مقایسه کاربردی امنیت داده و حفظ حریم خصوصی در داده کاوی

صمد سهراب
دانشگاه فناوری اطلاعات تعالی قم، دانشجوی کارشناسی ارشد فناوری اطلاعات (امنیت اطلاعات)

چکیده: امروزه استفاده از اینترنت و فناوری اطلاعات، در واقع در کنار مزایایی که دارد، خطر برملا شدن اسرار خصوصی را به همراه دارد. اتصال به اینترنت در سیستم ها و شبکه های کامپیوتری موجب شده است تا متخصصان امنیت، روزبه روز شاهد تهدیدات امنیتی جدیدی باشند. مکانیسم های رمزنگاری و احراز هویت در سیستم های مدیریت پایگاه داده به معنای در امان بودن داده ها از آسیب پذیری های احتمالی نبوده و این مکانیزم ها توانایی مقابله با انواع حملات را ندارند. باید سعی شود که امنیت داده ها و حفظ حریم خصوصی در هنگام داده کاوی، کشف روابط بین داده ها و ذخیره در پایگاه داده ها حفظ شود. هدف از داده کاوی عمومی سازی اطلاعات است، نه اینکه اطلاعات شخصی را در اختیار عموم قرار دهیم. لازم به ذکر است که استفاده از داده کاوی جهت ایجاد مدل و استخراج الگو، امری ضروری است. حفظ حریم خصوصی در داده کاوی یک موضوع مهم در زمینه داده کاوی و امنیت پایگاه داده می باشد. صاحبان داده ها به علت ترس از افشای اطلاعات شخصی و محرمانه خود توسط دیگران، چندان تمایلی جهت انجام داده کاوی روی داده های خود نداشته ولی این مطلب را نیز می دانند که بدون انجام داده کاوی به نتایج و دانش مفید از داده های یکدیگر دسترسی پیدا نمی کنند.
لذا با توجه به اهمیت حفظ حریم خصوصی و امنیت اطلاعات، در این مقاله به بررسی روشهای داده کاوی و مقایسه روشهای امنیت داده ها و حفظ حریم خصوصی در داده کاوی پرداخته می شود.

کلید واژه ها: داده کاوی، حفظ حریم خصوصی، امنیت ، روشهای امنیت داده کاوی

۱- مقدمه

داده ها مادامی که پردازش نشوند و اطلاعات مفید از آن ها جهت تصمیم گیری های راهبردی، مباحث رقابتی و رسیدن به سود بیشتر استخراج نگردد، برای مدیران و برنامه ریزان چندان کارساز نمی باشد. داده کاوی روی داده های مشترک باید طوری انجام شود که حریم شخصی صاحبان داده ها حفظ گردد. بنابراین مشکل اصلی بحث حفظ حریم شخصی در استخراج دانش، داده کاوی نیست، بلکه نحوه انجام داده کاوی است.

ابزار ها و روش های داده کاوی از حوزه های مختلفی همانند آمار، شناسایی الگو، یادگیری ماشین پایگاه داده بدست آمده است. در ادامه مقاله، ضمن تعریف یکسری مفاهیم مرتبط با موضوع مورد مطالعه، به بررسی دو گروه عمده از روش های داده کاوی (توصیفی وپیش بینی) پرداخته و از این میان به بررسی و مقایسه سه روش مختلف داده کاوی (روش خوشه بندی، طبقه بندی و کشف قوانین) با در نظر گرفتن راهکارهای امنیتی جهت بهبود روش و افزایش حریم خصوصی می پردازیم.

مشکل اصلی بحث حفظ حریم شخصی در استخراج دانش، داده کاوی نیست، بلکه نحوه انجام داده کاوی است. روش های موجود در زمینه داده کاوی با حفظ حریم شخصی به دو گروه اصلی تقسیم می شوند که عبارتند از تکنیک های محاسبات چند طرفه امن و تکنیک های تغییر شکل داده. روش های ارائه شده بر پایه محاسبات چند طرفه امن که بر پایه رمزنگاری و روش های ریاضی پیچیده می باشند، نسبت به روش ها و تکنیک های تغییر شکل داده، از دقت و امنیت بالاتری برخوردارند ولی پیچیدگی زمانی و محاسباتی و نرخ ارتباطات نیز در این روش ها بسیار بالاتر از روش های تغییر شکل داده است. همچنین اکثرا کارهای انجام شده در زمینه تکنیک های تغییر شکل داده شده مانند تصادفی سازی، پنهان سازی قوانین، گمنامی سطح K و ... داده های هر فرصت را مستقلانه تغییر شکل می دهند، لذا روابط و وابستگی های آماری بین داده ها به خوبی حفظ نمی شوند و دقت نتایج پایین می آید.

۲- تعاریف

۲-۱ داده کاوی: کشف ساختارهای جالب توجه، غیر منتظره و با ارزش از داخل مجموعه وسیعی از داده ها می باشد و فعالیتی است که اساسا با آمار و تحلیل های دقیق دادم ها منطبق است. داده کاوی یا استخراج و کشف سریع و دقیق اطلاعات با ارزش و پنهان از پایگاه های داده از جمله اموری است که در هر کشور، سازمان ها مظور توسعه علمی، فنی و اقتصادی خود به آن نیاز دارند. روش های داده کاوی شامل دو روش کلی می باشد:

2-1-1 داده کاوی پیش بینی کننده یا یادگیری با نظارت: مدلی از سیستم را ارائه می دهد که شامل بارگیری متغیر ها و صفات در انبار داده ها جهت پیشگوئی مقادیر ناشناخته می باشد. این مدل یک فرآیند دو مرحله ای می باشد. در گام اول ، یک مدل بر اساس مجموعه داده های آموزشی مانند رکوردها و نمونه های موجود در پایگاه داده ها ساخته می شود، مانند پایگاه داده ها شامل اطلاعات کار تهای اعتباری مشتریان. گام بعد که استفاده از مدل نام دارد به بررسی چگونگی استفاده از داده های نظارت شده و بررسی چگونگی استفاده از داده های موقت در یک مجموعه آزمایشی داده ها می پردازد. از تکنیک های داده کاوی پیشگوئی کننده، طبقه بندی، رگرسیون و سری های زمانی می باشند.

2-1-2 داده کاوی توصیف کننده یا یادگیری بدون نظارت: اطلاعات جدید و غیر بدیهی را بر اساس مجموعه داده های موجود ارائه می دهد. در این روش، هدف کلی به دست آوردن یک شناخت از سیستم های تجزیه و تحلیل شده توسط الگوها و روابط بین داده های انبار داده ها است و در این مدل برچسب کلاس هر نمونه آموزشی نا معلوم است که شامل روش های کشف الگوی ترکیبی، کشف قوانین وابستگی، خلاصه سازی و خوشه بندی می باشند. (دهکردی, ۲۰۱۱; کانتاردزیک, ۱۳۹۲)

۲-۲ حریم خصوصی: حفظ حریم خصوصی در داده کاوی، یعنی جلوگیری از ارتباط با هویت افراد و اطاعات افراد.) D.Aruna (1122 Kurmari,

۳- مروری بر انواع روش های داده کاوی پرکاربرد

۳-۱ خوشه بندی داده ها : در این روش هدف دسته بندی مجموعه ای از نقاط داده ای در یک یا چند گروه از اشیاء مشابه می باشد به طوری که اشیاء یک گروه مشابه ولی متفاوت با اشیاء دیگر گروه می باشند.

به طور کلی یک الگوریتم خوشه بندی داده ها باید دارای ویژگی های زیر باشد :

- مقیاس پذیر باشد.

- بتواند با انواع مختلف داده ها کار کند.
- قادر به تعیین خوشه هایی با اندازه و شکل های متفاوت باشد.
- نیازمندی به دانش محیط باری تعیین پارامترهای ورودی حداقل باشد.
- بتواند در محیط نویز دار با وجود outlier ها کار کند و آن ها را تشخیص دهد.
- خروجی آن به ترتیب رکورد های ورودی بستگی نداشته باشد.
- محدودیت های تعیین شده به وسیله کاربر را لحاظ کند.
- نتایج خروجی الگوریتم قابل استفاده و تفسیر باشد.

روش هایی که بر اساس آن ها برای حل مسئله خوشه بندی داده مطرح است، قطعی نیستند و تنها تقریبی از جواب بهینه را بدست می آورند. (آصفی, ۱۳۹۱)

الگوریتم های خوشه بندی را بطور کلی می توان در چهار دسته طبقه بندی نمود :

۳-۱-۱ الگوریتم های بخش بندی: روش های بخش بندی از این مزایا برخوردارند که می توانند با مجموعه وسیعی از داده ها کار کنند. اما مشکلی که الگوریتم های بخش بندی دارند آنست که تعداد خوشه های خروجی چگونه انتخاب شوند. روند تولید خوشه ها در تکنیک های بخش بندی بر این اساس است که یک تابع سنجش از قبل تعریف شده را بهینه کند. که این تابع ممکن است محلی باشد (بر روی زیر مجموعه ای از نمونه ها تعریف می شود) و یا سراسری (بر روی تمام نمونه ها).

۳-۱-۲ الگوریتم های سلسله مراتبی: این الگوریتم ها در دو دسته ادغام کننده و تقسیم کننده طبقه بندی می شوند. روش ادغام کننده (پایین به بالا) با تعداد زیادی خوشه ی داده آغاز می شود و در یک فرآیند تکراری، خوشه هایی که دارای بیشترین شباهت با یکدیگر هستند را ادغام می کند ، در حالی که در روش تقسیم کننده (بالا به پایین) در ابتدا تمامی داده ها در یک خوشه ی داده به دو بخش تقسیم می شود. فرآیند ادغام و یا تقسیم خوشه ها تا زمانی که یک شرط خاص بر آورده نشود، متوقف نمی گردد، که معمولا این شرط، رسیدن به تعداد خوشه های مورد نظر است.

۳-۱-۳ الگوریتم های مبتنی بر تراکم: الگوریتم های مبتنی بر تراکم قادرند تا خوشه هایی از داده ها با اشکال مختلف ایجاد کنند. در این الگوریتم ها در صورتی که تراکم یک ناحیه بالاتر از یک حد آستانه ی معین باشد، آن ناحیه به خوشه نسبت داده می شود و در غیر این صورت به عنوان نویز در نظر گرفته می شود.

۳-۱-٤ الگوریتم های مبتنی بر گرید: یک داده با n ویژگی را می توان به عنوان یک نقطه در فضای n بعدی در نظر گرفت. بنابراین یک مجموعه داده که دارای n ویژگی است، یک زیر فضا را در فضای n بعدی تشکیل می دهد. در الگوریتم های مبتنی برگرید این زیر فضا به تعدادی سلول جدا از هم که در ابتدا با یکدیگر مساوی هستند، تقسیم می شود، این ساختار داده گرید نامیده می شود. سلول های گرید به صورت پویا در حین اجرای الگوریتم خوشه بندی تغییر می کنند و سرانجام خوشه های داده از ادغام سلول های متراکم مجاور هم تشکیل می شود. (کانتاردزیک, ۱۳۹۲)

یک روش ساده برای داده کاوی از چندین منبع که داده ها را به اشتراک عموم نمی گذارند اینست که ابزارهای مربوط به داده کاوی را در هر سایت به صورت مستقل اجرا کرده و نتایج را یکجا جمع آوری کنیم.

۳-۲ طبقه بندی داده ها: طبقه بندی عبارتست از کاوش الگوهایی که می توانند داده های آتی را در طبقه های شناخته شده و معین رده بندی کند و برخلاف خوشه بندی داده ها در کلاس ها و طبقه های از پیش تعیین شده ای قرار می گیرند. این روش از تعمیم دنیای واقعی و قابلیت تطبیق داده های جدید با یک قالب کلی، استفاده میشود در این مورد می توان با تحلیل یک پایگاه داده ای موجود، خصوصیات مجموعه های داده را تعیین کرد.(حقیقی, ۱۳۸۵)

فرآیند طبقه بندی در واقع نوعی یادگیری نظارت شده می باشد که در طی دو مرحله آموزش و آزمایش انجام می گردد. در مرحله اول مجموعه ای از داده ها که در آن هر داده شامل تعدادی خصوصیت دارای مقدار و یک خصوصیت بنام خصوصیت کلاس می باشد، برای ایجاد یک مدل داده به کار می روند که این مدل داده در واقع توصیف کننده مفهوم و خصوصیات مجموعه داده- هایی است که این مدل از روی آن ها ایجاد شده است. مرحله دوم فرآیند طبقه بندی اعمال یا بکارگیری مدل داده ایجاد شده بر روی داده هایی است که شامل تمام خصوصیات داده هایی که برای ایجاد مدل داده بکار گرفته شده اند، می باشد. به جز خصوصیت کلاس این مقادیر که هدف از عمل طبقه بندی نیز تخمین مقدار این خصوصیت می باشد. از کاربرد های عمده این مدل می توان به مدیریت مشتریان ، تصویب اعتبار ، بازاریابی مستقیم در خرده فروشی اشاره کرد. از مدل های معروف طبقه بندی داده ها عبارتند از:

۳-۲-۱ درخت تصمیم گیری: درخت تصمیم گیری یکی از ابزارهای قوی و معمول برای طبقه بندی و پیش بینی می باشد که فرآیند تصمیم گیری برای تعیین طبقه یک نمونه ورودی را نشان می دهد. این روش یادگیری نظارت شده را ارائه می دهد. در واقع در هر درخت تصمیم گیری مثال آموزشی به این صورت طبقه بندی می شود که از ریشه درخت شروع می شود. آن گاه صفت و صفت مشخص شده توسط این گروه بررسی می گردد و سپس منطبق با ارزش صفت در طول شاخه ها حرکت رو به پائین انجام می دهد و این فرآیند برای گره های زیر درختان گره جدید تکرار می شود.

۳-۲-۲ روش های Bayesian Classifier: این روش یکی از روش های ساده یادگیری نظارت شده است که در آن فرض می شود که تمام متغیرهای ورودی به یک اندازه مهم و مستقل از هم می باشند و اگر یکی از شرایط هم برقرار نباشد، این روش در شرایطی کاربرد دارد و از طرفی دیگر اگر استقلال نظر نقض گردد، دقت از دست می رود. در این روش طبقه بندی ، ساخت و پیاده سازی بسیار ساده دارد و نیازی به برنامه های تخمین پارامتر ندارد. این روش معمولا فوق العاده عمل می کند و بسیار کارا می باشد و در بسیاری در کاربردهایی نظیر طبقه بندی متن و تشخیص پزشکی این روش کارایی قابل مقایسه ای با شبکه های عصبی و درخت تصمیم دارد.(دهکردی, ۲۰۱۱)

۳-۲-۳ طبقه بندی مبتنی بر قوانین: در این روش مدل ایجاد شده از روی داده ها به صورت مجموعه ای از قوانین می باشد. می توان گفت که هر قانون به صورت یک قانون اگر p آنگاه q می باشد که در آن p مجموعه از شرایط بوده و q نیز مشخص کننده برچسب یک طبقه خاص می باشد. اولویت دهی به قوانین به روش های مختلفی ممکن است انجام گردد. برای مثال ممکن است که ابتدا کلاسها اولویت دهی شوند و قوانین مربوط به هر کلاس نیز با تاثیر پذیری از این اولویت – دهی ، اولویت بگیرند. اولویت کلاسها نیز ممکن است بر اساس اهمیت کلاس یا تعداد داده های متعلق به آن کلاس مشخص گردد.(Stanley R. M. Oliveira, Osmar R. Za¨ıane, & Saygin, 1112؛بلر ای‌می, 1122)

بر خلاف یک مدل متمرکز ، مدل داده کاوی توزیع شده این طور فرض می کند که منابع اطلاعاتی در بین چندین سایت پخش شده اند. الگوریتم های تولید شده در این زمینه مشکل دست یابی به نتایج موثر را از همه اطلاعات موجود در چندین منبع مختلف حل کرده است. یک روش ساده برای داده کاوی از چندین منبع که داده ها را به اشتراک عموم نمی گذارند این است که ابزارهای مربوط به داده کاوی را در هر سایت به صورت مستقل اجرا کرده و نتایج را یک جا جمع آوری کنیم. به هر حال این روش اغلب در رسیدن به نتایج معتبر جهانی ناموفق عمل می کند.

۳-۲ قوانین وابستگی: یکی از روش های داده کاوی کشف قوانین وابستگی می باشد که عبارتست از کشف قوانینی به فرم X→Y با ضرایب پشتیبانی و اطمینان بیشتر از حد آستانه در تراکنش های موجود پایگاه داده. به هنگام داده کاوی در چند پایگاه ، داده ها در یک مکان مرکزی جمع آوری شده و الگوریتم هایی برای استخراج داده ها در آن به کار برده می شوند. (C.C.Aggarwal, 1112)

استحراج قوانین وابستگی از پایگاه های اطلاعاتی بزرگ در دو مرحله صورت می گیرد :

- شناسایی و تولید تمام مجموعه های (متوالی) یعنی شناسایی مجموعه ها با تکرار بیش از Q
- ایجاد قوانین وابستگی که با حداقل ضریب پشتیان و ضریب اطمینان متناسب باشد.

اگر مجموعه داده ای افراد دارای رکوردهایی با شناسه های یکسانی می باشند ولی هر مجموعه داده ای دارای صفت ها (ستون های) مربوط به خود می باشد. که این صفت ها در مجموعه های داده ای دیگر وجود ندارد به این مسئله، مسئله مشارکت ناهمگن می گویند. حال صاحبان (نگه دارندگان) این مجموعه داده های داده ای برای رسیدن به اهداف بالاتر در شرایطی که حریم شخصی و محرمانگی داده هایشان حفظ شود می خواهند با یکدیگر همکاری و مشارکت کنند و قوانین موجود بین اقلام (صفت ها) را بین این مجموعه ها داده ای استخراج کنند. به این مسئله، کشف قوانین انجمنی (وابستگی) با حفظ حریم شخصی بر روی پایگاه های داده پخش بندی شده عمودی گفته می شود.(G.Moro, 1112)

٤- روشهای ارائه شده برای حفظ حریم خصوصی

ارائه دهنده روش	توضیح در مورد روش
اولیورا و زایان (Stanley R. M. Oliveira, 1112)	روش های ارائه مبتنی بر شباهت شی ء و ارائه مبتنی بر کاهش بعد پذیری را برای خوشه بندی عمودی و متمرکز را پیشنهاد کرده اند. در هردوی این روش ها سعی شده است تا جنبه های عمده داده های خصوصی برای خوشه بندی حفظ شده و در عین حال ، داده ها به صورتی تغییر کنند که نیاز های اصلی در مورد حریم شخصی ، برآورده شود. البته این روش ها برای پایگاه های داده بخش بندی شده به صورت افقی قابل کاربرد نیستند. همچنین در این روش ممکن است که دقت داده کاوی کاهش یابد. زیرا بعد پذیری در داده های اصلی از بین می رود.
کلیفتون و وایدیا	محاسبات چند طرفه ایمن را بر روی الگوریتم ها K-means روی داده های بخش بندی شده افقی پیشنهاد می کنند. بیشتر کارهای اخیری که در بدان اشاره شده است و توسط کروگر و همکارانش انجام شده است، یک پروتکل K-means توزیعی با حفظ حریم شخصی بر روی داده های بخش بندی شده به صورت افقی را نشان می دهند. با این حال در این که پروتکل ها از رمزنگاری متقارن و ارزیابی توابع چند جمله ای به عنوان بلوکه ای سازنده مدل استفاده شده است که کارایی آن ها در زمانی مشخص می شود که از آن ها برای پایگاه های داده بزرگ استفاده می شود.
آتالا و همکارانش	یک پرووتکل محاسباتی ایمن را پیشنهاد کرده اند. این پروتکل از پیاده سازی برنامه نویسی پویا در میان چند نفر با داده های ورودی محرمانه استفاده می کند. به این طریق هر مسئله فرعی از طریق یک سری از رمزنگاری های متقارن انجام شده و پروتکل های جستجوگر کمتری نیز اجرا می شود. راه حل این مسائل فرعی بین افراد توزیع می شود تا زمانی که آخرین راه حل بتواند با رمزنگاری متقارن هزینه ارتباطاتی و محاسباتی بالایی را دارد. به عنوان یک پیشنهاد در این نوع مسائل استفاده از شخص ثالث نیمه مطمئن می باشد. مثلا در این پروتکل می توان برای مقایسه صفات عددی الفبایی از یک شخص ثالث نیمه صادق استفاده کرد که این کار منجر به کاهش هزینه ها می گردد.

جدول۱- روش های ارائه شده برای حفظ محرمانگی در خوشه بندی داده ها

ارائه دهنده روش	توضیح در مورد روش
آگراوال و همکارانش (C.C.Aggarwal, 1112)	یک تکنیک انحراف مقدار را برای حفظ حریم شخصی بوسیله نویز افزایشی تصادفی به واسطه توزیع گواسیان برای داده واقعی ارائه داده اند. آن ها نشان داده اند که این تکنیک در پنهان کردن داده در هنگام استخراج الگوهای اصلی مانند توزیع داده اصلی و مدل های درخت تصمیم دقت بالایی دارد. این تکنیک بعدا با ارائه الگوریتمی بر پایه به حداکثر رساندن امید ریاضی برای بازسازی بهتر توزیع گسترش یافت و هم چنین یک معیار علمی اطلاعات برای تضمین و اندازه گیری حریم شخصی نیز در آن مورد بحث قرار گرفته است. یکی از این طرح ها در روش تصادفی سازی داده ها تکنیک پاسخ تصادفی است که در آن قبل از فرستادن یک رکورد به فرد دیگر (یا به سرور)، یکی کاربر به صورت مستقل برای هر صفت، شیر یا خط می کند و تصمیم می گیرد که آیا در مورد همان حقیقت بر اساس نتایج شیر و خط کردن حقیقت را بگوید یا دروغ را .

Reiss و Dalenius	ایده اصلی از تعویض و جابجایی داده ها را مطرح کردند. در این طرح پایگاه داده ها به وسیله سوئیچ کردن یک زیر مجموعه از صفات (ستونهای جدول) بین جفت سطر های انتجاب شده، تغییر شکل داده می شود و باعث کاهش دفعات تکرار آیتمهای مرتب شده می شود وبا این کار قابلیت اعتماد و اطمینان داده ها به خطر نمی افتد

جدول۲- روش های ارائه شده برای حفظ محرمانگی در طبقه بندی داده ها

ارائه دهنده روش	توضیح در مورد روش
Lakshmi N. V. Muthu و همکارانش (SandhyaRani, 2012)	روش برای حفظ حریم شخصی در کشف قوانین وابستگی ارائه کرده اند که در الگوریتم خود از تکنیک های تغییر شکل داده و رمزنگاری استفاده کرده اند
Lambodar Jen. و همکارانش	به مطالعه تکنیک های تغییر شکل داده ای نظیر پنهان سازی قوانین، گمنامی سطح k و تکنیک هایی نظیر تعمیم داده و غیر پرداخته اند.
D.Aruna Kumari و همکارانش (D.Aruna Kurmari, ۲۰۱۱)	به بررسی تکنیک های رمزنگاری، تعمیم و انحراف داده ها در خوشه بندی داده ها و ارائه روشی کارا پرداخته اند.
Md. Golan Kaosar و همکارانش	ارائه راهکارهایی جهت کشف قوانین وابستگی در حالت دو نفره با استفاده از روش های رمزنگاری متقارن پرداخته اند.
Mahmoud Hussein	به بررسی تکنیک رمزنگاری متقارن در کشف قوانین وابستگی برای حالت توزیع شده پرداختند.
Rozenberg و همکاراانش (Ehud.Gudes, ۲۰۰۵)	یک الگوریتم برای حالت چند نفره ارئه شده اند. سپس به بررسی حریم شخصی پرداخته اند و روشی نیز در ادامه با توجه به یک معیار جدید و نسبی برای افزایش حریم شخصی ارائه داده اند. آن ها از شخص ثالث نیمه مطمئن برای کشف قوانین وابستگی با حفظ حریم شخصی بر روی پایگاه های داده افزار بندی شده عمودی استفاده کرده اند. در الگوریتم های آن ها یک شرکت کننده به عنوان سرویس دهنده و سایر شرکت کنندگان به عنوان سرویس گیرنده مشخص شده اند.

جدول۳- روش های ارائه شده برای حفظ محرمانگی در کشف قوانین وابستگی

۵- مقایسه کلی روش های ارائه شده

نام روش	هزینه استفاده	ویژگی ها	نقاط ضعف
گم نام سازی	یافتن راه حل کامل جزء NP-Hard می باشد اما استفاده از روش های هیورستیک هزینه را کاهش می دهد	عمومی سازی و حفظ محرمانگی، گسترش شیوه های جدید برای رفع نقص ها، کارایی برای داده های متنی و حرفی	درصورت موجود بودن دیدهای زیاد آسیب پذیر می باشد، کیفیت داده ها تنزل پیدا می کند
محاسبات امن	محاسبات زیاد، کاهش سربار در صورت ترکیب با سایر روش ها	امنیت، مستقل بودن پایگاه ها برای عملیات محلی از قبیل تصادفی سازی	پیچیدگی محاسبات، عدم حفاظت از نتایج میانی در برخورد موارد

جدول ۴- مقایسه روش های حفظ محرمانگی

٦- بحث و نتیجه گیری:

بسیاری از الگوریتم های موجود در روش محاسبات چند طرفه امن مانند تکنیک های رمزنگاری و اشتراک سری، سنکرون می باشند و نیاز به ارتباطات زیاد و چند مرحله ای بین شرکت کنندگان دارند و همچنین بسیاری مقیاس پذیر نبوده و افزایش تعداد شرکت کنندگان و به دنبال آن افزایش پایگاه های داده تاثیر سوء بر روند این الگوریتم ها دارد. با توجه به آنکه روش های مبتنی بر آشفته سازی اطلاعات نسبت به روش های مبتنی بر محاسبات امن، از امنیت و دقت پایینتر و در عوض از مرتبه زمانی و سربار ارتباطی کمتری نیز برخوردارند، باید سعی بر آن باشد الگوریتم هایی ارائه شوند که در عین دقت بالای نتایج، حریم شخصی را به خوبی حفظ کنند و میزان ارتباطات و سربار ارتباطی آن ها نیز پایین باشد و در حقیقت یک تعامل کارساز بین این فاکتور ها برقرار شود.

از آنجاییکه تا کنون هیچ معیار و استانداردی برای میزان حریم شخصی ارائه نشده است بهترین فاکتور و معیار بر اساس نظر صاحبان داده ها و بر اساس موضوع مشارکت می باشد. استفاده از تکنیک های یاد شده در این مقاله، و با به کار گیری رویکرد های ترکیبی، بهبود قابل توجهی در حفظ حریم خصوصی خواهیم داشت. چرا که با توسعه اینترنت و افزایش سرعت سیستم های کامپیوتری و افزایش حجم داده ها در پایگاه های داده، تامین امنیت اطلاعات امری مهم است.

٧- منابع

١. حقیقی, ع. (١٣٨٥). داده کاوی و کاربرد آن در کیفیت داده ها

٢. کانتاردزیک, م. (١٣٩٢). داده کاوی (Data Mining) (م. ا. علیخانی, Trans.): فرنگار رنگ.

٣. آصفی, م. م. و. م. آ. و. ه. (١٣٩١). بررسی روش های خوشه بندی داده ها با رعایت حریم خصوصی.

۴. دهکردی, م. ن. د. م. آ. (٢٠١١). مقایسه و تحلیل برترین روشهای حفاظت از حریم خصوصی در داده کاوی توزیع شده.

۵. ابراهیمی, ا. ق. ش. ل. م. پ. ر. (٢٠١٣). بررسی و مقایسه حفظ حریم خصوصی در استخراج قوانین انجمنی در داده کاوی

6. Stanley R. M. Oliveira, Osmar R. Za¨ıane, & Saygin, a. Y. u. (1112). Secure Association Rule Sharing
7. C.C.Aggarwal, D. A. a. (1112). On the design and quantification of privacy preserving data mining algoritm. from http://portal.acm.org/citation.cfm?id=2025?1
8. D.Aruna Kurmari, D. K. R. R., M.Suman. (1122). Privacy Preserving Clustering in DDM using Cryptography. *Research Journal of Computer System Engineering-RJCSE,*)12(11
9. Ehud.Gudes, B. R. a. (1112). association rules mining in vertically partitioned database.
10. G.Moro, M. K. a. S. L. a. (1112). Distributed clustering based on sampling local density estimates. *Proceedings of the* 21th *International Joint Conference on Artificial Intelligence*.
11. SandhyaRani, M. V. M. a. K. (1121). Privacy preserving association rule mining without trusted party for horizontaliy partitioned databasea. *International Jurnal of Data Mining & Knowledge Manamement Process(IJDKP),*