نشریه علمی- ترویجی مهندسی نقشه‌برداری و اطلاعات مکانی ، دوره هشتم، شماره ۲، خرداد ماه ۱۳۹۶

ارائه‌ی روش‌های جدید هرس کردن قوانین طبقه‌بندی انجمنی (موردکاوی: قوانین مرتبط با بهینگی مکان استقرار بانک‌های شهر تهران)

غدیر عشورنژاد [*1]، انیس عشورنژاد[2]، آرا تومانیان[3]

[1] دانشجوی دکتری تخصصی سنجش از دور و سیستم‌های اطلاعات مکانی - دانشکده جغرافیا - دانشگاه تهران
ashournejad@ut.ac.ir

[2] کارشناس ارشد مهندسی فناوری اطلاعات - موسسه‌ی آموزش عالی پویندگان دانش چالوس
anis.ashournejad@email.com

[3] استادیار دانشکده جغرافیا - دانشگاه تهران
a.toomanian@ut.ac.ir

(تاریخ دریافت آبان ۱۳۹۵، تاریخ تصویب دی ۱۳۹۵)

چکیده

با پیشرفت تکنولوژی در کشف داده، حجم داده‌ها بسیار افزایش یافته است. به همین دلیل روش‌های داده‌کاوی از جمله روش طبقه‌بندی انجمنی برای استخراج دانش از منابع داده‌ای حجیم بکار گرفته شد. در طبقه‌بندی انجمنی از قوانین انجمنی برای طبقه‌بندی داده‌ها استفاده می‌شود. بعد از تولید قوانین طبقه‌بندی به دلیل زیاد بودن تعداد آن‌ها، از روش‌های هرس کردن برای کنار گذاشتن قوانین اضافی استفاده می‌شود. در این تحقیق از قوانین طبقه‌بندی انجمنی برای تعیین رابطه‌ی بین مکان المان‌های شهری و بهینگی مکان استقرار شعب بانک‌ها وموسسات مالی و اعتباری شهر تهران استفاده می‌شود. زیرا مکان این المان‌های شهری تاثیر زیادی در تعیین یک مکان بهینه برای بانک‌ها دارند و بهینه بودن مکان بانک‌ها باعث سودآوری بیشتری برای آن‌ها خواهد شد. بعد از تولید قوانین طبقه‌بندی انجمنی، چهار روش جدید هرس کردن برای بهبود روش‌های قدیمی معرفی می‌گردد که روش‌های جدید هرس قوانین اضافی بدون کلاس، هرس قوانین اضافی بر حسب پوشش و هرس قوانین اضافی بر حسب پوشش بدون کلاس، به ترتیب موجب کاهش میانگین قوانین به تعداد ۱۲، ۴۳۵/۲ و ۴۳۷/۲ قانون نسبت به روش قدیمی هرس قوانین اضافی می‌شود و روش جدید پوشش با قوانین طولانی موجب بهبود میانگین دقت به میزان ۰/۲۰۹۱ درصد نسبت به روش قدیمی پوشش پایگاه داده شده است.

واژگان کلیدی: طبقه‌بندی انجمنی، هرس کردن قوانین طبقه‌بندی انجمنی، بانک‌ها و موسسات مالی و اعتباری

* نویسنده رابط

۱- مقدمه

پیشرفت تکنولوژی در کشف و ذخیره‌سازی داده، موجب رشد پایگاه داده‌های بزرگ شد. با زیاد شدن حجم داده‌ها، توانایی انسان در تفسیر و استفاده از آنها کاهش یافت. به همین دلیل کشف دانش و داده‌کاوی ظاهر شد. استخراج خودکار دانش از منابع داده‌ای حجیم طی یک فرآیند غیربدیهی مشخص، داده‌کاوی نامیده می‌شود [۱]. داده‌کاوی روش‌های مختلفی دارد که از آن جمله می‌توان به طبقه‌بندی[۱]، خوشه‌بندی[۲]، قوانین انجمنی[۳] و غیره اشاره کرد [۲].

روش طبقه‌بندی انجمنی[۴] هم روش جدیدی در داده‌کاوی است که از ادغام دو روش طبقه‌بندی و قوانین انجمنی به وجود آمده است [۳]. قوانین طبقه‌بندی ارتباط بین ویژگی‌های موجود در مجموعه داده و ویژگی کلاس را به صورت قانون نشان می‌دهند [۱-۴]. در قوانین انجمنی هم می‌توان گفت که وجود چه اشیایی بر وجود چه مجموعه اشیای دیگری اثرگذار است [۵-۶]. ترکیب این دو روش، روش طبقه‌بندی انجمنی را تشکیل می‌دهد. بکارگیری قوانین انجمنی برای طبقه‌بندی می‌تواند باعث بهبود دقت طبقه‌بندی شود [۷].

الگوریتم‌های طبقه‌بندی انجمنی به طور طبیعی مجموعه‌ی بزرگی از قوانین را استخراج می‌کنند، زیرا روش‌های کشف قوانین انجمنی که ترکیب همه‌ی مقادیر ویژگی‌ها را در پایگاه داده در نظر می‌گیرند برای کشف قوانین استفاده می‌شوند. به همین دلیل توجه زیادی به کاهش تعداد قوانین تولیده شده می‌شود. حذف قوانین اضافی که به آن هرس کردن[۵] قوانین می‌گویند، می‌تواند فرآیند طبقه‌بندی را دقیق‌تر و موثرتر نماید [۸]. به همین دلیل تاکنون روش‌های مختلفی برای هرس کردن قوانین طبقه‌بندی انجمنی[۶] و کنار گذاشتن قوانین اضافی ارائه شده است.

طبقه‌بندی و قوانین انجمنی در حوزه‌های مختلف از جمله کشاورزی، جرم شناسی، مدیریت شهری و غیره کاربرد دارد [۹-۱۰-۱۱]. در این تحقیق برای تعیین بهینه یا غیربهینه بودن مکان استقرار شعب بانک‌ها و موسسات مالی و اعتباری از روش طبقه‌بندی انجمنی استفاده خواهد شد. در جوامع امروزی بانک‌ها و موسسات مالی و اعتباری به عنوان مهمترین بخش اقتصادی جوامع بشری روز به روز در حال گسترش می‌باشند. یکی از مولفه‌های تاثیرگذار بر تجهیز منابع پولی در بانکداری نوین، توجه به مطلوبیت محل استقرار مکانی بانک‌ها و موسسات مالی و اعتباری می‌باشد. تصمیم‌گیری درباره‌ی مکان بهینه در موفقیت موسسه و نیل به اهداف آن و جلب مشتریان موثر می‌باشد [۱۲]. برای تحقق موفقیت در بانکداری می‌توان از روش‌های داده‌کاوی استفاده کرد [۱۳] و با بهره‌گیری از این روش‌هایی همچون قوانین انجمنی و طبقه‌بندی، قوانینی را که نشاندهنده‌ی ارتباط المان‌های شهری و مکان بهینه‌ی بانک می‌باشد تولید نمود. رویکرد سنتی که در تحقیقات گذشته برای مکان‌یابی شعب بانک‌ها استفاده شده است، مکانیابی نقاط بهینه با استفاده از نقشه‌ها و لایه‌های جغرافیایی موجود در محیط GIS می‌باشد [۱۴]. در این تحقیقات تک‌تک مکان‌های بهینه به صورت نقطه‌ای یا منطقه‌ای مشخص می‌شوند و برای شهر بزرگی مثل تهران، یک لیست طولانی از مناطق یا نقاط بهینه تشکیل می‌شود [۱۵-۱۶]. اما در این تحقیق سعی می‌شود که این امر با استفاده از قوانین طبقه‌بندی انجمنی برای کل شهر به صورت یکجا انجام شود.

در این مقاله در بخش ۲ مروری بر طبقه‌بندی انجمنی و روش‌های قدیمی هرس کردن قوانین طبقه‌بندی انجمنی می‌شود. در بخش ۳ چهار روش جدید برای هرس کردن قوانین طبقه‌بندی انجمنی ارائه می‌شود. در بخش ۴ هم این روش‌ها مورد ارزیابی قرار می‌گیرند. بخش ۵ هم شامل نتیجه‌گیری و پیشنهادات می‌باشد.

۲- مروری بر طبقه‌بندی انجمنی و روش‌های قدیمی هرس کردن

طبقه‌بندی انجمنی استفاده از روش قوانین انجمنی برای تولید قوانین طبقه‌بندی می‌باشد [۱۷]. تولید قوانین طبقه‌بندی انجمنی طی چند مرحله انجام می‌شود. ابتدا با استفاده از یکی از الگوریتم‌های قوانین انجمنی به نام‌های Apriori یا FP-Growth، مجموعه اشیای مکرر از مجموعه داده‌های آموزشی استخراج می‌شوند. مجموعه اشیایی که

---

۱ Classification
۲ Clustering
۳ Association Rule
٤ Associative Classification (AC)
٥ Pruning
٦ Associative Classification Rules

شرط حداقل مقدار پشتیبانی را رعایت کنند مجموعه اشیای مکرر نامیده می‌شوند [۱]. در مرحله‌ی بعد طبقه‌بندی برای مجموعه اشیای مکرر صورت می‌گیرد و قوانین طبقه‌بندی انجمنی تولید می‌شوند. سپس یک زیرمجموعه از این قوانین با استفاده از تکنیک‌های هرس کردن انتخاب می‌شود و در نهایت، کیفیت قوانین طبقه‌بندی تولید شده با داده‌های آزمایشی تعیین می‌شود [۳-۸].

برای تولید قوانین طبقه‌بندی انجمنی ابتدا مقدار پشتیبانی و اطمینان مجموعه اشیای مکرر برای هر کلاس محاسبه می‌شود. فرض کنید f یک مجموعه شی مکرر باشد، مقدار پشتیبانی و اطمینان آن از فرمول ۱ و ۲ محاسبه می‌شود:

$$support(f) = \frac{No\ of\ Occurrence\ of\ f\ with\ class}{Total\ number\ of\ data\ set} \quad (۱)$$

$$confidence(f) = \frac{No\ of\ Occurrence\ of\ f\ with\ class}{No\ of\ Occurrence\ of\ f} \quad (۲)$$

برای محاسبه مقدار پشتیبانی، تعداد وقوع مجموعه شی مکرر f با کلاس مورد نظر را بر تعداد کل نمونه‌های آموزشی موجود در مجموعه داده تقسیم می‌کنند. مقدار اطمینان هم با تقسیم تعداد وقوع مجموعه شی مکرر f با کلاس مورد نظر بر تعداد رخداد مجموعه شی f (بدون در نظر گرفتن کلاس آن) محاسبه می‌شود. برای هر تعداد مقدار ویژگی کلاس، هر مقداری که برای مجموعه شی مکرر مورد نظر دارای مقدار پشتیبانی و اطمینان بالاتری باشد به عنوان کلاس آن شناخته می‌شود و در سمت راست قانون قرار گیرد (در سمت چپ قانون، مجموعه شی مکرر قرار می‌گیرد) و بدین صورت قانون طبقه‌بندی برای آن مجموعه شی مکرر تولید می‌شود [۱۸].

همانطور که گفته شد برای حذف قوانین اضافی در طبقه‌بندی انجمنی باید قوانین را هرس کرد. برای هرس کردن قوانین اضافی، ابتدا باید آنها را رتبه‌بندی کرد. رتبه‌بندی قوانین به روش‌های مختلفی انجام می‌شود [۳-۱۹]. معمولا رتبه‌بندی قوانین به ترتیب بر حسب مقدار بیشتر اطمینان و سپس مقدار بیشتر پشتیبانی و در نهایت بر اساس طول کمتر قوانین (تعداد ویژگی‌های کمتر) انجام می‌شود [۲۰].

برای هرس کردن قوانین، روش‌های مختلفی ارائه شده است که یکی از آنها هرس کردن قوانین اضافی[۱] نام دارد. در این روش ابتدا فرآیند تولید قانون انجام می‌شود. سپس مرحله‌ی ارزیابی برای هرس کردن قوانین انجام می‌شود به طوری که قانونی مانند I'→C از مجموعه قوانین حذف می‌شود در جایی که برخی قوانین کلی با رتبه‌ی بالاتر مانند I→C وجود دارند و I زیرمجموعه‌ی I' باشد. این روش هرس کردن، قوانین اضافی را به حداقل می‌رساند [۸-۲۱].

یکی دیگر از روش‌های هرس کردن قوانین، پوشش پایگاه داده[۲] نام دارد. این روش قوانین تولید شده را در برابر مجموعه داده‌ی آموزشی تست می‌کند و فقط قوانین باکیفیتی را حفظ می‌کند که حداقل یک نمونه‌ی آموزشی را که توسط قوانین با رتبه‌ی بالاتر در نظر گرفته نشده‌اند پوشش دهد. اگر قانونی هم با هیچ یک از نمونه‌های آموزشی تطبیق نداشته باشد، آن قانون حذف خواهد شد [۸-۲۱].

۳- روش‌های پیشنهادی هرس کردن قوانین طبقه‌بندی انجمنی

همانطور که گفته شد طبقه‌بندی انجمنی با رشد نمایی قوانین روبرو است که موجب کاهش کارایی سیستم و افزایش زمان اجرا می‌شود و نیاز است که از روش‌های هرس کردن برای کنار گذاشتن قوانین اضافی و حفظ قوانین مهم استفاده شود. در این بخش هم چهار روش جدید برای هرس کردن قوانین طبقه‌بندی انجمنی ارائه می‌شود.

۳-۱- روش پوشش با قوانین طولانی

اولین روش جدیدی که برای هرس کردن قوانین طبقه‌بندی انجمنی معرفی می‌شود روش پوشش با قوانین طولانی نام دارد. همانطور که در توضیح روش پوشش پایگاه داده گفته شد، رتبه‌بندی قوانین به ترتیب بر حسب مقادیر بیشتر اطمینان و پشتیبانی آنها و در نهایت طول کوتاه‌تر قوانین انجام می‌گیرد. در این رتبه‌بندی بسیار اتفاق می‌افتد که قوانین کلی و کوتاه در رتبه‌ی بالاتری نسبت به قوانین خاص و طولانی قرار می‌گیرند و تعداد زیادی از نمونه‌ها را پوشش می‌دهند. در حالی که یک قانون خاص و طولانی که

۱ Redundant Rule Pruning
۲ Database Coverage

دارای تعداد ویژگی‌های مشابه بیشتری با یک نمونه‌ی آزمایشی باشد بهتر و مطمئن‌تر می‌تواند برای تعیین کلاس آن استفاده شود زیرا دارای شرایط نزدیکتری به نمونه‌ی آزمایشی می‌باشد. بنابراین در روش جدید هرس کردن، رتبه‌بندی قوانین به ترتیب بر حسب سه معیار طول، اطمینان و پشتیبانی به صورت نزولی انجام می‌شود. در واقع در ابتدا قوانین با تعداد ویژگی‌های بیشتر در رتبه‌های ابتدایی و قوانین کوتاه در انتها قرار می‌گیرند. اگر طول قوانین یکسان باشد قانونی در رتبه‌ی بالاتر قرار می‌گیرد که دارای مقدار اطمینان بیشتری باشد. اگر مقدار اطمینان قوانین هم یکسان باشد، قانون با مقدار پشتیبانی بیشتر در رتبه‌ی بالاتر قرار داده خواهد شد.

به عنوان یک نمونه جدول ۱ یک مجموعه قوانین طبقه‌بندی انجمنی را که شامل شش قانون تولید شده از مجموعه داده‌ی مکانیابی بانک‌ها و موسسات مالی و اعتباری می‌باشد و به ترتیب بر حسب طول، مقدار اطمینان و مقدار پشتیبانی به صورت نزولی مرتب شده‌اند نشان می‌دهد.

جدول۱- مرتب‌سازی قوانین به ترتیب بر حسب طول، اطمینان و پشتیبانی به صورت نزولی

شماره قانون	قانون	مقدار پشتیبانی	مقدار اطمینان
۱	DAFTARASNAD^ NEZAMI^ BIMARESTAN^ DARMANGA → 2	0.476	0.836
۲	SINAMA^PARK^ DABIRESTAN^ DANESHGA → 1	0.632	0.769
۳	PARK^ DABIRESTAN^ DANESHGA → 1	0.83	0.782
۴	DAROKHANE^ DARMANGA → 1	0.567	0.897
۵	DAROKHANE → 2	0.43	0.575
۶	DABIRESTAN → 1	0.288	0.575

بعد از مرتب‌سازی قوانین، پوشش نمونه‌های آموزشی توسط قوانین مورد بررسی قرار می‌گیرد. نمونه‌هایی که توسط قوانین پوشش داده شوند از مجموعه داده آموزشی حذف می‌شوند. اگر یک قانون هیچکدام از نمونه‌های آموزشی باقیمانده را پوشش ندهد، آن قانون از مجموعه قوانین طبقه‌بندی کنار گذاشته خواهد شد. خروجی این روش، مجموعه قوانین پوشش‌دهنده‌ی نمونه‌ها می‌باشد.

جدول۲ قوانین حاصل از اجرای روش پوشش با قوانین طولانی را بر روی جدول۱ نشان می‌دهد. همانطور که از جدول۲ نمایان است از قوانین مرتب شده در جدول۱، پنج قانون ۱، ۲، ۴، ۵ و ۶ پوشش دهنده‌ی نمونه‌ها می‌باشند و مقدار پوشش آنها غیرصفر می‌باشد اما قانون شماره‌ی۳ به دلیل پوشش ندادن هیچ نمونه‌ای، از مجموعه قوانین حذف می‌شود.

تعداد نمونه‌های پوشش داده شده توسط هر یک از قوانین باقیمانده تحت عنوان مقدار پوشش در جدول ۲ قابل مشاهده است.

جدول۲- قوانین حاصل از اجرای روش پوشش با قوانین طولانی

شماره قانون	قانون	مقدار اطمینان	مقدار پشتیبانی	مقدار پوشش
۱	DAFTARASNAD^ NEZAMI^ BIMARESTAN^ DARMANGA → 2	0.476	0.836	20
۲	SINAMA^PARK^ DABIRESTAN^ DANESHGA → 1	0.632	0.769	5
۳	DAROKHANE^ DARMANGA → 1	0.567	0.897	5
۴	DAROKHANE → 2	0.43	0.575	8
۵	DABIRESTAN → 1	0.288	0.575	6

۳-۲- روش هرس کردن قوانین اضافی بدون در نظر گرفتن کلاس قوانین

دومین روش جدید هرس کردن، روش قدیمی هرس کردن قوانین اضافی، این بار بدون در نظر گرفتن شرط یکسان بودن کلاس قوانین (سمت راست آنها) می‌باشد. مرتب‌سازی قوانین در این روش به ترتیب بر حسب مقادیر بالاتر اطمینان و سپس پشتیبانی و در نهایت بر حسب طول کمتر قوانین است. جدول۳ مرتب‌سازی قوانین را برحسب این سه معیار نشان می‌دهد. در این روش هنگام ارزیابی برای هرس کردن قوانین، قانونی مانند $I' \rightarrow C'$ از مجموعه قوانین حذف می‌شود در جایی که برخی قوانین کلی با رتبه‌ی بالاتر مانند $I \rightarrow C$ وجود دارند و I زیرمجموعه‌ی I' باشد. در این روش هرس کردن، ممکن است که مقدار C و C' یکسان نباشد. جدول۴ اجرای این روش هرس کردن را بر روی جدول۳ نشان می دهد.

جدول۳- مرتب‌سازی قوانین به ترتیب بر حسب مقادیر بالاتر اطمینان و سپس پشتیبانی و در نهایت بر حسب طول کمتر قوانین

شماره قانون	قانون	مقدار پشتیبانی	مقدار اطمینان
۱	DABIRESTAN → 1	0.288	0.975
۲	DAROKHANE → 2	0.43	0.955
۳	DAROKHANE^ DARMANGA → 1	0.567	0.897
۴	DAFTARASNAD^ NEZAMI^ BIMARESTAN^ DARMANGA → 2	0.476	0.836
۵	PARK^ DABIRESTAN^ DANESHGA → 1	0.83	0.782
۶	SINAMA^PARK^ DABIRESTAN^ DANESHGA → 1	0.632	0.769

جدول۴- قوانین حاصل از اجرای روش هرس کردن قوانین اضافی بدون در نظر گرفتن کلاس قوانین

شماره قانون	قانون	مقدار پشتیبانی	مقدار اطمینان
۱	DABIRESTAN → 1	0.288	0.975
۲	DAROKHANE → 2	0.43	0.955
۳	DAFTARASNAD^ NEZAMI^ BIMARESTAN^ DARMANGA → 2	0.476	0.836

۳-۳- هرس کردن قوانین اضافی بر حسب مقدار پوشش

همانطور که در بخش۲ گفته شد روش هرس کردن قوانین اضافی، قوانین کلی و با رتبه‌ی بالا را حفظ می‌کند و قوانین طولانی و با رتبه‌ی پایین را حذف می‌نماید به طوری که قوانین کلی زیرمجموعه‌ی قوانین طولانی می‌باشند. اما زمانی که در رتبه‌بندی قوانین، همه‌ی قانون‌های طولانی بالاتر از قوانین کوتاه و کلی قرار داشته باشند دیگر نمی‌توان این قوانین بلند و طولانی را حذف کرد. برای رفع این مشکل روش هرس کردن قوانین اضافی بر حسب مقدار پوشش معرفی می‌شود.

در این روش جدید هم ابتدا قوانین به ترتیب بر حسب تعداد ویژگی‌ها، مقدار اطمینان و مقدار پشتیبانی آنها به صورت نزولی مرتب می‌شوند. در واقع در این روش هم به قوانین طولانی اولویت داده می‌شود. سپس تعداد نمونه‌های پوشش داده شده توسط هر قانون همانند روش پوشش پایگاه داده تعیین می‌شود. در ادامه اولین قانون با قوانین بعد خود مقایسه می‌شود تا به یک قانون زیرمجموعه‌ی خود برسد. اگر مقدار پوشش قانون زیرمجموعه از مقدار پوشش قانون موردنظر بیشتر باشد و مقدار کلاس آنها هم یکی باشد، قانون موردنظر که دارای رتبه‌ی بالاتری هم می‌باشد حذف می‌شود و تعداد نمونه‌های تحت پوشش آن به مقدار پوشش قانون زیرمجموعه اضافه می‌گردد. اگر مقدار پوشش قانون زیرمجموعه کمتر از مقدار پوشش قانون طولانی باشد جستجو برای یافتن سایر قوانین زیرمجموعه تا آخرین قانون، ادامه خواهد یافت. این روند برای همه‌ی قوانین رتبه‌بندی شده انجام می‌شود و قوانینی که دارای قانون زیرمجموعه‌ای با تعداد پوشش بیشتر بوده‌اند حذف شده و سایر قوانین باقی می‌مانند.

فرض کنید جدول ۵ نتیجه‌ی مرتب‌سازی قوانین و تعیین تعداد نمونه‌های پوشش داده شده توسط هر قانون باشد. طبق روش هرس کردن قوانین اضافی بر حسب مقدار پوشش، قانون ۲ دارای یک زیرمجموعه می‌باشد و آن قانون ۵ می‌باشد، مقدار کلاس هر دو قانون یکی می‌باشد و مقدار پوشش قانون ۵ از مقدار پوشش قانون ۲ بیشتر است. بنابراین قانون ۲ که قانون طولانی و خاص می‌باشد حذف می‌شود و مقدار پوشش آن به قانون ۵ اضافه خواهد شد. حاصل این تغییرات در جدول ۶ قابل مشاهده است. در جدول ۵، قانون ۴ هم زیرمجموعه‌ی قانون ۳ می‌باشد و مقدار پوشش آن هم از قانون ۳ بیشتر است اما به دلیل آنکه مقدار کلاس دو قانون یکسان نمی‌باشد، قانون ۴ نمی‌تواند باعث حذف قانون ۳ شود.

جدول۵- مرتب‌سازی قوانین به ترتیب بر حسب طول، اطمینان و پشتیبانی به صورت نزولی و تعیین قوانین پوشش‌دهنده‌ی نمونه‌ها به همراه مقدار پوشش آنها

شماره قانون	قانون	مقدار اطمینان	مقدار پشتیبانی	مقدار پوشش
۱	DAFTARASNAD^ NEZAMI^ BIMARESTAN^ DARMANGA → 2	0.476	0.836	20
۲	SINAMA^ PARK^ DABIRESTAN^ DANESHGA → 1	0.632	0.769	5
۳	DAROKHANE^ DARMANGA → 1	0.567	0.897	5
۴	DAROKHANE → 2	0.43	0.575	8
۵	DABIRESTAN → 1	0.288	0.575	6

جدول۶- قوانین حاصل از اجرای روش هرس کردن قوانین اضافی بر حسب مقدار پوشش

شماره قانون	قانون	مقدار اطمینان	مقدار پشتیبانی	مقدار پوشش
۱	DAFTARASNAD^ NEZAMI^ BIMARESTAN^ DARMANGA → 2	0.476	0.836	20
۲	DAROKHANE^ DARMANGA → 1	0.567	0.897	5
۳	DAROKHANE → 2	0.43	0.575	8
۴	DABIRESTAN → 1	0.288	0.575	11

۴-۳- هرس کردن قوانین اضافی بر حسب مقدار پوشش بدون در نظر گرفتن کلاس قوانین

روش هرس کردنی که در این بخش معرفی می‌شود همانند روش هرس کردن قوانین اضافی بر حسب مقدار پوشش (بخش۳-۳) می‌باشد ولی شرط یکسان بودن کلاس (سمت راست) قوانین در این روش حذف می‌شود. در این روش هم ابتدا قوانین به ترتیب و به صورت نزولی بر حسب طول، مقدار اطمینان و مقدار پشتیبانی مرتب می‌شوند و مقدار پوشش هر قانون تعیین می‌گردد. سپس اولین قانون با قوانین بعد آن برای یافتن قانون زیرمجموعه مورد بررسی قرار می‌گیرند. اگر مقدار پوشش یک قانون زیرمجموعه از مقدار پوشش قانون موردنظر بیشتر باشد، قانون طولانی حذف می‌شود و تعداد نمونه‌های تحت پوشش آن به مقدار پوشش قانون زیرمجموعه اضافه می‌گردد. در این روش دیگر یکسان بودن کلاس قانون موردنظر و قانون زیرمجموعه، مهم نمی‌باشد.

دوباره جدول۵ را به عنوان نتیجه‌ی مرتب‌سازی و تعیین قوانین پوشش‌دهنده‌ی نمونه‌ها در نظر بگیرید. طبق روش هرس کردن که در این بخش معرفی شد قانون ۵، قانون ۲ را به دلیل بیشتر بودن مقدار پوشش خود حذف می‌کند. در ادامه قانون زیرمجموعه‌ی ۴ هم قانون ۳ را حذف خواهد کرد، زیرا مقدار پوشش آن از مقدار پوشش قانون ۳ بیشتر است. مقدار پوشش قانون ۳ هم به قانون ۴ اضافه می‌شود. در این روش دیگر مقدار کلاس قوانین مقایسه نمی‌شود. جدول۷ حاصل اجرای این روش بر روی جدول۵ می‌باشد.

جدول۷- قوانین حاصل از اجرای روش هرس کردن قوانین اضافی بر حسب مقدار پوشش بدون در نظر گرفتن کلاس قوانین

شماره قانون	قانون	مقدار اطمینان	مقدار پشتیبانی	مقدار پوشش
۱	DAFTARASNAD^ NEZAMI^ BIMARESTAN^ DARMANGA → 2	0.476	0.836	20
۲	DAROKHANE → 2	0.43	0.575	13
۳	DABIRESTAN → 1	0.288	0.575	11

۴- ارزیابی روش‌های جدید هرس کردن قوانین طبقه‌بندی انجمنی بر روی مجموعه داده‌ی بانک‌ها و موسسات مالی و اعتباری شهر تهران

در ادامه می‌خواهیم روش‌های جدیدی را که برای هرس کردن قوانین انجمنی معرفی شدند بر روی مجموعه داده‌ی بانک‌ها و موسسات مالی و اعتباری شهر تهران و برای فواصل مختلف پیاده کنیم. داده‌هایی که در این تحقیق استفاده می‌شوند مربوط به بهینه یا غیربهینه بودن مکان استقرار بانک‌ها و موسسات مالی و اعتباری شهر تهران می‌باشند. در اطراف هر بانک، مراکز و المان‌های شهری مختلف واقع شده‌اند که در این تحقیق تعداد ۱۷ المان شهری در قالب ۱۷ ویژگی در نظر گرفته می‌شوند. مقدار هر ویژگی به صورت فاصله‌ی آن المان شهری از شعبه‌ی بانک بر حسب متر می‌باشد. این المان‌های شهری شامل شرکت تعاونی، فروشگاه زنجیره‌ای، درمانگاه، مرکز خرید، اداره، سینما، دبیرستان، بیمارستان، هتل، پارک، موزه، مرکز نظامی، مسافرخانه، رادیولوژی، داروخانه، دفتر اسناد و دانشگاه می‌باشند. ویژگی هدف هم بهینه بودن (مقدار ۱) یا بهینه نبودن (مقدار ۲) مکان استقرار بانک یا موسسه مالی و اعتباری را نشان می‌دهد. این ۱۷ ویژگی و ویژگی هدف، ۲۵۵۱ رکورد مجموعه داده را تشکیل می‌دهند.

برای آغاز کار ابتدا باید مجموعه اشیای مکرر از مجموعه داده تولید شود. پیش از این کار، ۷۰ درصد داده‌ها که شامل ۱۷۸۶ رکورد می‌باشد به طور تصادفی انتخاب و به عنوان نمونه‌های آموزشی و ۳۰ درصد باقیمانده هم که ۷۶۵ رکورد است، به عنوان نمونه‌های آزمایشی برای مرحله‌ی تست قوانین، کنار گذاشته می‌شود. در ادامه نمونه‌های آموزشی به عملگر FP-Growth در RapidMiner داده می‌شود تا مجموعه اشیای

مکرر با استفاده از این عملگر تولید شوند. تولید مجموعه اشیای مکرر برای فاصله‌های ۲۵۰ متر، ۳۰۰ متر، ۳۵۰ متر، ۴۰۰ متر و ۴۵۰ متر اجرا می‌شود. حداقل مقدار پشتیبانی برای این عملگر، مقدار 0.01 در نظر گرفته می‌شود. بعد از تولید مجموعه اشیای مکرر، قوانین طبقه‌بندی انجمنی برای این مجموعه اشیای مکرر در هر شعاع اثرگذاری، طبق مطالب بخش ۲ تولید خواهند شد. این قوانین رابطه‌ی بین المان‌های شهری قرار گرفته در اطراف بانک‌ها و بهینه بودن یا نبودن مکان استقرار بانک را نشان می‌دهند.

بعد از تولید قوانین چهار روش جدید و دو روش قدیمی هرس کردن بر روی قوانین تولید شده اجرا می‌شود و میزان دقت و تعداد قوانین هر یک از این روش‌ها بر روی نمونه‌های آزمایشی که توسط برنامه و به زبان C# نوشته شده است محاسبه می‌گردد. تعداد قوانین و میزان دقت حاصل از هر یک از این روش‌ها برای فواصل ۲۵۰، ۳۰۰، ۳۵۰، ۴۰۰ و ۴۵۰ متر به صورت جدول ۸ و ۹ می‌باشد.

اولین روش جدید در جدول ۸، روش پوشش با قوانین طولانی می‌باشد. همانطور که در جدول دیده می‌شود روش پوشش با قوانین طولانی به دلیل قرار دادن قوانین طولانی در رتبه‌های بالاتر، تعداد قوانین بیشتری را نسبت به روش پوشش پایگاه داده شامل می‌شود.

طبق جدول ۸ دومین روش جدید، هرس کردن قوانین اضافی بدون در نظر گرفتن کلاس قوانین می‌باشد. این روش توانسته است تعداد قوانین را نسبت به روش هرس قوانین اضافی کاهش دهد. زیرا شرط یکسان بودن کلاس بین قانون کلی و قوانین بعد آن برداشته شده است.

جدول۸- تعداد قوانین حاصل از اجرای روش‌های ارائه شده و روش‌های قدیمی هرس کردن

روش جدید	روش جدید	روش جدید		روش جدید			
هرس قوانین اضافی بر حسب پوشش، بدون کلاس	هرس قوانین اضافی بر حسب پوشش	هرس کردن قوانین اضافی، بدون کلاس	هرس قوانین اضافی	پوشش با قوانین طولانی	پوشش پایگاه داده	تعداد قوانین اولیه	فاصله (متر)
23	23	217	218	386	171	710	250
61	65	427	435	573	283	1978	300
71	74	473	483	638	279	3743	350
98	100	625	639	666	287	6081	400
144	145	781	808	707	285	9426	450
79.4	81.4	504.6	516.6	594	261	4387.6	میانگین

برای دو روش هرس قوانین اضافی بر حسب مقدار پوشش و روش هرس قوانین اضافی بر حسب مقدار پوشش بدون در نظر گرفتن مقدار کلاس، همانطور که در جدول ۸ دیده می‌شود قوانین به مقدار قابل توجهی کاهش یافته‌اند. زیرا در این دو روش، ملاک حفظ یک قانون کلی و حذف قانون طولانی، بیشتر بودن تعداد رکوردهای آموزشی تحت پوشش قانون کلی نسبت به یک قانون طولانی و خاص می‌باشد. بنابراین با اجرای این دو روش جدید هرس کردن قوانین اضافی می‌توان تعداد قوانین را نسبت به روش قدیمی هرس قوانین اضافی به طور قابل ملاحظه‌ای کاهش و با کاهش قوانین نیز سرعت سیستم را افزایش داد.

البته باید به این نکته هم توجه کرد که تعداد قوانین حاصل از روش هرس کردن قوانین اضافی بر حسب مقدار پوشش بدون در نظر گرفتن مقدار کلاس، نسبت به روش هرس قوانین اضافی بر حسب مقدار پوشش، دارای کاهش بیشتری می‌باشد زیرا شرط یکسان بودن کلاس قوانین برداشته شده است.

حال در ادامه جدول ۹ میزان دقت حاصل از روش‌های جدید و قدیمی هرس کردن قوانین طبقه‌بندی انجمنی را به درصد نشان می‌دهد.

جدول۹- درصد دقت حاصل از روش‌های ارائه شده و روش‌های قدیمی هرس کردن

روش جدید	روش جدید	روش جدید				
هرس قوانین اضافی بر حسب پوشش، بدون کلاس	هرس قوانین اضافی بر حسب پوشش	هرس کردن قوانین اضافی، بدون کلاس	هرس قوانین اضافی	پوشش با قوانین طولانی	پوشش پایگاه داده	فاصله (متر)
73.7255	73.7255	73.0719	73.0719	73.4640	73.0719	250
74.5098	74.2484	73.8562	73.8562	74.3791	73.8562	300
73.3333	73.0718	74.1176	74.1176	73.8562	74.1176	350
73.4640	73.0719	74.1176	74.1176	73.9869	73.9869	400
74.2484	74.3791	74.1176	74.1176	74.2483	73.8562	450
73.8562	73.6993	73.8562	73.8562	73.9869	73.7778	میانگین

همانطور که از مقدار میانگین دقت روش پوشش با قوانین طولانی می‌توان فهمید این روش دارای میانگین دقت بالاتری نسبت به روش قدیمی پوشش پایگاه داده می‌باشد. زیرا در این روش سعی می‌شود که با در اولویت قرار دادن قوانین طولانی، ابتدا از این قوانین که دارای شرایط مشابه بیشتری با رکوردها می‌باشند و دانش

بیشتری را دربرمی‌گیرند برای تعیین مقدار کلاس رکوردهای آزمایشی استفاده شود.

روش هرس کردن قوانین اضافی بدون در نظر گرفتن کلاس، هم توانسته است با وجود کاهش تعداد قوانین، میزان دقت را نسبت به روش هرس قوانین اضافی حفظ نماید.

در روش هرس کردن قوانین اضافی بر حسب مقدار پوشش هم در مواردی مقدار دقت آن نسبت به روش هرس کردن قوانین اضافی کاهش یافته است، در حالی که تعداد قوانین آن نسبت به روش هرس قوانین اضافی دارای کاهش چشم‌گیری بوده است. زمانی که کاهش تعداد قوانین و افزایش سرعت سیستم اهمیت زیادی داشته باشد می‌توان از این کاهش دقت در هرس قوانین اضافی بر حسب مقدار پوشش، چشم‌پوشی کرد. روش هرس کردن قوانین اضافی بر حسب مقدار پوشش بدون در نظر گرفتن کلاس، هم علاوه بر کاهش بسیار قابل توجه تعداد قوانین، مقدار میانگین دقت آن با روش قدیمی هرس قوانین اضافی یکسان می‌باشد.

البته به این نکته هم باید توجه کرد که روش هرس کردن قوانین اضافی بر حسب مقدار پوشش بدون در نظر گرفتن کلاس، نسبت به روش هرس کردن قوانین اضافی بر حسب مقدار پوشش، علاوه بر کاهش بیشتر تعداد قوانین، دارای میانگین دقت بیشتری هم می‌باشد.

۵- نتیجه گیری

با استفاده از قوانین طبقه‌بندی انجمنی می‌توان ارتباط بین المان‌های شهری و مکان بهینه یا غیربهینه‌ی بانک‌ها و موسسات مالی و اعتباری را نشان داد. زمانی که قوانین تولید می‌شوند تعدادی از آنها با روش‌های هرس کردن کنار گذاشته می‌شوند. بنابراین تعداد قوانین نسبت به تعداد اولیه‌ی آنها بسیار کاهش می‌یابد. چهار روش جدید هرس کردن که در این تحقیق معرفی شدند موجب افزایش دقت یا کاهش قابل توجه تعداد قوانین می‌شوند. بنابراین می‌توان گفت که روش‌های جدید موجب بهبود روش‌های قدیمی خواهند شد. سپس با همین قوانین می‌توان مکان بهینه یا غیربهینه بانک را برای کل شهر تعیین کرد.

شکل ۱(a) نمودار تغییرات مربوط به میانگین تعداد قوانین هر یک از روش‌های قدیمی و جدید هرس کردن را نشان می‌دهد. همانطور که از شکل نمایان است روش جدید هرس قوانین اضافی بر حسب پوشش بدون در نظر گرفتن کلاس قوانین و بعد آن روش هرس قوانین اضافی بر حسب مقدار پوشش، دارای کمترین تعداد قوانین می‌باشند. زمانی که سرعت سیستم امر مهمی باشد، این روش‌ها می‌توانند مناسب باشند.

شکل ۱(b) هم نمودار تغییرات مربوط به میانگین دقت هر یک از روش‌های قدیمی و جدید هرس کردن را نشان می‌دهد. همانطور که در شکل نمایان است روش پوشش با قوانین طولانی دارای بالاترین مقدار میانگین دقت می‌باشد. زمانی که بالا بودن دقت امری ضروری است می‌توان از این روش استفاده کرد.

(a)

(b)

شکل ۱- (a) نمودار میانگین تعداد قوانین و (b) نمودار میانگین دقت، در هر یک از روش‌های قدیمی و جدید هرس کردن قوانین طبقه‌بندی انجمنی

پیشنهادات

با توجه به اینکه هرس کردن قوانین طبقه‌بندی انجمنی تاثیر مهمی در میزان کارایی سیستم و دقت کار خواهد داشت، بررسی تغییرات در تکنیک‌های مختلف هرس کردن می‌تواند در بالا رفتن میزان کارایی موثر باشد. به همین دلیل یکی از پیشنهادات این تحقیق تغییر در

ترتیب معیارهای رتبه‌بندی قوانین می‌باشد. در نظر گرفتن معیارهای دیگری همانند مقدار کلاس اکثریت، کاردنالیتی قوانین و یا انتخاب تصادفی یک قانون را هم می‌توان در رتبه‌بندی قوانین لحاظ کرد. همچنین می‌توان سه معیار طول، مقدار اطمینان و مقدار پشتیبانی را برای رتبه‌بندی قوانین در روش‌های دیگر هرس کردن بررسی کرد و تغییرات دقت و تعداد قوانین آنها را مورد بررسی قرار داد.

مراجع

[۱] صنیعی آباده، محمد، محمودی، سینا، طاهرپرور، محدثه (۱۳۹۱). داده‌کاوی کاربردی، تهران: انتشارات نیاز دانش، چاپ اول.

[2] Tan, P. N., Steinbach, M., Kumar, V. (2005). Introduction to Data Mining, Boston: Addison-Wesley Longman Publishing Co.

[3] Thabtah, F. (2007) “A review of associative Classification mining” Knowledge Engineering Review, Vol. 22, No. 1, pp. 37-65. ISSN 0269-8889.

[4] Phyu, Th. N. (2009) “Survey of Classification Techniques in Data Mining” Proceedings of the International MultiConference of Engineers and Computer Scientists, IMECS 2009, March 18 - 20Hong Kong.

[5] Guo, D., Mennis, J. (2009) “Spatial data mining and geographic knowledge discovery-An introduction” Computers, Environment and Urban Systems, Vol. 33, Issue 6, PP. 403-408.

[6] Zhao, Q., & Bhowmick, S. S. (2003) “Association rule mining: A survey” Nanyang Technological University, Singapore.

[7] Tang, Zh., Liao, Q. (2007) “A New Class Based Associative Classification Algorithm” IAENG International Journal of Applied Mathematics, Vol. 3, No. 2, pp. 685-689.

[8] Thabtah, F. (2005) “Rules pruning in associative classification mining” In Proceedings of the IBIMA Conference, pp. 7-15.

[9] Ding, Q., Ding, Q., Perrizo, W. (2008) “PARM-An Efficient Algorithm to Mine Association Rules From Spatial Data” IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS- PART B: CYBERNETICS, VOL. 38, NO. 6, pp. 1513-1524.

[10] Koperski, K., Han, J. (1995) “Discovery of Spatial Association Rules in Geographic Information Databases” Advances in Spatial Databases, pp 47-66.

[11] Sathyaraj, S. R., Thangavelu, A., Balasubramanian, S., Sridhar, R., Chandran, M., Prashanthi Devi, M. (2010) “C LUSTERED SPATIAL ASSOCIATION RULE TO EXPLORE LARGE VOLUMES OF GEOREFERENCED CRIME TO CRIME DATA” 3th INTERNATIONAL CONFERENCE ON CARTOGRAPHY AND GIS.

[۱۲] بامداد، ناصر، رفیعی مهرآبادی، نگار (۱۳۸۷). «بررسی رضایت مشتریان از کیفیت خدمات خودپرداز بانک‌ها» پژوهشنامه‌ی علوم انسانی و اجتماعی مدیریت، شماره‌ی ۳۱، صص ۵۸-۳۹.

[۱۳] سپهری، محمد مهدی، نوروزی، اشرف، تیمورپور، بابک، چوبدار، سروناز (۱۳۹۰). «کشف دلایل رویگردانی مشتری از خدمات بانکداری با ترکیب روش‌های داده‌کاوی و تحقیق پیمایشی» پژوهش‌های مدیریت در ایران، دوره ۱۵، شماره ۴، صص ۱۲۶-۹۷.

[۱۴] عشورنژاد، غدیر، فرجی سبکبار، حسنعلی، علوی پناه، سید کاظم، نامی، محمدحسن (۱۳۹۰). مکانیابی شعب جدید بانک‌ها و موسسات مالی و اعتباری با استفاده از فرآیند تحلیل شبکه‌ای فازی، مجله پژوهش و برنامه‌ریزی شهری، سال دوم، شماره‌ی هفتم، صص ۲۰-۱.

[۱۵] موسوی، ناصر (۱۳۸۰). «اولویت‌بندی و انتخاب مکان مناسب شعب بانک کشاورزی با استفاده از تکنیک تجزیه و تحلیل سلسله مراتبی (AHP) » دانشگاه تهران، دانشکده مدیریت.

[16] Boufounou, P. V. (1995). “Evaluating Bank Branch Location and Performance: A case Study” European Journal of Operational Research, Vol. 87, No. 2, pp. 389-402.

[17] Liu, B., Hsu, W., Ma, Y. (1998) “Integrating Classification and Association Rule Mining” International Conference on Knowledge Discovery and Data Mining, 1998, pp. 80-86.

[18] Pandey, K., Asati, Sh., Gagged, Gh., Jaisankar, N. (2013) “Associative Classification Approach for the Analysis of the Geospatial Data” The International Journal of Computer Science & Applications (TIJCSA), Vol. 2, No. 06, PP. 43-50.

[19] Wedyan, S. (2014) “Review and Comparison of Associative Classification Data Mining Approaches” World Academy of Science, Engineering and Technology, International Journal of Computer, Control, Quantum and Information Engineering Vol:8, No:1, pp. 34-45.

[20] Balaji, B. V., Rao, V. V. (2013) “Improved Classification Based Association Rule Mining” International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 5, pp. 2211-2221.

[21] Abu Mansour H. Y. (2012) “Rule pruning and prediction methods for associative classification approach in data mining” Doctoral thesis, University of Huddersfield.