



فرم ایده های خلاقانه

تاریخ: ۷ مردادماه ۱۳۹۱

شماره:

صفحه ۱ از ۳

عنوان ایده		
اعمال الگوریتم های داده کاوی روی مجموعه داده های استاندارد نفتی در جهت استخراج دانش مفید برای تصمیم سازی مدیران شرکت ملی نفت ایران و شرکت های تابعه		
مخاطب اصلی در وزارت نفت		وزیر محترم نفت
مخاطبان دیگر		مدیران بلند پایه شرکت ملی نفت ایران و شرکت نفت فلات قاره
نام و نام خانوادگی		اشکان جعفری شورباخورلو
پیشنهاد دهنده		اسمت سازمانی طرح نظام، نفت فلات قاره، منطقه سیری
		پست الکترونیک Ajafari35@gmail.com
بیان مسئله و چالش <p>داده کاوی، پایگاه ها و مجموعه های حجیم داده ها را در پی کشف و استخراج دانش، مورد تحلیل و کند و کاوهای ماشینی (و نیمه ماشینی) قرار می دهد. این گونه مطالعات و کاوش ها را به واقع می توان همان امتداد و استمرار دانش کهن و همه جا گیر آمار دانست. تفاوت عمده در مقیاس، وسعت و گوناگونی زمینه ها و کاربردها، و نیز ابعاد و اندازه های داده های امروزی است که شیوه های ماشینی مربوط به یادگیری، مدل سازی، و آموزش را طلب می نماید.</p> <p>اصطلاح Data Mining همان طور که از ترجمه آن به معنی داده کاوی مشخص می شود به مفهوم استخراج اطلاعات نهان و یا الگوها و روابط مشخص در حجم زیادی از داده ها در یک یا چند بانک اطلاعاتی بزرگ است.</p> <p>بسیاری از شرکت ها و موسسات دارای حجم انبوهی از اطلاعات هستند. تکنیک های داده کاوی به طور تاریخی به گونه ای گسترش یافته اند که به سادگی می توان آنها را بر ابزارهای نرم افزاری امروزی تطبیق داده و از اطلاعات جمع آوری شده بهترین بهره را برد. در صورتی که سیستم های Data Mining بر روی سکوها Client/Server قوی نصب شده باشد و دسترسی به بانک های اطلاعاتی بزرگ فراهم باشد، می توان به سوالاتی از قبیل: کدامیک از مشتریان ممکن است خریدار کدامیک از محصولات آینده شرکت باشند، چرا، در کدام مقطع زمانی و بسیاری از موارد مشابه پاسخ داد.</p> <p>مسئله این است که اعمال الگوریتم های داده کاوی روی مجموعه داده های نفتی تا چه حد می تواند به مدیران در جهت اخذ تصمیمات دقیق و پیش بینی شده کمک کند؟</p>		
شرح پیشنهاد بهبود <p>داده کاوی به بهره گیری از ابزارهای تجزیه و تحلیل داده ها به منظور کشف الگوها و روابط معتبری که تا کنون ناشناخته بوده اند اطلاق می شود. این ابزارها ممکن است مدل های آماری، الگوریتم های ریاضی و روش های یاد گیرنده (Machine Learning Method) باشند که کار این خود را به صورت خودکار و بر اساس تجربه ای که از طریق شبکه های عصبی (Neural Networks) یا درخت های تصمیم گیری (Decision Trees) به دست می آورند بهبود می بخشد. داده کاوی منحصر به گردآوری و مدیریت داده ها نبوده و تجزیه و تحلیل اطلاعات و پیش بینی را نیز شامل می شود برنامه های کاربردی که با بررسی</p>		



فرم ایده های خلاقانه

تاریخ: ۷ مردادماه ۱۳۹۱

شماره:

صفحه ۲ از ۳

فایل‌های متن یا چند رسانه‌ای به کاوش داده‌ها می‌پردازند پارامترهای گوناگونی را در نظر می‌گیرد که عبارت‌اند از:

*رابطه (Association): الگوهایی که بر اساس آن یک رویداد به دیگری مربوط می‌شود مثلاً خرید قلم به خرید کاغذ.

- ترتیب (Sequence) الگویی که به تجزیه و تحلیل توالی رویدادها پرداخته و مشخص می‌کند کدام رویداد، رویدادهای دیگری را در پی دارد مثلاً تولد یک نوزاد و خرید پوشاک.
- دسته‌بندی (Classification) شناسایی الگوهای جدید مثلاً همزمانی خرید چسب و پوشه
- خوشه‌بندی (Clustering) کشف و مستند سازی مجموعه‌ای از حقایق ناشناخته مثلاً موقعیت جغرافیایی خرید محصولی با مارک خاص
- پیش‌بینی (Forecasting) کشف الگوهایی که بر اساس آنها پیش‌بینی قابل قبولی از رویدادهای آتی می‌شود، مثلاً رابطه عضویت در یک باشگاه ورزشی با شرکت در کلاسهای ورزشی.

برنامه‌های کاربردی که در زمینه تجزیه و تحلیل اطلاعات به کار می‌روند از امکاناتی چون پرس و جوی ساخت یافته (Structured query) که در بسیاری از بانک‌های اطلاعاتی یافت می‌شود و از ابزارهای تجزیه و تحلیل آماری برخوردارند اما برنامه‌های مربوط به داده کاوی در عین برخورداری از این قابلیت‌ها از نظر نوع با آنها تفاوت دارند. بسیاری از ابزارهای ساده برای تجزیه و تحلیل داده‌ها روشی بر پایه راستی‌آزمایی (verification) را به کار می‌برند که در آن فرضیه‌ای بسط داده شده آنگاه داده‌ها برای تایید یا رد آن بررسی می‌شوند. به طور مثال ممکن است این نظریه مطرح شود که فردی که یک چکش خریده حتماً یک بسته میخ هم خواهد خرید. کارایی این روش به میزان خلاقیت کاربر برای ارائه فرضیه‌های متنوع و همچنین ساختار برنامه بکار رفته بستگی دارد. در مقابل در داده کاوی روشهایی برای کشف روابط بکار برده می‌شوند و به کمک الگوریتم‌هایی روابط چند بعدی بین داده‌ها تشخیص داده شده و آنهایی که یکتا (unique) یا رایج هستند شناسایی می‌شوند. به طور مثال در یک فروشگاه سخت‌افزار ممکن است بین خرید ابزار توسط مشتریان با تملک خانه شخصی یا نوع خودرو، سن، شغل، میزان درآمد یا فاصله محل اقامت آنها با فروشگاه رابطه‌ای برقرار شود.

در نتیجه قابلیت‌های پیچیده‌اش برای موفقیت در تمرین داده کاوی دو مقدمه مهم است یکی فرمول واضحی از مشکل که قابل حل باشد و دیگری دسترسی به داده متناسب. بعضی از ناظران داده کاوی را مرحله‌ای در روند کشف دانش در پایگاه داده‌ها می‌دانند. (KDD) مراحل دیگری در روند KDD به صورت تساعدی شامل، پاکسازی داده، انتخاب داده انتقال داده، داده کاوی، الگوی ارزیابی، و عرضه دانش

فرم ایده های خلاقانه

تاریخ: ۷ مردادماه ۱۳۹۱

شماره:

صفحه ۳ از ۳

می باشد. بسیاری از پیشرفت ها در تکنولوژی و فرآیندهای تجاری بر رشد علاقه مندی به داده کاوی در بخش های خصوصی و عمومی سهمی داشته اند. بعضی از این تغییرات شامل:

- رشد شبکه های کامپیوتری که در ارتباط برقرار کردن پایگاهها داده مورد استفاده قرار می گیرند.
- توسعه افزایش تکنیکهایی بر پایه جستجو مثل شبکه های عصبی و الگوریتم های پیشرفته.
- گسترش مدل محاسبه کلاینت سروری که به کاربران اجازه دسترسی به منابع داده های متمرکز شده را از روی دسک تاپ می دهد.
- و افزایش توانایی به تلفیق داده از منابع غیر متناجس به یک منبع قابل جستجو می باشد.

علاوه بر پیشرفت ابزارهای مدیریت داده، افزایش قابلیت دسترسی به داده و کاهش نرخ نگهداری داده نقش ایفا می کند. در طول چند سال گذشته افزایش سریع جمع آوری و نگه داری حجم اطلاعات وجود داشته است. با پیشنهاد های برخی از ناظران مبنی بر آنکه کمیت داده های دنیا به طور تخمینی هر ساله دوبرابر می گردد. در همین زمان هزینه ذخیره سازی داده ها بطور قابل توجهی از دلار برای هر مگابایت به پنی برای مگابایت کاهش پیدا کرده است. مطابقاً قدرت محاسبه ها در هر ۱۸ - ۲۴ ماه به دوبرابر ارتقاء پیدا کرده است این در حالی است که هزینه قدرت محاسبه رو به کاهش است.

با توجه به توضیحات بالا، بکارگیری الگوریتم های داده کاوی بر روی مجموعه داده های استاندارد نفتی می تواند در تصمیم سازی مدیران بسیار نقش فعال داشته باشد. همچنین از حجم داده های ذخیره شده غیر مفید در پایگاه های داده کم شده و سرعت محاسبات و نرم افزارهایی که از این داده ها استفاده می کنند افزایش خواهد یافت. این پیشنهاد می تواند بطور قابل توجهی در شناسایی مشتریان خوب، نحوه برخورد با مشتریان، مقایسه شرکت ها و کشورهایی که از ما نفت خریداری می کنند و پیش بینی رفتار آن ها در مقابل برخی تغییرات و همچنین پیش بینی میزان نفت قابل تولید و یا قابل فروش و همچنین در شرکت های پالایش و پخش و میزان تولید نفت و گاز و یا پیش بینی یک مورد خاص در حین تولید یا فروش نفت و گاز و بسیاری موارد دیگر مفید واقع شود.

متخصصان نخبه در زمینه مهندسی کامپیوتر و رشته های مرتبط با کامپیوتر میتوانند با تشکیل کارگروهی بنام کارگروه داده کاوی نفتی، با همکاری نخبگانی در زمینه نفت تخصص دارند، الگوریتم های داده کاوی را روی داده های نفتی اعمال کنند و نتایج را به مقامات بالاتر طی یک بازه زمانی بصورت دوره ای گزارش دهند.