مروری بر روش‌های حفظ حریم خصوصی در امنیت پایگاه داده تراکنشی

زهرا شیخی نژاد[1]، محمد نادری دهکردی[2]، حمید رستگاری[3]

[1]دانشجوی کارشناسی ارشد کامپیوتر نرم‌افزار دانشگاه آزاد اسلامی واحد نجف‌آباد، دانشکده مهندسی کامپیوتر;S_sheykhinezhad@yahoo.com
[2]استادیار دانشگاه آزاد اسلامی واحد نجف‌آباد، دانشکده مهندسی کامپیوتر;Naderi@iaun.ac.ir
[3]استادیار دانشگاه آزاد اسلامی واحد نجف‌آباد، دانشکده مهندسی کامپیوتر; Rastegari@iaun.ac.ir

چکیده

حفظ حریم خصوصی در داده‌کاوی موجب افزایش بهره‌وری از داده‌کاوی شده است. حفظ حریم خصوصی داده‌کاوی به دنبال یافتن مسیری برای استفاده از داده‌کاوی، بدون نگرانی‌های حاصل از افشای اطلاعات حساس می‌باشد. برای برآورده شدن این هدف، نیاز به اعمال تغییرات روی پایگاه داده اصلی می‌باشد. به‌طوری‌که اطلاعات حساس و محرمانه در اثر تکنیک‌های داده‌کاوی استخراج نشوند. در این مقاله تمرکز بر روی پنهان‌سازی دانش حساس، برای جلوگیری از به خطر افتادن اطلاعات محرمانه می‌باشد. این مقاله به بررسی روش‌های موجود جهت حفظ حریم خصوصی پرداخته است و بابیان کارایی و محدودیت‌های هرکدام از تکنیک‌ها، بستر مناسبی برای محققان فراهم ساخته است تا بتوانند با توجه به کارایی و کاربرد هر روش، بهترین را جهت ایمن‌سازی پایگاه داده تراکنشی موردنظر به کار ببندند.

کلمات کلیدی

پنهان‌سازی قواعد انجمنی، داده‌کاوی، حفظ حریم خصوصی در داده‌کاوی، الگوریتم پنهان‌سازی

۱- مقدمه

داده‌کاوی، انجام عملیات استخراج اطلاعات از بین انبوه داده‌ها می‌باشد، طوری که اطلاعات حساس موجود در پایگاه داده افشاء نشود. روش‌ها و راه‌حل‌های بسیاری در این زمینه بیان‌شده است و محققان همچنان به دنبال ارائه راه‌حل‌های جدید و بهبود روش‌های قبلی می‌باشند.

در همین راستا به دنبال تکمیل مقاله [1] به بررسی دقیق‌تر تکنیک‌های داده‌کاوی پرداخته‌شده است و در این مقاله تمرکز بیشتر بر روی انتخاب تکنیک و اصلاح برخی الگوریتم‌های بیان‌شده می‌باشد.

تکنیک‌های داده‌کاوی به دودسته کلی پیشگویی[2] و توصیفی[3] تقسیم‌بندی شده‌اند[2] . یکی از پرکاربردترین تکنیک‌های موجود در دسته توصیفی، قواعد انجمنی[4]می‌باشد.

تمرکز اصلی این مقاله بر روی قوانین انجمنی می‌باشد. این قوانین ارتباط‌های موجود بین داده‌ها را کشف می‌کند[3]. ادامه این مقاله به‌صورت زیر ارائه‌شده است: در قسمت دوم تعریف کاملی از داده‌کاوی و نمودار درختی از شاخه‌های داده‌کاوی و کاربردهای مهم آن ذکرشده است.در بخش سوم فاکتورهای ارزیابی الگوریتم‌ها بررسی‌شده است. سپس در بخش چهارم طبقه‌بندی الگوریتم‌های پنهان‌سازی قواعد انجمنی و بررسی هر طبقه و مقایسه‌ای بین روش‌ها بیان می‌شود. و درنهایت در بخش پنجم نتیجه‌گیری ارائه‌شده است.

۲- تعریف و کاربرد داده‌کاوی

[1] زهرا شیخی نژاد

[2]Predictive
[3]Descriptive
[4]Association Rules

داده‌کاوی اشاره به استخراج و کشف دانش از بین مقادیر زیادی داده خام دارد[4]. کشف دانش، فرآیندی می‌باشد که دو مرحله اصلی را شامل می‌شود:

۱- پیش‌پردازش داده،که در این مرحله داده‌ها برای استخراج دانش آماده‌سازی می‌شوند. و شامل گام‌های زیر می‌باشد:

- یکپارچه‌سازی[5] :چندین منبع داده با هم ترکیب می‌شوند و منبع داده یکسانی را ایجاد می‌کنند.
- پاک‌سازی داده[6] : داده‌های ناسازگار و نویزهای موجود در داده‌ها حذف می‌شود.هدف از این پاک‌سازی تهیه‌ی داده‌های باکیفیت است.
- انتخاب داده[7] :داده‌های موردنیاز انتخاب می‌شوند همچنین داده‌های مرتبط با آنالیز پایگاه داده بازیابی می‌شوند.
- تبدیل داده[8] : داده‌ها به‌گونه‌ای تبدیل یا همسان‌سازی می‌شوند تا مناسب عملیات داده‌کاوی گردند.

۲- داده‌کاوی،که این مرحله شامل گام‌های زیر می‌باشد:

- داده‌کاوی: فرآیند اصلی استخراج الگو از بین داده‌های موجود.
- ارزیابی الگو[9] : بررسی الگو به‌وسیله معیارهای اندازه‌گیری، ازنظر سودمندی
- ارائه دانش[10] : نمایش دانش استخراج‌شده به کاربران با استفاده از تکنیک‌های بصری

روش‌های مختلفی برای داده‌کاوی وجود دارد که این روش‌ها با بررسی داده‌های خام و تعیین مناسب‌ترین مدلی که به ویژگی‌های آن‌ها شباهت بیشتری دارد، داده‌ها را در قالب مدلی مناسب ارائه دهند.

۲-۱ روش‌های داده‌کاوی

برای بیان روش‌ها و الگوریتم‌های کاربردی ابتدا به داده‌کاوی اشاره‌کرده‌ایم و سپس روش‌های داده‌کاوی ذکر می‌شود که عبارت‌اند از:

۱- روش پیش‌بینی: مقدار داده را با استفاده از مقادیر قبلی و بر پایه استنتاج از نتایج، حدس می‌زند.

۲- روش توصیفی: به دنبال کشف راهی برای آگاهی از خصوصیات داده می‌باشد.

۲-۱-۱ روش‌های مربوط به داده‌کاوی پیش‌بینی

- کلاس‌بندی[11] : موجب شناسایی ویژگی‌های گروهی که هر مورد به آن تعلق دارد می‌شود[5]. از این داده‌ها برای پیش‌بینی نوع رفتار نمونه‌های جدید استفاده می‌شود و هم باعث فهم بیشتر داده‌های موجود در هر گروه می‌شود. داده‌کاوی مدل طبقه‌بندی را با توجه به داده‌های دسته‌بندی‌شده‌ی پیشین ایجاد می‌کند و به‌صورت استقرایی الگوی پیش‌بینی کننده را می‌یابد. تکنیک‌های داده‌کاوی مورداستفاده در طبقه‌بندی شامل تکنیک‌های درخت تصمیم‌گیری و شبکه‌های عصبی می‌باشند. کاربرد ویژگی طبقه‌بندی در بررسی تقلب و تصویب اعتبار می‌باشد.
- رگرسیون[12] : با استفاده از مقادیر موجود، مقادیر دیگر را پیش‌بینی می‌کند. رگرسیون معمولاً از تکنیک‌های آماری استاندارد مثلاً رگرسیون خطی استفاده می‌کند. ولی در دنیای واقعی همه مقادیر به‌صورت تصویر خطی از مقادیر قبلی نیستند و ممکن است تکنیک‌های پیچیده‌تری برای پیش‌بینی موردنیاز باشد. تکنیک‌هایی چون درخت تصمیم، شبکه‌های عصبی و رگرسیون منطقی، می‌توانند در این موارد به کار گرفته شود.
- تحلیل زمانی[13] : مقادیر ناشناخته را با استفاده از پیش‌بینی کننده‌های متغیر بازمان، مثل رگرسیون، پیش‌بینی می‌کند.

---

[5]Data cleaning
[6]Data integration
[7]Data selection
[8]Data transformation
[9]Pattern evaluation
[10]Knowledge presentation
[11]Classification
[12]Regression
[13]Time series

۲-۱-۲ روش‌های مربوط به داده‌کاوی توصیفی

- کشف توالی[14] :کشف مواردی که اتفاق افتادن آن‌ها در وجود یا عدم وجود موارد دیگر نقش دارد[5].
- قواعد انجمنی: یافتن قواعد در مواردی که باهم اتفاق می‌افتند را کشف قواعد وابستگی می‌نامند. مثلاً اجناسی که احتمال باهم خریده شدن آن‌ها در فروشگاه زیاد است، کنار هم قرار بگیرند.
- خوشه‌بندی[15] : در این روش داده‌های موجود در یک گروه بسیار شبیه به هم می‌باشند و داده‌های گروه‌های مختلف بیشترین تفاوت را باهم دارند. در خوشه‌بندی گروه‌ها از پیش تعین شده نمی‌باشند و مشخص نیست گروه‌ها بر اساس کدام خصوصیات می‌باشند. بنابراین بعد از انجام خوشه‌بندی نیاز می‌باشد توضیحاتی توسط فرد خبره راجع به خوشه‌ها و فاکتورهای طبقه‌بندی آورده شود. بعدازاینکه داده‌ها به گروه‌های توجیه‌پذیر تقسیم شدند از آن‌ها می‌توان برای کسب اطلاعات در مورد داده جدید استفاده کرد. از مهم‌ترین الگوریتم‌های مورداستفاده در خوشه‌بندی می‌توان به k-means و Kohnen اشاره نمود[6].
- خلاصه‌سازی[16] : برای شناخت دقیق داده‌ها و انجام داده‌کاوی روی آن‌ها از ابزارهایی برای خلاصه‌سازی نتایج استفاده می‌شود. ابزارهایی چون انحراف معیار و میانگین در این زمینه مفید می‌باشند. و برای شناخت داده‌ها می‌توان از ابزارهایی چون گراف سازی و تصویرسازی داده‌ها می‌توان استفاده کرد و توزیع مقادیر مختلف داده‌ها را یک نمودار مشاهده کرد و به‌طور تقریبی میزان داده‌های معیوب را حدس زد. ولی برای تحلیل پارامترهای مرتبط باهم نیاز به فرد خبره می‌باشد[7].

۲-۲ ماهیت الگوریتم‌های پنهان‌سازی

- روش دقیق[17] : به دنبال یافتن بهترین جواب برای برآوردن کلیه اهداف می‌باشد. و راه‌حلی بهینه با کمترین عوارض جانبی ارائه می‌دهد. در این رویکرد محدودیت زمانی مطرح نمی‌باشد و از ابزار برنامه‌ریزی دودویی جهت انتخاب تراکنش‌های مناسب برای تغییر، استفاده می‌شود و برای یافتن پاسخ از فرمول ریاضی استفاده می‌شود. متأسفانه، این امکان وجود دارد که رویکرد دقیق به جواب نرسد بنابراین با حذف برخی اهداف مواجه می‌شویم.
- روش مبتنی بر مرز[18] : با استفاده از مفهوم مرز، قواعد حساس ورودی را هرس می‌کند و با کاهش تعداد قواعد مخفی‌شده، موجب کاهش تغییرات و افزایش کیفیت پایگاه داده ایمن شده می‌شود. در این رویکرد از طریق اصلاح مرزهای اصلی، پنهان‌سازی دانش حساس ممکن می‌گردد.
- روش اکتشافی[19] : الگوریتم‌های روش اکتشافی فضای حالت مسئله را به‌صورت کامل پیمایش نمی‌کنند و سعی دارند با اجرای یک یا چند سیاست هوشمندانه این فضا را بسیار محدود کنند تا بتوانند بهترین راه‌حل را در آن فضا بیابند و سرعت پیدا کردن جواب را افزایش دهند. بنابراین الگوریتم‌های این روش سریع‌تر و کارآمدتر می‌باشند. در حقیقت در این رویکرد از سیاست تقسیم و غلبه استفاده می‌شود.بسیاری از الگوریتم‌های جدید به این مسیر تعلق دارند[6].

۲-۳ انواع الگوریتم‌های روش اکتشافی

- الگوریتم‌های تغییر داده[20] : داده‌های موجود در پایگاه داده متناسب با کاربرد آن دستخوش تغییراتی می‌شود[8].
- الگوریتم‌های نوسازی داده[21]: در عملیات نوسازی داده مجموعه‌ای از عناصر فراوان از پایگاه داده اصلی استخراج می‌شوند. تمامی تغییرات لازم، بر روی این مجموعه از عناصر استخراج‌شده انجام می‌شود. سپس پایگاه داده ایمن شده از روی مجموعه عناصر تغییریافته ایجاد می‌شود. انتظار می‌رود از چنین پایگاه داده‌ای، امکان استخراج قواعد حساس وجود نداشته باشد.

---

[14] Sequence discovery
[15] Clustering
[16] Summery
[17] Exact Approach
[18] Border-based Approach
[19] Heuristic Approach
[20] Data Modification
[21] Data Reconstruction

۲-۳-۱ انواع تغییر داده

- تغییر داده مستقیم: پنهان‌سازی قواعد انجمنی با کاهش ضریب اطمینان و پشتیبانی قانون.

روش تحریف داده‌ها_ روش مبتنی برConfidence: اگر الگوریتم پنهان‌سازی ازConfidenceاستفاده کرده باشد-Confidence basedمی‌باشد.

روش تحریف داده‌ها_ روش مبتنی برSupport: اگر الگوریتم پنهان‌سازی ازSupportاستفاده کرده باشدSupport-basedمی‌باشد.

- تغییر داده غیرمستقیم: با استفاده از روش‌های پیچیده‌تری نسبت به روش مستقیم، قواعد انجمنی را پنهان می‌کند.

۲-۴ روش‌های حفظ حریم خصوصی از دیدگاه توزیع‌شدگی منابع داده

- روش‌های مبتنی بر داده‌های متمرکز: در این روش عملیات داده‌کاوی به روی یک پایگاه داده واحد که حاوی تمامی موجودیت‌ها و تمامی صفات آن‌ها می‌باشد انجام می‌گیرد.
- روش‌های مبتنی بر داده‌های توزیع‌شده: برای حفظ حریم خصوصی در داده‌کاوی می‌توان داده‌ها را بین منابع متفاوت توزیع نمود. بخش‌بندی[22] به تقسیم‌بندی منطقی پایگاه داده یا عناصر تشکیل‌دهنده آن، بین اعضای مجزا و مستقل اشاره دارد.

۲-۵ انواع تغییرات در الگوریتم‌های تغییر داده

در این بعد، چگونگی توزیع داده‌ها بررسی می‌شود[9]. برخی از رویکردها برای پایگاه داده‌های متمرکز و برخی دیگر برای پایگاه داده‌های توزیع‌شده کاربرد دارند. از طرفی داده‌های توزیع‌شده به دودسته افقی[23] و عمودی[24] تقسیم می‌شوند. در داده‌های توزیع‌شده عمودی، مقادیر صفات[25] مختلف پایگاه داده در مکان‌های مختلف استقرار دارند درحالی‌که در داده‌های توزیع‌شده افقی، رکوردهای[26] مختلف پایگاه داده، در مکان‌های مختلف هستند.

۲-۵-۱ در داده‌های متمرکز

روش تحریف [27] داده‌ها با تکنیک آشفته سازی[28] : یکی از معروف‌ترین و پرکاربردترین تکنیک‌های موجود در عرصه پنهان‌سازی قواعد انجمنی است که با جایگزینی ۱ به‌جای۰ یا جایگزینی ۰ به‌جای ۱، عناصر موجود در پایگاه داده دچار تغییراتی می‌شوند،که در اثر این تغییرات عناصری که از حساسیت بالا برخوردار بودند از پایگاه داده حذف یا به‌گونه‌ای کم‌اهمیت جلوه می‌نمایند. این روش سبب ایجاد ناهماهنگی در پایگاه داده می‌شود. 2][، 10][، 5][

- روش مسدودسازی داده‌ها[29] : جایگزینی بعضی ارزش‌های اصلی با علامت سؤال یا null. 2][، 10][، 5][

در حقیقت عناصری که می‌خواهیم از دسترسی‌های غیرمجاز در امان باشد را نشان نداده و ؟ را جایگزین آن خواهیم کرد. پرواضح است که چنین پایگاه داده‌ای را می‌توان مورداستفاده قرارداد، ولی افراد سودجو توانایی استخراج قواعد حساسی که با این روش پنهان‌شده‌اند را نخواهند داشت. این روش در پایگاه داده‌های دارویی و پزشکی و امثال آن‌که داده در آن‌ها از حساسیت‌های خاص برخوردار است قابل‌استفاده نیست، زیرا ممکن است عواقب جبران‌ناپذیری را به همراه داشته باشد.

- تکنیک متراکم سازی و تجمیع[30] : در این تکنیک تعدادی از مقادیر داده‌ها با یکدیگر ترکیب می‌شوند.
- تکنیک تعویض کردن[31] : در این روش تعدادی از مقادیر رکوردهای اصلی با یکسری مقادیر منحصربه‌فرد تعویض می‌شود.

---

[22] Partitioning
[23] Horizontal
[24] Vertical
[25] Attribute
[26] Records
[27] Data Distortion
[28] Perturbation
[29] Data Blocking
[30] Aggregation

- تکنیک نمونه‌گیری[32] : انتشار داده برای تنها یک نمونه از یک جمعیت صورت می‌گیرد.
- تکنیک رمزنگاری[33] : در این روش مقادیر اصلی با استفاده از یکسری روش‌های رمزنگاری مثل کلیدهای متقارن، نامتقارن و ... رمزنگاری می‌شوند به‌گونه‌ای که مقادیر اصلی تغییر نمی‌کنند.

۲-۵-۲ در داده‌های توزیع‌شده

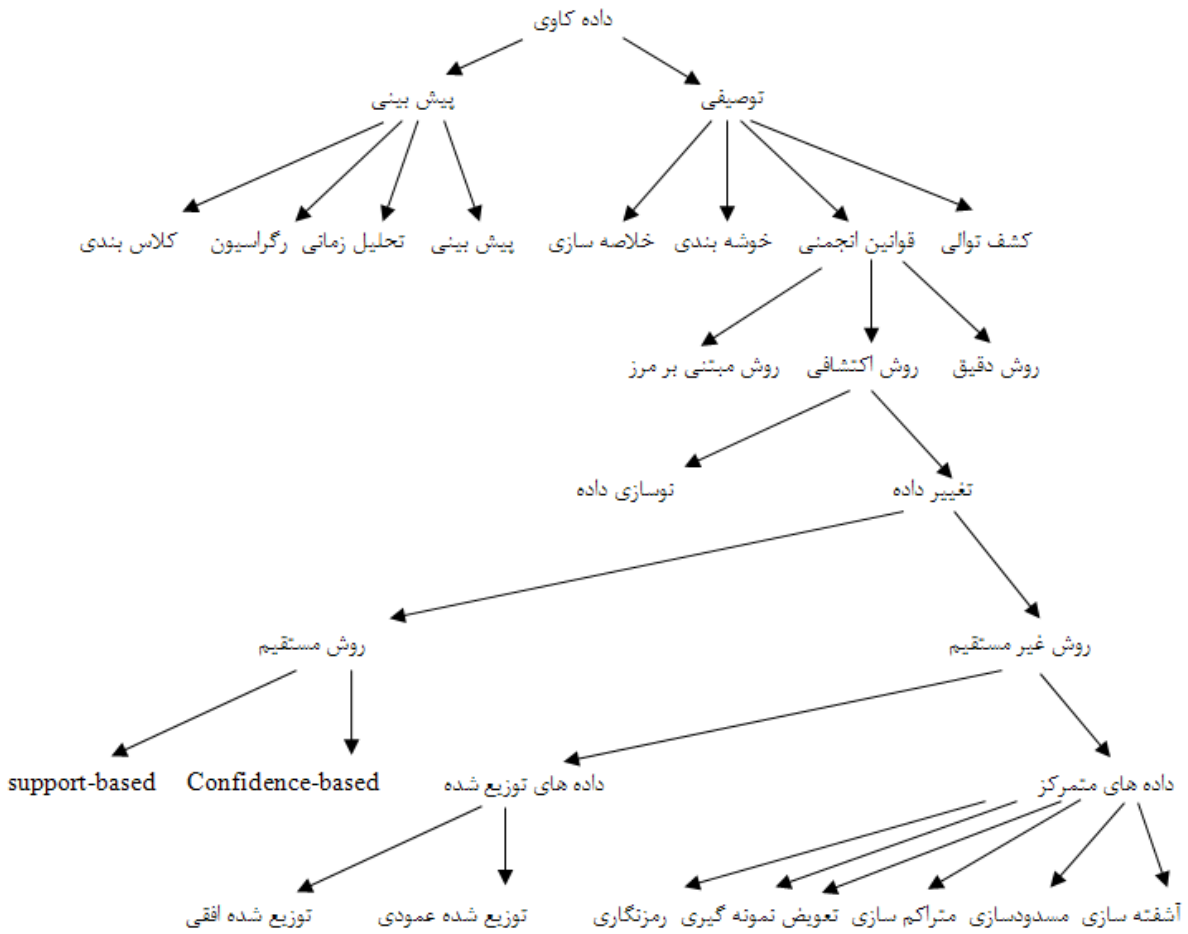
- بخش‌بندی افقی: هر پایگاه داده زیرمجموعه‌ای از موجودیت‌ها را شامل می‌شود و رکوردها کاملاً شبیه می‌باشند11][. یعنی هر پایگاه داده شامل اطلاعات کاملی درباره مجموعه‌ای از موجودیت‌هاست.
- بخش‌بندی عمودی: هر پایگاه داده زیرمجموعه‌ای از صفات همه موجودیت‌ها را شامل می‌شود. یعنی هر پایگاه داده بخشی از اطلاعات همه موجودیت‌ها را دارا می‌باشد.

برای حفظ حریم خصوصی در پایگاه داده‌های توزیع‌شده از روش‌های مبتنی بر رمزنگاری استفاده می‌شود. این روش‌ها بر اساس محاسبات امن کار می‌کنند. در محاسبات امن چند عضوی یا SMC(Secure Multi Computation)هیچ‌یک از اعضا به‌جز ورودی‌های خود و خروجی، اطلاعات اضافه‌تری ندارند و از ورودی سایر اعضا بی‌اطلاع‌اند. چهار روش برای PPDMبر پایه SMCبه شرح زیر می‌باشد:12][

۱- مجموع امن Secure sum
۲- اتحاد مجموعه امن Secure sum union
۳- اندازه امن تقاطع مجموعه Secure size of intersection
۴- حاصل‌ضرب عددیScalar product

در شکل (۲) دسته‌بندی کلی الگوریتم‌های حفظ حریم خصوصی ارائه‌شده است.

[31]Swapping
[32]Sampling
[33]Encryption technique



شکل (۲) نمودار درختی روش‌های پنهان‌سازی

۳- فاکتورهای ارزیابی

۱- کارایی[34] : الگوریتم‌ها و روش‌های پیشنهادی باید ازنظر زمان اجرا جهت پنهان نمودن قواعد حساس کارا باشند شکست در پنهان‌سازی(Hiding Failure) نداشته باشند.

۲- سودمندی[35] : اطلاعات ازدست‌رفته (میزان قواعد گم‌شده) حداقل باشد.Lost rule, Lost Item

۳- سطح عدم قطعیت: اطلاعات پنهان‌شده، تا حد ممکن قابل پیش‌بینی نباشند.

۴- استحکام: مفید ماندن اطلاعات غیر حساس

۵- اصالت پایگاه داده[36] : پایگاه داده جدید باید آشفتگی کمی داشته باشد و به پایگاه داده قبلی شباهت داشته باشد.

۶- تولید الگوی مصنوعی[37] : میزان قواعد غیرواقعی حداقل باشد[13]Ghost rule.

فرمول‌هایی برای سنجش هرکدام از فاکتورها در[10] آورده شده است.

طبقه‌بندی الگوریتم‌ها بر اساس ویژگی‌های آن‌ها در این قسمت آورده شده است [2]،[10]،[5]. به‌طوری‌که برای هر روش از پنهان‌سازی مثالی در شکل(۳) ذکرشده و نحوه پنهان‌سازی الگوریتم‌ها باهم مقایسه شده است. پیش از آن به معرفی مفاهیم زیر می‌پردازیم:

[34] Performance
[35] Efficiency
[36]Dissimilarity
[37] Artificial Patterns

- Data sharing: ابتدا ایمن‌سازی انجام می‌شود سپس قوانین استخراج می‌شوند. پایگاه داده ایمن‌سازی شده به‌عنوان خروجی می‌باشد.
- Pattern sharing: ابتدا قوانین را استخراج می‌کنند سپس ایمن‌سازی را انجام می‌دهند. پایگاه داده به‌طور مستقیم در اختیار کاربر قرار نمی‌گیرد بلکه فقط قوانینی که ایمن‌سازی روی آن‌ها انجام‌شده در اختیار کاربر قرار می‌گیرد.
- آستانه افشا[38] (φ): محدوده‌ای بین ۰ تا ۱ می‌باشد، اگر ۰ باشد به معنی ایمن‌سازی کل تراکنش‌های دارای قانون حساس است. و اگر ۱ باشد به این معنی است که هیچ‌کدام از تراکنش‌ها ایمن‌سازی نشوند. اگر ۰,۵ باشد نیمی از قوانین حساس را شامل می‌شود و احتمال شکست در پنهان‌سازی وجود دارد. بنابراین عددی که در نظر می‌گیریم روی فاکتورهای ارزیابی چون lost rule, hiding failure,...تأثیرگذار می‌باشد.

۴- خلاصه و نتیجه‌گیری

این مقاله به مقایسه الگوریتم‌های مختلف ارائه‌شده در زمینه پنهان‌سازی قوانین انجمنی پرداخته است. در ارزیابی الگوریتم‌ها اولویت با نداشتن شکست در پنهان‌سازی می‌باشد سپس میزان قواعد ازدست‌رفته مورد ارزیابی قرارمی گیرد و بعدازآن میزان قواعد مصنوعی بررسی می‌شود. در روش‌های اکتشافی معمولاً فاکتور زمان را در نظر نمی‌گیرند روش اکتشافی به‌طور ذاتی در این مورد بهینه عمل می‌کند و به روش مبتنی بر مرز و روش دقیق، ترجیح داده می‌شود. از بین الگوریتم‌های موجود درروش اکتشافی، الگوریتم‌هایی که بر مبنای تغییر داده‌ها، پنهان‌سازی را انجام می‌دهند عملکرد بهتری نشان داده‌اند زیرا در هر دو حوزه داده‌های متمرکز و داده‌های توزیع‌شده قابل‌اجرا می‌باشند.

درنتیجه محققان و پژوهشگران می‌توانند بامطالعه دقیق این مقاله، رویکرد و روش مناسبی جهت پنهان‌سازی انتخاب کنند.
جهت کار در آینده می‌توان در مورد الگوریتم‌های اکتشافی تدبیری اندیشید تا علاوه بر پنهان‌سازی قواعد حساس، به‌طور همزمان میزان قواعد گم‌شده و قواعد جدید کمتری داشته باشد و همچنین اصالت پایگاه داده نیز حفظ شود. با تغییراتی در الگوریتم RRLRو انتخاب بهترین عنصر برای حذف و انتخاب بهترین تراکنش برای درج، می‌توان به اهداف ذکرشده دست‌یافت. اگر هنگام حذف عنصر فاکتور جدیدی دخالت داده شود موجب کاهش قواعد ازدست‌رفته خواهد شد. فاکتورهایی از قبیل درجه برخورد و یا انتخاب قوانین نماینده در هنگام حذف عنصر.

[38]Disclosure threshold

نام الگوریتم	قانون ازدست‌رفته	قانون جدید	شکست پنهان‌سازی	اثر جانبی	اسکن	Mst یا Mct	Sanitization algorithm	Dissimilarity
2a	دارد	دارد	ندارد	مطرح نشده	>۲	Mct	Data sharing	مطرح نشده
2b	دارد	دارد	مطرح نشده	مطرح نشده	مطرح نشده	Mst	Data sharing	مطرح نشده
2c	دارد	٪۰	مطرح نشده	٪۰	مطرح نشده	Mst	Data sharing	مطرح نشده
WSDA	دارد	٪۰	ندارد	مطرح نشده	مطرح نشده	هر دو	Data sharing	مطرح نشده
DSRRC	دارد	٪۲۷/۲۸	ندارد	٪۰	مطرح نشده	مطرح نشده	Data sharing	٪۵/۴
ADSRRC	دارد	٪۰	ندارد	٪۰	مطرح نشده	هر دو	Data sharing	٪۵/۴
RRLR	دارد	٪۰	به‌تازگی کشف‌شده	٪۰	مطرح نشده	هر دو	Data sharing	٪۰
Round Robin	وابسته به φ	٪۰	مطرح نشده	وابسته به φ	مطرح نشده	وابسته به φ	Data sharing	وابسته به φ تا ۶۰
Random	وابسته به φ	٪۰	مطرح نشده	مطرح نشده	۲ مرتبه	وابسته به φ	Data sharing	وابسته به φ تا ۶۰
Min FIA	وابسته به φ	٪۰	٪۳/۲	وابسته به φ	۲ مرتبه	وابسته به φ	Data sharing	وابسته به φ تا ٪۱۶/۴۱
Max FIA	وابسته به φ	٪۰	مطرح نشده	وابسته به φ	۲ مرتبه	وابسته به φ	Data sharing	وابسته به φ تا ٪۶/۳۵
Hidden-First	دارد	٪۰	دارد	مطرح نشده	مطرح نشده	Mst	Data sharing	مطرح نشده
Non-Hidden-First	کمتر از HF	٪۰	دارد	مطرح نشده	مطرح نشده	Mst	Data sharing	مطرح نشده
HPCME	وابسته به φ	٪۰	دارد	وابسته به φ	مطرح نشده	Mst	Data sharing	وابسته به φ
MDSRRC	٪۰	٪۰	مطرح نشده	مطرح نشده	مطرح نشده	هر دو	Data sharing	کمتر از ٪۵
قوانین نماینده	٪۰	دارد	مطرح نشده	٪۰	>۲	Mct	Data sharing	٪۰
Menon	دارد	دارد	ندارد	مطرح نشده	مطرح نشده	Mst	Data sharing	دارد
Bee Colony	دارد	دارد	ندارد	مطرح نشده	مطرح نشده	Mct	Data sharing	دارد
Aggregate	دارد	دارد	ندارد	بسیارکم	مطرح نشده	Mst	مطرح نشده	دارد
Dis Aggregate	دارد	دارد	ندارد	مطرح نشده	مطرح نشده	Mst	مطرح نشده	بسیارکم

جدول (۳) جدول ارزیابی الگوریتم‌های پنهان‌سازی

۵- مراجع

[1] شیخی نژاد زهرا، نادری دهکردی محمد، رستگاری حمید, "مقایسه تحلیلی روش‌های حفظ حریم خصوصی در استخراج قواعد انجمنی," همایش ملی مهندسی برق و کامپیوتر, اردیبهشت ۱۳۹۳.

[2] M. K. J. Han, "Data Mining Concepts and Techniques," *Elsevier Inc, San Francisco, CA*, 2006.

[3] E. B. V. S. Verykios, "State of the Art in Privacy Preserving Data Mining," *SIGMOD Rec*, vol. Vol. 33, No. 1, 2004.

[4] R. S. A. Evfimievski, R. Agrawal, J. Gehrke, "Privacy Preserving Mining of Association Rules," *Proc. Int. Conf. on knowledge discovery and data mining*, 2002.

[5] H. W. S. Wu1, "Research On The Privacy Preserving Algorithm Of Association Rule Mining In Centralized Database," *Proc. of the International Symposiums on Information*, 2008.

[6] A. T. S. Vijayarani, M. Sampoorna, "Analysis of Privacy Preserving K-Anonymity Methods and Techniques," *Proc. of Int. Conf. on Communication and Computational Intelligence*, 2010.

[7] S. R. M. Oliveira, Zaïane, Osmar R., "Protecting Sensitive Knowledge By Data Sanitization," 2005, 2005.

[8] V. S. V. E. Dasseni, A. Elmagarmid, E. Bertino, "Hiding Association Rules by Using Confidence and Support," *Proc. of the 4th Information Hiding Workshop*, 2001.

[9] G. Y, "Reconstruction-based association rule hiding.," *Proceedings of SIGMOD*, pp. 51–56, 2007.

[10] A. G.-D. Vassilios S. Verykios, "A Survey of Association Rule Hiding Methods for Privacy," 2008.

[11] J. C. D. Yitao, J. Zhan, "Efficient Privacy Preserving Association Rule Mining: P4P Style," *In Proc. of Int. Conf. on Computational Intelligence and Data Mining*, pp. 654 – 660, 2007.

[12] A. C. Yao, "How to Generate and Exchange Secrets," *Proc. of the 27th IEEE Symposium on Foundations of Computer Science*, pp. 162-167 1986.

[13] G. L. E. T. Wang "An Efficient Sanitization Algorithm for Balancing Information Privacy and Knowledge Discovery in Association Patterns Mining," *Data & Knowledge Eng*, vol. Vol. 65, No. 3, 2008.