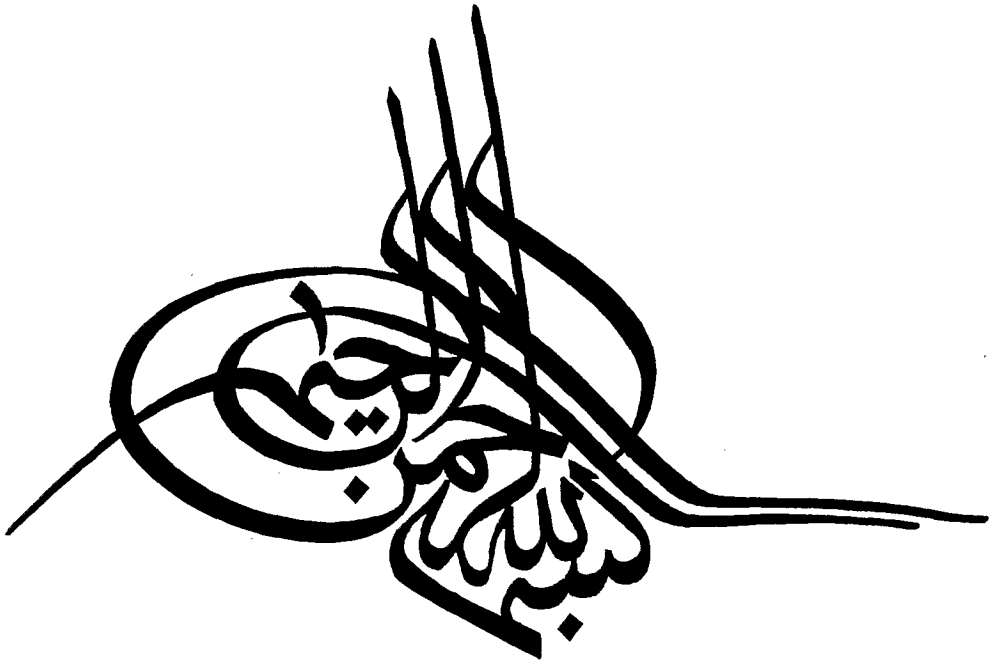

دانشکده مهندسی

پایان نامه کارشناسی ارشد در رشته ی مهندسی کامپیوتر- نرم افزار

استخراج مجموعه آیتم های فراوانی در محیط توزیع شده

بوسیله‌ی:
الهام پرنیان

استاد راهنما:
دکتر محمد هادی صدرالدینی

شهریورماه ۱۳۸۷

به نام خدا

استخراج مجموعه آیتم های فراوانی در محیط توزیع شده

به وسیله‌ی

الهام پرنیان

پایان نامه

ارائه شده به تحصیلات تکمیلی دانشگاه به عنوان بخشی
از فعالیت های تحصیلی لازم برای اخذ درجه کارشناسی ارشد

در رشته‌ی

مهندسی کامپیوتر (نرم افزار)

از دانشگاه شیراز

شیراز

جمهوری اسلامی ایران

ارزیابی شده توسط کمیته پایان نامه با درجه: **عالی**

دکتر محمد هادی صدرالدینی، استادیار بخش مهندسی و علوم کامپیوتر (رییس کمیته).....

دکتر غلامحسین دستغیبی فرد، استادیار بخش مهندسی و علوم کامپیوتر.....

دکتر فریبرز سبحان منش، استاد یار بخش مهندسی و علوم کامپیوتر.....

شهریور ۱۳۸۷

تقدیم به تمامی آموزگاران الفبای زندگی

به مادر، باران فداکاری

و به پدر، آسمان شکیبایی

سپاسگزاری

اکنون که به یاری خداوند متعال موفق به اتمام این پایان نامه شده‌ام، از زحمات اساتید بزرگوار جناب آقای دکتر محمد هادی صدرالدینی، دکتر غلامحسین دستغیبی فرد و دکتر فریبرز سبحان منش سپاسگزارم.

از تمامی دوستان و خانواده عزیزم که مشوقم بودند و همچنین از تمامی کادر بخش مهندسی و علوم کامپیوتر که در انجام این کار با اینجانب همکاری داشته‌اند، صمیمانه سپاسگزارم.

و از مرکز تحقیقات مخابرات ایران به دلیل حمایت مالی از این پایان نامه، کمال تشکر را دارم.

چکیده

استخراج مجموعه آیتم های فراوانی در محیط توزیع شده

بوسیله‌ی:

الهام پرنیان

حجم روزافزون داده‌ها در فایلها، پایگاه‌های داده و دیگر مخازن داده‌ای، توسعه ابزارهای تجزیه تحلیل و تفسیر داده‌ها و استخراج اطلاعات جالب از این داده‌ها را ایجاب می‌نماید. این اطلاعات در پروسه‌های تصمیم گیری سازمانها مورد استفاده قرار می‌گیرند. داده‌کاوی، کشف الگوهای پنهان و اطلاعات مفید از پایگاه داده‌هاست و یکی از قدم های مهم در پروسه کشف دانش است.

ارتباطات پنهان بین داده‌ها در بانک اطلاعاتی تراکنشی، قوانین وابستگی نامیده می‌شود. کشف قوانین وابستگی یکی از مهمترین شاخه‌های داده کاوی است که زمینه‌های کاربردی فراوانی دارد. درابتدا الگوریتمهای زیادی در زمینه کشف این قواعد به صورت مرکزی، ارایه شد. اما با ظهور سازمانهایی که به صورت جغرافیایی توزیع شده هستند و با توجه به حجم بالای ارتباطات شبکه ای و لزوم حفظ محرمانگی داده ها، کاوش قوانین وابستگی در محیط های توزیع شده هم مد نظر قرار گرفت.

کاوش مجموعه اقلام فراوان اولین قدم در کشف قوانین وابستگی است. به طور معمول این کار باعث ایجاد یک مجموعه بزرگ از مجموعه قلم‌ها می‌شود که تعداد زیادی از آنها اضافی هستند. در سال های اخیر، روشهایی در زمینه کاوش مجموعه قلم های تکرار شونده بسته پیشنهاد شده است. این مجموعه قلم ها، یک نمایش خلاصه و کامل از مجموعه قلم های تکرارشونده در پایگاه داده ارایه می‌دهند.

حجم ارتباطات و زمان لازم برای اجرا، از پارامترهای مهم در کشف قوانین وابستگی در محیط توزیع شده هستند. از آنجایی که ارسال مجموعه اقلام فراوان در یک محیط توزیع شده، حجم بالایی از ارتباطات شبکه ای را سبب می شود، ما در این پایان نامه با استفاده از تکنیک فشرده سازی مجموعه اقلام در یکی از معروفترین الگوریتم‌های توزیع شده به نام او. دی. ای. ام، یک روش جدید ایجاد کرده ایم و آن را با داده‌های واقعی مورد آزمایش قرارداده‌ایم. با استفاده از روش جدید حجم ارتباطات در الگوریتم او. دی. ای. ام به نحو قابل قبولی کاهش می‌یابد. نتایج ارزیابی صورت گرفته بر روی دو الگوریتم او. دی. ای. ام و الگوریتم جدید، کارایی برتر الگوریتم جدید را از دید حجم ارتباطات شبکه ای و زمان اجرا نشان می‌دهد. این روش را می توان در هر الگوریتم توزیع شده به کار برد.

فهرست مطالب

فهرست شکل ها

عنوان صفحه

فهرست جداول

فصل اول
داده کاوی و کاوش قواعد وابستگی

۱- داده کاوی و کاوش قواعد وابستگی

۱-۱ مقدمه:

امروزه حجم زیادی از اطلاعات در پایگاه های داده مربوط به شرکت ها، مراکز تجاری و دولتی ذخیره می شود. استفاده معمول از این داده ها انجام عملیات گزارش گیری برای کاربران و مدیران است. استفاده دیگری که امروزه از حجم انبوه داده های ذخیره شده در پایگاه های داده و انبار های داده می شود انجام عملیات داده کاوی است. در عملیات داده کاوی ما به دنبال الگوهای پنهان و احتمالاً سودمند هستیم. برخی از این الگوها در انجام تصمیم گیری ها می توانند به مدیران کمک کرده و یا برای کاربران و مشتریان مفید باشند. الگوهایی که در عملیات مختلف داده کاوی پیدا می شوند انواع گوناگونی دارند. یک نوع پرکاربرد و معروف از این الگوها قوانین یا قواعد وابستگی[1] هستند. معروف ترین کاربرد قوانین وابستگی در تحلیل سبد خرید برای فروشگاه ها و مراکز تجاری است. به عنوان مثال پس از کاوش پایگاه داده مربوط به یک فروشگاه زنجیره‌ای ممکن است مشخص شود که مشتریانی که از این فروشگاه شیر می خرند به احتمال ۶۰٪ کره نیز خواهند خرید. یافتن چنین قواعدی می تواند در چیدن قفسه ها، راهنمایی مشتریان و مسائل مدیریتی سودمند باشد. عملیات یافتن قواعد وابستگی را کشف یا کاوش قواعد وابستگی گویند.

امروزه استفاده از پایگاه داده های توزیع شده[2] به ویژه برای سازمان ها و مراکزی که در یک گستره ی جغرافیایی قرار دارند، گسترش روز افزونی پیدا کرده است. با ذخیره شدن حجم زیادی از داده ها در این نوع پایگاه داده ها همانند پایگاه داده های متمرکز بحث داده کاوی در آنها مطرح شده است. کاوش قواعد وابستگی به عنوان یکی از پرکاربردترین روش های داده کاوی در پایگاه های داده توزیع شده از اهمیت خاصی برخوردار است. ما در این پایان نامه در ابتدا مروری بر چگونگی انجام عملیات داده کاوی خواهیم داشت. از میان عملیات مختلف داده کاوی کاوش قواعد وابستگی را به دقت بررسی خواهیم کرد. روش ها و الگوریتم های ارائه شده برای کاوش قواعد وابستگی در حالت متمرکز را بررسی خواهیم نمود. سپس الگوریتمهای کاوش قواعد وابستگی فشرده و در ادامه انجام عملیات داده کاوی در محیط توزیع شده را مورد مطالعه قرار داده و تفاوت های آن با حالت متمرکز را بررسی می کنیم. کاوش قواعد وابستگی در محیط توزیع شده به عنوان هدف اصلی این پایان نامه مدنظر قرار گرفته است و روش ها و الگوریتم های ارائه شده در این زمینه مورد مطالعه دقیق قرار گرفته و با هم مقایسه می شوند. در نهایت ما الگوریتم جدیدی برای

---

[1] Association Rules Mining
[2] Distributed DataBases

کاوش قوانین وابستگی در محیط توزیع شده ارائه خواهیم داد که از کارائی بهتری نسبت به الگوریتم های قبلی برخوردار باشد. الگوریتم جدید را پیاده سازی کرده و بر روی داده های توزیع شده ارزیابی خواهیم کرد.

۱-۲ داده کاوی

اخیراً سرعت تولید و جمع آوری داده ها در پایگاههای داده به صورت روزافزونی زیاد شده است. استفاده گسترده از بارکد در فروش تولیدات، کامپیوتری شدن تعداد زیادی از کارهای تجاری، اداری و دولتی و پیشرفت در زمینه ابزار جمع آوری داده ها ما را با حجم زیادی از داده ها مواجه کرده است. امروزه پایگاه داده ها در زمینه های تجاری، اداری، علمی، مهندسی و زمینه های دیگر استفاده می شوند. تعداد چنین پایگاههای داده ای به دلیل نیاز مبرم به جمع آوری و گزارش گیری از داده ها و همچنین وجود سیستم های پایگاه داده قدرتمند در حال افزایش است. این چنین رشد فزاینده ای در داده ها و پایگاههای داده یک نیاز ضروری برای ابزار جدیدی که بتوانند به طور هوشمند و خودکار این داده ها را پردازش کرده و به اطلاعات و به دانش های سودمند تبدیل کنند، بوجود آورده است. در نتیجه داده کاوی به یک زمینه تحقیقاتی با اهمیت فراوان تبدیل شده است.

از داده کاوی همچنین به عنوان کشف دانش در پایگاههای داده یاد می شود، که به معنی فرآیند استخراج اطلاعات غیر صریح و احتمالاً سودمندی از پایگاههای داده است که در گذشته نا شناخته و پنهان بوده اند. با انجام عملیات داده کاوی دانش های جالب و گاه غیر منتظره، نظم ها و الگوهای پنهان، یا اطلاعات سطح بالا می توانند از مجموعه ای از داده های مرتبط در پایگاه داده استخراج شوند و از زوایای مختلف مورد بررسی قرار گیرند. بنابراین پایگاههای داده حجیم را می توان به عنوان منبعی غنی و قابل اطمینان برای تولید و وارسی برخی دانش ها و اطلاعات در نظر گرفت.

کاوش اطلاعات و دانش از پایگاههای داده حجیم به عنوان یک موضوع کلیدی برای محققینی که در زمینه پایگاههای داده و یادگیری ماشین کار می کنند و به فرصتی برای کسب درآمد های بیشتر توسط شرکت های صنعتی و تجاری تبدیل شده است. دانش های کشف شده توسط داده کاوی می توانند در مدیریت اطلاعات، پردازش گزارش ها، انجام تصمیم گیری ها و بسیاری زمینه های دیگر استفاده شوند. به علت وجود گسترده داده ها در حجم زیاد و نیاز مبرم به تبدیل این داده ها به اطلاعات و دانش مفید برای کاربرد های مختلف، داده کاوی در سال های اخیر توجه زیادی را به خود جلب کرده است[1,2].

داده کاوی موضوعی وابسته به کاربرد است و کاربردهای مختلف نیازمند روش ها و تکنیک های داده کاوی مختلفی هستند. کاوش قواعد وابستگی، دسته بندی[۳]، خوشه بندی[۴]، پیش بینی[۵] و تحلیل سری های زمانی[۶] از جمله مهمترین روش ها و تکنیک های داده کاوی به شمار می آیند. در ادامه هر کدام از این روش ها به صورت خلاصه توضیح داده می شوند.

کشف یا کاوش قواعد وابستگی در پایگاه داده های رابطه ای[۷] یا تراکنشی [8] که موضوع اصلی این مطالعه نیز است، اخیراً جذابیت زیادی را در انجمن های مربوط به پایگاههای داده بوجود آورده است. در این تکنیک داده کاوی وابستگی ها و ارتباطات بین داده های موجود در یک پایگاه داده بدست می آیند. نتیجه این عملیات داده کاوی دسته ای از قواعد است که به آنها قواعد وابستگی گفته می شود.

یکی دیگر از روش های مهم داده کاوی توانایی انجام عملیات دسته بندی در حجم زیاد داده هاست. این عملیات کاوش قوانین دسته بندی نیز نامیده می شود. در این روش اشیاء موجود در یک پایگاه داده بر اساس مقادیر چند خصوصیت از آنها، به دسته های مجزا تقسیم می شوند. در دسته بندی داده ها مجموعه ای از داده های آزمایشی تحلیل می شوند. برای مجموعه داده های آزمایشی برچسب کلاس ها مشخص است. برای هر کلاس داده های آزمایشی مدلی بر اساس خصوصیات داده ها ساخته می شود. حاصل عملیات دسته بندی می تواند یک درخت تصمیم یا مجموعه ای از قوانین دسته بندی باشد که برای فهم بهتر داده های موجود در پایگاه داده و همچنین دسته بندی داده هایی که در آینده به پایگاه داده اضافه می شوند به کار می رود. دسته بندی داده ها ارتباط تنگاتنگی با کاوش قواعد وابستگی دارد، به طوری که گاهی دسته بندی را به کمک قواعد وابستگی انجام می دهند.

به عنوان مثال برای فروشنده‌ی ماشین دسته بندی مشتریانش بر اساس تمایل و علاقه آنها به انواع مختلف ماشین مطلوبست به طوری که بتواند به مشتریانش خدمات بهتری ارائه دهد و کاتالوگ های محصولات جدید مورد نظر آنها را برایشان بفرستد و درآمد حاصل از فروش خود را افزایش دهد.

خوشه بندی داده ها از مهمترین روش های داده کاوی به شمار می آید. درخوشه بندی مجموعه ای از داده ها گروهبندی می شوند. فرق خوشه بندی با دسته بندی داده ها در این است

---

[3] Classification
[4] Clustering
[5] Prediction
[6] Time Series Analysis
[7] Relational
[8] Transactional

که در خوشه بندی برخلاف دسته بندی تعداد کلاس ها در ابتدا مشخص نیستند .خوش بندی داده ها بر اساس اصل مفهومی زیر صورت می گیرد:

"حداکثر کردن شباهت های بین اعضای هر کلاس و حداقل کردن شباهت ها بین اعضای مربوط به کلاس های مختلف".

به عنوان مثالی از خوشه بندی، مجموعه ای از کالاها را می توان در ابتدا به صورت مجموعه ای از کلاس های مختلف خوشه بندی کرده و سپس مجموعه ای از قوانین را بر اساس این چنین دسته بندی نتیجه گیری کرد. پیش بینی یکی دیگر از تکنیک های داده کاوی محسوب می شود که در آن مقادیر ممکن برای متغیرهای نامعلوم پیش بینی می شوند. در پیش بینی ابتدا داده هایی که به متغیر نامعلوم مربوط هستند بوسیله‌ی برخی تحلیل های آماری پیدا می شوند. سپس از برخی روش های هوشمند مانند شبکه های عصبی و الگوریتم ژنتیک برای انجام پیش بینی استفاده می شود. برای مثال مقدار حقوقی که یک کارمند می‌گیرد را می توان با استفاده از چگونگی توزیع حقوق کارمندان مشابه در آن اداره، پیش بینی کرد. روش های دیگری از جمله تحلیل رگرسیون[9]، تحلیل وابستگی[10]، درخت تصمیم [11] در انجام یک پیش بینی با کیفیت مؤثرند.

تحلیل سری های زمانی از دیگر روش های کاربردی داده کاوی است. در این روش حجم زیادی از داده های سری زمانی برای یافتن خصوصیات جالب توجه و نظم های مشخص، تحلیل می شوند. رخداد وقایع متوالی، مجموعه وقایعی که بعد از یک واقعه مشخص به وقوع می پیوندند، روند ها و انحراف ها از جمله این نظم ها و پدیده های جالب توجه هستند. برای مثال می توان روند تغییر قیمت برای یک کالای مشخص را در یک کارخانه با استفاده از داده های گذشته، شرایط تجاری و رقبای تجاری، پیش بینی کرد.

موارد ذکر شده در بالا تعدادی از تکنیک ها و روش های مهم در داده کاوی بودند روش های دیگری نیز وجود دارند که در این نوشته مجالی برای بحث در مورد آنها نیست.

۱-۳ مراحل داده کاوی

به داده کاوی عملیات اکتشاف دانش در پایگاه های داده هم گفته می شود، اگرچه برخی محققین داده کاوی را بخشی از عملیات کلی تر اکتشاف دانش[12] می دانند. به طور کلی عملیات اکتشاف دانش شامل اجرای تکراری مراحل زیر است:

---

[9] Regression Analysis
[10] Correlation Analysis
[11] Decision Tree
[12] Knowledge Discovery

- خالص کردن داده ها[13]: در این مرحله داده ها از وجود خطاها، داده های مشکوک و داده های غیر مرتبط پاک می شوند.
- سرهم سازی داده ها[14]: در این مرحله داده هایی که از منابع مختلف گردآوری شده اند در یک منبع واحد مجتمع می شوند.
- گزینش داده ها [15] : در این مرحله داده هایی را که عملیات تحلیل روی آنها انجام می شود از پایگاه داده بازیابی می شوند.
- تبدیل داده ها[16]: در این مرحله داده ها به قالب و فرم مناسب برای انجام عملیات داده کاوی تبدیل می شوند.
- داده کاوی: مرحله ای ضروری است که در آن روش های هوشمند به منظور استخراج الگوها، به داده های تبدیل یافته اعمال می شوند.
- ارزیابی الگوها[17] : در این مرحله الگوهای بدست آمده توسط یک سری معیارهای مقبولیت سنجیده می شوند و در نهایت از میان الگوهای بدست آمده آنهایی که واقعا با ارزش هستند تشخیص داده می شوند.
- ارائه دانش[18] : در این مرحله نتایج بدست آمده به شکل معنا داری به وسیله ی عملیات گرافیکی و سایر روش های نمایش دانش به کاربر عرضه می گردند.

با گسترش روز افزون استفاده از پایگاه های داده و انبار های داده چهار مرحله ی نخست شامل خالص کردن، سرهم سازی، گزینش و تبدیل، با ساختن انبارهای داده و انجام برخی عملیات گزارش گیری مربوط به انبارهای داده قابل انجام هستند. عملیات داده کاوی، ارزیابی الگوها و ارائه دانش گاهی در یک مرحله کلی بنام داده کاوی انجام می گیرد.

۱-۴ روشهای داده کاوی

داده، انبارهای داده، آمار، یادگیری ماشین، نمایش داده، بازیابی اطلاعات و محاسبات سریع سرچشمه گرفته است، انجام عملیات مختلف داده کاوی نیازمند آگاهی از این زمینه هاست. آگاهی از فنون مختلف هوش مصنوعی مانند شبکه های عصبی، محاسبات فازی و الگوریتم ژنتیک برای انجام برخی عملیات داده کاوی ضروری است.

---

[13] Data Cleaning
[14] Data integration
[15] Data selection
[16] Data transformation
[17] Pattern evaluation
[18] Knowledge presentation

مجموعه وسیعی از عملیات مربوط به آمار که در طول سال ها توسعه یافته اند در انجام برخی عملیات تحلیل داده ها استفاده می شوند. یادگیری ماشین در عملیاتی چون دسته بندی داده ها استفاده می شود. شبکه های عصبی در عملیات دسته بندی، خوشه بندی و پیش بینی کاربرد فراوانی دارد. به هر حال از آنجایی که در عملیات داده کاوی ما با حجم زیاد داده هایی که در پایگاه های داده ذخیره می شوند سر و کار داریم در استفاده از این روش ها به مشکل کارائی و مقیاس پذیری روبرو می شویم.

استفاده از ساختمان داده های مناسب، روش های اندیس گذاری و دستیابی به داده ها که در بحث پایگاه های داده مطرح شده اند تأثیر بسزایی در انجام سریع عملیات داده کاوی دارند. روش های تحلیل داده ها که در آمار و یادگیری ماشین مطرح هستند برای روبرو شدن با حجم زیاد داده ها در داده کاوی، بازنگری می شوند. برای انجام داده کاوی مؤثر باید الگوریتم های مقیاس پذیر و مبتنی بر مجموعه ها داشته باشیم.

برخلاف عملیات تحلیل داده ها که همیشه یک فرض وجود دارد، در عملیات داده کاوی هدف ما اکتشاف است و سعی می کنیم الگوهایی ناشناخته از داده ها را بدست آوریم و نیازمند جستجوهای سنگین و وقت گیر در داده ها هستیم. بنابراین انجام محاسبات سریع نقش مهمی را در داده کاوی بازی می کند. روش های موازی، توزیع شده و افزایشی بیشتری در زمینه داده کاوی به این منظور باید توسعه یابند.

۱-۵ کاوش قواعد وابستگی

یکی از مهمترین و پرکاربردترین روش های داده کاوی کاوش قواعد وابستگی است، که در آن هدف پیدا کردن قوانینی است که در پایگاه داده های با حجم بالا وجود دارند. به عنوان مثال پایگاه داده مربوط به یک فروشگاه زنجیره ای کالا را در نظر بگیرید که در آن اطلاعات مربوط به اجناس خریداری شده توسط هر مشتری ذخیره می شود، در طول سالهای کار این فروشگاه حجم داده های ذخیره شده در پایگاه داده انباشته و انباشته تر می شود. استفاده معمول از این داده ها گزارش گیری های مربوط به وضعیت فروش و مشتریان این فروشگاه است. اما استفاده مفید دیگری که از این داده ها می توان کرد یافتن یک سری قوانین است که درآن وابستگی و ارتباط بین کالای خریداری شده توسط مشتریان نمایان می شود. به عنوان مثال جدول ۱-۱ را به عنوان پایگاه داده یک فروشگاه در نظر بگیرید، ستون سمت چپ شماره تراکنش هایی است که در این پایگاه داده ذخیره شده اند، و ستون ستون سمت راست معرف مجموعه اقلام مربوط به هر تراکنش است.

جدول ۱-۱ یک پایگاه داده نمونه

Tid	Items
1	{bread, coke, milk}
2	{bread, butter, milk, ice cream}
3	{ice cream, coke}
4	{battery, bread, butter, milk}
5	{bread, butter, milk}
6	{battery, ice cream, bread, butter}

اگر چه در پایگاه داده مثال ما تنها شش تراکنش وجود دارد، ولی در پایگاههایی که روی آنها عملیات کاوش قواعد وابستگی صورت می گیرد حجم زیادی از تراکنش ها وجود دارد. یکی از این قوانین موجود در پایگاه داده جدول ۱-۱ می تواند به این صورت باشد که هرکسی که از این فروشگاه کره خریده است به احتمال٪۸۰ نان هم خریده است، که این قاعده در ٪۶۶ موارد خرید یا تراکنش های این فروشگاه دیده شده است. این نوع قوانین را در اصطلاح قواعد وابستگی می نامند و مجموعه اقلامی که به طور مکرر در پایگاه داده تکرار می شوند و ما قوانین مورد نظر را از روی آنها تولید می کنیم مجموعه اقلام تکراری[۱۹] یا مجموعه الگوهای مکرر[۲۰] نام دارند مانند مجموعه {bread, butter} در پایگاه داده مثال بالا. احتمال مطرح شده را درصد اطمینان[۲۱] و درصدی از تراکنش ها که این قاعده را پوشش می دهند، درصد پشتیبانی[۲۲] این قاعده می نامند. عملیات مربوط به پیدا کردن این قوانین را کاوش قواعد وابستگی یا به اختصارARM می نامند[3]. درصد پشتیبانی و درصد اطمینان، دو معیار مربوط به مقبولیت قوانین کشف شده هستند. معیار های دیگری نیز در این زمینه وجود دارند.

معمولاً برای یافتن قوانین حداقلی برای اندازه معیار های پشتیبانی و اطمینان توسط کاربر در نظر گرفته می شود، که به آنها به ترتیب حداقل حد آستانه[۲۳] پشتیبانی و حداقل حد آستانه اطمینان گفته می شود. عملیات اصلی در مورد کاوش قواعد وابستگی یافتن مجموعه اقلام تکراری با استفاده از حداقل حد آستانه پشتیبانی است. در واقع با داشتن این مجموعه اقلام تولید قواعد وابستگی بسیار سر راست خواهد بود. بنابراین مسئله کاوش قواعد وابستگی به مسئله یافتن مجموعه اقلام تکراری کاهش می یابد. از آنجایی که کاوش مجموعه اقلام تکراری نیاز به خواندن از

[19] Frequent itemset
[20] Frequent pattern
[21] Confidence
[22] Support
[23] Thereshold