

مقایسه روش‌های تناظریابی عوارض چندضلعی در داده‌های برداری چندنمایشی

مژگان چمنی^۱، رحیم علی‌عباسپور^{۲*}، علیرضا چهرقان^۳

^۱ کارشناس ارشد سیستم‌های اطلاعات مکانی - دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی - پردیس
دانشکده‌های فنی - دانشگاه تهران
m.chamani@ut.ac.ir

^۲ استادیار دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی - پردیس دانشکده‌های فنی - دانشگاه تهران
abaspour@ut.ac.ir

^۳ دکتری سیستم‌های اطلاعات مکانی - دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی - پردیس دانشکده‌های فنی
- دانشگاه تهران
achehrehghan@ut.ac.ir

(تاریخ دریافت فروردین ۱۳۹۶، تاریخ تصویب آبان ۱۳۹۶)

چکیده

با افزایش کاربرد داده‌های مکانی در زندگی روزمره، تولید این داده‌ها از منابع اطلاعاتی مختلف با دقت و مقیاس‌های متفاوت رشد وسیعی داشته است. در این مجموعه داده‌های متنوع در برخی از کاربردها، نیاز به شناسایی عوارض یکسان تحت عنوان مسائل تناظریابی می‌باشد. به عبارتی دیگر در مسائل تناظریابی، هدف شناسایی عوارض مشابه و یکسان در دنیای واقعی است که در مجموعه داده‌های متفاوت ممکن است با مقیاس، دقت، هندسه و رابطه مکانی متفاوتی نمایش داده شده باشند. کاربرد مسائل تناظریابی در حوزه‌های مختلف مانند ادغام داده‌ها و یا غنی‌سازی داده‌های معتبر، نشان‌دهنده اهمیت بالای این مسائل در زمینه‌های مختلف علوم مکانی می‌باشد. مسائل تناظریابی در مجموعه داده‌های برداری از منظر نوع عارضه تناظریابی شده به سه دسته نقطه‌ای، خطی و چندضلعی تقسیم‌بندی می‌شوند که اکثر مطالعات انجام شده در گذشته بر روی بررسی عوارض خطی می‌باشد. از این رو در این تحقیق مطالعات بر روی عوارض چندضلعی بررسی شده است که دارای پیچیدگی‌های به مراتب بیشتری نسبت به عوارض خطی هستند و بیشتر مطالعات در سال‌های اخیر به این دسته معطوف شده است. در این مقاله سعی شده است با بررسی دقیق و جامع مطالعات پیشین، مسائل تناظریابی عوارض چندضلعی از سه منظر معیارهای اندازه‌گیری شباهت، الگوریتم‌های تناظریابی و روابط متناظر بین عوارض بررسی شوند. همچنین دسته‌بندی‌ها در هر دیدگاه و مزایا و معایب روش‌های مختلف ارائه شده است.

واژگان کلیدی: تناظریابی، عوارض چندضلعی، شباهت عوارض، روابط عوارض متناظر، چندنمایشی

۱- مقدمه

امروزه با رشد سریع فناوری بویژه در حوزه اطلاعات و ارتباطات، تولید مجموعه داده‌های مکانی مختلف از منابع متفاوت در حال افزایش است. در بین مجموعه داده‌های مختلف می‌توان به محبوبیت مجموعه داده‌های مکانی داوطلبانه (VGI) در کنار مجموعه داده‌های معتبر و همچنین گسترش پیاده‌سازی زیرساخت اطلاعات مکانی (SDI) اشاره نمود [۱]. در این میان با توجه به تنوع منابع تولید مجموعه داده‌های مکانی، ممکن است یک عارضه یکسان در دنیای واقعی در مجموعه داده‌های مختلف با روابط مکانی، معنایی، مقیاس و هندسه متفاوتی نمایش داده شود. به عبارت دیگر، مجموعه داده‌های مختلف از دنیای واقعی نشان‌دهنده عوارض یکسان در مقیاس‌ها و دقت‌های مختلفی هستند؛ از این رو نیاز به شناسایی این عوارض مشابه و کشف شباهت بین آن‌ها تحت عنوان مسائل تناظریابی می‌باشد که در کاربردهای متنوعی استفاده می‌شود. این کاربردها را می‌توان در چهار حوزه ادغام داده‌ها، غنی‌سازی داده‌های مکانی معتبر، ارزیابی کیفیت و مدیریت داده‌ها دسته‌بندی نمود.

حوزه یکپارچه‌سازی و ادغام داده‌های مکانی ناهمگن شامل تلفیق، به‌اشتراک‌گذاری و اتصال می‌باشد. به کمک ادغام داده‌ها می‌توان اطلاعات مجموعه داده‌های مختلف را ترکیب نمود و اطلاعات جدیدی را که تنها از یک منبع داده قابل استخراج نمی‌باشد، استخراج کرد. همچنین مطالعات با هدف ادغام بر روی حفظ داده‌های صحیح، حذف افزونگی داده و کاهش تناقضات بین مجموعه داده‌ها تمرکز دارند که یکی از مراحل مهم در این فرآیندها، تناظریابی اشیاء مکانی یکسان می‌باشد [۲-۴].

در حوزه دوم که اخیراً مورد توجه قرار گرفته است، هدف غنی‌سازی و بروزرسانی داده‌های مکانی معتبر به کمک داده‌های مکانی داوطلبانه است. داده‌های مکانی داوطلبانه معمولاً توسط شهروندان داوطلب جمع‌آوری می‌شوند و یا محصول جانبی یک برنامه کاربردی با هدفی متفاوت هستند. این داده‌های داوطلبانه در مقایسه با داده‌های معتبر، دارای دقت هندسی پایین‌تر، ساختار ضعیف و فاقد فراداده‌های مرتبط هستند؛ با وجود این، تغییرات دنیای واقعی را به سرعت منعکس می‌کنند و هزینه اخذ پایینی دارند. بدین ترتیب استفاده از داده‌های

مکانی معتبر و داوطلبانه که بتوانند مکمل یکدیگر باشند و تصویری غنی‌تر، به‌روزتر و کامل‌تر از داده‌های مکانی را ارائه دهند، مطلوبتر است [۵-۷].

کاربرد مسائل تناظریابی در حوزه ارزیابی کیفیت شامل شناسایی ناسازگاری، بررسی صحت موقعیت، صحت شکل، کامل بودن داده‌ها و صحت موضوعی است [۸، ۱]. در این حوزه، مجموعه داده‌هایی همانند داده‌های مکانی داوطلبانه که دارای دقت به نسبت پایین‌تری هستند با یک مجموعه داده مرجع با دقت بالاتر مقایسه می‌شوند و پس از شناسایی عوارض مشابه به کمک تناظریابی، پارامترهای مختلف ارزیابی کیفیت برای آن‌ها محاسبه می‌شود.

کاربرد مسائل تناظریابی در حوزه مدیریت داده‌ها شامل مطالعات انتشار به‌روزرسانی و شناسایی تغییرات می‌باشد [۱]. انتشار به‌روزرسانی یکی از روش‌های کارآمد در بروزرسانی مجموعه داده‌های چندمنظمی است که به‌روزرسانی خودکار یا نیمه‌خودکار از مجموعه داده با جزئیات بیشتر به مجموعه داده با جزئیات کمتر انتشار می‌یابد [۹]. کلید اصلی در فرآیند انتشار به‌روزرسانی، ارتباط عوارض مشابه می‌باشد که به کمک الگوریتم‌های تناظریابی شناسایی می‌شود [۱۰، ۱۱].

در تمام یا بخشی از فرآیند کاربردهای ذکر شده، نیاز به حل مسائل تناظریابی می‌باشد. تناظریابی داده‌های مکانی به عنوان فرآیند شناسایی عوارض مشابه بین مجموعه داده‌های مختلف تعریف می‌شود. این عوارض مشابه نشان‌دهنده یک موجودیت یکسان در دنیای واقعی هستند، اما به دلایل مختلفی همچون اشتباهات ناگزیر در تولید نقشه و یا تنوع مراکز تولید نقشه با دقت‌های مختلف، دارای موقعیت و شکل متفاوتند. به عبارت دیگر، یک عارضه یکسان در دنیای واقعی ممکن است در مجموعه داده‌های مختلف با روابط مکانی، معنایی، مقیاس و هندسه متفاوتی نمایش داده شود [۱۱]. در نتیجه نیاز به استفاده از راه‌کارهایی جهت شناسایی عوارض مشابه در مجموعه داده‌های مختلف می‌باشد.

تناظریابی نقشه‌ها به دو گروه مجزای رستری و بُرداری تقسیم‌بندی می‌شود که در این مقاله تمرکز کامل بر روی حوزه بُرداری می‌باشد و مباحث رستری که بیشتر در رابطه با حوزه فتوگرامتری و سنجش از دور است بررسی نمی‌شود. از طرفی تناظریابی عوارض در مدل داده بُرداری بر اساس نوع عارضه منطبق شده به سه دسته نقطه‌ای، خطی و چندضلعی دسته‌بندی می‌شود [۱۱].

شناسایی روابط متناظر عوارض چندضلعی بین دو یا چند مجموعه داده متفاوت است که به عنوان خروجی الگوریتم‌های تناظریابی مطرح می‌شود.

۲- معیارهای اندازه‌گیری شباهت

در الگوریتم‌های تناظریابی عوارض چندضلعی به مطالعه شباهت‌ها (عدم شباهت‌ها) بین دو چندضلعی نیاز است. در بررسی میزان شباهت دو عارضه از معیارهای مختلف اندازه‌گیری شباهت کمک گرفته می‌شود که این معیارها با توجه به نوع عارضه تناظریابی شده و در مطالعات مختلف در دسته‌بندی‌های متفاوتی قرار دارند. به عنوان نمونه در مطالعات Samal و همکاران [۲] این معیارها به دو دسته وابسته به زمینه^۱ و مستقل از زمینه تقسیم‌بندی شده‌اند. در این مقاله معیارهای اندازه‌گیری شباهت دو عارضه چندضلعی بر اساس مطالعات Du و همکاران [۷] و Wang و همکاران [۴] در سه دسته هندسی، توصیفی (معنایی) و روابط مکانی قرار دارند. شکل ۲ دسته‌بندی در نظر گرفته شده در این تحقیق را نشان می‌دهد.



شکل ۲- دسته‌بندی معیارهای اندازه‌گیری شباهت

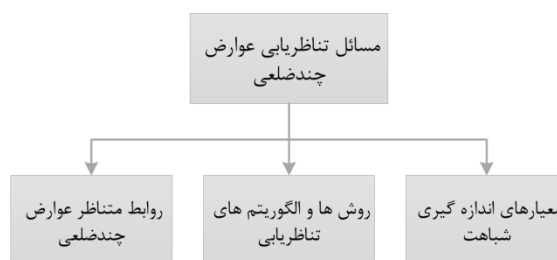
در معیارهای شباهت هندسی از عوامل مربوط به هندسه و شکل عارضه چندضلعی مانند مساحت و محیط عارضه، مساحت هم‌پوشانی دو عارضه، فاصله بین مراکز ثقل آن‌ها و کوچکترین مستطیل محیطی^۲ (MBR) استفاده می‌شود. در معیارهای توصیفی میزان شباهت اطلاعات توصیفی و غیرهندسی دو عارضه مکانی مانند نام یا نوع عارضه به کمک تناظریابی معنایی^۳ بررسی می‌شود. در معیارهای روابط مکانی نیز روابط توپولوژی و زمینه جغرافیایی ملاک بررسی قرار می‌گیرد. در توپولوژی رابطه مکانی دو عارضه بررسی می‌شود و در زمینه جغرافیایی

^۱ Context-dependent

^۲ Minimum Bounding Rectangle

^۳ Ontology matching

در این تحقیق بررسی روش‌های تناظریابی عوارض چندضلعی مورد نظر می‌باشد که در حالت کلی می‌توان از سه دیدگاه مختلف این مسائل را حل نمود. در دیدگاه اول می‌توان برای تناظریابی عوارض چندضلعی، نقاط روی مرز آن‌ها را خروجی گرفت و بر اساس الگوریتم‌های تناظریابی عوارض نقطه‌ای عمل کرد. تناظریابی عوارض نقطه‌ای از دو نوع دیگر (عوارض خطی و چندضلعی) ساده‌تر می‌باشد و معمولاً بر اساس فاصله بین موقعیت‌ها محاسبه می‌شود ولی به دلیل وجود اطلاعات بسیار اندک در عارضه نقطه‌ای، این روش توصیه نمی‌شود [۷، ۱۱]. در دیدگاه دوم، می‌توان مرزهای خطی عارضه چندضلعی را استخراج و بر اساس الگوریتم‌های تناظریابی عوارض خطی، دو چندضلعی را انطباق داد. اگرچه بر روی الگوریتم‌های تناظریابی عوارض خطی مطالعات بسیاری شده است و دارای قدمت بیش‌تری است، ولی به دلیل وجود حجم زیاد داده‌ها به هنگام خروجی گرفتن عارضه خطی به جای چندضلعی و عدم توجه به نقاط درون و خارج مرز چندضلعی‌ها و مساحت آن‌ها، استفاده از این الگوریتم‌ها نیز بهینه نیست [۷]. در دیدگاه سوم که یکی از ضرورت‌های این تحقیق را نشان می‌دهد، نیاز به استفاده از الگوریتم‌های خاص تناظریابی عوارض چندضلعی است که در مطالعات کمتری به آن‌ها پرداخته شده و می‌تواند معایب دو دیدگاه قبلی را پوشش دهد.



شکل ۱- دسته‌بندی مسائل تناظریابی از منظرهای مختلف

هدف این مقاله بررسی تناظریابی عوارض چندضلعی از سه منظر (۱) معیارهای اندازه‌گیری شباهت، (۲) روش‌ها و الگوریتم‌های تناظریابی و (۳) روابط متناظر عوارض می‌باشد (شکل ۱). در معیارهای اندازه‌گیری شباهت (منظر اول)، معیارهایی که جهت مقایسه نزدیکی دو عارضه چندضلعی وجود دارند، بررسی می‌شوند. در منظر دوم، هدف بررسی انواع روش‌ها و الگوریتم‌های تناظریابی عوارض چندضلعی می‌باشد و در منظر سوم، هدف

(dilation) شی A به ازای آن ترسیم شود، شامل کل شی B گردد. منظور از dilation که یکی از عملگرهای پایه‌ای در مورفولوژی ریاضی است، گسترش یک شکل به کمک حرکت دایره‌ای با شعاع معین بر روی مرز آن می‌باشد.

$$\begin{aligned} h(A, B) &= \min \{ \varepsilon : A \subset B \oplus S(\varepsilon) \} \\ h(B, A) &= \min \{ \varepsilon : B \subset A \oplus S(\varepsilon) \} \end{aligned} \quad (1)$$

$$d_{Hausdorff}(A, B) = \max\{h(A, B), h(B, A)\} \quad (2)$$

در رابطه (۱)، $h(A, B)$ و $h(B, A)$ به ترتیب برابر با فاصله هاسدورف چندضلعی A از B و B از A می‌باشد. علامت \oplus نشان‌دهنده عملگر dilation در مورفولوژی ریاضی است و همچنین $S(\varepsilon)$ برابر با دایره‌ای به شعاع ε می‌باشد. فاصله هاسدورف نهایی دو چندضلعی A و B طبق رابطه (۲) برابر با حداکثر مقدار بین $h(A, B)$ و $h(B, A)$ می‌باشد. در مطالعات Huh و همکاران [۱۳] از فاصله هاسدورف برای محاسبه فاصله دو چندضلعی در مسئله تناظریابی استفاده شده است. همچنین Min و همکاران [۲۷] فاصله هاسدورف میانه را به عنوان تعمیم فاصله هاسدورف ارائه دادند که می‌توان جهت تناظریابی استفاده نمود.

جدول ۱- بررسی مطالعات پیشین از منظر معیارهای اندازه‌گیری شباهت عوارض چندضلعی

مطالعات پیشین	معیارهای اندازه‌گیری شباهت
[۱۴], [۱۳], [۱۲], [۱۰], [۴], [۲], [۱۹], [۱۸], [۱۷], [۱۶], [۱۵]	فاصله
[۱۹], [۲۱], [۲۰], [۱۲], [۸], [۶]	مساحت هم‌پوشانی
[۱۶], [۱۱], [۱۲], [۴], [۱۰], [۲۲], [۱۹], [۱۸], [۱۷]	ویژگی‌های هندسی
[۱۷], [۲۴], [۴], [۲], [۲۳], [۸], [۱۸]	شکل
[۱۶], [۲۱], [۲۵], [۲], [۱۰]	توصیفی
[۲۲], [۳], [۲], [۲۳]	روابط مکانی

۲-۱-۲- معیار مساحت هم‌پوشانی

مساحت هم‌پوشانی دو چندضلعی یکی از معیارهایی است که در اکثر مطالعات تناظریابی در مرحله شناسایی شباهت دو عارضه یا در مرحله پیش‌پردازش جهت انتخاب عوارض کاندید استفاده شده است [۲۰]. همچنین به

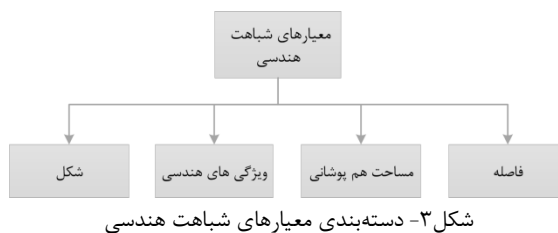
رابطه عارضه با عوارض همسایه از جهات مختلف بررسی می‌شود. در جدول ۱ به اختصار، مطالعاتی که در هر یک از معیارها انجام شده، به تفکیک نوع معیار بیان شده است.

۲-۱- معیارهای شباهت هندسی

این معیارها موقعیت و شکل عارضه چندضلعی را در نظر می‌گیرند. این معیارها را می‌توان بر اساس مطالعات Xavier و همکاران [۱] در چهار دسته فاصله، مساحت هم‌پوشانی، ویژگی‌های هندسی و شکل تقسیم‌بندی نمود. شکل ۳ این دسته‌بندی را نشان می‌دهد. در ادامه هر یک از این دسته‌ها و مطالعات مربوط به آن‌ها ذکر شده است.

۲-۱-۱- معیار فاصله

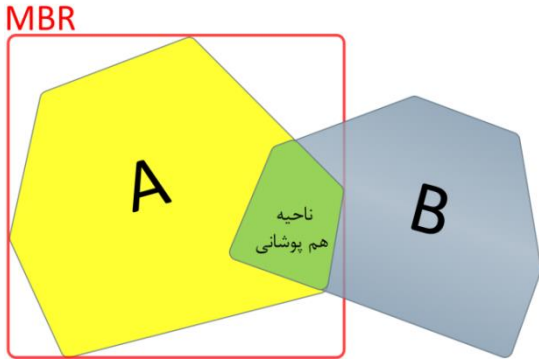
فاصله مکانی یکی از معیارها در اندازه‌گیری موقعیت نسبی بین اشیا مکانی است که درجه‌ای از شباهت آن‌ها را نشان می‌دهد. معروف‌ترین فاصله مورد استفاده در GIS، فاصله اقلیدسی است که فاصله بین دو نقطه منفرد را اندازه‌گیری می‌کند. جهت محاسبه فاصله اقلیدسی بین سایر اشیا مکانی (خط/چندضلعی)، حداقل/حداکثر فاصله اقلیدسی بین نقاط دو شی و یا فاصله بین مراکز ثقل دو شی محاسبه می‌شود. در اکثر مطالعات تناظریابی عوارض چندضلعی از فاصله اقلیدسی بین مراکز ثقل عوارض استفاده شده است [۲, ۴, ۱۰, ۱۲].



شکل ۳- دسته‌بندی معیارهای شباهت هندسی

از دیگر فواصل مورد بررسی، فاصله هاسدورف^۱ است که فاصله بین دو مجموعه نقاط را با انتخاب حداکثر مقدار از بین حداقل فاصله یک مجموعه از مجموعه دیگر محاسبه می‌کند [۲۶]. برای دو چندضلعی A و B، این فاصله بر اساس روابط (۱) و (۲) محاسبه می‌شود. مطابق با شکل ۴، فاصله هاسدورف چندضلعی B از A ($h(B, A)$) برابر با حداقل شعاع (ε) دایره‌ای است که اگر گسترش

^۱ Hausdorff distance



شکل ۵- نمایش MBR چندضلعی A و مساحت هم‌پوشانی بین دو چندضلعی A و B

۲-۱-۴- معیار شکل

یکی از معیارهای مهم در عوارض چندضلعی، معیار شکل عارضه می‌باشد که ترجیحاً باید نسبت به دوران، انتقال و مقیاس مستقل باشد. یکی از توابع مهم در تعریف شکل عارضه چندضلعی، تابع چرخش^۱ است که توسط Arkin و همکاران [۲۸] ارائه شده است. همان‌طور که در شکل ۶ ملاحظه می‌شود، تابع چرخش برای یک چندضلعی در هر نقطه، زاویه پادساعتگرد خط مماس بر شکل (ϕ_i) را به عنوان تابعی از طول تجمیعی نرمال شده شکل اندازه‌گیری می‌کند. به عبارتی دیگر، زاویه تجمیعی خط مماس بر شکل با چرخش اضلاع شکل به سمت چپ افزایش می‌یابد و با چرخش به سمت راست کاهش می‌یابد.

تابع چرخش به‌طور مستقل و یا در رابطه ترکیبی با سایر معیارهای هندسی ذکر شده، در مسائل تناظریابی عارضه چندضلعی استفاده شده است [۸]. فاصله بین توابع چرخش دو چندضلعی A و B جهت اندازه‌گیری میزان شباهت آن‌ها طبق رابطه (۴) محاسبه می‌شود.

$$d(A, B) = \left(\int_0^1 (T_A(l) - T_B(l)) \right)^{\frac{1}{2}} \quad (4)$$

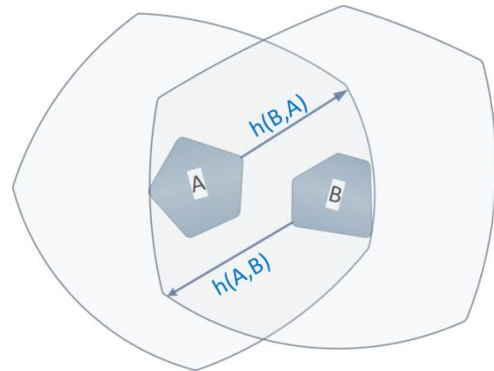
در این رابطه $T_A(l)$ و $T_B(l)$ به ترتیب برابر با تابع چرخش چندضلعی A و B می‌باشند.

در کنار توابع مختلفی که جهت بررسی شکل عوارض با جزئیات و دقت بالا وجود دارند، یکسری معیارهای ساده‌تر نیز معرفی شده‌اند؛ به عنوان نمونه Fan و همکاران

صورت ترکیبی نیز از این معیار استفاده می‌شود؛ به عنوان نمونه از نسبت مساحت هم‌پوشانی بر روی میانگین مساحت دو عارضه یا بر روی کمترین مساحت دو عارضه طبق رابطه (۳) استفاده می‌شود [۶، ۸].

$$\text{Sim}(A, B) = \frac{\text{Area}(A \cap B)}{\min(\text{Area}(A), \text{Area}(B))} \quad (3)$$

در این رابطه $\text{Area}(A \cap B)$ مساحت ناحیه هم‌پوشانی دو چندضلعی A و B می‌باشد و $\min(\text{Area}(A), \text{Area}(B))$ برابر با حداقل مقدار بین مساحت دو چندضلعی A و B است. علاوه بر مساحت هم‌پوشانی چندضلعی‌ها، از مساحت هم‌پوشانی MBR دو عارضه نیز جهت شناسایی عوارض کاندید کمک گرفته می‌شود [۱۲]. شکل ۵ مساحت هم‌پوشانی دو عارضه چندضلعی A و B را نشان می‌دهد و MBR عارضه A نیز مشخص شده است.



شکل ۴- نمایش فاصله هاسدورف بین دو چندضلعی A و B

۲-۱-۳- معیار ویژگی‌های هندسی

می‌توان از روی ویژگی‌های هندسی دو عارضه چندضلعی میزان شباهت آن‌ها را بررسی نمود. ویژگی‌های هندسی معمول یک عارضه چندضلعی شامل محیط، مساحت، طول قطر MBR و جهت شکل می‌باشد. به عنوان نمونه Wang و همکاران [۱۰] به صورت ترکیبی از مساحت، مساحت هم‌پوشانی، طول قطر MBR و فاصله مراکز ثقل استفاده کرده‌اند. در برخی از مطالعات از معیار جهت قطر MBR در ترکیب با سایر معیارها استفاده شده است [۴، ۱۲]. همچنین Wang و همکاران [۱۱] از معیارهای طول، مساحت، جهت و فاصله استفاده کرده‌اند.

^۱ Turning function

اختلاف بین دو رشته حرفی با طول برابر را به کمک شمارش حروف متفاوت محاسبه می‌کند [۳۱].

۲-۳- معیارهای شباهت رابطه مکانی

روابط مکانی قابل تعریف برای یک عارضه شامل روابط توپولوژی و زمینه جغرافیایی می‌باشد. در روابط توپولوژی رابطه مکانی عوارض نزدیک بررسی می‌شود و بیشتر در تناظریاتی شبکه کاربرد دارد [۱]. به عنوان نمونه در یک شبکه، درجه گره که معرف تعداد یال‌های متصل به آن گره است، معیاری جهت تناظریاتی می‌باشد [۳۲] و یا همبندی یال‌ها به عنوان معیار رابطه توپولوژی انتخاب می‌شود [۳۳]. بر اساس مطالعات Zhang و همکاران [۲۳] شباهت عوارض مکانی به هنگام شرایط پیچیده وابسته به زمینه جغرافیایی آن‌ها می‌باشد. طبق تعریف Samal و همکاران [۲]، زمینه جغرافیایی یک عارضه به رابطه مکانی عارضه با عوارض همسایه آن گفته می‌شود که از جهات مختلف مانند طول یا جهت قابل بررسی است و عوارض همسایه نیز شامل عوارض مهم جغرافیایی در ناحیه می‌باشند که به آن‌ها landmark گفته می‌شود. جهت تناظریاتی عوارض چندضلعی در برخی از مطالعات زمینه جغرافیایی لحاظ شده که در ادامه توضیح داده می‌شود.

Samal و همکاران [۲] جهت تناظریاتی عوارض چندضلعی از گراف نزدیکی به عنوان معیار زمینه جغرافیایی استفاده کردند. این گراف، گرافی ستاره‌ای است که زمینه عارضه در مرکز را نشان می‌دهد. در این نمودار، سایر عوارض دیگر به عنوان گره‌های کناری به گره مرکزی توسط یال متصل می‌باشند و هر یال بر اساس فاصله و زاویه ما بین دو گره متصل به آن وزندهی می‌شود. گره‌های کناری که landmarkها هستند به کمک معیارهای هندسی و توصیفی انتخاب می‌شوند. در انتها نمودار نزدیکی دو عارضه با یکدیگر مقایسه می‌شود و ملاک شباهت آن‌ها قرار می‌گیرد.

Kim و همکاران [۳] به کمک مفاهیم رشد حریم، نمودار ورونوی و مثلث‌بندی، روش ارائه‌شده توسط Samal و همکاران را توسعه دادند. در این روش تنها از landmarkهایی استفاده می‌شود که چندضلعی ورونوی آن‌ها در همسایگی چندضلعی ورونوی شی ورودی می‌باشد. در انتها مثلث‌بندی تشکیل شده برای چندضلعی

[۸] معیار میزان مستطیلی بودن^۱ شکل که نسبت مساحت شکل به مساحت MBR آن می‌باشد را ارائه داده‌اند. Zhang و همکاران [۲۳] میزان کشیدگی^۲ شکل را که نسبت عرض به طول MBR شکل است، بررسی می‌کنند. Samal و همکاران [۲] از روش حریم^۳ ساده کمک گرفته‌اند که میزان درصدی از یک عارضه را که در حریم عارضه دیگر قرار دارد، ملاک بررسی شکل قرار داده‌اند. Wang و همکاران [۴] از نسبت مساحت به محیط عارضه جهت بررسی شکل کمک گرفته‌اند. علاوه بر این معیارها، Ai و همکاران [۲۴] از تبدیل فوریه به عنوان یک معیار پیچیده اندازه‌گیری شباهت عارضه استفاده کرده‌اند.

۲-۲- معیارهای شباهت توصیفی

اطلاعات توصیفی (اطلاعات معنایی) شامل اطلاعات غیرهندسی یک شی مکانی است. معیارهای مربوط به تناظریاتی اطلاعات توصیفی پیچیدگی‌های خاص خود را دارند و به آن‌ها مسائل تناظریاتی معنایی نیز گفته می‌شود. این مسائل بر اساس نوع داده تناظریاتی شده به سه دسته داده‌های عددی، دامنه‌های لیستی و رشته‌های حرفی تقسیم‌بندی می‌شوند [۱]. در مطالعات Wang و همکاران [۱۰] برای هر سه دسته از اطلاعات توصیفی، روابط ریاضی مربوطه استفاده شده است.

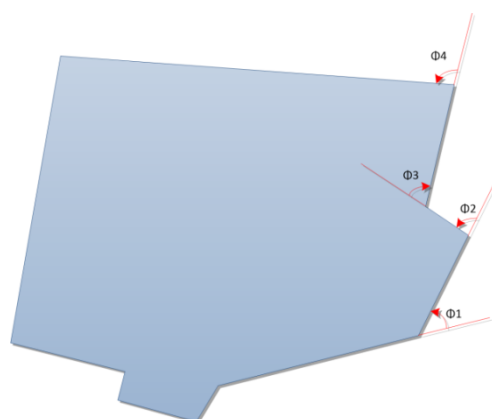
شباهت دو رشته حرفی که به عنوان نمونه بررسی شباهت نام عوارض می‌باشد، نیاز به مقایسه تک به تک حروف آن‌ها دارد و پیچیده‌تر از دو دسته دیگر است. یکی از فواصل پرکاربرد در این مقایسه، فاصله Levenshtein است. این فاصله اختلاف بین دو دنباله از حروف را با شمارش تعداد عملگرهای ویرایش (الحاق، حذف یا جانشینی) محاسبه می‌کند که اگر این عملگرها به یکی از دنباله‌ها اضافه شود، مشابه دیگری می‌گردد [۲۹]. یک راه‌حل مشابه که حداقل تعداد عملگرهای ویرایش را محاسبه می‌کند، فاصله Damerau-Levenshtein می‌باشد [۳۰]. این فاصله در مطالعات Samal و همکاران [۲] استفاده شده است. همچنین از دیگر فواصل مورد استفاده در مقایسه دو رشته حرفی، فاصله Hamming است که

^۱ Rectangularity

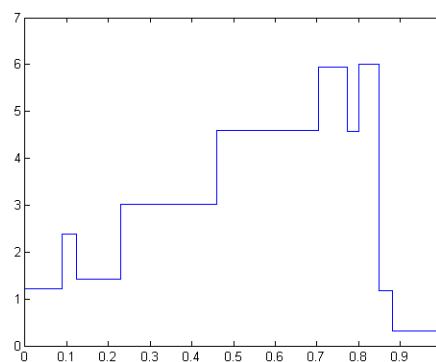
^۲ Elongation

^۳ Buffer

ورودی و landmarkها با مثلث‌بندی چندضلعی‌های کاندید و landmarkهای مربوطه مقایسه می‌شوند.



(الف)



(ب)

شکل ۶- تابع چرخش یک چندضلعی با نمایش زاویه پادساعتگرد خط مماس بر شکل (الف) و نمودار آن (ب)

۳- روش‌های تناظریابی عوارض چندضلعی

دو راه‌کار اصلی جهت تناظریابی عوارض چندضلعی وجود دارد. در راه‌کار اول مسئله تناظریابی عوارض چندضلعی تبدیل به مسئله تناظریابی عوارض نقطه‌ای یا خطی می‌گردد که به این روش‌ها، تناظریابی بر مبنای کاهش بعد نیز گفته می‌شود [۱۱]. در راه‌کار دوم روش‌های خاص تناظریابی عوارض چندضلعی بررسی می‌شوند که در آن‌ها خصوصیات این عوارض مانند مساحت یا مساحت هم‌پوشانی مد نظر قرار می‌گیرد. روش‌های خاص تناظریابی عوارض چندضلعی خود به دو دسته کلی الگوریتم‌های ماتریس‌مبنا و احتمالی‌مبنا تقسیم‌بندی می‌شوند. همچنین الگوریتم‌های احتمالی‌مبنا بر اساس نحوه وزندهی به معیارها خود به دو دسته احتمالاتی و بهینه‌سازی دسته‌بندی می‌شوند. شکل ۷

دسته‌بندی ارائه شده در این تحقیق را نشان می‌دهد. در جدول ۲، مجموعه‌ای از مطالعات پیشین به تفکیک روش‌های تناظریابی مورد استفاده در عوارض چندضلعی آمده است. توضیحات تکمیلی هر روش با تشریح برخی از مطالعات انجام شده، در ادامه بیان می‌شود.

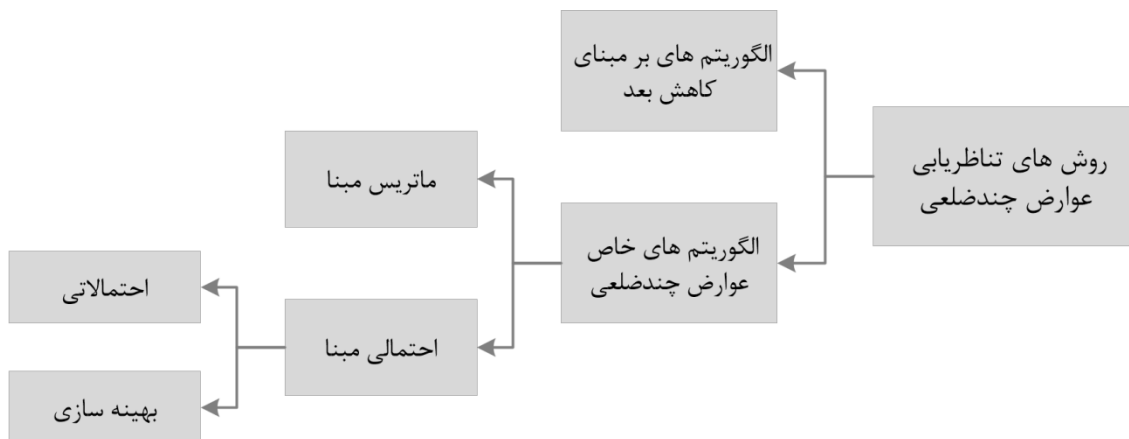
جدول ۲- مطالعات پیشین از منظر روش‌های تناظریابی عوارض چندضلعی

مطالعات پیشین	روش‌های تناظریابی عوارض چندضلعی
[۳۵]، [۳۴]	بر مبنای کاهش بعد
[۱۳]، [۳]، [۲]	ماتریس‌مبنا
[۲۰]، [۳۶]، [۱۰]، [۸]، [۶]	احتمالی
[۱۶]، [۱۸]، [۱۲]، [۲۲]	احتمالی‌مبنا
[۴]، [۱۱]	بهینه‌سازی

۳-۱- الگوریتم‌های بر مبنای کاهش بعد

همان‌طور که ذکر شد، در الگوریتم‌های بر مبنای کاهش بعد، به جای استفاده از الگوریتم‌های تناظریابی خاص عوارض چندضلعی، عارضه چندضلعی تبدیل به عارضه خطی یا نقطه‌ای می‌شود و سپس از الگوریتم‌های تناظریابی خطی یا نقطه‌ای استفاده می‌شود. به عنوان نمونه نقاط روی مرز چندضلعی‌ها و یا خطوط مرز این عوارض خروجی گرفته می‌شوند و سپس مسئله تناظریابی حل می‌شود. همان‌طور که در بخش مقدمه ذکر شد، این راه‌کار به نسبت بهینه نمی‌باشد.

در مطالعات Kieler و همکاران [۳۴] جهت تناظریابی مجموعه داده‌های چندمقیاسی رودخانه‌ها، از تناظریابی بر مبنای کاهش بعد استفاده شده است. در این روش ابتدا خطوط مرکزی چندضلعی‌های رودخانه‌ها به کمک الگوریتم اسکلت‌بندی استخراج می‌شوند به طوری که روابط توپولوژی حفظ شود. سپس تناظریابی گره‌ها و یال‌های شبکه خطوط استخراج شده پیاده‌سازی می‌شود. همچنین Safra و همکاران [۳۵] روشی جهت تناظریابی مکان‌های مختلف ارائه دادند که مکان‌ها را به صورت نقاط فرض کرده‌اند. در نتیجه در این روش، مرکز جرم عوارض چندضلعی به جای کل عارضه در نظر گرفته شده و عملیات تناظریابی مکان‌ها برای مراکز جرم انجام شد.



شکل ۷- دسته‌بندی روش‌های تناظریابی عوارض چندضلعی

۳-۲- الگوریتم‌های ماتریس مبنا

احتمال انطباق دو عارضه چندضلعی بر اساس میزان شباهت یا فاصله بین معیارهای مختلف تناظریابی که در بخش دوم بیان شد، محاسبه می‌شود. به عبارتی دیگر همان‌طور که در رابطه (۵) ملاحظه می‌شود، در الگوریتم‌های احتمالی مبنا، ترکیب خطی از سه نوع معیار شباهت هندسی $(S_{geo}(A, B))$ ، توصیفی $(S_{att}(A, B))$ و رابطه مکانی $(S_{rel}(A, B))$ با وزن‌های مربوطه که به ترتیب با W_{geo} ، W_{att} و W_{rel} نشان داده شده‌اند، به عنوان میزان شباهت دو عارضه چندضلعی محاسبه می‌شود. به هنگام محاسبه این رابطه، وزندهی به هر معیار به دو روش نظر کارشناس و آموزش هوشمند به کمک الگوریتم‌های هوش مصنوعی انجام می‌شود. در نتیجه می‌توان الگوریتم‌های موجود در این دسته را به دو گروه الگوریتم‌های احتمالی و بهینه‌سازی دسته‌بندی نمود.

$$S(A, B) = W_{geo} \times S_{geo}(A, B) + W_{att} \times S_{att}(A, B) + W_{rel} \times S_{rel}(A, B) \quad (5)$$

Fu و Wu [۲۰] الگوریتم‌های احتمالاتی جهت تناظریابی عوارض خطی و چندضلعی ارائه دادند. در الگوریتم احتمالاتی عارضه چندضلعی از معیارهای شباهت مساحت عارضه، مساحت هم‌پوشانی و فاصله باری‌سنتر^۱ استفاده شده است. Tong و همکاران [۳۶] یک الگوریتم احتمالاتی جهت تناظریابی عوارض نقطه‌ای، خطی و چندضلعی ارائه دادند که در این روش، احتمال انطباق میانگین وزن‌داری از معیارهای شباهت موقعیت، شکل، جهت و روابط توپولوژی می‌باشد. Wang و همکاران [۱۰] یک روش تناظریابی احتمالاتی به منظور برزورسانی

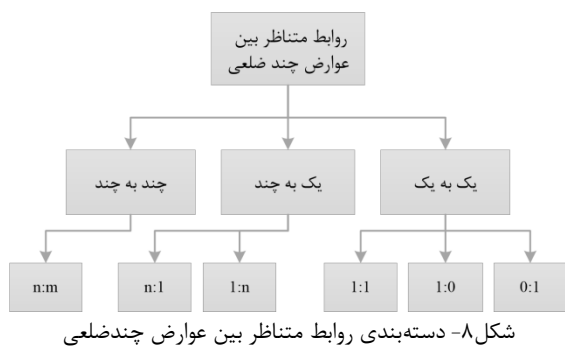
در الگوریتم‌های ماتریس مبنا، ماتریس اولیه‌ای بر مبنای معیارهای شباهت چندضلعی‌ها تولید می‌شود و سپس این ماتریس در یک روند تکراری بر اساس سایر معیارها بروز می‌شود. در این روش، تمام روابط شامل یک به یک، یک به چند و چند به چند به کمک عملیات ماتریسی قابل شناسایی است؛ اما انتخاب حد آستانه‌ها تأثیر بسیاری در نتیجه نهایی انطباق دارد؛ بنابراین دارای انعطاف‌پذیری کمتری در مقایسه با سایر روش‌ها می‌باشد. Samal و همکاران [۲] یک روش ماتریس مبنا جهت تلفیق مجموعه داده‌های مکانی ارائه دادند. در این روش ماتریس مشابهت اولیه با محاسبه شباهت‌های مستقل زمینه مقداردهی شده است. سپس این ماتریس در یک روند تکراری به کمک معیار شباهت زمینه محاسبه شده توسط نمودار نزدیکی، بروز می‌شود. Kim و همکاران [۳] به کمک نمودار ورونوی و مثلث‌بندی، محاسبات شباهت زمینه ارائه شده توسط Samal و همکاران را بهبود بخشیدند. Huh و همکاران [۱۳] یک روش ماتریس مبنا متفاوتی ارائه دادند. در این روش یک ماتریس باینری به کمک روابط هم‌پوشانی بین دو مجموعه داده عوارض چندضلعی مقداردهی اولیه می‌شود و سپس در یک روند تکراری با ضرب در خودش بروز می‌شود.

۳-۳- الگوریتم‌های احتمالی مبنا

گروه دوم الگوریتم‌های تناظریابی عوارض چندضلعی، الگوریتم‌های احتمالی مبنا می‌باشند که اکثر مطالعات انجام شده در این گروه قرار می‌گیرند. در این الگوریتم‌ها

^۱ Barycentre distance

تقسیم‌بندی را نشان می‌دهد. تناظریابی یک به یک یکی از ساده‌ترین دسته‌ها می‌باشد که از اولین مطالعات وجود داشته است. در این گروه هر سه رابطه $1:1$ ، $1:0$ و $0:1$ قابل شناسایی است. در دسته روابط یک به چند، دو نوع رابطه $1:n$ و $n:1$ قابل استخراج است که در مجموعه داده‌های چندمقیاسی با سطح جزئیات متفاوت بسیار کاربرد دارد. استخراج روابط چند به چند نیز جزو روش‌های پیچیده‌تری می‌باشد که تمام شرایط ممکن را بررسی می‌کند. در جدول ۳ دسته‌بندی الگوریتم‌های تناظریابی عوارض چندضلعی بر اساس نوع روابط قابل تشخیص بیان شده است.



جدول ۳- بررسی مطالعات پیشین از منظر روابط متناظر بین عوارض چندضلعی

روابط متناظر بین عوارض چندضلعی	مطالعات پیشین
یک به یک، یک به هیچ و هیچ به یک	[۱۴]، [۱۷]، [۲۱]، [۳۷]، [۲]، [۸]
یک به چند	[۳۶]، [۲۷]، [۳]، [۲۵]
چند به چند	[۱۵]، [۱۶]، [۱۳]، [۲۳]، [۸]، [۲۰]
	[۱۱]، [۴]، [۶]، [۱۲]، [۱۹]، [۲۲]

در شکل ۹ انواع روابط متناظر بین عوارض چندضلعی از دو مجموعه داده مختلف نشان داده شده است. چندضلعی‌های سبز رنگ و آبی رنگ به ترتیب مربوط به مجموعه داده اول و دوم می‌باشند. در حالت $1:1$ ، یک عارضه از مجموعه داده اول با یک عارضه از مجموعه داده دوم متناظر است. در حالت‌های $1:0$ یا $0:1$ ، برای یک عارضه از یک مجموعه داده هیچ عارضه متناظری از مجموعه داده دوم پیدا نشده است. در حالت‌های $1:n$ یا $n:1$ ، یک عارضه از مجموعه داده‌ای با n عارضه از مجموعه داده دوم متناظر است. در حالت $n:m$ که پیچیده‌ترین حالت می‌باشد، n عارضه از یک مجموعه داده با m عارضه از مجموعه داده دیگر متناظر است.

انتشار یافته مجموعه داده‌های چندنمایی ارائه دادند. این روش بر مبنای محاسبه احتمال شباهت‌های هندسی و توصیفی دو عارضه چندضلعی می‌باشد.

Fan و همکاران [۸] جهت ارزیابی کیفی داده‌های OSM روشی برای تناظریابی ساختمان‌ها ارائه دادند. در این روش احتمالی، معیارهای مساحت هم‌پوشانی و تابع چرخش برای محاسبه مشابهت عوارض استفاده شده است. همچنین Fan و همکاران [۶] روشی جهت تناظریابی شبکه راه‌های OSM با شبکه راه‌های معتبر ارائه دادند. در این روش، ابتدا بلوک‌های شهری که توسط راه‌های اطراف احاطه شده‌اند استخراج شده، سپس با انطباق بلوک‌ها و به دنبال آن، انطباق اضلاع بلوک‌ها، راه‌های اختصاص یافته به هر ضلع منطبق می‌شوند. جهت انطباق بلوک‌ها از نسبت هم‌پوشانی دو بلوک با احتمال بیشتر از ۵۰ درصد به عنوان شرط انطباق دو بلوک بر یکدیگر استفاده شد.

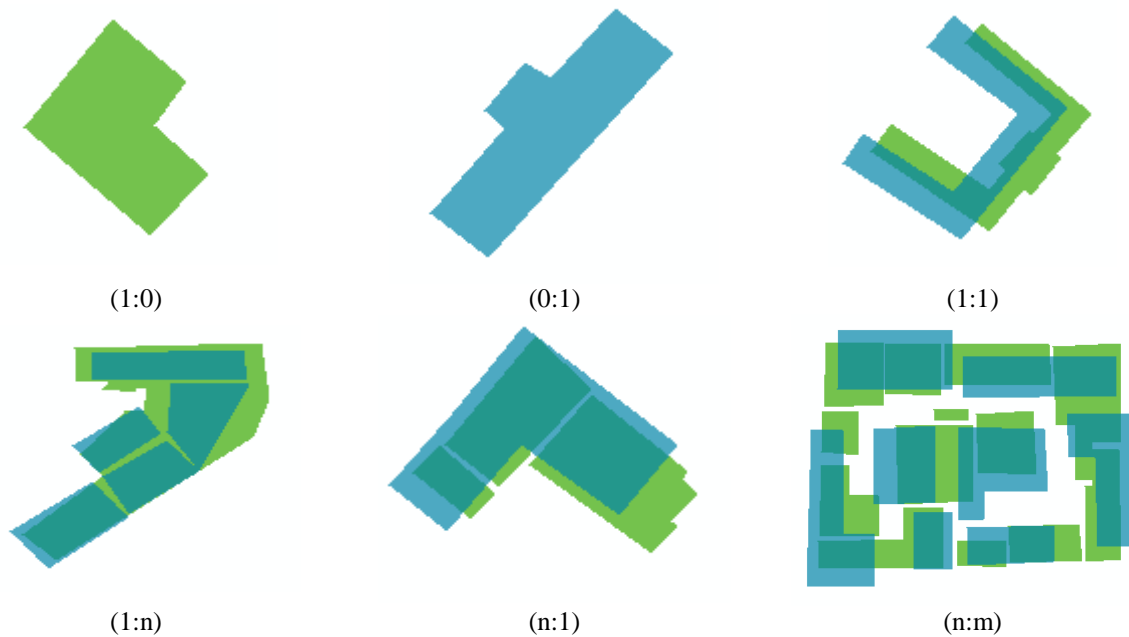
در تناظریابی عوارض چندضلعی با الگوریتم‌های بهینه‌سازی دو مطالعه انجام شده است. Wang و همکاران [۱۱] الگوریتم بهینه‌سازی جهت تناظریابی عوارض چندضلعی در انتشار روزرسانی مجموعه داده‌های چندنمایی ارائه دادند. در این روش از الگوریتم تناظریابی احتمال‌مبنا با وزن‌دهی به کمک شبکه عصبی پس انتشار^۱ (BPNN) استفاده شده است. در تحلیلی مقایسه‌ای، نتایج نشان دهنده کارایی و صحت بالای این روش در مقایسه با الگوریتم‌های احتمالاتی می‌باشد. همچنین Wang و همکاران [۴] الگوریتم بهینه‌سازی تناظریابی عوارض چندضلعی جهت ادغام داده‌ها ارائه دادند که در آن مشابه با روش قبلی از یک الگوریتم تناظریابی احتمال‌مبنا استفاده شده است؛ با این تفاوت که وزن‌دهی بر اساس شبکه عصبی بهینه‌سازی ازدحام ذرات^۲ (PSO) پیاده‌سازی شده است.

۴- روابط متناظر بین عوارض چندضلعی

در فرآیند تناظریابی، انواع روابط قابل استخراج بین عوارض چندضلعی شامل شش رابطه $1:n$ ، $1:1$ ، $0:1$ ، $1:0$ ، $n:m$ و $n:1$ می‌باشد که می‌توان در سه دسته یک به یک، یک به چند و چند به چند تقسیم‌بندی نمود. شکل ۸ این

^۱ Back Propagation Neural Network

^۲ Particle Swarm Optimization Neural Network



شکل ۹- انواع روابط متناظر بین عوارض چندضلعی از دو مجموعه داده مختلف (چندضلعی‌های مجموعه داده اول با رنگ سبز و چندضلعی‌های مجموعه داده دوم با رنگ آبی مشخص شده‌اند)

۵- بحث و نتیجه‌گیری

در این مقاله مروری بر روش‌های تناظریابی عوارض چندضلعی در کاربردهای مختلف علوم کامپیوتری و GIS انجام شده است. مطالعات موجود از سه دیدگاه معیارهای اندازه‌گیری شباهت عوارض چندضلعی، روش‌ها و الگوریتم‌های تناظریابی عوارض چندضلعی و روابط متناظر بین این عوارض بررسی شده‌اند. معیارهای شباهت عوارض خود به سه دسته معیارهای هندسی، توصیفی و رابطه مکانی تقسیم‌بندی شده‌اند. در مطالعات مختلف با توجه به منطقه مورد بررسی و میزان پیچیدگی شناسایی عوارض مشابه، ترکیبی از این معیارها استفاده شده است.

در الگوریتم‌های تناظریابی عوارض چندضلعی، دو دسته کلی الگوریتم‌های ماتریس‌مبنا و احتمالی‌مبنا بررسی شده‌اند که دارای کارایی‌های متفاوتی می‌باشند. در پایان، انواع روابط متناظر عوارض به سه دسته یک به یک، یک به چند و چند به چند تقسیم‌بندی شده‌اند که هدف از اجرای الگوریتم‌ها شناسایی این روابط بین عوارض چندضلعی از دو یا چند مجموعه داده مختلف است.

معیارهای مختلفی جهت ارزیابی روش‌های مختلف تناظریابی وجود دارد که دو معیار بسیار متداول جهت ارزیابی کارایی این مسائل، Precision و Recall می‌باشند که در اکثر مطالعات به آن‌ها اشاره شده است [۲، ۴، ۵،

[۲۳]. معیار Precision نسبت تعداد عوارض صحیح منطبق شده به کل عوارضی را که صحیح یا اشتباه منطبق شده‌اند، بررسی می‌کند. معیار Recall نسبت عوارض صحیح منطبق شده به کل عوارضی را که صحیح منطبق شده‌اند و یا منطبق نشده‌اند، ارزیابی می‌کند. به عبارتی دیگر در معیار Precision تعداد عوارضی که اشتباه منطبق شده‌اند و در معیار Recall تعداد عوارضی که منطبق نشده‌اند، در مقایسه دو عدد تاثیرگذار می‌باشد.

روش‌های ارائه شده در مطالعات مختلف، به کمک معیارهای فوق، ارزیابی شده‌اند که با بررسی آن‌ها، می‌توان به نتایج مفیدی در رابطه با روش‌های مختلف تناظریابی عوارض چندضلعی رسید که در اینجا به مهم‌ترین آن‌ها اشاره می‌شود. در رابطه با معیارهای اندازه‌گیری شباهت مورد استفاده، Samal و همکاران [۲] نشان دادند که استفاده از شباهت زمینه بر پایه دو معیار جهت و فاصله (شباهت رابطه مکانی)، عملکرد الگوریتم را نسبت به حالتی که تنها از شباهت هندسی استفاده شود، بهبود می‌بخشد. همچنین در روش Kim و همکاران [۳] که جهت بهبود روش Samal و همکاران [۲] ارائه شده است، استفاده از نمودار ورونوی دو مزیت عمده در عملکرد الگوریتم دارد؛ (۱) تنها از landmarkهایی در نمودار نزدیکی استفاده می‌شود که چندضلعی‌های ورونوی آن‌ها به صورت پیوسته و در همسایگی چندضلعی شی ورودی

بر اساس مطالعات انجام شده و معایب و مزایای هر روش پیشنهاد می‌گردد در آینده مطالعات مقایسه‌ای بر روی روش‌های ارائه شده تا به امروز انجام گردد و بر اساس دو معیار کارایی Precision و Recall بر روی مجموعه داده‌های نمونه یکسان، نتایج با یکدیگر مقایسه شوند. همچنین میزان پیچیدگی الگوریتم‌های ارائه شده بر اساس زمان اجرا یا سایر پیچیدگی‌های متداول بررسی نشده‌اند و بهتر است در آینده مطالعاتی در این زمینه‌ها انجام گردد.

قرار دارند و تاثیر مهم‌تری در این فرآیند دارند و (۲) نمودار ورونوی کمک می‌کند تا تعداد محدودی از اشیا کاندید در مجموعه داده هدف انتخاب شوند.

در مقایسه الگوریتم‌های بهینه‌سازی با الگوریتم‌های احتمالاتی، Wang و همکاران [۴] نشان دادند که وزن‌دهی با الگوریتم‌های هوشمند، کارایی مسائل تناظریابی را بهبود می‌بخشد. با توجه به اینکه در این روش‌ها وزن‌دهی به-جای نظر کارشناس با الگوریتم‌های هوشمند است، تایید شد که وزن‌های اندازه‌گیری‌های یکسان با عوض شدن زمینه جغرافیایی، تغییر می‌کنند.

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, p. 39, 2016.
- [2] A. Samal, S. Seth, and K. Cueto 1, "A feature-based approach to conflation of geospatial sources," *International Journal of Geographical Information Science*, vol. 18, pp. 459-489, 2004.
- [3] 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.
- [4] 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. 1.۲۱۹-۸۹
- [5] H. Du, N. Alechina, M. Jackson, and G. Hart, "A Method for Matching Crowd-sourced and Authoritative Geospatial Data," *Transactions in GIS*, 2016.
- [6] 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.
- [7] H. Du, "Matching disparate geospatial datasets and validating matches using spatial logic," *University of Nottingham*, 2015.
- [8] H. Fan, A. Zipf, Q. Fu, and P. Neis, "Quality assessment for building footprints data on OpenStreetMap," *International Journal of Geographical Information Science*, vol. 28, pp. 700-719, 2014.
- [9] L. Harrie and A. Hellstrom, "A case study of propagating updates between cartographic data sets," in *Proceedings/Actes, 19th International Cartographic Conference, 11th General Assembly of ICA, Ottawa, 1999*.
- [10] 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*, ed: Springer, 2014, pp. 139-153.
- [11] 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, 2015.
- [12] L. Huang, S. Wang, Y. Ye, B. Wang, and L. Wu, "Feature matching in cadastral map integration with a case study of Beijing," in *2010 11th International Conference on Geoinformatics*, 2010, pp. 1-4.
- [13] Y. Huh, 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.
- [14] L. Li and M. Goodchild, "Automatically and accurately matching objects in geospatial datasets," *Adv. Geo-Spat. Inf. Sci*, vol. 10, pp. 71-79, 2012.
- [15] M. Butenuth, G. v. Gösseln, M. Tiedge, C. Heipke, U. Lipeck, and M. Sester, "Integration of heterogeneous geospatial data in a federated database," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 62, pp. 328-346, 2007.

- [16] S. YING, L. LI, Y. GAO, and Y. MIN, "PROBABILISTIC MATCHING OF MAP OBJECTS IN MULTI-SCALE SPACE".
- [17] X. Zhang, X. Zhao, M. Molenaar, J. Stoter, M.-J. Kraak, and T. Ai, "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, August-September 2012, pp. 19-24, 2012.
- [18] T. Wenjing, H. Yanling, Z. Yuxin, and L. Ning, "Research on areal feature matching algorithm based on spatial similarity," in 2008 Chinese Control and Decision Conference, 2008, pp. 3326-3330.
- [19] J. Kim and K. Yu, "AREAL FEATURE MATCHING BASED ON SIMILARITY USING CRITIC METHOD," ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 1, pp. 75-78, 2015.
- [20] F. Zhonglianga and W. Jianhuaa, "Entity matching in vector spatial data," in Proceedings of the XXI ISPRS Congress, 2008, pp. 1467-1472.
- [21] J. Hastings, "Automated conflation of digital gazetteer data," International Journal of Geographical Information Science, vol. 22, pp. 1109-1127, 2008.
- [22] H. Yue, X. Zhu, D. Chen, and L. Liu, "A Multi-Scale Settlement Matching Algorithm Based on ARG," ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, pp. 139-143, 2016.
- [23] X. Zhang, T. Ai, J. Stoter, and X. Zhao, "Data matching of building polygons at multiple map scales improved by contextual information and relaxation," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 92, pp. 147-163, 2014.
- [24] T. Ai, X. Cheng, P. Liu, and M. Yang, "A shape analysis and template matching of building features by the Fourier transform method," Computers, Environment and Urban Systems, vol. 41, pp. 219-233, 2013.
- [25] M. Al-Bakri and D. Fairbairn, "Assessing similarity matching for possible integration of feature classifications of geospatial data from official and informal sources," International Journal of Geographical Information Science, vol. 26, pp. 1437-1456, 2012.
- [26] W. Rucklidge, Efficient visual recognition using the Hausdorff distance vol. 1173: Springer Heidelberg, 1996.
- [27] 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.
- [28] E. M. Arkin, L. P. Chew, D. P. Huttenlocher, K. Kedem, and J. S. Mitchell, "An efficiently computable metric for comparing polygonal shapes," DTIC Document 1989.
- [29] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions and reversals," in Soviet physics doklady, 1966, p. 707.
- [30] F. J. Damerau, "A technique for computer detection and correction of spelling errors," Communications of the ACM, vol. 7, pp. 171-176, 1964.
- [31] R. W. Hamming, "Error detecting and error correcting codes," Bell System technical journal, vol. 29, pp. 147-160, 1950.
- [32] B. Rosen and A. Saalfeld, "Match criteria for automatic alignment," in Proceedings of 7th international symposium on computer-assisted cartography (Auto-Carto 7), 1985, pp. 1-20.
- [33] V. Walter and D. Fritsch, "Matching spatial data sets: a statistical approach," International Journal of Geographical Information Science, vol. 13, pp. 445-473, 1999.
- [34] B. Kieler, W. Huang, J.-H. Haurert, and J. Jiang, "Matching river datasets of different scales," in Advances in GIScience, ed: Springer, 2009, pp. 135-154.
- [35] E. Safra, Y. Kanza, Y. Sagiv, C. Beerli, and Y. Doytsher, "Location-based algorithms for finding sets of corresponding objects over several geo-spatial data sets," International journal of geographical information science, vol. 24, pp. 69-106, 2010.
- [36] 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.
- [37] C. B. Jones, J. M. Ware, and D. R. Miller, "A probabilistic approach to environmental change detection with area-class map data," in Integrated Spatial Databases, ed: Springer, 1999, pp. 122-136.