

[34] Zhang, X., et al., "Pattern classification approaches to matching building polygons at multiple scales", in ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume I-2, XXII ISPRS Congress, International Society for Photogrammetry and Remote Sensing, pp. 19-24, 2012.

[35] Yue, H., et al., "A Multi-Scale Settlement Matching Algorithm Based on ARG", International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, Vol.41, 2016.

[36] Arkin, E.M., et al., "An efficiently computable metric for comparing polygonal shapes", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.13, 1991.

[37] Chamani, M., R. Ali Abbaspour, and A.R. Chehreghan, "Matching of Polygon Objects Based on Geometric Measures in a Multi-Scale Dataset", Journal of Geomatics Science

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

Journal of Geospatial Information Technology
Vol.9 No.3, Autumn 2021

Research Paper

Polygon Objects Matching by Optimizing Geometric Criteria

Ali Moeini Roudbali[1], Rahim Ali Abbaspour [2]*, Alireza Chehreghan [3]

1- MSc Student of GIS, School of Surveying and Geospatial Information Engineering, College of Engineering, University of Tehran, Tehran, Iran.
2- Associate professor, School of Surveying and Geospatial Information Engineering, College of Engineering, University of Tehran, Tehran, Iran.
3- Assistant professor, Mining Engineering Faculty, Sahand University of Technology, Tabriz, Iran.

Abstract

Despite the semantic criteria, geometric criteria have different performances on polygon feature matching in different vector datasets. By using these criteria for measuring the similarity of the two polygons in all matchings, the same results would not have been obtained. In order to achieve the best matching results, determination of optimal geometric criteria for each dataset is considered to be necessary. In the previous researches, the most used geometric criteria are the overlapping area and the Euclidian distance between the two features, the orientation difference and the shape similarity of the two features. In addition to determining the impact factor of each criterion in the best result, the best geometric criteria combination should be specified. In this study, unlike the previous ones which have considered objects matching as a unique issue in all datasets, objects matching is considered as a separate issue in each dataset and by converting the problem into an optimization one, an approach is proposed to define optimal weights of criteria for different datasets using a genetic algorithm. In each dataset, the best corresponding weights are distinguished that lead to the best matching result. To evaluate the proposed approach, a variety of spatial datasets of residential buildings have been used including a part of Bandar Abbas city in 1:25000, 1:50000, and 1:100000 scales; a part of district 6 of Tehran city in 1:25000 and 1:50000 scales; and a part of Rasht city in 1:25000, 1:50000, and 1:100000 scales. The results showed that the proposed approach has done a good performance in both polygon feature matching and identifying six corresponding relationship classes in all study areas. Moreover, matching results have been improved by an average of 28.61% compared to the case where all criteria are considered with equal weights and an average of 9.13% compared to the case that criteria are assessed according to the expert opinions.

Key words: *Polygon Feature Matching, Geometric Criteria, Optimization, Genetic Algorithm.*

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

[DOR: 20.1001.1.20089635.1400.9.3.5.6]
[DOI: 10.52547/jgit.9.3.1]