مفاهیم و روشهای داده کاوی فضایی در سامانه اطلاعات مکانی

نویسنده : سجاد علی پور اشلیکی، دانشجوی کارشناسی ارشد شهرسازی دانشگاه علم و صنعت ایران

چکیده :
امروزه حجم داده های گردآوری شـده در سـامانه اطلاعات مکانی به میزان فزاینده ای افزایش یافته اسـت، به گونه ای که پردازش، تجزیه و تحلیل و درک روند های موجود در آنها به روشهای معمول امکان پذیر نمی باشد. جهت حل این مشکل رهیافت داده کاوی فضایی ارائه شـده اسـت که مبتنی بر تعمیم تکنیکهای آماری، فن آوری پایگاه داده و نیز نمایش توصیفی اطلاعات می باشـد. در این مقاله به بررسی مفاهیم و روشهای داده کاوی فضایی به عنوان ابزاری در جهت تحلیل در یک سامانه پشتیبان تصمیم گیری مکانی و در یک مفهوم گسترده تر با عنوان «اکتشاف دانش» از پایگاه اطلاعاتی پرداخته می شود. همچنین توانمندی این روشها در شرایطی که تجزیه و تحلیل اطلاعات و استخراج الگوها از پایگاه های اطلاعاتی بزرگ مد نظر است مورد ارزیابی قرار می گیرد.

کلمات کلیدی: سامانه اطلاعات جغرافیایی(GIS)، مدل تصمیم گیری مکانی، اکتشاف دانش از پایگاه اطلاعاتی، تحلیل فضایی

مقدمه :

بکارگیری روشهای داده کاوی در سامانه های اطلاعاتی به عنوان یک نیاز ضروری در بسیاری از شاخه های علوم کاربردی مطرح شده است، از آن جهت که پیاده سازی این روشها منجر به افزایش کارایی در پردازش داده ها و تحلیل اطلاعات می گردد، این امر در شرایطی که بررسی و استخراج نتایج از منابع داده ای با حجم زیاد و گستردگی ابعاد پایگاه اطلاعاتی مورد نظر باشد نمود بیشتری می یابد. با توسعه سامانه های اطلاعات جغرافیایی و استفاده بیش از پیش داده های مکانی در مدلهای برنامه ریزی و تصمیم گیری، بکارگیری روشهای داده کاوی فضایی در این سامانـه ها مورد توجـه قرار گرفـت. دو عـامـل، یـکی افزایـش پیچیدگی در ساختارهای اطلاعات فضایی که منتج از گسترش سطوح تحلیلی بود و دیگری ساخت نیافته بودن مسائل فضایی که منجر به تعمیم ناپذیری در حل آنها می شد،تجزیه و تحلیل و اکتشاف دانش از آنها را در قالب یک فرآیند تکرارپذیر و کارا با مشکلاتی مواجه کرد. در واقع با توسعه سامانه های سنجش از دور (RS)، تصاویر ماهواره ای و نیز مطرح شدن نیازهای جدید که در ارتباط با توسعه مدلهای تصمیم‌گیری مکانی ایجاد شد، یکپارچه سازی روند تحلیل اطلاعات و تمرکز بر مجموعه ای عظیم از داده ها به آسانی امکان پذیر نبود. تا قبل از دهه ۹۰ میلادی بسیاری از روشهای به کار رفته در برنامه ریزی و تصمیم سازی بر پایه رویکردهای آماری استوار بود، اما این روشها در

ارتباط با پایگاه های اطلاعاتی بزرگ (نظیر پایگاه های داده مکانی) از قابلیتهای لازم برخوردار نبودند. علاوه بر این، استفاده از آنها نیاز به سطح وسیعی از دانش آمار داشته و جهت اجرای آن وجود پیش فرضهای فراوان (که اکثرا" در مورد داده های فضایی با ابهاماتی مواجه است) الزامی بود.

در مقایسه با مدلهای آمار کلاسیک، روشهای به کار رفته در سامانه های داده کاوی فضایی کارکرد بهتری را در تحلیلهای فضایی ارائه می نمایند، از آن جهت‌که علاوه بر سازگاری با مدلهای تصمیم گیری و فن آوری پردازش اطلاعات در GIS، قابلیت تجزیه و تحلیل داده های حجیم و کسب اطلاعات ارزشمند از آنها را دارند و لذا توسعه آنها در جهت افزایش توانمندی سامانه های اطلاعات مکانی مورد نظر است.

در بخش دوم این مقاله، پس از بیان تعاریف و مفاهیم داده کاوی فضایی و ضرورت بکارگیری آن در مجموعه های داده بزرگ، این رهیافت در قالب مفهوم استخراج دانش از پایگاه های اطلاعاتی مورد ارزیابی قرار می گیرد. در بخش سوم و چهارم پس از معرفی انواع داده ها در داده کاوی فضایی، سه نمونه از مهمترین روشهای داده کاوی فضایی بررسی می گردد و در پایان، در بخش پنجم مهمترین نتایج مباحث مطرح شده ارائه می شود.

تعاریف و مفاهیم داده کاوی فضایی

در یک تعبیر کلی روشهای داده کاوی[1] در جهت خودکارسازی تشخیص الگوهای پنهان در پایگاه های اطلاعاتی ایجاد شده است (Pamela Jean Pratt et al, 2003). در ارتباط با پایگاه های داده فضایی به جهت وجود متغیرهایی که ویژگی های مکانی را در خود ذخیره می کنند و نیز همبستگی های فضایی میان آنها سطح پیچیدگی تحلیل افزایش می یابد. به عنوان نمونه در برنامه ریزی شهری می توان از سامانه های اطلاعات مکانی جهت تعیین مکان بهینه یک کاربری خاص در شهر استفاده نمود. برای این منظور لازم است تا اطلاعات مربوط به خصوصیات مکانی محدوده مورد نظر به همراه سایر اطلاعات توصیفی جمع آوری شده و در قالب یک مدل تحلیل شود. خروجی چنین مدلی اغلب چندین آلترناتیو خواهد بود که انتخاب نهایی در آنها بر اساس شاخصه هایی از پیش تعیین شده صورت می پذیرد. حال چنانچه تعداد متغیرها و یا اطلاعات مربوط به آنها در مدل افزایش یابد میزان پردازش اطلاعات به صورت فزاینده ای افزایش خواهد یافت. به این علت که باید ارتباط متغیرهای جدید با متغیرهای موجود به صورت متقابل ارزیابی شود و لذا تشخیص روندها و تعیین الگوهای موجود در آنها که در نهایت در ایجاد مدلهای تصمیم گیری کمک کننده است با مشکل مواجه می شود. در این شرایط روشهای داده کاوی فضایی توانایی تشخیص این الگوها را که در پیکره ارتباطات پیچیده میان متغیرها پنهان شده است فراهم می سازند. مزیت استفاده از داده کاوی فضایی نسبت به سایر روشها این است که جهت اجرای آنها به هیچ فرضیه خاصی نیاز نیست و معمولا این روشها خود به تعیین الگوها و تشخیص روندها می پردازند. علاوه بر این الگوریتمهای طراحی شده در آنها به گونه ای است که قابلیت جستجو و پردازش حجم عظیمی از اطلاعات را دارا می باشند. داده کاوی فضایی در ارتباط با مفهوم گسترده تری به نام اکتشاف دانش از پایگاه های اطلاعاتی یا به اختصار KDD[2] قراردارد. KDD عبارت است از فرآیند پردازش داده ها به منظور اکتشاف الگوهای قابل درک، معتبر و به طور بالقوه مفید از یک پایگاه اطلاعاتی (Martin, Ester et al, 2001) تفاوت روشهای داده کاوی فضایی نسبت به روشهای معمول در داده کاوی که در ابتدا به وجود آمد آن‌است که این روشها متغیرهای با مختصات مکانی را نیز در تجزیه و تحلیل و مدلسازی حل مساله دخالت می دهند. به عبارتی داده کاوی فضایی عبارت است از فرآیند اکتشاف دانش و استخراج اطلاعات ارزشمند، روابط فضایی و سایر الگوهایی که به طور ضمنی در پایگاه های اطلاعاتی ذخیره شده است . (kopersky et al,1996, Jukka M. Krisp, 2008, P2) منظور

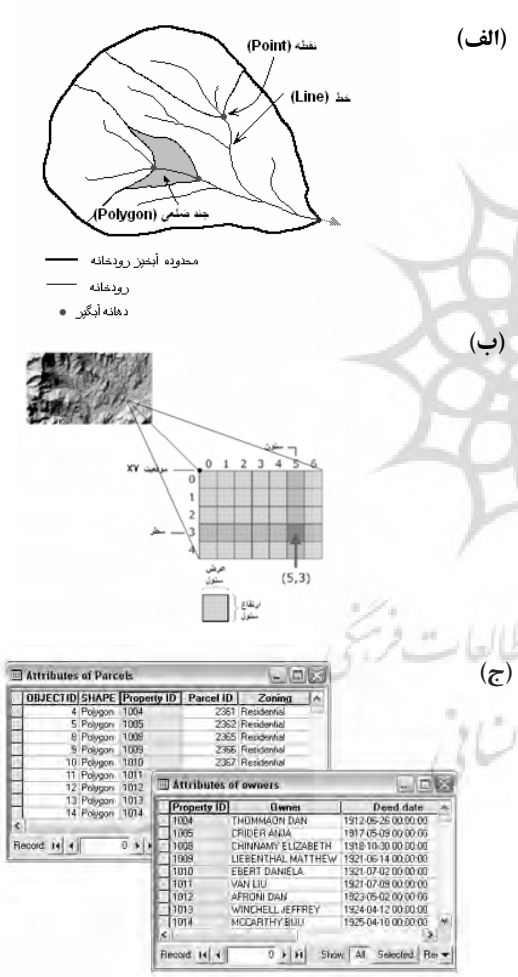
از الگوهای ضمنی جستجوی اطلاعات عمیق و پنهان در یک منبع داده است که به تولید دانش استراتژیک منجر می شود (Xiaobai Yao, 2003, P4).

گرچه توسعه روشها و ساخت مدلهای جدید در داده کاوی فضایی موضوع اصلی دانشهایی نظیر شهرسازی و یا مهندسی سامانه اطلاعات جغرافیایی نیست اما پیاده سازی این روشها در قالب مدلهای تصمیم گیری و یا تجزیه و تحلیل اطلاعات به وسیله آنها به عنوان یکی از موضوعات اساسی و به عنوان دانشی بین رشته ای محسوب می گردد، و بنابر این درک مفاهیم و روشهای به کار رفته در آنها اهمیت فراوانی خواهد داشت. روشهای به کار رفته در داده کاوی فضایی تعمیم روشهای موجود در آمار کلاسیک، یادگیری ماشین[3]، نمایش توصیفی اطلاعات و تجزیه و تحلیل چند بعدی داده ها[4] است که پس از توصیف ساختار داده های فضایی مورد بررسی قرار می گیرد.

انواع داده ها در داده کاوی فضایی

انواع داده ها در داده کاوی فضایی عبارتست از داده های فضایی[5] و داده های توصیفی غیرفضایی[6]. داده های نوع اول خود به دو دسته برداری و رستری تقسیم بندی می شوند. داده های برداری جهت نمایش عوارض جغرافیایی به کار می روند و در سه شکل نقاط، خطوط و چندضلعیها نمایش داده می شوند (verbila,2002,P27). دسته دوم شامل یک شبکه منظم از سلولهای مربعی شکل می باشند که هم اندازه و تقسیم ناپذیر بوده و اطلاعات مربوط به عوارض درهر یک از آنها به صورت جداگانه ذخیره می شود (verbila,2002,P119). داده های توصیفی غیرمکانی به توصیف خصوصیات عوارض جغرافیایی می پردازند و معمولا در جداول پایگاه اطلاعاتی در قالب سطرها و ستونها ذخیره می شوند.

در یک سامانه اطلاعات مکانی داده های فضایی و داده های توصیفی غیر فضایی در مجموعه های داده[7] ذخیره و بازیابی و به شکل لایه های اطلاعاتی نمایش داده می شوند. بنابراین یک مجموعه داده شامل ترکیبی از اطلاعات مربوط به موقعیت جغرافیایی عوارض، شکل و نیز اطلاعات توصیفی مربوط به آنهاست. الگوریتمهای محاسبات و جستجو در روشهای داده کاوی فضایی بر روی این مجموعه های داده عمل می کنند و نتایج حاصل از این روشها می تواند به شکل نقشه، ،جدول و یا نمودارهایی باشد که به توصیف و یا تحلیل اطلاعات می پردازد.



شکل شماره ۱: انواع داده ها در داده کاوی فضایی: الف) برداری، ب) رستری، ج) توصیفی غیر مکانی
مأخذ:ArcGIS Desktop Help 9.2

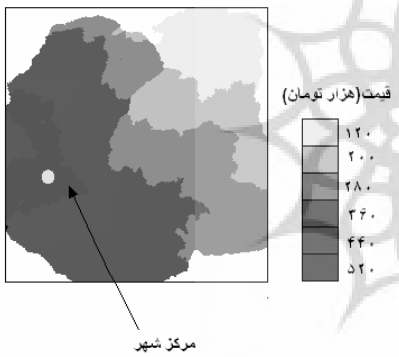
دو خاصیت مهم در ارتباط با داده های فضایی وجود دارد که عبارتست از ژئومتری و توپولوژی. ژئومتری داده های فضایی در واقع همان خواص هندسی آنها محسوب می گردد همانند مساحت، محیط، موقعیت فضایی و... . خواص توپولوژی عناصر فضایی بیانگر روابط فضایی آنهاست که در مدلسازی و تحلیل داده ها کاربرد فراوانی دارد. از انواع خواص توپولوژی عناصر می توان روابط همسایگی میان عوارض[۸]، عضویت در یک محدوده[۹] و یا مجاورت با عوارض دیگر[۱۰] را نام برد (Vera Karasova, 2005, P7).

روشهای داده کاوی فضایی

مدلهای تصمیم گیری مکانی که در گذشته استفاده می شد اغلب شامل روشهایی بود که فضا را به عنوان عاملی تعیین کننده در مدل بکار نمی گرفت و لذا اثر یک متغیر در فضا به شکل متوسط و یا مجموع آن به تمام ناحیه تعمیم داده می شد. این فرضیه که ناحیه مورد مطالعه از نظر فضایی کاملا همگن و یکنواخت است منطقی به نظر نمی رسد (Sharifi, 2004, P1). مزیت استفاده از روشهای داده کاوی فضایی نسبت به روشهای پیشین این است که آنها ویژگی های مکانی را به همراه سایر عوامل غیرمکانی در یک مدل تصمیم گیری وارد می کنند. این روشها در مواردی چون پیش بینی، کلاسه بندی و نیز کشف الگوهای فضایی بکار می روند. همچنین آنها توانایی تحلیل چندبعدی داده ها و بررسی روابط متقابل میان آنها را به شکلی کارا دارند. در این بخش به معرفی بعضی از مهمترین روشهای داده کاوی فضایی پرداخته می شود. به جهت وجود ساختار متفاوت داده های فضایی، الگوریتمهای به کار رفته در این روشها با روشهای معمول در داده کاوی متفاوت است. در اینجا بدون پرداختن به جزئیات، تعاریف و مفاهیم مربوط به هر روش ارائه خواهد شد.

تعیین روند فضایی[۱۱]

در این روش تغییرات منظم خواص یک پدیده در طی جابجایی در فضا نسبت به نقطه شروع بررسی می گردد (Vera Karasova,2005 ,P15) و نتایج آن در قالب مدلهای کمی و کیفی قابل ارائه است. نمونه ای از روشهای معمول در تعیین روند فضایی مدل رگرسیون خطی فضایی است که رابطه بین خواص یک پدیده و موقعیت آنرا در فضا در قالب رابطه خطی بیان می کند. ضریب همبستگی فضایی (نظیر موران و جری) نمونه هایی دیگر از این روش محسوب می شوند که درجه وابستگی میان متغیرهای فضایی را نشان می دهند. در شکل شماره ۲ نمونه ای از الگوی تعیین روند فضایی نشان داده شده است که تغییر اجاره بهای کاربری تجاری را با فاصله گرفتن از مرکز شهر مشخص می کند.



شکل شماره ۲.خروجی یک مدل تعیین روند فضایی که تغییر اجاره بهای کاربری تجاری را در یک شهر نشان می دهد.
ماخذ :نگارنده

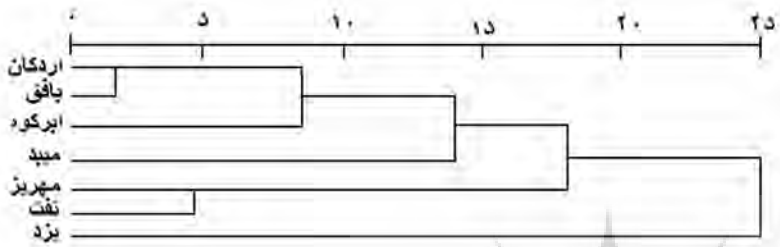
نمونه ای دیگر از به کارگیری این روشها در ارتباط با داده های مکانی-زمانی[۱۲] است که خواص منظم تغییرات یک پدیده را با توجه به پارامترهای مکان و زمان تحلیل می کند؛ لذا این روشها قابلیت پیش بینی در یک بازه زمانی را نیز خواهند داشت. برای نمونه مدلهای تحلیل جرم [۱۳] معمولا از داده های مکانی-زمانی

استفاده می نمایند و به وسیله آن اطلاعات مکانی و زمانی مربوط به رخداد جرم و بزه کاری را بررسی و تحلیل می نمایند.

خوشه بندی[14]

خوشه بندی فرآیند تقسیم مجموعه ای از داده ها به کلاسهایی به نام خوشه است که توسط این رویه فهم ساختار داده ها و گروهبندی آنها ساده تر می گردد(جعفر محمدی,32,1382). هر خوشه شامل یک سری از داده های مشابه است که به صورت یک گروه رفتار می کنند. نمونه ای از روشهای خوشه بندی مدل K-means است. در این روش مجموعه داده ها براساس خواص تعریف شده برای آنها به K خوشه تقسیم می شوند. اعضای درون هر خوشه بیشترین شباهت را با یکدیگر و بیشترین تفاوت را با اعضای خوشه های دیگر دارند. نوع دیگر روش خوشه بندی، خوشه بندی سلسله مراتبی[15] است که نسبت به روش K-means ، روشی نظارت نشده محسوب می شود؛ به این مفهوم که تعداد خوشه ها از قبل مشخص نیست و داده ها بر اساس تشابهشان در یک سلسله مراتب از زیر مجموعه های همگن توزیع می گردند. در شکل شماره ۳ نمونه ای از مدل سلسله مراتبی خوشه بندی ارائه گردیده است.

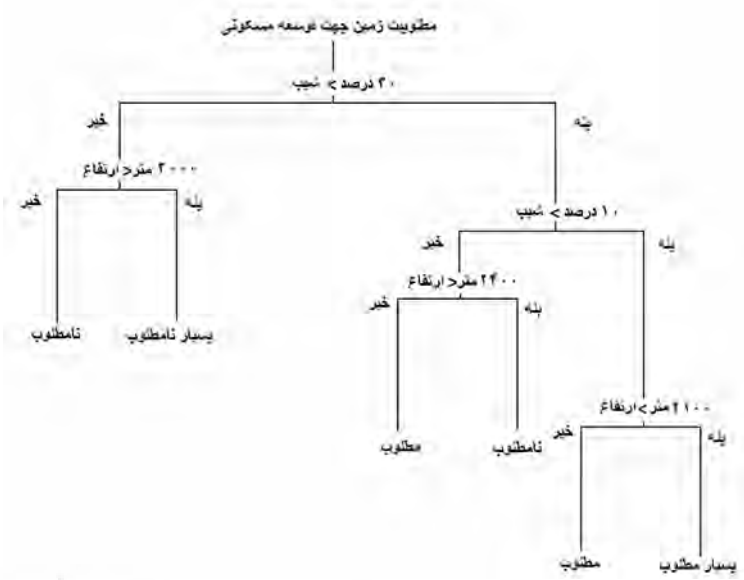
در این نمودار شهرهای منتخب استان یزد بر اساس برخی شاخصهای توسعه یافتگی در یک سلسله مراتب از زیرمجموعه های همگن نشان داده شده اند. برای نمونه شهر اردکان بیشترین شباهت را با شهر بافق دارد ، و یا شهر یزد بیشترین تفاوت را با سایر شهر ها داشته و لذا یک شاخه منفرد را به خود اختصاص داده است. هر چه اعضای مجموعه از نظر شاخصهای مورد بررسی همگنتر باشند تعداد خوشه ها کمتر و اختلاف میان اعضا بر روی محور افقی کوچک تر است. از خروجیهای دیگر روش خوشه بندی ماتریس مجاورت[16] است که میزان تفاوت میان اعضا را در یک ماتریس متقارن به شکل کمی نشان می دهد. روشهای خوشه بندی در مدلهای برنامه ریزی منطقه ای و ناحیه ای قابلیت استفاده زیادی دارند. برای مثال در پیشنهاد مربوط به تعیین حوزه های همگن و هم پیوند شهری و روستایی می توان از روشهای خوشه بندی سلسله مراتبی استفاده نمود.



شکل شماره ۳ : نمودار خوشه بندی سلسله مراتبی شهرستانهای استان یزد،۱۳۷۵(حکمت نیا ،حسن،۲۳۹،۱۳۸۲)

درخت تصمیم گیری[17]

این روش توسعه یافته روش خوشه بندی محسوب می گردد، چرا که این امکان را فراهم می سازد تا یک متغیر وابسته، بر اساس متغیرهای پیشگو کلاسه بندی و یا پیش بینی شود. اصولی که برای اجرای الگوریتم کلاسه بندی به کار می رود به قواعد تصمیم گیری[18] معروف است که به وسیله آن گره ها و شاخه های یک درخت تصمیم گیری تعیین می گردد(جعفر محمدی,20,1382). از انواع درختان تصمیم گیری می توان به مدل CART[19] اشاره کرد که نوعی درخت تصمیم گیری خطی محسوب می شود. مزیت مدل CART نسبت به روشهای دیگر درخت تصمیم گیری این است که قابلیت به کارگیری متغیرها در انواع مقیاسها (اسمی، رتبه ای و فاصله ای) را دارد. در شکل شماره ۴ نمونه ای از مدل CART نشان داده شده است.



شکل شماره۴: تعیین مطلوبیت زمین جهت توسعه مسکونی توسط دو عامل شیب و ارتفاع، ماخذ :نگارنده

در این مدل میزان مطلوبیت زمین جهت توسعه مسکونی در یک ناحیه کوهستانی مورد ارزیابی قرار گرفته و میزان این مطلوبیت در قالب یک متغیر کیفی نشان داده شده است. متغیرهای پیشگو (مستقل) در این مدل میزان شیب و ارتفاع هستند. با دنبال کردن سرشاخه ها به سمت ریشه درخت می توان خصوصیات زمینهای مطلوب و یا نامطلوب را مشاهده نمود. برای مثال زمین بسیار مطلوب از نظر توسعه مسکونی در این ناحیه زمینی است که در شیب کمتر از ۱۰ درصد و ارتفاع کمتر از ۲۱۰۰ متر قرار داشته باشد. روش درخت تصمیم گیری یکی از مهمترین ابزارهای پشتیبان تصمیم گیری مکانی در ارتباط با مدلهای چند معیاره محسوب می شود. امروزه با توسعه مدلهای تصمیم سازی و ارائه رهیافتهای ترکیبی[20] در جهت افزایش قابلیتها، استفاده از روشهایی نظیر درخت تصمیم گیری در ترکیب با سایر روشها نظیر تصمیم گیری بر اساس منطق فازی، و یا مدل AHP صورت می پذیرد.

نتیجه

در این مقاله تعاریف و مفاهیم داده کاوی فضایی بیان گردیده و بعضی روشهای معمول در آن با ذکر مثالهایی ارائه شد. بدون شک یکی از اهداف اساسی در استفاده از چنین ابزارهایی استخراج اطلاعات ارزشمند و دستیابی به دانش لازم در یک فرآیند تصمیم سازی است. لذا آنچه که در داده کاوی فضایی مورد تاکید است انطباق آن با مدلهای تصمیم سازی، بهبود کارکرد مفهومی و در نهایت تاکید بر جنبه های کاربردی و قابل تفسیر است. با توسعه سامانه های پردازش اطلاعات مکانی و فراگیر شدن آن در زنجیره تصمیم سازی انتظار می رود که بکارگیری روشهای داده کاوی فضایی به عنوان یک نیاز ضروری بیش از پیش معمول گردد.

منابع

- حکمت نیا، حسن (۱۳۸۵)کاربرد مدل در جغرافیا با تاکید بر برنامه ریزی شهری و ناحیه ای، انتشارات علم نوین
- محمدی، جعفر (۱۳۸۲)طراحی یک زبان تولید پرس‌وجوی انعطاف‌پذیر برای داده‌کاوی(FlexQG),دانشگاه علم وصنعت ایران
- Jukka M. Krisp, (2008) Visual Spatial Data Mining, München, Germany,
- Martin Ester, Hans-Peter Kriegel, Jörg Sander (2001) Algorithms and Applications for Spatial Data Mining, University of Munich.Germany.
- MohammedA.Sharifi and Vasilios Retsios (2004) Site selection for waste disposal through spatial multiple criteria decision analysis, Journal of Telecommunication and Information Technology, Netherlands.
- Pamela Jean Pratt, David H.Olsen, (2003) An approach to targeted marketing using spatial data mining at Camp Bell Scientific, Inc, Utah State University, USA.
- Vera Karasova (2005) Spatial data mining as a tool for improving geographical models, Helsinki university of technology.
- Verbila David L,(2001) Practical GIS Analysis, London, England.
- Xiaobai Yao,(2003) Research Issues in Spatio-temporal Data Mining, Department of Geography, University of Georgia

[1] -Data Mining
[2] -Knowledge Discovery from Database
[3] - Machine Learning
[4] -Multidimensional Data Analysis
[5] - Spatial Data
[6] - Non Spatial Data
[7] -Data set
[8] - Adjacency
[9] -Containment
[10] - proximity
[11] - spatial Trend Detection
[12] - Spatio-Temporal
[13] - Crime Analysis
[14] -Clustering
[15] -Hierarchical Clustering
[16] - Proximity Matrix
[17] - Decision Tree
[18] - Decision Rules
[19] - Classification and Regression Tree
[20] - Hybrid Approach