فصلنامه رویکردهای نوین در تحقیقات علوم پایه، فنی و مهندسی
سال پنجم، شماره ۱۸، پاییز ۱۴۰۱
ISSN: ۲۶۴۵-۵۴۳۹
www.jonarbset.ir

تجزیه و تحلیل روشها و برنامه های کاربردی داده کاوی

سید علی محمدیه[۱] ، محمدکاظم بشکنی[۲]

[۱] گروه ریاضی محض،دانشکده ریاضیات،دانشگاه کاشان
[۲] دانشجوی کارشناسی ارشد مدیریت بازرگانی-آموزش عالی حکیم نظامی قوچان (نویسنده مسئول)

چکیده

داده کاوی به مفهوم استخراج اطلاعات نهان یا الگوها و روابط مشخص در حجم زیادی از داده‌ها در یک یا چند بانک اطلاعاتی بزرگ گفته می‌شود.در عصر فناوری اطلاعات، اطلاعات نقش حیاتی در هر حوزه زندگی انسانی دارند .جمع آوری داده از منابع مختلف داده، ذخیره و نگهداری داده ، تولید اطلاعات، تولید دانش و انتشار داده، اطلاعات و دانش به هر ذینفع بسیار مهم است. با توسعه سریع فن آوری اطلاعات، مجموعه پایگاه داده های بزرگ در شبکه های داده ای جهانی درحال رشد هستند جهت استفاده از حجم زیاد داده ها در پایگاه داده های بزرگ نیاز به ابزاری هات تجزیه و تحلیل و تفسیر داده ها وجود دارد. داده کاوی مجموعه ای از فعالیت ها برای پیدا کردن الگوهای هدید، مخفی و یا غیر مدتظره از داده ها و یا الگوهای غیر معمول است. در عصر فناوری اطلاعات،اطلاعات نقش حیاتی در هر حوزه زندگی انسانی دارد. جمع آوری داده از منابع مختلف داده، ذخیره و نگهداری داده، تولید اطلاعات، تولید دانش و انتشار داده، اطلاعات و دانش به هز ذینفع بسیار مهم است.با گسترش سیستمهای بایگانی و حجم بالای داده های ذخیره شده در این سستم ها به ابزاری نیاز است تا بتوان ای داده ها را پردازش کرد و اطلاعات حاصل از آن را در اختیار کاربران قرار داد.در این مقاله خلاصه ای سیستمهای داده کاوی و برخی از برنامه های کاربردی آن ارائه شده است.

واژه‌های کلیدی: داده کاوی، برنامه های کاربردی داده کاوی، روشهای داده کاوی.

۱. مقدمه

برای تولید اطلاعات به مجموعه گسترده ای از داده نیاز است.داده میتواند ارقام عددی ساده و اسناد متنی در اطلاعات پیچیده تر مانند داده فضایی،داده چند رسانه ای و اسناد فرامتن باشد.برای استفاده کامل از داده ، بازیابی داده واقعا کافی نیست لازم است ابزاری برای خلاصه سازی خودکار داده ها،استخراج ماهیت اطلاعات ذخیره شده و کشف الگوها در داده های خام باشد.با توجه به مقدار زیاد داده ذخیره شده در فایل ها،پایگاه داده ها و مخازن دیگر،خیلی مهم است که ابزار قدرتمندی برای تجزیه و تحلیل و تفسیر چنین داده هایی و استخراج دانش جالب توجه که میتواند در تصمیم گیری کمک کند توسعه یابد.تنها پاسخ به تمام موارد فوق"داده کاوی" است. داده کاوی استخراج اطلاعات پنهان قابل پیش بینی از پایگاه داده های بزرگ است. یک فناوری قدرتمند با توانایی بالایی برای کمک به سازمانها متمرکز بر مهمترین اطلاعات موجود در انبارهای داده است.ابزارهای داده کاوی روندها و رفتارهای آینده را پیش بینی می کنند، به سازمانها برای تصمیم گیری های دانش محور فعال کمک میکنند.تجزیه و تحلیل های خودکار و موثر در آینده توسط داده کاوی فرای تجزیه و تحلیل رویدادهای گذشته فراهم شده توسط ابزارهای معمولی گذشته نگر سیستم های پشتیبانی تصمیم گیری ارائه شد.ابزارهای داده کاوی می توانند به سوالاتی پاسخ دهند که حل و فصلشان بطور سنتی بسیار وقت گیر است.پایگاه داده های را برای یافتن الگوهای پنهان آماده می کنند،اطلاعات قابل پیش بینی را که ممکن است کارشناسان از دست بدهند،فرای انتظاراتشان می یابد. داده کاوی،با عنوان کشف دانش در پایگاه داده(KDD) شناخته شده است،استخراج غیر مستقیم از اطلاعات ضمنی،قبلا مجهول و اطلاعات بالقوه مفید از داده ها در پایگاه داده ها است.اگرچه داده کاوی و کشف دانش در پایگاه داده(KDD) اغلب به عنوان مترادف هستند،داده کاوی در واقع بخشی از فرایند کشف دانش است.

۲. وظایف داده کاوی

وظایف داده کاوی وابسته به کاربرد نتیجه داده کاوی انواع مختلف هستند، وظایف داده کاوی به صورت زیر طبقه بندی می شوند:

تجزیه و تحلیل اکتشافی داده: به سادگی داده ها را بدون هیچ آگاهی روشنی از آنچه ما دنبال آن هستیم کاوش می کند. این تکنیک ها تعاملی و بصری هستند.

مدل سازی توصیفی: همه داده ها را توصیف می کند، شامل مدل هایی برای توزیع احتمال داده ها، تقسیم فضای p بُعدی به گروه ها و مدل های توصیف روابط بین متغیرها می باشد.

مدل سازی پیش بینی: این مدل به مقدار یک متغیر اجازه می دهد تا از مقادیر شناخته شده متغیرهای دیگر پیش بینی شود.

کشف الگوها و قوانین: با تشخیص الگو مرتبط است هدف، کشف رفتار جعلی با شناسایی مناطق فضایی که انواع مختلف معاملات را تعیین می کند که نقاط دادهها به طور قابل توجهی متفاوت از سایر نقاط است.

بازیابی محتوا: الگوی مشابه به الگوی مورد علاقه در مجموعه داده را می یابد. این کار معمولا برای مجموعه داده های متنی و تصویری مورد استفاده قرار می گیرد[۱].

۳.انواع سیستم های داده کاوی

سیستم های داده کاوی را می توان براساس معیارهای مختلف طبقه بندی کرد به شرح زیر:

طبقه بندی سیستم های داده کاوی با توجه به نوع منبع داده استخراجی: این طبقه بندی بر اساس نوع داده هایی که بکار می برند مانند داده فضایی، داده چند رسانه ای، داده سری های زمانی، داده متنی، وب جهان گستر و غیره است.

طبقه بندی سیستم های داده کاوی با توجه به مدل داده ها: این طبقه بندی براساس مدل داده ای شامل پایگاه داده رابطه ای، پایگاه داده شیء گرا، انبار داده، پایگاه داده تراکنشی و غیره است.

طبقه بندی سیستم های داده کاوی با توجه به نوع دانش کشف شده: این طبقه بندی بر اساس نوع دانش کشف شده یا ویژگی های داده کاوی، مانند توصیف خصوصیت، فرق، ارتباط، طبقه بندی، خوشه بندی، و غیره است .

برخی از سیستم ها تمایل به سیستم های جامع ارائه دهنده چندین ویژگی داده کاوی با هم دارند.

طبقه بندی سیستم های داده کاوی با توجه به تکنیک های کاوش بکار رفته : این طبقه بندی بر اساس رویکرد تجزیه و تحلیل داده های بکار رفته در یادگیری ماشینی، شبکه های عصبی، الگوریتم های ژنتیک، آمار، تجسم، پایگاه داده گرا یا انبار داده گرا و غیره است.

طبقه بندی همچنین می تواند به میزان تعامل کاربر در فرآیند داده کاوی نظیر سیستم های مبتنی بر پرس و جو)کوئری(، سیستم های اکتشافی تعاملی یا سیستم های مستقل اشاره کند. یک سیستم جامع، طیف گسترده ای از تکنیک های داده کاوی را در اختیار موقعیت های مختلف و گزینه های مختلف قرار می دهد و درجه های مختلف تعامل کاربر را ارائه می دهد[۲].

۴.چرخه حیات داده کاوی

چرخه حیات پروژه داده کاوی شامل شش مرحله است. توالی مرحله ها سخت نیست. حرکت به جلو و عقب بین مراحل مختلف همیشه ضروری است. به نتیجه هر مرحله بستگی دارد. مرحله های اصلی عبارتند از:

۱,۴.درک کسب و کار

این مرحله تمرکز بر درک اهداف و الزامات پروژه از دیدگاه تجاری است، سپس این دانش را به یک تعریف مسئله داده کاوی و یک برنامه اولیه طراحی شده برای رسیدن به اهداف تبدیل می کند.

۲,٤. درک داده

با جمع آوری داده های اولیه شروع می شود برای آشنا شدن با داده ها، شناسایی مسائل مربوط به کیفیت داده، کشف بینش های اولیه در مورد داده ها یا تشخیص زیر مجموعه های جالب به منظور ایجاد فرضیه ها برای اطلاعات نهان.

۳,۴.آماده سازی داده ها

همه فعالیت ها را برای ساختن مجموعه داده نهایی از داده های خام اولیه پوشش می دهد.

۴،۴.مدلسازی

در این مرحله، تکنیک های مختلف مدل سازی انتخاب و مورد استفاده قرار می گیرند و پارامترهای آنها برای مقادیر بهینه کالیبره می شوند.

۵،۴. ارزیابی

در این مرحله مدل به طور کامل ارزیابی و بررسی می شود. مراحل انجام شده برای ساختن مدل برای اطمینان از اینکه به درستی به اهداف کسب و کار برسد. در پایان این مرحله، باید به تصمیمی برای استفاده از نتایج داده کاوی رسید.

۶،۴.موضع گیری

هدف از مدل این است که دانش داده را افزایش دهد؛ دانش به دست آمده باید سازماندهی و ارائه شود تا مشتری بتواند از آن استفاده کند. مرحله استقرار می تواند به اندازه ی تولید یک گزارش ساده باشد یا به پیچیدگی پیاده سازی یک فرآیند داده کاوی قابل تکرار در سراسر شرکت باشد[۳].

۵.مدلهای داده کاوی

مدل های داده کاوی دو نوع هستند: پیش بینی و توصیفی است.

مدل پیش بینی با استفاده از مقادیر داده شناخته شده، داده های مجهول را پیش بینی می کند. برای مثال طبقه بندی، رگرسیون، تجزیه و تحلیل سری های زمانی، پیش بینی و غیره.

مدل توصیفی، الگوها یا روابط را در داده ها شناسایی می کند و خواص داده های آزمایش شده را بررسی می کند. برای مثال خوشه بندی، خلاصه سازی، قانون وابستگی ، کشف توالی و غیره.

بسیاری از برنامه های کاربردی داده کاوی با هدف پیش بینی وضعیت آینده داده ها هستند. پیش بینی فرایند تجزیه و تحلیل وضعیت های فعلی و گذشته ویژگی و پیش بینی حالت آینده اش است. طبقه بندی یک روش نگاشت داده های هدف به گروه ها یا کلاس های از پیش تعریف شده است یک یادگیری نظارتی است زیرا کلاس ها قبل از بررسی داده هدف، از پیش تعریف شده اند. رگرسیون شامل یادگیری تابعی است که آیتم داده را به متغیر پیش بینی سنجیده شده واقعی نگاشت می کند. در تجزیه و تحلیل سری های زمانی ارزش یک ویژگی بررسی می شود زیرا در طول زمان تغییر می کند. در تجزیه و تحلیل سری های زمانی از روش اندازه گیری فاصله برای تعیین شباهت بین سری های مختلف زمانی استفاده می شود، ساختار خطی برای تعیین رفتار آن مورد بررسی قرار می گیرد و از جدول سری های زمانی تاریخی برای پیش بینی مقادیر آینده متغیر استفاده می شود.

خوشه بندی شبیه به طبقه بندی است به جز اینکه گروه ها از پیش تعریف نشده اند بلکه تنها توسط داده ها تعریف می شوند. همچنین به عنوان یادگیری یا بخش بندی بدون نظارت اشاره می شوند. تقسیم بندی یا بخش بندی داده ها به گروه ها یا خوشه ها است. خوشه ها با مطالعه رفتار داده ها توسط متخصصین حوزه تعریف می شوند. اصطلاح بخش بندی در زمینه بسیار خاص استفاده می شود؛ این یک فرایند تقسیم بندی پایگاه داده به گروه بندی پیوسته چندتایی های مشابه است.

خلاصه سازی روش ارائه خلاصه اطلاعات از داده ها است. قانون وابستگی پیوند بین صفات مختلف را پیدا می کند. کاوش قانون وابستگی یک فرایند دو مرحله ای است: یافتن تمامی مجموعه های مکرر آیتم، ایجاد قوانین وابستگی قوی از مجموعه های مکرر آیتم . کشف توالی فرایند یافتن الگوهای توالی در داده ها است. این توالی می تواند برای درک روند استفاده شود[۴].

۶.فرایند کشف دانش

داده کاوی یکی از وظایف در فرآیند کشف دانش از پایگاه داده است. مراحل در فرآیند KDD شامل:

تمیز کردن داده ها: همچنین به عنوان پاکسازی داده ها شناخته شده است. در این مرحله داده نویزی و داده نامناسب از مجموعه حذف می شوند.

ادغام داده ها: در این مرحله، منابع متعدد داده، اغلب ناهمگن، در یک منبع مشترک ترکیب می شوند.

انتخاب داده ها: داده مرتبط با تجزیه و تحلیل تصمیم گرفته می شود و از مجموعه داده ها بازیابی می شود.

تبدیل داده ها: همچنین به عنوان تثبیت داده شناخته می شود؛ در این مرحله داده های انتخاب شده به اشکال مناسب برای روش استخراج تبدیل می شوند.

داده کاوی: این مرحله مهمی است که در آن تکنیک های هوشمندانه برای استخراج الگوهای بالقوه مفید استفاده می شود.

ارزیابی الگو: در این مرحله، الگوهای جالب ارائه دانش بر اساس اندازه گیری های معین شناسایی می شوند.

ارائه دانش: مرحله نهایی است که در آن دانش کشف شده به صورت بصری به کاربر ارائه می شود. این مرحله ضروری از تکنیک های تجسم برای کمک به کاربران در درک و تفسیر نتایج داده کاوی استفاده می کند.

۷.روشهای داده کاوی

روش های داده کاوی به طور گسترده ای دسته بندی می شوند: پردازش تحلیلی(On-Line) روی خط (OLAP)، طبقه بندی، خوشه بندی، کاوش قانون وابستگی، داده کاوی موقتی، تحلیل سری های زمانی، کاوش فضایی، کاوش وب و غیره. این روش ها از الگوریتم ها و داده های مختلف استفاده می کنند . منبع داده می تواند انبار داده، پایگاه داده، فایل مسطح یا فایل متنی باشد. الگوریتم ها ممکن است الگوریتم های آماری، مبتنی بر درخت تصمیم، نزدیک ترین همسایه، مبتنی بر شبکه عصبی ، مبتنی بر الگوریتم های ژنتیکی ، مبتنی بر قوانین، ماشین بردار پشتیبانی و غیره. انتخاب الگوریتم داده کاوی عمدتاً به نوع داده های مورد استفاده برای کاوش و نتیجه مورد انتظار فرایند کاوش بستگی دارد . کارشناسان حوزه نقش مهمی در انتخاب الگوریتم برای داده کاوی بازی می کنند[۵].

یک فرآیند کشف دانش (KD) شامل داده های پیش پردازش ، انتخاب یک الگوریتم داده کاوی و پس پردازش نتایج کاوش است. گزینه های بسیار زیادی برای هر یک از این مراحل وجود دارد و تعاملات بی قید و شرط بین آنها وجود دارد.
بنابراین هر دو، تازه کارها و متخصصان داده کاوی نیاز به کمک در فرآیندهای کشف دانش دارند.

دستیارهای کشف هوشمند (IDA) ، به کاربران در استفاده از پروسه های کشف دانش معتبر کمک می کنند. IDA می تواند سه مزیت را برای کاربران فراهم کند: شمارش سیستماتیک فرآیندهای کشف دانش معتبر، رتبه بندی موثر فرایندهای معتبر با معیارهای مختلف که به انتخاب میان گزینه ها کمک می کند، زیرساخت برای به اشتراک گذاشتن دانش، که منجر به بیگانگی شبکه می شود.

چندین تلاش دیگر برای خودکار کردن این فرآیند و طراحی یک ابزار داده کاوی تعمیم یافته، انجام شده است که برای انتخاب داده ها و الگوریتم های داده کاوی و تا حدودی کشف دانش دارای هوش است.

۸.برنامه های کاربردی داده کاوی

برنامه های کاربردی داده کاوی می تواند عمومی یا دامنه خاص باشند. برنامه های عمومی نیازمند یک سیستم هوشمند است که به تنهایی می تواند تصمیمات خاصی مانند انتخاب داده ها، انتخاب روش داده کاوی، ارائه و تفسیر نتیجه را بگیرند. برخی از برنامه های کاربردی عمومی داده کاوی نمی توانند این تصمیمات را به خود اختصاص دهند اما کاربران را برای انتخاب داده ها، انتخاب روش داده کاوی و برای تفسیر نتایج راهنمایی می کنند. برنامه کاربردی داده کاوی مبتنی بر چند عامل دارای قابلیت انتخاب اتوماتیک تکنیک داده کاوی است تا انجام شود. سیستم چند عاملی در سطوح مختلف استفاده شده : اول، در سطح مفهوم تعریف سلسله مراتبی و سپس در سطح نتیجه تا بهترین تصمیم سازگار برای کاربر را ارائه دهد. این تصمیم در پایگاه دانش برای استفاده در تصمیم گیری بعدی ذخیره می شود. ابزار سیستم چند عاملی مورد استفاده برای توسعه کلی سیستم داده کاوی عمومی از عوامل مختلف برای انجام وظایف مختلف استفاده می کند[۶].

برنامه های کاربردی داده کاوی می تواند عمومی یا دامنه خاص باشند. برنامه های عمومی نیازمند یک سیستم هوشمند است که به تنهایی می تواند تصمیمات خاصی مانند انتخاب داده ها، انتخاب روش داده کاوی، ارائه و تفسیر نتیجه را بگیرند. برخی از برنامه های کاربردی عمومی داده کاوی نمی توانند این تصمیمات را به خود اختصاص دهند اما کاربران را برای انتخاب داده ها، انتخاب روش داده کاوی و برای تفسیر نتایج راهنمایی می کنند. برنامه کاربردی داده کاوی مبتنی بر چند عامل دارای قابلیت انتخاب اتوماتیک تکنیک داده کاوی است تا انجام شود. سیستم چند عاملی در سطوح مختلف استفاده شده : اول، در سطح مفهوم تعریف سلسله مراتبی و سپس در سطح نتیجه تا بهترین تصمیم سازگار برای کاربر را ارائه دهد. این تصمیم در پایگاه دانش برای استفاده در تصمیم گیری بعدی ذخیره می شود. ابزار سیستم چند عاملی مورد استفاده برای توسعه کلی سیستم داده کاوی عمومی از عوامل مختلف برای انجام وظایف مختلف استفاده می کند.

یک سیستم چند سطحی داده کاوی برای افزایش کارایی فرآیند داده کاوی پیشنهاد شده است. دارای اجزای اساسی مانند رابط کاربری، خدمات داده کاوی، خدمات دسترسی به داده ها و داده ها است. سه معماری متفاوت برای سیستم داده کاوی، یعنی معماری One-tire،Two-tire و Three-tire وجود دارد.

سیستم عمومی مورد نیاز تا جاییکه ممکن است بسیاری الگوریتم یادگیری کامل می کندو تصمیم می گیرد که بهترین الگوریتم برای استفاده است.کوربا (CORBA) معماری واسط درخواست شی عمومی دارای ویژگی هایی مانند:

یکپارچه سازی برنامه های مختلف کد گذاری شده در هر زبان برنامه نویسی بسیار آسان است.

اجازه می دهد تا قابل استفاده مجدد در یک راه قابل اجرا باشد و در نهایت امکان ساخت سیستم بزرگ و مقیاس پذیر را فراهم می کند.

معماری سیستم داده کاوی مبتنی بر کوربا توسط گروه مدیریت شی ارائه شده است و دارای تمام خصوصیات برای انجام محاسبات توزیع شده و شی گرا است.

تمرکز داده محور و متدولوژی های خودکار باعث می شود که داده کاوی برای غیر کارشناسان دسترس پذیر باشد.
استفاده از رابط های سطح بالا می تواند متدولوژی های خودکار را اجرا کند که مفاهیم داده کاوی را از کاربران پنهان کند.
طراحی داده محور تمام جزئیات متدولوژی کاوش را پنهان می کند و آنها را از طریق وظایف سطح بالا که هدف گرا هستند، به نمایش می گذارد. این وظایف هدف گرا با استفاده از API های داده محور اجرا می شوند. این طراحی وظایف داده کاوی را مانند سایر انواع کوئری ها ایجاد می کند که کاربران بر روی داده ها انجام می دهند.

در داده کاوی اگر داده بزرگ در دسترس باشد نتایج بهتری می تواند بدست آید. منجر به ادغام و اتصال پایگاه های محلی می شود. معماری جدید داده کاوی مبتنی بر تکنولوژی اینترنت این مشکل را حل کرد[۷].

عامل زمینه نقش حیاتی در موفقیت داده کاوی دارد. اهمیت و معنای داده های مشابه در زمینه های مختلف متفاوت است. داده ها در یک زمینه بسیار مهم هستند و ممکن است در سایر زمینه ها اهمیت زیادی نداشته باشند. چارچوب داده کاوی زمینه آگاه ، عوامل زمینه ای مفید و جالب را فیلتر کرد و می تواند پیش بینی دقیق و صریح با استفاده از این عوامل ایجاد کند.

برنامه های کاربردی خاص دامنه برای استفاده از داده های خاص دامنه و الگوریتم داده کاوی که مقصود خاصی را هدف قرار داده ، متمرکز شده اند. برنامه های مورد مطالعه در این زمینه با هدف تولید دانش خاص است. در حوزه های مختلف، منابع تولید داده ها انواع مختلفی از داده ها را تولید می کنند. داده ها می توانند از یک متن ساده، تعدادی داده های صوتی تصویری پیچیده تر باشند. برای استخراج الگوها و در نتیجه دانش از این داده ها، انواع مختلفی از الگوریتم های داده کاوی استفاده می شود. بنابراین جمع آوری و انتخاب داده های زمینه خاص و استفاده از الگوریتم داده کاوی برای تولید دانش زمینه خاص ، یک کار مهارانه است. در بسیاری از برنامه های داده کاوی خاص دامنه، کارشناسان دامنه نقش مهمی را در یادگیری دانش کاوش ایفا می کنند.

در علوم پزشکی محدوده وسیعی برای کاربرد داده کاوی وجود دارد. تشخیص زودرس، مراقبت های بهداشتی، نمایه سازی بیمار و تولید تاریخ و غیره، نمونه های کمی هستند. ماموگرافی روشی است که در تشخیص سرطان پستان مورد استفاده قرار می گیرد. رادیولوژیست ها در تشخیص تومورها با مشکلات زیادی مواجه هستند. روش های متداول کامپیوتری می تواند به کارکنان پزشکی کمک کند و دقت تشخیص را بهبود بخشند. شبکه های عصبی با انتشار معکوس و کاوش قانون وابستگی برای طبقه بندی تومور در ماموگرافی استفاده می شود. داده کاوی به طور موثر در تشخیص اختلال ریه که ممکن است سرطانی یا خوش خیم باشد استفاده می شود. الگوریتم های داده کاوی به طور قابل توجهی خطرات بیمار و هزینه های تشخیص را کاهش می دهد. با استفاده از الگوریتم های پیش بینی، دقت پیش بینی مشاهده شده در٪ ۳.۹۱ موارد ۱۰۰٪ بود. استفاده از

داده کاوی در مراقبت های بهداشتی بطور گسترده از برنامه های کاربردی داده کاوی استفاده می کند. داده های پزشکی پیچیده و تجزیه و تحلیلشان دشوار است.

یک (REMIND) استخراج قابل اطمینان داده و استنتاج معنادار از داده های غیر ساختار یافته داده های بالینی ساخت یافته و بدون ساختار در سوابق بیمار را به طور خودکار داده بالینی ساختار یافته با کیفیت بالا ایجاد می کند. کیفیت بالا ساختار اجازه می دهد تا پرونده های موجود بیمار برای حمایت از انطباق دستورالعمل ها و بهبود مراقبت از بیمار استخراج شوند.

داده کاوی در آموزش از راه دور به طور خودکار اطلاعات مفیدی برای افزایش فرایند یادگیری بر اساس مقدار گسترده ای از داده های تولید شده توسط معلمان و تعامل دانش آموزان با محیط آموزش از راه دور مبتنی بر وب تولید کرد. برنامه های داده کاوی، داده را به اطلاعات و بازخورد به محیط یادگیری الکترونیکی انتقال می دهند. این راه حل مقدار زیادی از داده های بی فایده را به یک سیستم نظارت و توصیه هوشمند که در فرآیند یادگیری اعمال می شود تبدیل می کند.

در آموزش مبتنی بر وب روش های داده کاوی برای بهبود آموزش استفاده می شود. روابط در میان داده های مصرفی بهبود یافته در طول جلسات دانشجویان کشف شده است. این دانش برای معلم یا نویسنده دوره بسیار مفید است که می تواند تصمیم بگیرد کدام تغییرات برای بهبود اثربخشی دوره مناسب باشد.

روش های داده کاوی نیز برای ارائه شاگردان با بازخورد تطبیقی بلادرنگ در مورد ماهیت و الگوهای ارتباطات آنلاین خودشان در هنگام یادگیری بطور اشتراکی استفاده می شود. این امر باعث افزایش آگاهی دانش آموزان می شود. استفاده از روش های داده کاوی در گپ های آموزشی هم امکان پذیر است و می تواند بهبود در محیط یادگیری را به ارمغان بیاورد.

داده کاوی مهندسان نگهداری نرم افزار را برای درک ساختار یک سیستم نرم افزاری و ارزیابی قابلیت نگهداری آن، تسهیل می کند.الگوریتم خوشه بندی به طور موثر برای تولید نظریات سیستم ها با ایجاد متقابل گروه های انحصاری کلاس ها، داده یا روش های عضوگیری را براساس شباهت های آنها استفاده می شوند و از این رو زمان لازم برای درک کلی سیستم را کاهش می دهد. این روش همچنین در کشف الگوهای برنامه نویسی و موارد غیر عادی و یا داده های خارج از محدوده که ممکن است مورد توجه قرار گیرند، کمک می کند.

تشخیص ناهنجاری در شبکه بسیار دشوار است و نیاز به یک تماسی بسیار نزدیک در ترافیک داده دارد. تشخیص نفوذ نقش مهمی در امنیت کامپیوتر دارد. روش طبقه بندی داده کاوی برای طبقه بندی ترافیک شبکه معمولی یا ترافیک غیرعادی استفاده می شود. اگر هر هدر TCP متعلق به هیچ یک از خوشه های هدر موجود TCP نیست، آن را می توان به عنوان ناهنجاری در نظر گرفت.

اجرایی مخرب، تهدیدی برای امنیت سیستم است به سیستم آسیب می رساند یا اطلاعات حساس بدون مجوز کاربر به دست می آورد. روش های داده کاوی برای دقت شناسایی اجرایی مخرب قبل از اجرای آنها استفاده شده اند. الگوریتم های طبقه بندی RIPPER، بیزهای ساده و یک سیستم چند طبقه بندی برای شناسایی اجرای جدید مخرب استفاده می شود. این طبقه بندی نرخ تشخیص ۷۶.۹۷٪ را نشان داد.

تجارت الکترونیک نیز قابل پیش بینی ترین حوزه برای داده کاوی است. ایده آل است زیرا بسیاری از عناصر لازم برای موفقیت داده کاوی به راحتی در دسترس هستند: سوابق داده ها فراوان است، مجموعه الکترونیکی داده های قابل اطمینان فراهم می کند ، بینش به راحتی می تواند تبدیل به عمل شود، و بازگشت سرمایه گذاری می تواند اندازه گیری شود. ادغام تجارت الکترونیک و داده کاوی نتایج قابل توجهی را بهبود می بخشند و کاربران را در ایجاد دانش و تصمیم گیری درست کسب و کار هدایت می کند. این ادغام به طور موثر چندین مشکل عمده مرتبط با ابزارهای ترازی داده کاوی را حل می کند از جمله تلاش های فراوانی که در پیش پردازش داده ها قبل از اینکه بتوان آن را برای کاوش استفاده کرد و و ایجاد نتایج کاوش قابل اجرا است[۸].

بازیابی کتابخانه دیجیتال ، جمع آوری، ذخیره و حفظ داده دیجیتال است. ظهور منابع الکترونیکی و افزایش استفاده از آنها در کتابخانه ها تغییرات قابل توجهی را در کتابخانه به وجود آورده است. داده ها و اطلاعات در قالب های مختلف در دسترس هستند. این فرمت ها عبارتند از متن، عکس ها، ویدئو، صوتی، تصویر، نقشه ها و غیره. بنابراین کتابخانه دیجیتال یک دامنه مناسب برای کاربرد داده کاوی است.

۹.نتیجه گیری

اغلب مطالعات قبلی در زمینه کاربردهای داده کاوی در زمینه های مختلف از انواع مختلف داده ها متنوع از متن تا تصاویر استفاده کردند و در انواع پایگاه های داده و ساختارهای داده ذخیره کردند. روش های مختلف داده کاوی برای استخراج الگوهای و در نتیجه دانش از این پایگاه داده های مختلف استفاده می شوند. انتخاب داده ها و روش های استخراج داده ها یک کار مهم در این فرآیند است و نیاز به دانش حوزه دارد. تلاش های متعددی برای طراحی و توسعه سیستم داده کاوی عمومی صورت گرفته است اما هیچ سیستمی به طور کلی عمومی نیست. بنابراین، برای هر حوزه دستیار متخصص حوزه اجباری است. کارشناسان حوزه باید سیستم را هدایت کنند تا دانش خود را برای استفاده از سیستم های داده کاوی به منظور ایجاد دانش مورد نیاز به طور موثر اعمال کنند. کارشناسان حوزه نیازمند تعیین انواع داده هایی که باید در حوزه خاص مسئله جمع آوری شوند هستند، انتخاب داده های خاص برای داده کاوی، تمیز کردن و تحول داده ها، استخراج الگوها برای تولید دانش و در نهایت تفسیر الگوهای و تولید دانش است.

اکثر برنامه های داده کاوی حوزه خاص دقتی بیش از ۹۰٪ را نشان می دهند. برنامه های داده کاوی عمومی دارای محدودیت هستند. با مطالعه برنامه های مختلف داده کاوی مشاهده شده است که هیچ نرم افزاری به نام نرم افزار عمومی ۱۰۰٪ عمومی نیست. رابطهای هوشمند و عوامل هوشمند تا حدودی نرم افزار را عمومی می کند اما دارای محدودیت هایی است. کارشناسان حوزه در مراحل مختلف داده کاوی نقش مهمی ایفا می کنند. تصمیمات در مراحل مختلف تحت تاثیر عواملی مانند دامنه و جزئیات داده، هدف از داده کاوی و پارامترهای زمینه است. برنامه های خاص دامنه با هدف استخراج دانش خاصی انجام می شود. کارشناسان دامنه با در نظر گرفتن نیازهای کاربر و سایر پارامترهای زمینه سیستم را هدایت می کنند. نتایج حاصل از برنامه های خاص دامنه دقیق تر و مفید تر است. بنابراین نتیجه گیری می شود که برنامه های کاربردی خاص برای داده کاوی خاصتر هستند. از مطالعه فوق به نظر می رسد طراحی و توسعه یک سیستم داده کاوی که می تواند به صورت پویا برای هر دامنه کار کند بسیار دشوار است.

اکثر برنامه های خاص داده کاوی دامین دقت بیش از ۹۰٪ را نشان می دهند. برنامه های داده کاوی عمومی دارای محدودیت هستند. از مطالعه برنامه های مختلف داده کاوی مشاهده شده است که هیچ نرم افزاری به نام نرم افزار عمومی ۱۰۰٪ عمومی نیست. اینترفیس های هوشمند و عوامل هوشمند تا حدودی نرم افزار را عمومی می کند اما دارای محدودیت هایی است. کارشناسان دامنه در مراحل مختلف داده کاوی نقش مهمی ایفا می کنند. تصمیمات در مراحل مختلف تحت تاثیر عوامل مانند دامنه و جزئیات داده، هدف از داده کاوی و پارامترهای زمینه است. برنامه های خاص دامنه با هدف استخراج دانش خاصی انجام می شود. کارشناسان دامنه با در نظر گرفتن نیازهای کاربر و سایر پارامترهای زمینه سیستم را هدایت می کنند. نتایج حاصل از برنامه های خاص دامنه دقیق تر و مفید تر است. بنابراین نتیجه گیری می شود که برنامه های کاربردی خاص خاص برای داده کاوی است. از مطالعه فوق به نظر می رسد بسیار دشوار است برای طراحی و توسعه یک سیستم داده کاوی، که می تواند به صورت پویا برای هر دامنه کار می کنند.

۱۰.مراجع

[1] Adderley, R., Townsley, M., & Bond, J. (2012). Use of data mining techniques to model crime scene investigator performance. Knowledge- based Systems.

[2] Ahn, H., Ahn, J. J., Oh, K. J., & Kim, D. H. (2015). Facilitating cross-selling in a mobile telecom market to develop customer classification model based on hybrid data mining techniques.

[3] Assous, F., & Chaskalovic, J. (2010). Méthodes de data mining pour l'analyse d'approximations numériques: Le cas de solutions asymptotiques des équationsde Vlasov–Maxwell= data mining techniques for numerical approximations analysis: A test case of asymptotic solutions to the Vlasov–Maxwell equations.

[4] Assous, F., & Chaskalovic, J. (2011). Data mining techniques for scientific computing: Application to asymptotic paraxial approximations to model ultrarelativistic particles.

[5] Bhramaramba, R., Allam, A. R., Kumar, V. V., & Sridhar, G. R. (2011). Application of data mining techniques on diabetes related proteins. International Journal of Diabetes in Developing Countries.

[6] Chen, S. C., & Huang, M. Y. (2011). Constructing credit auditing and control & management model with data mining technique. Expert Systems with Applications.

[7] Dakheel, F. I., Smko, R., Negrat, K., & Almarimi, A. (2011). Using data mining techniques for finding cardiac outlier patients. Proceedings of World Academy of Science, Engineering and Technology.

[8] Fesharaki, M., Shirazi, H., & Bakhshi, A. (2011). Knowledge acquisition from database of information management and documentation softwares by data mining techniques.