ششمین همایش بین‌المللی افق های نویـن در
مهندسی برق، کامپیوتـر و مکانیک
6th International Conference on the New Horizons in
Electrical Engineering, Computer and Mechanical
www.mhconf.ir

بررسی روش‌های استخراج الگوهای پرتکرار پیشرفته

نی نا پیران, ۱، اصغر درویشی۲

دانشجوی دکتری تخصصی نرم افزار دانشگاه آزاد اسلامی واحد تهران جنوب)، Piran.neyna@gmail.com

استادیار و عضو هیئت علمی گروه مهندسی نرم افزار دانشگاه ا زاد اسلامی واحد تهران جنوب)،A_Darvishy@azad.ac.ir

چکیده

استخراج الگوی پرتکرار یک فرایند تحلیلی است که هدف آن کشف الگوها و ارتباطات تکراری از داده‌ها است. با توجه به تنوع مجموعه داده‌های تراکنش حاصل از آزمایشات، برخی از نقاط ضعف روش‌های استخارج الگوی سنتی را تحدید می‌کند. از این رو روشهای استخراج الگوهای پیشرفته به منظور رفع نقاط ضعف روش‌های کاوش الگوهای پرتکرار سنتی ارائه شده است. در این مقاله ابتدا روش‌های استخراج الگوهای پرتکرار پیشرفته به چهار دسته کلی استخراج الگوهای چند سطحی، استخراج الگوی مبتنی بر محدودیت، استخراج الگوهای با ابعاد بالا و الگوهای عظیم و استخراج الگوهای فشرده یا تقریبی تقسیم شد. استخراج الگوهای چند سطحی خود به چهار دسته استخراج قوانین انجمنی چند سطحی، استخراج قوانین انجمنی چند بعدی، استخراج قوانین انجمنی کمی و استخراج قوانین انجمنی کمیاب تقسیم می‌شود. در مقاله حاضر دسته ای و زیر دسته‌های مربوط به استخراج الگوی پیشرفته معرفی شده و برخی از تحقیقات هر کدام از این دسته‌ها و زیر دسته‌ها به تفکیک بررسی شده است

واژه‌های کلیدی

داده کاوی، استخراج الگو، استخراج الگوهای پرتکرار، استخراج الگوهای چندسطحی، استخراج الگوی مبتنی بر محدودیت، استخراج الگوهای الگوهای عظیم، استخراج الگوهای تقریبی

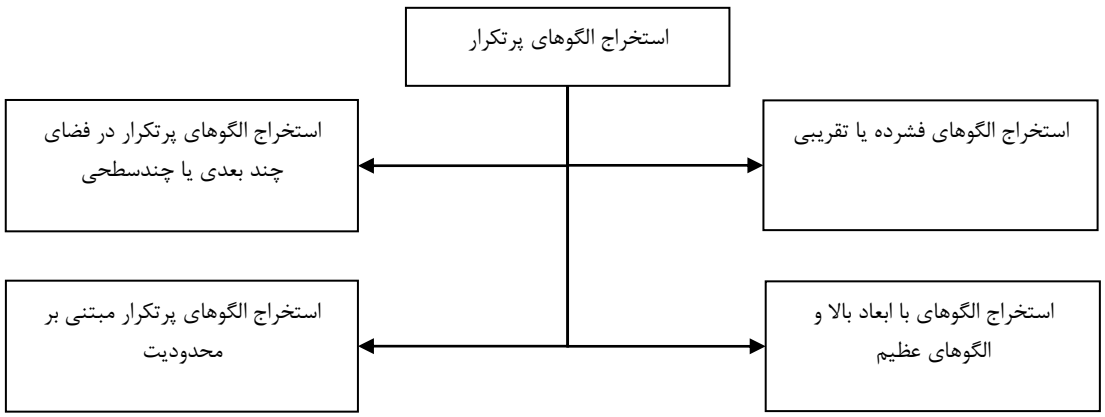
۱. متن مقاله

استخراج الگو یک حوزه تحقیقاتی کلیدی در داده کاوی است که شامل استفاده از الگوریتم‌هایی برای شناسایی الگوهای جالب در داده‌ها است. به طور کلی، اگر الگویی اطلاعات جدیدی را کشف کند که برای درک گذشته یا پیش بینی آینده مفید هستند، می‌تواند جالب تلقی شود. در طول سال‌ها، تکنیک‌هایی برای استخراج الگوها از انواع مختلفی از داده‌ها مانند داده‌های معاملاتی، سری زمانی، گزارش روند کار، داده‌های مکانی، توالی‌ها و نمودارها طراحی شده اند. بسته به برنامه‌های کاربردی، الگوی جالب متفاوت است. بنابراین، الگوریتم‌هایی ارائه شده اند که معیارهای مختلفی را برای شناسایی الگوهایی مانند یافتن آیتم‌هایی با فرکانس و اطمینان بالا وقوع هستند، به کار می‌برند. وظایف استخراج الگو می‌تواند بسیار چالش برانگیز باشد، زیرا یک الگوریتم باید تعداد بالقوه زیادی از الگوهای ممکن را برای کشف الگوی مورد نظر در نظر بگیرد. در نتیجه، الگوریتم‌های کارآمد بر اساس ساختار داده‌های کارآمد و استراتژی‌های هرس فضای جستجو طراحی شده اند[۱].

استخراج الگوهای پرتکرار[1] (FPM) یکی از رایج ترین مسائل در داده کاوی است و در چندین موضوع فرعی از جمله استخراج قوانین انجمنی، استخراج الگوهای ترتیبی و طبقه بندی نقش مهمی دارد. استخراج برنامه‌های کاربردی از پایگاه‌های اطلاعاتی دقیق باعث شده است که بسیاری از محققان در سراسر جهان مورد تحقیق قرار گیرند[۲]. استخراج الگوی پرتکرار به دلیل تحقیقات قابل توجه، گسترش بسیاری از دامنه مسئله و مطالعات گسترده کاربرد بسیار فراتر از اصول اولیه رسیده است. از این رو روش‌های بسیاری برای استخراج الگوهای پرتکرار در نشریات اراده شده است. در این مقاله ابتدا به دسته بندی کلی روش‌های کشف الگو می‌پردازیم و سپس روش‌هایی که اخیراً در زمینه استخراج الگو تلاش کرده اند مورد بررسی قرار خواهیم داد.

۱-۲ دسته بندی تکنیک‌های استخراج الگوی پیشرفته

استخراج الگوی پرتکرار یک فرایند تحلیلی است که هدف آن کشف الگوها و ارتباطات تکراری از داده‌ها است. FPM یک روش مهم داده کاوی است که برای استخراج بینش‌های معنی دار، قابل استفاده و اطلاعات از حجم زیادی از داده‌ها استفاده می‌شود. علاوه بر این، FPM همچنین روند کشف مجموعه‌های مکرر موارد تکراری را در نمودارهای بزرگ، براساس آستانه فرکانس مشخص شده توسط کاربر، توصیف می‌کند[۳]. در راستای استخراج الگوهای پرتکرار روش‌های زیادی ارائه شده است که در شکل ۱ دسته بندی روش‌های استخراج الگوهای پرتکرار نشان داده شده است.



شکل ۱. دسته بندی کلی روش‌های استخراج الگو

[1] Frequent Pattern mining

همانطور که در شکل ۱ نشان داده شده است، روش‌های استخراج الگوها به ۴ دسته کلی استخراج الگوهای پرتکرار در فضای چند بعدی یا چندسطحی، استخراج الگوهای پرتکرار مبتنی بر محدودیت، استخراج الگوهای پرتکرار فشرده یا تقریبی و استخراج و کاربرد الگو تقسیم می‌شود که هر یک از این دسته‌ها خود به شاخه‌های مختلف تقسیم می‌شوند. در ادامه این مقاله به بررسی تحقیقات در هر یک از دسته‌های بالا خواهیم پرداخت.

۲. کارهای مرتبط

با افزایش حجم داده‌های حاصل از آزمایشات و تحقیقات علمی، نیاز به استخراج الگوهای با ارزش از میان حجم انبوهی از داده‌ها به یک چالش اساسی در داده کاوی تبدیل شده است. الگوهای معنی دار علاوه بر اطلاعات ارزشمندی که در مورد داده‌های ذخیره شده در اختیار ما قرار می‌دهند، می‌توانند در راستای بهبود عملکرد آتی کاربردهای مورد استفاده از داده‌ها نیز موثر واقع شوند. از این رو استخراج الگوها همه روزه محبوبیت بیشتری اخذ می‌کند و توسط محققان بیشتری مورد توجه قرار می‌گیرد. در ادامه این بخش از مقاله به بررسی روش‌های کشف الگو به تفکیک دسته‌های کلی استخراج الگو خواهیم پرداخت.

۲-۱ استخراج الگوهای چند سطحی

استخراج الگوهای چند سطحی شامل وقوع آیتم‌های پرتکرار در رده بندی‌های مختلفی از آیتم‌های موجود در مجموعه داده‌های تراکنشی موجود است. به عبارت دیگر تقسیم یک آیتم به انواع تشکیل دهنده آن موجب ایجاد چندین آیتم متفاوت است که ریشه آن‌ها با هم مشترک می‌باشد. این شاخه از استخراج الگوها در توجه زیادی در مجموعه داده‌های تراکنشی دریافت کرده اند. استخراج الگوهای چند سطحی خود به چهار شاخه استخراج قوانین انجمنی چند سطحی، استخراج قوانین انجمنی چند بعدی، استخراج قوانین انجمنی کمی و استخراج قوانین انجمنی کمیاب تقسیم شده است که در ادامه در مورد هر شاجه چندین روش بررسی خواهد شد.

۲-۱-۱ استخراج قوانین انجمنی چند سطحی

استخراج قوانین انجمنی چند سطحی یک حوزه مهم برای کشف روابط جالب بین عناصر داده با چندین سطح انتزاع است. بیشتر الگوریتم‌های موجود در رابطه با این مسئله مبتنی بر روش‌های جستجوی طاقت فرسا مانند Apriori و FP-growth هستند. با این حال، وقتی این روش‌ها در برنامه‌های داده بزرگ استفاده شوند، در جستجوی قوانین انجمن، هزینه محاسباتی زیادی متحمل می‌شوند. به ویژه برای موارد رویکردهای استخراج قوانین انجمنی چند سطحی، کشف روابط جالب بین عناصر داده با چندین سطح انتزاع است. برنامه‌های موفق شامل تجزیه و تحلیل داده‌های مکانی، تجزیه و تحلیل رویداد اضطراری، داده کاوی شبکه حسگر و استخراج هستی شناسی ژن هستند. با این حال، بیشتر الگوریتم‌های استخراج معیارهای ارتباط چندسطحی موجود برای یافتن الگوهای پرتکرار در سطح مختلف انتزاع، به بررسی جامع پایگاه داده متکی هستند. هنگامی که مجموعه داده مقیاس بندی شود، این الگوریتم‌ها به دلیل هزینه محاسبات بیش از حد متضرر می‌شوند و سیستم به دلیل جستجوی سنگین پایگاه داده بزرگ، عقب می‌افتد. وقتی از الگوریتم‌ها در برنامه‌های کلان داده استفاده می‌شود، این مسئله به عنوان یک گلوگاه، برجسته تر می‌شود.

Zhiguo Qu و همکاران در سال ۲۰۱۶، یک معماری استخراج الگوی چند سطحی برای پشتیبانی از مدیریت خودکار شبکه با کشف الگوهای جالب از داده‌های نظارت بر شبکه مخابراتی ارائه داده اند. این معماری از اکتشافات متداول آیتم‌های موجود در جریان داده‌ها، کاهش قوانین انجمنی، استخراج الگوهای پرتکرار ترتیبی و تکنیک‌های استخراج الگوهای پرتکرار زمانی استفاده کرده است و در عین حال از سیستم عامل‌های پردازش توزیع شده برای دستیابی به توان تولید با حجم بالا بهره برده است[۴].

Yang Xu و همکاران در سال ۲۰۱۴، برای تسریع در جستجوی قوانین انجمنی چند سطحی و جلوگیری از سربار محاسباتی، یک روش مبتنی بر ژنتیک جدید با سه نوآوری اساسی ارائه داده اند. ابتدا، از درخت دسته بندی برای توصیف مجموعه داده‌های برنامه چند

سطحی به عنوان دامنه دانش استفاده شده است. سپس، برای ایجاد الگوریتم استخراج قوانین انجمنی چند سطحی اکتشافی، یک طرح ویژه کدگذاری درخت بر اساس درخت دسته بندی ارائه شده است. در نهایت، الگوریتم ژنتیک بر اساس طرحواره رمزگذاری درخت پیشنهاد شده است که فضای جستجو برای قوانین انجمنی را تا حد زیادی کاهش دهد. این روش به ویژه در استخراج قوانین انجمنی چند سطحی در برنامه‌های بزرگ مربوط به داده‌های بزرگ مفید است [۵].

J. H. DUAN و همکاران در سال ۲۰۱۷، الگوریتم استخراج قوانین انجمنی چندسطحی بهبود یافته را با استفاده از پایگاه داده چند شاخصی پیشنهاد کرده اند. همچنین مجموعه آیتم‌های پرتکرار در قوانین انجمنی تداعی مجموعه ای از ویژگی‌های مورد نظر را ایجاد می‌کند. علاوه بر ای تنگ کردن فضای زیر ویژگی مجموعه آیتم‌های نامزد، این روش از فشرده سازی مقیاس بر اساس اسکن مکرر شاخص پایگاه داده استفاده می‌کند و یک فهرست از آیتم‌ها با توجه به حداقل آستانه پشتیبانی و قوانین فیلتر شرایط محدودیت، با توجه به قوانین انجمنی لیست ساختار داده برای دستیابی به توزیع بازآرایی و بهبود بهره وری شاخص، ارائه می‌دهد[۶].

Prajapati و همکاران در سال ۲۰۱۷، یک الگوریتم استخراج قوانین انجمنی چند سطحی ارائه کرده اند که در آن در ابتدا قوانین انجمنی چندسطحی شامل قوانین بین سطح برای هر منطقه با استفاده از الگوریتم استخراج الگوهای پرتکرار چندسطحی توزیع شده (DMFPM[۲]) استخراج شده است. این قوانین انجمنی چند سطحی تولید شده به قدری بزرگ است که تحلیل آن پیچیده می‌شود. بنابراین، الگوریتم تشخیص قوانینی سازگار و ناسازگار چندسطحی مبتنی بر MapReduce (MR-MCIRD[۳]) برای شناسایی قوانین چندسطحی سازگار و ناسازگار از داده‌های بزرگ سلسله مراتبی ارائه می‌شود که دانش مفید و عملی را به کارشناسان این حوزه ارائه می‌دهد. این قوانین جالب هرس شده نیز دانش مفیدی برای استراتژی بهتر بازاریابی می‌دهد. قوانین سازگار و ناسازگار چند سطحی استخراج شده بر اساس معیارهای مختلف جذابیت همراه با نتایج تجربی که منجر به نتیجه گیری نهایی می‌شوند، ارزیابی و مقایسه شده اند[۷].

۲-۱-۲ استخراج قوانین انجمنی چند بعدی

تکنیک‌های استخراج دانش بسیاری بررسی و پیشنهاد شده است که به آنها داده کاوی می‌گویند. به دلیل شهودی بودن، به استخراج قوانین انجمنی توجه زیادی شده است. نگرانی‌هایی مربوط به استخراج قوانین انجمنی در مورد سوابق معاملات مرتبط با هم در یک تراکنش، اما تعداد کمی در مورد تراکنش‌های متعدد، وجود دارد. از این رو شاخه ای از تحقیق با عنوان استخراج قانون انجمنی چند بعدی، معرفی شده است. واضح است که ابعاد بزرگتر از آیتم‌ها، اطلاعات مفیدتری را می‌تواند استخراج کند. اما مسئله در مورد پیش بینی‌های متعدد سخت می‌شود، زیرا به زمان و فضای پیچیده ای نیاز دارد.

Yokobayashi و همکاران در سال ۲۰۱۸، یک چارچوب مناسب برای داده کاوی چند بعدی بر اساس تانسور پیشنهاد می‌کنیم. یک مدل داده ای تنسور[۴] (TDM) ساختار داده ای با نظم بالا و توصیف ساده برای بازیابی اطلاعات را در اختیار ما قرار می‌دهد. در این روش در مورد استخراج قانون چند بعدی بحث شده است. به طور کلی، استخراج قوانین انجمنی (یا استخراج قوانین تداعی) نگرانی در مورد سوابق معاملات مرتبط با محمول واحد است و بررسی موارد بیش از چندین گزاره دشوار است زیرا پیچیدگی‌های زمان و فضا به طول می‌انجامد. در این روش نشان داده شده است که TDM اجازه می‌دهد چندین عملیات خاص برای داده کاوی چند بعدی مدلسازی شود اما هنوز مقدار توصیف کاهش نمی یابد[۸].

Yuan Tian و همکاران در سال ۲۰۲۰، مقاله ای ارائه کرد اند که در آن فن آوری داده کاوی صنعت نیرو برای تحقق بخشیدن به داده واقعی استخراج قوانین انجمنی چند بعدی در زمان واقعی سیستم شبکه برق بر اساس فناوری تعامل اطلاعات تجسم اطلاعات و فناوری داده کاوی، به طور عمیق تحلیل شده است. به طوری که می‌توان رابطه متقابل و ارتباط داخلی در داده‌های عظیم قدرت را عمیقا کاوش کرد و تجسم با استفاده از نمودارها و روش‌های دیگر انجام می‌شود. نظارت بر زمان واقعی بار، نظارت بر بار زیاد و نظارت بر عدم تعادل

---

[2] Distributed Multilevel Frequent Pattern Mining
[3] MapReduce based multilevel consistent and inconsistent rule detection
[4] Tensor Data Model

نیرو در سال ۲۰۲۰ به عنوان نمونه‌های تحقیق انجام شده است. ارتباط چند بعدی شاخص‌های مختلف داده‌های قدرت از طریق مطالعه موردی و تجزیه و تحلیل نشان داده شده است[۸].

Pradhan و همکاران در سال ۲۰۱۷، داده‌های سری زمان واقعی فعالیت‌های عضلانی شرکت کنندگان انسانی را که از چندین حسگر الکترومیوگرام (EMG[5]) بدست آمده اند، در نظر گرفته و الگوهایی را در این داده‌های سری زمانی EMG کشف کرده اند. هر داده سری EMG با ویژگی‌های کمی مانند انرژی سیگنال و زمان شروع در ارتباط است که لازم است همراه با الگوهای سری زمانی EMG استخراج شود. در این مقاله یک روش دو مرحله ای برای این منظور پیشنهاد شده است: در مرحله اول، تأکید بر کشف الگوهای پرتکرار چندبعدی در چندین سری زمانی با انجام استخراج متوالی در برش‌های زمانی است و در مرحله بعدی، فقط به ویژگیهای کمی فقط در سریهای زمانی توجه شده است که در الگوهای کشف شده در مرحله اول وجود دارد. ارزیابی این مقاله با مجموعه زیادی از داده‌های سری زمانی از چندین حسگر EMG نشان می‌دهد که رویکرد دو مرحله ای روند یافتن قوانین انجمنی در چنین محیط چند بعدی را در مقایسه با سایر روش‌ها سرعت می‌بخشد و از نظر تعداد سری‌های زمانی درگیر به صورت خطی تنظیم می‌شود[۹].

۲-۱-۳ استخراج قوانین انجمنی کمی[6]

در داده کاوی، کشف روابط جالب در داده‌ها یک تکنیک مکرر مورد استفاده است که به عنوان قوانین انجمنی ارتباط شناخته می‌شود. قوانین انجمنی را می‌توان به صورت $A \rightarrow C$ بیان کرد، جایی که A و C اقلام زوجین هستند که از نظر صفت ارزش دارند و $A \cap C = \emptyset$ را تأمین می‌کنند. پیشنهادهای زیادی را می‌توان در نشرات یافت که به کشف قوانین انجمنی در مجموعه داده‌ها با مقادیر عددی، معروف به قوانین ارتباط کمی (QAR) بپردازد. بسیاری از آنها بر اساس الگوریتم‌های تکاملی بنا به عملکرد خوب ارائه شده در مشکلات فضاهای جستجوی پیچیده، بنا نهاده شده اند. با این حال، فرآیند تطبیق قوانین انجمنی از نظر زمان اجرا، پرهزینه ترین مرحله این الگوریتم‌ها است زیرا تعداد دفعات زیادی نیاز به پردازش عملکردهای گران تابع تناسب، دارد. ارزیابی هر راه حل مستلزم پردازش تمام سوابق موجود در مجموعه داده است. متناوباً، ضعف دیگری که معمولاً وجود دارد، مصرف حافظه است. این مسئله به ویژه هنگامی مورد تأکید قرار می‌گیرد که این الگوریتم‌ها مجبورند مجموعه داده‌های مقیاس بزرگ را کنترل کنند. به طور کلی، فرآیند استخراج دانش به یک کار دشوار و پیچیده تبدیل می‌شود زیرا در بسیاری از موارد، مقدار داده‌های تولید شده بیش از توانایی پردازش سیستم‌های معمولی است. طراحی الگوریتم‌های کارآمد که قادر به پردازش و تجزیه و تحلیل این مقدار داده است، در حال تبدیل شدن به یک چالش بزرگ برای محققان است.

در این راستا D. Martín و همکاران در سال ۲۰۱۸، MRQAR[7] را پیشنهاد داده اند که یک چارچوب موازی عمومی جدید برای کشف قوانین انجمنی کمی در مقدار زیادی از داده‌ها است که به دنبال الگوی MapReduce با استفاده از Apache Spark طراحی شده است. MRQAR یک یادگیری افزایشی را انجام می‌دهد که قادر به اجرای هر الگوریتم قانون انجمنی کمی متوالی در مشکلات Big Data بدون نیاز به طراحی مجدد چنین الگوریتم‌هایی است. به عنوان یک مطالعه موردی، الگوریتم تکاملی چند هدفه MOPNAR[8] در MRQAR ادغام شده است تا چارچوب عمومی MapReduce پیشنهادی در این کار تأیید شود. نتایج به دست آمده در یک مطالعه تجربی که روی پنج مسئله داده‌های حجیم انجام شده است، توانایی MRQAR را برای بدست آوردن مجموعه ای از قوانین با کیفیت بالا در زمان مناسب ثابت می‌کند[۱۰].

۲-۱-۴ استخراج قوانین انجمنی کمیاب

---

[5] Electromyogram
[6] Quantitative Association Rules
[7] Map Reduce Quantitative Association Rules
[8] Multi Objective Positive Negative Association Rule

هدف اصلی استخراج قوانین انجمنی کشف روابط بین مجموعه آیتم‌ها در پایگاه داده تراکنشی است. استخراج قاعده انجمین، برای استخراج همبستگی‌های جالب، الگوهای مکرر، ارتباطات یا ساختارهای گاه به گاه در میان مجموعه موارد موجود در پایگاه داده تراکنشی یا مخازن داده‌های دیگر، ارائه شده است. این روابط براساس خواص ذاتی داده‌ها نیست بلکه بر اساس همزمانی موارد موجود در پایگاه داده است. این ارتباطات بین موارد به عنوان قوانین انجمنی نیز شناخته می‌شوند. دو نوع قانون عمده را می‌توان در پایگاه داده یافت: قوانین پرتکرار و کمیاب. هر دو قانون انجمنی پرتکرار و کمیاب اطلاعات مختلفی را درباره پایگاه داده ای که در آن یافت می‌شود ارائه می‌دهند. قوانین پرتکرار بر الگوهایی تمرکز دارند که به طور مکرر اتفاق می‌افتند، در حالی که قوانین کمیاب بر الگوهایی تمرکز می‌کنند که به ندرت اتفاق می‌افتند. در بسیاری از دامنه‌ها، حوادثی که به طور مکرر رخ می‌دهند ممکن است کمتر از رویدادهایی که به ندرت اتفاق می‌افتد جالب باشد، زیرا الگوهای پرتکرار نمایانگر موارد شناخته شده و مورد انتظار هستند، در حالی که الگوهای کمیاب ممکن است ارتباطات غیر منتظره یا ناشناخته ای را نشان دهند که برای متخصصان دامنه مفید است. نمونه‌هایی از استخراج اقلام کمیاب شامل شناسایی بیماریهای نسبتاً کمیاب، پیش بینی خرابی تجهیزات مخابراتی و یافتن ارتباط بین اقلام سوپرمارکت‌هایی است که به ندرت خریداری می‌شوند. در حوزه پزشکی، پاسخهای متناوب و مورد انتظار، نسبت به پاسخهای نادر و استثنایی که نشانگر واکنشهای جانبی یا تداخلات دارویی است، کمتر جالب است[۱۱].

از سوی دیگر الگوهای ترتیبی منفی[9] (NSP)، که بر رفتارهای غیرتکراری اما جالب توجه تمرکز دارند (به عنوان مثال سوابق مصرف از دست رفته)، دیدگاه ویژه ای برای تجزیه و تحلیل الگوهای ترتیبی فراهم می‌کنند. تاکنون روشهای کمی برای حل مسئله استخراج NSP پیشنهاد شده است و این روشها فقط NSP را از الگوهای متوالی مثبت (PSP) استخراج می‌کنند. با این حال، از آنجا که بسیاری از قوانین ارتباط منفی مفید از مجموعه‌های نادر مورد استخراج قرار می‌گیرند، بسیاری از NSPهای معنی دار را نیز می‌توان از توالی‌های مثبت نادر[10] (IPS) یافت. چالش استخراج NSP از IPS چگونگی محدود کردن IPS است که می‌تواند در طی فرآیند NSP استفاده شود زیرا اگر بدون محدودیت باشد، تعداد IPS برای رسیدگی بسیار زیاد است.
NuritGal-Oz و همکاران در سال ۲۰۱۹، یک الگوریتم ارائه کرده اند که از اطلاعات اجازه جلسه استفاده می‌کند و با از بین بردن نقش‌ها با استفاده بسیار کم، بر نویزهای ذاتی غلبه می‌کند. با این وجود برخی از قوانین ذاتاً نادر هستند، یا برای گروه بسیار کمی از کاربران تعریف می‌شوند یا حاوی مجوزهایی برای داده‌هایی هستند که به ندرت استفاده می‌شوند. آن‌ها در مقاله دیگری، یک الگوریتم ارائه داده اند که بین نویز و قوانین نادر تفاوت قائل می‌شود. این روش در حالی که کنترل تراکنش بین معیارهای حداقل و توانایی کشف قوانین نادر را کنترل می‌کند، امکان استخراج قوانین را نیز فراهم می‌کند. الگوریتم با شبیه سازی مبتنی بر سلسله مراتب قوانین واقعی آزمایش می‌شود و اثر بخشی آن از طریق شناسایی قوانین صحیح، حتی اگر نادر باشد، در سطوح مختلف نویز نشان داده می‌شود[۱۲].
. Yongshun Gong و همکاران در سال ۲۰۱۷، ابتدا یک استراتژی را پیشنهاد داده اند تا محدودیت IPS را برای استخراج NSP مورد استفاده قرار دهند. سپس یک روش بهینه سازی ذخیره سازی برای نگهداری این اطلاعات IPS ارائه داده اند. سرانجام، یک الگوریتم کارآمد به نام الگوی متوالی منفی کارآمد استخراج از دو الگوی متوالی مثبت مکرر و نادر (e-NSPFI[11]) برای استخراج NSP پیشنهاد شده است. نتایج تجربی نشان می‌دهد که e-NSPFI می‌تواند به طرز موثری الگوهای منفی بسیار جالب تری از e-NSP پیدا کند.

۲-۲ استخراج الگوی مبتنی بر محدودیت

یک فرایند داده کاوی ممکن است هزاران قانون از یک مجموعه داده را کشف کند، که در نهایت بیشتر آنها برای کاربران غیر مرتبط یا جالب نیستند. غالباً، کاربران درک خوبی از اینکه "جهت" استخراج ممکن است به الگوهای جالب توجه و "شکل" الگوها یا قوانینی که می‌خواهند بیابند، منجر می‌شود. آنها همچنین ممکن است احساس "شرایط" قوانین را داشته باشند، که باعث می‌شود کشف برخی

---

[9] Negative sequential patterns
[10] Infrequent Positive Sequences
[11] Efficient Mining Negative Sequential Pattern

قوانین که می‌دانند مورد علاقه آنها نیست، از بین برود. بنابراین، یک ابتکار خوب این است که کاربران چنین شهود یا انتظاراتی را به عنوان محدودیت‌هایی برای محدود کردن فضای جستجو مشخص کنند. این استراتژی به عنوان استخراج الگوی مبتنی بر محدودیت شناخته می‌شود[۱۳]. استخراج الگوی متوالی مبتنی بر محدودیت با هدف شناسایی الگوهای مکرر در یک پایگاه داده متوالی از آیتم‌ها در حالی که محدودیت‌های مشخص شده بر روی ویژگی‌های مورد را مشاهده می‌کند.

Hosseininasab و همکاران در سال ۲۰۱۹، تکنیک‌های جدیدی را برای استخراج الگوی پرتکرار ترتیبی مبتنی بر محدودیت معرفی کرده اند که متکی به نمایش نمودار چند مقداری (MDD[۱۲]) است. به طور خاص، این مقاله می‌تواند چندین ویژگی آیتم و انواع محدودیت‌های مختلف، از جمله تعدادی از محدودیت‌های غیر یکنواخت را در خود جای دهد. برای ارزیابی کاربرد رویکرد، یک الگوریتم پیش بینی طرح ریزی مبتنی بر MDD ایجاد شده و عملکرد آن را با یک نوع تولید الگوی معمولی بررسی شده است. همچنین با یک الگوریتم استخراج الگوی متوالی مبتنی بر محدودیت مقایسه شده است. نتایج نشان می‌دهد که این روش از نظر مقیاس پذیری و کارایی با روش‌های دیگر قابل رقابت یا برتر است[۱۴].

روشهای مبتنی بر محدودیت برای الگوهای استخراج در سالهای اخیر توسعه یافته است. آنها بر اساس روش بالا به پایین برای هرس الگوهای نامزد انتخاب می‌شوند. با این حال، برای استخراج الگوی عظیم، روش‌های از پایین به بالا روش‌های کارآمدی هستند، اما رویکردهای قبلی برای هرس الگوهای نامزد بر اساس روش بالا به پایین نمی توانند در استفاده از الگوی عظیم با محدودیت هنگام استفاده از روش پایین به بالا اعمال شوند. Thanh-Long Nguyen وهمکاران در سال ۲۰۱۷، مشکل استخراج الگوی عظیم با محدودیت‌های الگوی را بیان کرده اند. در این مقاله یک قضیه برای هرس کارآمد الگوهای نامزد با روش پایین به بالا ایجاد شده است. در آخر، یک الگوریتم کارآمد برای استخراج الگوهای مبتنی بر محدودیت‌های عظیم بر اساس این قضیه پیشنهاد شده است[۱۵].

Sofya Titarenko و همکاران در سال ۲۰۱۹، دو الگوریتم جدید، c-FaRPaM1 و c-FaRPaM2 ارائه داده اند که می‌تواند در طیف گسترده ای از مجموعه داده‌های زمانی اعمال شود. هر دو آنها عدم قطعیت زمانی و محدودیت‌های زمانی و آیتم را در الگوهای پرتکرار مجاز می‌دانند. الگوریتم‌ها برای ایستگاه‌های کاری چند هسته ای بسیار بهینه شده اند. آنها از چند رشته و بردارسازی مجاز در معماری‌های مدرن بهره می‌برند. یک ساختار ذخیره سازی اطلاعات جدید (جمع و جورتر) سرعت محاسبات را بهبود می‌بخشد. الگوریتم دوم (c-FaRPaM2) برای بدست آوردن پیشرفتهای بیشتر در سرعت، از دانش قبلی در مورد ساختار داده بهره می‌برد[۱۵].

۲-۳ استخراج الگوهای با ابعاد بالا و الگوهای عظیم

با توجه به افزایش حجم داده‌های تولید شده و به خصوص داده‌های تراکنشی، استخراج مجموعه‌های پرتکرار از یک مجموعه داده بزرگ یک اشکال عمده برای استخراج الگوهای پرتکرار در داده کاوی است. تعداد زیادی از اقلام تولید شده که حداقل پشتیبانی را دارند، حافظه کامپیوتر را برای ذخیره و پردازش بی مورد به هدر می‌دهند. بنابراین، مفاهیم آیتم‌های پرتکرار مشابه(CFI[۱۳]) و حداکثر آیتم‌های پرتکرار (MFI[۱۴]) معرفی شدند. CFI مجموعه مواردی با پشتیبانی است که حداقل پشتیبانی را برآورده می‌کند و مجموعه‌های جانبی مناسب با پشتیبانی مشابه را ندارد. از طرف دیگر، MFI مجموعه ای از اقلام است که حداقل پشتیبانی را برآورده می‌کند و هیچ یک از مجموعه‌های برتر آن حداقل پشتیبانی را برآورده نمی کند. اگرچه MFI نسبت به CFI جمع و جورتر است، اما منجر به از دست دادن اطلاعات می‌شود. در عوض، کل مجموعه آیتم‌های پرتکرار را می‌توان بر اساس اطلاعات CFI بدست آورد. بنابراین، می‌توان نتیجه گرفت که CFI حاوی اطلاعات کاملی از مجموعه‌های آیتم‌های پرتکرار است. بنابراین، تولید الگوریتم استخراج را می‌توان با محدود کردن آن در مجموعه موارد آیتم‌های پرتکرار مشابه، کاهش داد[۱۶].

---

[12] Multi Dimension Diagram
[13] Closed Frequent Itemset
[14] Maximum Frequent Itemset

Nguyen و همکاران در سال ۲۰۱۷، الگوریتم‌های کارآمد و جدیدی برای استخراج الگوهای عظیم ارائه داده اند. در مرحله اول، درخت CP[15] (الگوی عظیم) طراحی شده است. در مرحله بعدی، دو قضیه برای محاسبه سریع الگوها و هرس الگوها بدون از دست دادن اطلاعات در الگوهای عظیم ایجاد می‌کند. براساس درخت CP و این قضیه‌ها، الگوریتمی (به نام CP-Miner) برای حل مسئله الگوهای عظیم استخراج پیشنهاد شده است. بنابراین یک استراتژی مرتب سازی برای استخراج موثر الگوهای عظیم توسعه یافته است. این استراتژی به کاهش تعداد نامزدهای مهم و زمان مورد نیاز برای بررسی زیرمجموعه‌ها و زیر مجموعه‌ها کمک می‌کند. سپس الگوریتم PCP-Miner که از این استراتژی استفاده می‌کند، پیشنهاد شده است و همچنین آزمایشاتی را برای نشان دادن کارایی این الگوریتم‌ها انجام یافته است[۱۷].

Vanahalli و همکاران در سال ۲۰۱۹، الگوریتمی برای هرس مجموعه کاملی از ویژگی‌های بی ربط و آیتم‌های بی ربط ارائه کرده اند و یک الگوریتم استخراج کارآمد تعویض پویای مجموعه آیتم‌های نزدیک عظیم (DSFCCIM[16]) تأکید کرده اند. الگوریتم پیشنهادی DSFCCIM به طور کارآمد بین روش‌های شمارش آیتم‌ها و ویژگی‌ها بر اساس ویژگی‌های داده در طول فرآیند استخراج جابجا می‌شود. علاوه بر این، الگوریتم DSFCCIM با یک جدول جدید کاردینالیتی آیتم‌ها، جدول حمایت مجموعه آیتم، دو روش کارآمد برای بررسی نزدیکی مجموعه آیتم‌ها و دو استراتژی هرس کارآمد برای کاهش فضای ادغام شده است. الگوریتم پیشنهادی DSFCCIM اولین الگوریتم سوئیچینگ دینامیکی برای استخراج مجموعه آیتم‌های پرتکرار نزدیک عظیم از مجموعه داده‌های متشکل از تعداد زیادی ردیف و تعداد زیادی ویژگی است[۱۸].

آنها همچنین در سال ۲۰۲۰، الگوریتم استخراج آیتم‌های پرتکرار عظیم موازی توزیع شده متعادل (BDPFCCIM) برای مجموعه داده‌های با ابعاد بالا طراحی کرده اند. یک روش بررسی نزدیکی کارآمد برای بررسی نزدیکی ردیف و یک استراتژی هرس کارآمد برای شکستن فضای جستجوی استخراج شمارش سطر با الگوریتم پیشنهادی BDPFCCIM محصور شده است. الگوریتم پیشنهاد شده BDPFCCIM اولین الگوریتم متعادل کننده بار توزیع شده برای استخراج آیتم‌های پرتکرار نزدیک عظیم از مجموعه داده‌های بیولوژیکی با ابعاد بالا است. نتایج تجربی عملکرد کارآمد الگوریتم BDPFCCIM پیشنهادی را در مقایسه با الگوریتم‌های پیشرفته ارائه می‌دهد[۱۹].

Sureshan و همکاران در سال ۲۰۱۷، یک الگوریتم موازی جدید برای استخراج الگوهای پرتکرار نزدیک عظیم از مجموعه داده‌های با ابعاد بالا پیشنهاد داده اند. این الگوریتم در چارچوب Hadoop برای بهره برداری موازی توزیع شده ذاتی خود با استفاده از مدل برنامه نویسی MapReduce پیاده سازی شده است. نتایج آزمایش نشان می‌دهد که الگوریتم موازی پیشنهادی در چارچوب Hadoop عملکردی کارآمد از نظر زمان اجرا در مقایسه با الگوریتم‌های موجود ارائه می‌دهد[۲۰].

۲-۴ استخراج الگوهای فشرده یا تقریبی

یک چالش اساسی در استخراج مکرر الگوی، تعداد بسیار زیادی از الگوهای کشف شده است. استفاده از حداقل آستانه پشتیبانی برای کنترل تعداد الگوهای یافت شده تأثیر محدودی دارد. مقدار بسیار کم می‌تواند منجر به تولید تعداد زیادی از الگوهای خروجی شود، در حالی که مقدار بسیار زیاد می‌تواند منجر به کشف تنها الگوهای متغیر شود. برای کاهش مجموعه عظیمی از الگوهای مکرر تولید شده در کاوش ضمن حفظ الگوهای با کیفیت بالا، می‌توانیم یک مجموعه فشرده یا تقریبی از الگوهای مکرر را استخراج کنیم. بالاترین الگوهای نزدیک مرتبه بالا برای ایجاد تمرکز فرآیند استخراج فقط روی مجموعه ای از الگوهای متداول ارائه شد. اگرچه جالب است، آنها معمولاً به دلیل توزیع ناهموار فرکانس در میان اقلام، نمونه‌های k را نشان نمی دهند[۱۳].

Baek و همکاران در سال ۲۰۲۰، برای غلبه بر محدودیت آیتم‌های زیاد و قوانین انجمنی نامرتبط و خطهای موجود در میان الگوها، یک روش جدید با در نظر گرفتن نویز به منظور استخراج الگوهای تقریبی، پیشنهاد کرده اند. روش پیشنهادی دامنه ابزارهای قابل اعتماد برای

---

[15] Colossal Pattern
[16] Dynamic Switching Frequent Colossal Closed Itemset Mining

الگوها را با استفاده از یک عامل تحمل اطمینان محاسبه می‌کند. با استفاده از این فاکتور، می‌توان الگوهای قدرتمند پرتکرار را که از آن به عنوان تقریبی ازالگوهای سودمند مرتبط استفاده می‌شود، از یک پایگاه داده نویزی استخراج کرد. برای ارزیابی عملکرد الگوریتم پیشنهادی، آزمایشات مختلفی از نظر زمان اجرا، استفاده از حافظه و مقیاس پذیری طراحی و انجام شده است. نتایج تجربی نشان می‌دهد که الگوریتم پیشنهادی نسبت به رقبا، بهتر عمل می‌کند[۲۱].

Acosta-Mendoza و همکارن در سال ۲۰۲۰، یک الگوریتم کارآمد برای استخراج زیر مجموعه ای از زیر گراف‌ها تقریبی پر تکرار در مجموعه‌های نمودارهای چند سطحی پیشنهاد کرده اند. الگوریتم پیشنهادی با محاسبه زیر مجموعه ای از زیر گراف‌ها تقریبی پر تکرار به گزینه ای برای کاهش تعداد زیرگراف‌هاس تقریبی پرتکرار استخراج شده تبدیل می‌شود. مهم است که برجسته شود، تا آنجا که دانش ماست ؛ این روش اولین الگوریتم استخراج زیر گراف‌ها تقریبی پر تکرار در مجموعه‌های گرافیکی چند سطحی است. این روش در مقایسه با سایر راه حل‌های گزارش شده مقایسه شده و از طریق نمودارهای چند سطحی مصنوعی و دنیای واقعی ارزیابی شده است. علاوه بر این، مفید بودن الگوهای استخراج شده توسط این روش برای طبقه بندی تصویر نشان داده شده است[۲۲].

۳. خلاصه

استخراج الگوی پرتکرار به دلیل تحقیقات قابل توجه، گسترش بسیاری از دامنه مسئله و مطالعات گسترده کاربرد بسیار فراتر از اصول اولیه رسیده است. در این مقاله، روش‌های استخراج الگوی پیشرفته بررسی شده است. مقاله حاضر روش‌هایی را برای استخراج انواع مختلف الگوها معرفی کرده است. در این مقاله روشهای استخراج انواع مختلف الگوها در دسته‌های زیر تسیم بندی شده اند: الگوهای چند سطحی، الگوهای چند بعدی، الگوهای پرتکرار مبتنی بر محدودیت، الگوهای پرتکرار در داده‌های با ابعاد بالا، الگوهای عظیم و الگوهای تقریبی. توجه داشته باشید که استخراج الگو اصطلاح عمومی تری نسبت به استخراج الگوی پرتکرار است، زیرا نمونه اول نیز الگوی نادر و منفی را در بر می‌گیرد. با این حال، وقتی هیچ ابهامی وجود ندارد، این دو اصطلاح به جای هم استفاده می‌شوند. در جدول ۱ خلاصه ای از روش‌های بررسی شده در مقاله حاضر نشان داده شده است.

جدول ۱. خلاصه روش‌های بررسی شده

تکنیک استخراج الگو		نویسنده و سال چاپ	روش مورد استفاده در مقاله
استخراج الگوهای چند سطحی	استخراج قوانین انجمنی چند سطحی	Zhiguo Qu ۲۰۱۶	معماری استخراج الگوی چند سطحی برای پشتیبانی از مدیریت خودکار شبکه
		Yang Xu ۲۰۱۴	روش مبتنی بر الگوریتم ژنتیک و درخت دسته بندی
		J. H. DUAN ۲۰۱۷	الگوریتم استخراج قوانین انجمنی چندسطحی بهبود یافته با استفاده از پایگاه داده چند شاخصی
		Prajapati ۲۰۱۷	استخراج قوانین انجمنی چندسطحی شامل قوانین بین سطح
	استخراج قوانین انجمنی چند بعدی	Yokobayashi ۲۰۱۸	چارچوب مناسب برای داده کاوی چند بعدی بر اساس تانسور
		Yuan Tian ۲۰۲۰	استخراج قوانین انجمنی چند بعدی در زمان واقعی سیستم شبکه برق
		Pradhan ۲۰۱۷	استخراج الگوهای داده‌های سری زمان واقعی فعالیت‌های عضلانی شرکت کنندگان انسانی
	استخراج قوانین انجمنی کمی	D. Martín ۲۰۱۸	چارچوب موازی عمومی جدید برای کشف قوانین انجمنی کمی در مقدار زیادی از داده‌ها

	استخراج قوانین انجمنی کمیاب	NuritGal-Oz ۲۰۱۹	استخراج الگوهای کمیاب برای غلبه بر نویز
		Yongshun Gong ۲۰۱۷	استخراج الگوهاس منفی برای رفع محدودیت توالیهای مثبت نادر
استخراج الگوی مبتنی بر محدودیت		Hosseininasab ۲۰۱۹	ستخراج الگوی پرتکرار ترتیبی مبتنی بر محدودیت متکی به نمایش نمودار جند مقداری
		Thanh-Long ۲۰۱۷ Nguyen	استخراج الگوی عظیم با محدودیت‌های الگو برای هرس کارآمد الگوهای نامزد با روش پایین به بالا
		Sofya Titarenko ۲۰۱۹	استخراج الگوهای پرتکرار بر اساس محدودیت‌های عدم قطعیت زمانی و محدودیت‌های زمانی و آیتم در الگوهای پرتکرار مجاز
استخراج الگوهای با ابعاد بالا و الگوهای عظیم		۲۰۱۷ Nguyen	الگوریتم‌های کارآمد برای استخراج الگوهای عظیم با استفاده از درخت CP (الگوی عظیم)
		۲۰۱۹ Vanahalli	الگوریتمی برای هرس مجموعه کاملی از ویژگی‌های بی ربط و آیتم‌های بی ربط
		۲۰۲۰ Vanahalli	استخراج آیتم‌های پرتکرار عظیم موازی توزیع شده متعادل
		Sureshan ۲۰۱۷	الگوریتم موازی جدید برای استخراج الگوهای پرتکرار نزدیک عظیم
استخراج الگوهای فشرده یا تقریبی		Baek ۲۰۲۰	غلبه بر محدودیت آیتم‌های زیاد و قوانین انجمنی نامرتبط و خطهای موجود در مین الگوها
		Acosta-Mendoza ۲۰۲۰	استخراج زیر مجموعه ای از زیر گراف‌ها تقریبی پر تکرار در مجموعه‌های نمودارهای چند سطحی

همانطور که در جدول ۱ نشان داده شده است، روش‌های استخراج الگوهای پرتکرار با توجه به به زیردسته مربوطه بررسی شده است. با توجه به کاربردهای متنوع استخراج الگوها در مجموعه داده‌های تراکنشی، تنوع روش‌های استخراج الگوهای پیشرفته روز به روز بیشتر می‌شود. از این رو در این مقاله به معرفی زیر شاخه‌های استخراج الگوهای پیشرفته پرداختیم و برای هر یک از روش‌ها موجود در زیرشاخه‌ها، برخی از تحقیقات انجام شده را بررسی کردیم.

۴. نتیجه‌گیری

الگوهای مکرر الگوهایی هستند که اغلب در یک مجموعه داده ظاهر می‌شوند. ایتم‌های موجود در مجموعه داده‌های تراکنشی در صورتی که به طور مکرر در یک پایگاه داده تراکنشی رخ دهد، به عنوان یک الگوی پرتکرار است. یافتن الگوهای مکرر در ارتباطات استخراج، همبستگی و بسیاری از روابط جالب دیگر بین داده‌ها نقش اساسی دارد. بنابراین، الگوی کاوی مکرر به یک کار مهم داده کاوی و به یک موضوع متمرکز در تحقیقات داده کاوی تبدیل شده است. استخراج الگوهای پیشرفته فراتر از استخراج الگوهای پرتکرار است که سعی در بهبود نقاط ضعف روش‌ها استخراج الگوهای پرتکرار دارد. در این مقاله ابتدا روش‌های استخراج الگوهای پرتکرار به چهار دسته کلی استخراج الگوهای چند سطحی، استخراج الگوی مبتنی بر محدودیت، استخراج الگوهای با ابعاد بالا و الگوهای عظیم و استخراج الگوهای فشرده یا تقریبی تقسیم شد. استخراج الگوهای چند سطحی خود به چهار دسته استخراج قوانین انجمنی چند سطحی، استخراج قوانین

انجمنی چند بعدی، استخراج قوانین انجمنی کمی و استخراج قوانین انجمنی کمیاب تقسیم می‌شود. در مقاله حاضر دسته ای و زیر دسته‌های مربوط به استخراج الگوی پیشرفته معرفی شده و برخی از تحقیقات هر کدام از این دسته‌ها و زیر دسته‌ها به تفکیک بررسی شده است.

۵. منابع

۱. Fournier-Viger, P., et al., *A survey of sequential pattern mining.* Data Science and Pattern Recognition, 2017. **1**(1): p. 54-77.

۲. Le, T., et al., *Mining top-k frequent patterns from uncertain databases.* Applied Intelligence, 2020: p. 1-11.

۳. Chaki, J ,.et al., *Pattern mining approaches used in sensor-based biometric recognition: A review.* IEEE Sensors Journal, 2019. **19**(10): p. 3569-3580.

۴. Qu, Z., et al., *Multilevel pattern mining architecture for automatic network monitoring in heterogeneous wireless communication networks.* China communications, 2016. **13**(7): p. 108-116.

۵. Xu, Y., et al., *A genetic algorithm based multilevel association rules mining for big datasets.* Mathematical Problems in Engineering, 2014. **2014**.

۶. DUAN, J. and M. YUAN, *Multilevel Index Algorithm Based on Improved Association Rule Mining.* DEStech Transactions on Computer Science and Engineering, 2017(cii.(

۷. Prajapati, D.J., S. Garg, and N.C. Chauhan, *MapReduce based multilevel consistent and inconsistent association rule detection from big data using interestingness measures.* Big data research, 2017. **9**: p. 18-27.

۸. Tian, Y. and W. Ma. *Real-time Multidimensional Data Mining and Analysis Technology Based on Big Data*. in *2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)*. 2020. IEEE.

۹. Pradhan, G.N. and B. Prabhakaran, *Association rule mining in multiple, multidimensional time series medical data.* Journal of Healthcare Informatics Research, 2017. **1**(1): p. 92-118.

۱۰. Martín, D., et al., *MRQAR: A generic MapReduce framework to discover quantitative association rules in big data problems.* Knowledge-Based Systems, 2018. **153**: p. 176-192.

۱۱. Koh, Y.S. and S.D. Ravana, *Unsupervised rare pattern mining: a survey.* ACM Transactions on Knowledge Discovery from Data (TKDD), 2016. **10**(4): p. 1-29.

۱۲. Gal-Oz, N., Y. Gonen, and E. Gudes, *Mining meaningful and rare roles from web application usage patterns.* Computers & Security, 2019. **82**: p. 296-313.

۱۳. Han, J., J. Pei, and M. Kamber ,*Data mining: concepts and techniques*. 2011: Elsevier.

۱۴. Hosseininasab, A., W.-J. van Hoeve, and A.A. Cire. *Constraint-Based Sequential Pattern Mining with Decision Diagrams*. in *Proceedings of the AAAI Conference on Artificial Intelligence*. 2019.

۱۵. Nguyen, T.-L., et al. *Constraint-based method for mining colossal patterns in high dimensional databases*. in *International Conference on Information Systems Architecture and Technology*. 2017. Springer.

۱۶. Zaki, F.A.M. and N.F. Zulkurnain, *Frequent Itemset Mining in High Dimensional Data: A Review*, in *Computational Science and Technology*. 2019, Springer. p. 325-334.

۱۷. Nguyen, T.-L., B. Vo, and V. Snasel, *Efficient algorithms for mining colossal patterns in high dimensional databases.* Knowledge-Based Systems, 2017. **122**: p. 75-89.

۱۸. Vanahalli, M.K. and N. Patil, *An efficient dynamic switching algorithm for mining colossal closed itemsets from high dimensional datasets.* Data & Knowledge Engineering, 2019. **123**: p. 101721.

۱۹. Vanahalli, M.K. and N. Patil, *Distributed load balancing frequent colossal closed itemset mining algorithm for high dimensional dataset.* Journal of Parallel and Distributed Computing, 2020.

۲۰. Sureshan, S., et al., *Mining closed colossal frequent patterns from high-dimensional dataset : Serial versus parallel framework*, in *Progress in Intelligent Computing Techniques: Theory, Practice, and Applications*. 2018, Springer. p. 317-326.

۲۱. Baek, Y., et al., *Approximate high utility itemset mining in noisy environments.* Knowledge-Based Systems. **212**: p. 106596.

۲۲. Acosta-Mendoza, N., et al., *Mining clique frequent approximate subgraphs from multi-graph collections.* Applied Intelligence, 2020. **50**(3): p. 878-892.