مطالعات مدیریت و کارآفرینی

دوره ۳، شماره ۱/۲، بهار ۱۳۹۶

صفحات ۲۵۷- ۲۴۹

Online ISSN: ٢٤٧٦-٤٣٧X

Print ISSN: ٢٥٣٨-٢٠٦٣

www.irijournals.com

تحلیل رفتار مشتری با استفاده از تحلیل های Big Data

علیرضا جباری رودی

استاد دانشگاه پیام نور مشهد- رشته مدیریت بازرگانی

چکیده

اگر جه سیستم های بسیاری وجود دارند که جهت تحلیل رفتار مشتری پیادسازی شده اند، ولی همچنان بـازار کشـف نشـده و آینده داری با پتانسیل زیادی برای پیشـرفت بهتـر در ایـن زمینـه وجـوددارد.Big Data یکـی از رو بـه رشـدترین رونـدهای تکنولوژی است که به طور قابل توجهی توانایی تغییر روش سازمان های کسب و کاری را دارد که از رفتار مشتری جهت تحلیل و تبدیل آن به بینش های ارزشمندی اسستفاده می کنند. حتی درخت های تصمیم گیری نیز می توانند به طور مـوثری بـرای تحلیل داده ها مورد استفاده قرار گیرند. در انتهای این مقاله، یک پیاده سازی Map Reduce با استفاده از طبقه بندی آماری شناخته شده الگوریتم ذرخت تصمیم C4.5 ارائه شده است. جدای از این سیستم قصد دارد نمنایش بصری داده های مشـتری را با استفاده از اسناد مشتق شده از داده ها[1] (d3.js) پیاده سازی کند که امکان ایجاد گرافیک هـای سفارشـی شـده خـوبی را فراهم می‌آورد.

واژه‌های کلیدی: تحلیل big data، الگوریتم C4.5، D3.j، نمایش بصری داده ها، Hadoop، map reduce.

۱. مقدمه

در اینجا منظور از Big Data در واقع مجموعه ای از داده ها بدون ساختار است که حجم بسیار زیادی دارند، از منابع متنوعی مانند وب، سازمان های تجاری و غیره با فرمت های مختلف بدست می آیند و با سرعت زیادی به دست ما می‌رسند که پردازش آنها با استفاده از ابزار معمولی مدیریت پایگاه داده را پیچیده و خسته کننده می کنند. آن را می توان یک جریان شدید روبه رشد نامید؛ بنابراین مسائل اصلی مورد تقاضا در پردازش[1] Big Data شامل ذخیره سازی، جستجو، توزیع، انتقال، تحلیل و نمایش بصری هستند (ریچمن،۲۰۱۱).

پیش از این، اصطلاح "تحلیل" در واقع مطالعه داده های موجود جهت پژوهش در مورد روندهای بالقوه و تحلیل تاثیر تصمیم گیری ها یا رویدادهای خاصی را نشان میداد که می توانند برای هوش تجاری جهت کسب بینش های باارزشی مورد استفاده قرار می گیرند. بزرگترین چالش امروزه این است که چگونه تمام اطلاعات پنهان از مقدار زیاد اطلاعات جمع آوری شده از مجموعه منابع متنوع کشف شود. اینجاست که تحلیل Big Data وارد صحنه میشود. یکی ازآنها تحلیل رفتاری مشتری است که به عنوان تحلیل مشتری اشاره شده است (هیلبرت، ۲۰۱۱).

تحلیل مشتری کمک می کند تا داده های بزرگ (یعنی Big Data) به ارزش های بزرگی تبدیل شوند که به سازمان ها اجازه ی پیش بینی رفتار خریداران را می دهد که در نتیجه ی آن باعث بهبود فروش خود، بهینه سازی بازار، برنامه ریزی موجودی، تشخیص تقلب و بسیاری کاربردهای دیگر خواهد شد. طیف گسترده ای از رویکردها موجود هستند و می توانند پیاده سازی شوند ولی یکی از این روش ها که جدای از آنها قرار می گیرد در واقع استفاده ا ز درختهای تصمیم گیری برای هدف طبقه بندی است و می توانند به طور موثری در تحلیل مصرف کننده استفاده شوند (اشنایدر، ۲۰۱۲).

در طول دوره ای از زمان الگوریتم های درخت تصمیم متعددی با افزایش درکارایی و توانایی جهت رسیدگی به انواع مختلف داده ها توسعه داده شده اند. یکی از الگوریتم های شناخته شده درخت تصمیم C4.5 است که بسط داده شده ی الگوریتم درخت تصمیم ID3 پایه است (ماگولاس، ۲۰۰۹). تحلیل مشتری بدون به نمایش درآوردن بصری داده ها ناقص است. علاوه بر طبقه بندی داده ها با استفاده از درخت[2] تصمیم، نمایش بصری داده ها نیز مهم است چرا که یک دید بصری به سازمان ها می دهد تا تغییرات الگوهای مصرف مشتری را درک کنند (ابراهیم، ۲۰۱۵).

۲. مروری بر کارهای مرتبط

سیستم های تحلیل معمولی برای رفتار مشتری عبارتند از (درینک، ۲۰۱۶):

در اواخر دهه ۱۹۷۰، دو رویکرد برای ساخت تصمیم های مدیریت پایگاه داده [1](DBMS ها) وجود داشت. رویکرد اول بر اساس مدل داده های سلسله مراتبی بود، نمونه ای که بوسیله (سیستم های مدیریت اطلاعات) از IBM در پاسخ به نیازهای ذخیره سازی اطلاعات عظیم تولید شده توسط برنامه فضایی آپولو ایجاد شده بود. رویکرد دوم بر اساس مدل سلسله مراتبی از قبیل ناتوانی آن د رنمایش روابط پیچیده ی DBMS ها را نیز حل نمود. با این حال، این دو مدل تعدادی معایب اساسی داشتند به عنوان مثال برنامه های پیچیده ای جهت پاسخ به کوئرهای حتی ساده باید نوشته می شد. همچنین حداقل عدم وابستگی داده ها وجود داشت (دی مارو ۲۰۱۴).

با ظاهر شدن محصولات تجاری د ردهه ۱۹۷۰ و اوایل دهه ۱۹۸۰، بسیاری از DBMS های رابطه ای تجربی پیاده سازی شده اند. DBMSرابطه ای که بطور گسترده ای در دهه های ۸۰ و۹۰ استفاده شده بود تنها برای برخورد با موجودیت

[1] Data management system(DBMS)
Object –relation database management system
Object- orinted data base management
[2] Ven diagram

پیچیده تر و داده های مورد نیاز شرکت ها محدود شده بودند، چنانکه عملیات و برنامه های کاربردی آنها بـه طـور فزاینـده ای پیچیده شد. دو مدل داده ای جدید در پاسخ به افزایش پیچیدگی برنامه های کاربردی پایگاه داده به وجود آمد، سیسـتم هـای مدیریت پایگاه داده ی شئ – رابطه ای [۲](ORDBMS) و سیستم های مدیریت پایگاه داده ی شئ گرا [۳](OODBMS) کـه به ترتیب مدل های داده ای رابطه ای و شئی ای را بیان می کنند.OODMS و ORDBMS جهت ارائه نسل سـوم سیسـتم های مدیریت پایگاه داده ترکیب شده اند (مایر، ۲۰۱۳).

طلوع تحلیل Big Data:

داده ها وقتی تبدیل به Big Data شدند که حجم، سرعت، یا تنوع آنها به سمتی فراتر از توانایی های سیستم هـای عملیـاتی فناوری اطلاعات جهت جمع آوری، دخیره سازی، تحلیل و پردازش آنها رفتند (لی، ۲۰۱۳).

بسیاری از سازمانها توانایی پرداختن به مقدار فراوانی از داده های بدون ساختار را با استفاده از ابزار و تجهیزات متنـوعی دارنـد ولی با سرعت حجم رو به رشد و جریان سریع داده ها، آنها نیز قابلیت داده کاوی و استنتاج بینش های لازم را در مـدت زمـان مناسبی ندارند.

Big Data از عرصه پروژه های علمی در شرکت های در حال ظهور است تا به غول مخابراتی کمک کنند که بفهمد کدام یـک از مشتری ها از سرویس خود خشنود هستند و چه فرایندهایی باعث ناخشنودی آنها می شود و پیش بینی اینکه کـدام یـک از مشتری ها قصد تغییر سرویس را دارند. برای بدست آوردن این اطلاعات نیاز است که میلیاردها بایت داده با ساختار بی ربط در مکان های مختلف پردازش شوند تا اطلاعات موردنیاز کسب شود. این نوع تحلیل، مدیریت اجرایی را قادر می سازد تا فراینـدها و افراد معیوب را رفع کنند یا ممکن است قادر به ارتباط برقرار کردن جهت حفظ مشتری های د رمعـرض خطـر باشـند. Big Data د رحال تبدیل به یکی از مهمترین گرایش های فناوری اسـت کـه پتانسـیلی بـرای تغییـر چشـمگیر روش سـازمان هـا دراستفاده از رفتار مشتری جهت تحلیل و تبدیل آن به بینش های ارزشمند را دارد (اودونوگو ۲۰۱۲)

مفاهیم کلیدی در تجزیه و تحلیل مشتری:

مروری بر تحلیل مشتری، مفاهیم کلیدی زیر را آشکار ساخته است:

۱) نمودار ون [۱] – کشف روابط پنهان:
بخش های مختلف را برای کشف اتصالات، روابط یا تفاوت ها ترکیب کنیـد. مشـتریانی کـه محصـولاتی از دسـتههای مختلـف خریده اند مورد کاوش قراردهید و به سادگی فرصت های متقابل فروش را شناسایی کنید.

۲) اطلاعات پروفایل – شناسایی ویژگی های مشتری:
رکوردهایی را از درخت داده های خود انتخاب کرده و پروفایل مشتری را وری تولید کنید که ویژگی ها و رفتار مشـترک وی را نشان می دهد. از پروفایل مشتری جهت اطلاع موثر استراتژی فروش و بازاریابی استفاده کنید (بریجیان، ۲۰۱۴).

۳) پیش بینی تحلیل سری های سازمانی:
پیش بینی شمار ار قادر می سازد تا با تغییرات، گرایش روندها و الگوهای فصلی انطباق داشته باشید. شما مـی توانیـد بـا دقـت حجم فروش ماهیانه یا تعداد شفارشات مورد انتظار را در هر ماه را پیش بینی کنید (وی، ۲۰۱۴).

۴) نقشه برداری – شناسایی مناطق جغرافیایی:

نقشه برداری از کدبندی رنگی بای نشان دادن رفتار مشتری استفاده می کند که در سراسر مناطق جغرافیایی تغییر می کند. یک نقشه به چند ضلعی هایی تقسیم میشود که نمایش دهنده مناطق جغرافیایی است که به شما نشان می دهد فروش شما در کدام منطقه متمرکز شده یا محصولات خاص در کدام مناطق بیشترین فروش را داشته اند.

۵) قوانین انجمنی – علت / اثر – تحلیل سبد:
این روش رابطه یا وابستگی الگوها را در سراسر داده ها تشخیص داده و مجموعه ای از قوانین را تولید می کند. این روش به طور خودکار قوانینی را انتخاب می کند که برای پیش بینی های تجاری مفیدترین باشند: چه محصولاتی را مشتریان به طور همزمان خریده اند وچه زمانی این کار را انجام داده اند؟ کدام مشتری ها در حال خرید نیستند و چرا؟ چه فرصت های متقابل فروش جدیدی وجود دارد؟

۶) درخت تصمیم – طبقه بندی و پیش بینی رفتار:
درختان تصمیم یکی از محبوب ترین روش ها برای طبقه بندی در برنامه های کاربردی مختلف داده کاوی هستند و به فرآیند تصمیم گیری کمک می کنند. طبقه بندی به شما کمک می کند تا انجام کارهایی مانند انتخاب محصولات صحیحی را جهت توصیه به مشتریان خاص انتخاب کنید و مسیر بالقوه را پیش بینی کنید. عمده ترین الگوریتم های درخت تصمیم که مورد استفاده قرار می گیرند شامل ID3,C4.5,CART هستند (بوید، ۲۰۱۲).

ابزاری برای نمایش بصری داده ها

۱) Polymaps: polymaps یک کتابخانه رایگان جاوا اسکریپت و یک پروژه مشترک از simpleGeo و Stamen است. این ابزار پیچیده ی پوشش نقشه می تواند داده ها را در طیف وسیعی بارگذاری کرده و قابلیت های چندین بار زوم را در سطوح مختلف از تمام راههای کشور گرفته تا نمای خیابان پیشنهاد دهد (وی، ۲۰۱۴).
۲) Flot: یک کتابخانه رسم نمودار جاوا اسکریپت برای جی کوئری است، flot یک برنامه کاربردی مبتنی بر مرورگر است که با اکثر مرورگرهای رایج از جمله اینترنت اکسپلورر، کروم، فایرفاکس، سافاری و اپرا سازگار است. flot از گزینه های متنوع نمایش بصری برای نقاط داده ها، نمودارهای تعاملی، نمودارهای پشته ای، چرخش و زوم و دیگر قابلیت های از طریق پلاگین های متنوع برای هر قابلیت خاص پشتیبانی میکند (وی، ۲۰۱۴).
۳) D3.js: یک کتابخانه جاوا اسکریپت برای ایجاد نمایش های بصری داده ها با تاکیدی بر استانداردهای وب با استفاده از html,svg,css است و از این اسناد با رویکرد داده محور جهت د[1] ستکاری DOM-با تمام قابلیت های مرورگر بدون محدودیت های چهارچوب های اختصاصی جهت استفاده مجدد دانش استفاده کند (وی، ۲۰۱۴).
۴) تحلیل بصری SAS[1] :تحلیل بصری SAS ابزاری برای کاوش مجموعه داده هایی با تمام اندازه ها بصورت بصری برای تحلیل جامع تر است. با یک بستر حسی و ابزار پیش بینی اتوماتیک، تحلیل بصری SAS حتی به کاربران غیر فنی نیز اجازه کشف روابط عمیق تر در پشت داده ها و پرده برداشتن از فرصت های پنهان را می دهد (وی، ۲۰۱۴).

۳. تکنولوژی های مرتبط:

۳-۱ APACHE HADOOP

APACHE HADOOP یک چارچوب نرم افزاری متن باز است.HADOOP شامل دو جز اصلی است: یک چارچوب پردازشی توزیع شده به نام Mapreduce و یک سیستم فایل توزیع شده که با عنوان سیستم فایل توزیع شده Hadoop یا

[1] Sas visual 'analytics
Hadoop distributed file system

HDFS[2] شناخته می شود. پردازش مقدار زیادی از داده ها یکی از مهمترین دلایل استفاده ا زاین چهارچوب در این پروژه است وانجام تحلیل آنها که با دیگر سیستم ها امکان پذیر نیست. ذخیره سازی به وسیله HDFS فراهم می شود و تحلیل توسط MapReduce انجام میشود. اگرچه Hadoop بیشتر به خاطر MapReduce و سیستم فایل توزیع شده اش شناخته شده است، با این حال پروژه های فرعی دیگری نیز سرویس های مکملی را فراهم می کنند یا بر روی هسته ایجاد شده اند تا انتزاعی از سطح بالا فراهم آورند (لی، ۲۰۱۳).

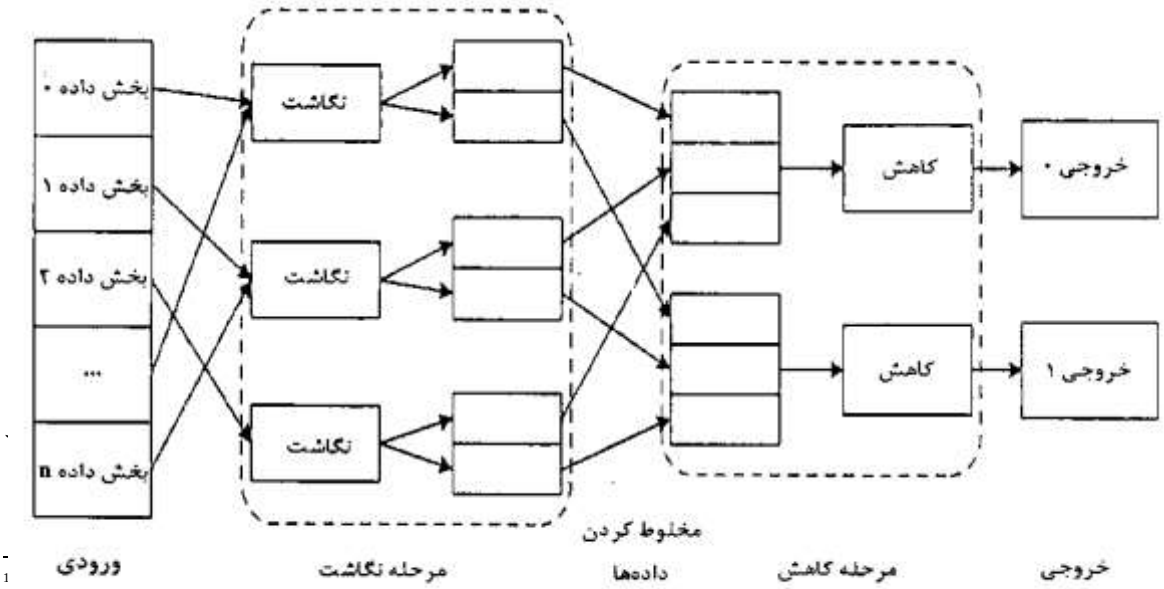
۳-۲ سیستم فایل توزیع شده Hadoop

سیستم فایل توزیع شده Hadoop (HDFS) [15] جزء ذخیره سازی است. بطور خلاصه،HDFS یک معماری توزیع شده را برای ذخیره سازی در مقیاس بسیار بزرگ فراهم می کند که می تواند به سادگی با گسترش مقیاس توسعه داده شود. هنگامی که یک فایل در hdfs ذخیره می شود، آنگاه فایل به بلوک هایی با اندازه های مساوی تقسیم می شود. اندازه بلوک ها می تواند سفارشی انتخاب شود یا از مقادیر پیش فرض استفاده شود. در این پروژه، مجموعه دادگان مشتری درhdsf ذخیره شده است.

مجموعه دادگان شامل رکوردهایی از مشتری با توجه به خریدها است. همچنین، خروجی فایل که شامل قوانین تصمیم است نیز بر روی HDFS نوشته می شوند (لی، ۲۰۱۳).

۳-۳ مدل نگاشت کاهش (MapReduce)

MapReduce یک مدل برنامه نویسی برای پردازش و تولید مجموعه دادگان بزرگ با الگوریتمی موازی و توزیع شده بر روی یک خوشه است. mapreduce به وسیله شکستن پردازش به دو مرحله کار میکند که مرحله نگاشت و مرحله کاهش. هر فازی دارای جفت های مقدار – کلید به عنوان ورودی و خروجی است، انواع هر یک ممکن است توسط برنامه نویس انتخاب شود. همچنین برنامه نویس دو تابع را مشخص می کند: تابع نگاشت وتابع کاهش. داده های خام مشتری ها ورودی به مرحله نگاشت ما هستند. یک فرمت متن ورودی را انتخاب کردیم که هر خط در مجموعه دادگان را به عنوان یک مقدار متن به ما می دهد. نکته کلیدی آفست[1] ابتدای خط از ابتدای فایل است. خروجی از تابع نگاشت پیش از ارسال به تابع کاهش توسط چارچوب MapReduce پردازش می شود. این پردازش جفت های مقدرا – کلید را براساس کلید مرتب و گروه بندی می کند (لی، ۲۰۱۳).

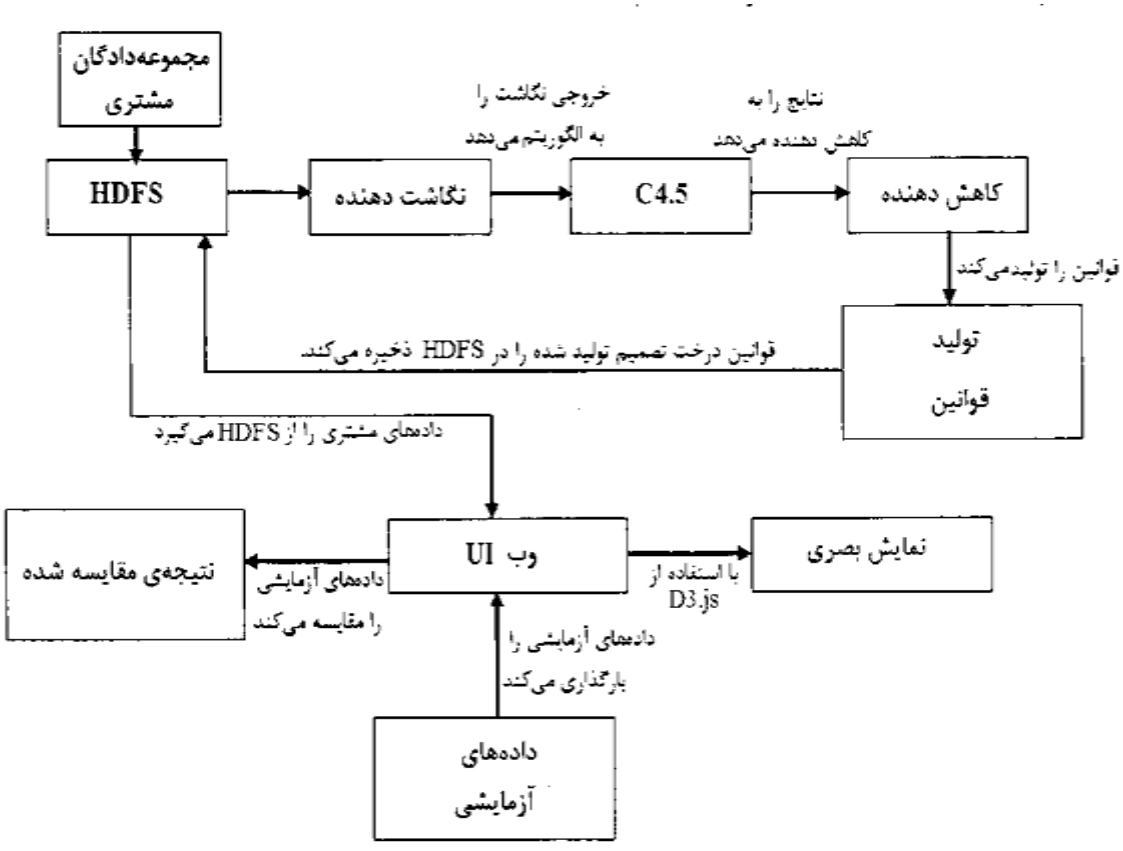


شکل ۱: مدل برنامه‌نویسی MapReduce

۴- روش تحقیق:

جریان سیستم به شرح زیر است:

۱) مجموعه داده مشتری از HDFS به عنوان ورودی برای برای الگوریتم بارگذاری کنید.
۲) از نمونه کلاس C4.5 استفاده کنید.
۳) با استفاده از چارچوب MapReduce ازhadoop، تابع نگاشت برای بررسی اینکه این نمونه به گره ی جاری تعلق دارد یا خیر مورد استفاده قرار می گیرد. این تابع برای همه ویژگی های کشف شده اندیس و مقدار آن و برچسب کلاس نمونه را به عنوان خروجی می دهد.
۴) تابع کاهش تعداد پیش آمدهای ترکیب (اندیس و مقدار آن و برچسب کلاس) را شمرده و شمارش را در مقابل آن چاپ میکند.
۵) آنتروپی، سود اطلاعات و نرخ سود ویژگی ها را محاسبه کنید.
۶) مجموعه دادگان ورودی از HDFS را مطابق تعریف شده ی داده کاوی درخت تصمیم C4.5 در چارچوب MapReduceپردازش کنید.
۷) قوانین تصمیم را تولید و آن را در HDFS ذخیره کنید.
۸) داده های آزمایشی جدید را از وب UI دریافت کنید.
۹) با دسترسی به قوانین و بر اساس آنها، در مورد دسته داده های جدید تصمیم گیری کنید.
۱۰) نمایش بصری مجموعه دادگان HDFS را بر روی ui د رفرمت نمودار های میله ای، نمودار پای و غیره با استفاده از از D3.js فراهم کنید.



شکل ۲: فلوچارت سیستم پیشنهادی

شده توسط C4.5 می توانند برای طبقه بندی استفاده شوند و به همین دلیل است که اغلب C4.5 به عنوان یک طبقه بندی آماری خوانده می شود. الگوریتم C4.5 از سود اطلاعات به عنوان معیار تقسیم استفاده می کند. این الگوریتم می تواند داده ها با مقادیر قطعی یا عددی بپذیرد. این الگوریتم برای رسیدگی به مقادیر بالای آستانه و مقادیر مساوی آستانه یا کمتر از آستانه تقسیم می کند. الگوریتم C4.5 می تواند به سادگی مقادیر از دست رفته را اداره کند. چنانکه مقادیر ویژگی از دست رفته توسط C4.5 در محاسبات سود استفاده نمی شود.

اجازه دهید C نمایش دهنده تعداد کلاس ها باشد. در این مورد، دو کلاس وجود دارد که قرار است رکوردها به این دو کلاس طبقه بندی شوند. کلاسهابصورت بله و خیر هستند.P(S,J) نسبت نمونه های موجود در S می باشد که به کلاس J ام نسبت داده شده است؛ بنابراین، آنتروپی ویژگی S بصورت زیر محاسبه می شود:

$$ENTROPY\ (S) = -\sum J = 1\ c\ p\ (s,j) * \log p\ (s,j)$$

آنتروپی برای هر رکورد یک ویژگی خاص محاسبه می شود:

بر اساس سود اطلاعات به وسیله مجموعه دادگان آموزشی T به صورت زیر تعریف می شود:

$$gain(s,j) = entropy\ (s) - \sum V\ \epsilon\ values \frac{(Ts)|T(s,y)|}{T(s)} * \log p\ (s,j)$$

که values (Ts)مجموعه ای از مقادیر s در T است، Ts زیر مجموعه ای از T که بوسیله S و Ts القا شده است، V زیرمجموعه ای از T است که در آن ویژگی S مقدار v دارد

۲-۲ نمایش بصری داده ها با استفاده از D3.js

D3.js یک کتابخانه جاوا اسکریپت برای دستکاری اسناد مبتنی بر داده ها است D3 به شما کمک می کند تا به داده ها با استفاده از css,cvg.html حیات ببخشید تاکید D3 بر روی استاندارد ها ی وبی است که به شما قابلیت های کامل مرورگرهای مدرن را میدهد بدون آنکه خودرا در گیر یک چارچوب خاص نمایید و این امر را با استفاده از اجزا نمایش بصری قدرتمند و یک رویکرد داده محور دستکاری داده ها ممکن میسازد (لی، ۲۰۱۳).

ویژگی های کلیدی D3.js

- چسباندن داده های دلخواه به dom
- ایجاد نمودارهای میله ای svj تعاملی
- ایجادجدول های html از مجموعه دادگان
- تنوع اجزا و پلاگین ها جهت ارتقا قابلیت ها
- اجزای ساخته شده قابل استفاده مجدد برای سهولت در برنامه نویسی

۵- نتیجه گیری:

این مقاله یک سیستم پیشنهادی برای پیاده سازی توزیع شده الگوریتم c4.5 با استفاده از چارچوب mapreduce به همراه نمایش بصری داده های مشتری را ارائه می دهد با افزایش توسعه محاسبات ابری و BigData الگوریتم های درخت تصمیم معمولی نمی توانند متناسب عمل کنند از این رو ما پیاده سازی mapreduce الگوریتم درخت تصمیم C4.5 را معرفی نمودیم. نمایش بصری با استفاده از D3.js انجام میشود که سریع و قابل استفاده مجدد است. چرا که D3.js از عناصر html

معمولی بهمراه گرافیک برداری مقیاس پذیر (SVG) استفاده می کند. در کارهای آینده استفاده از سیستم های سریع و زمان واقعی مانند mongodb,apache,hbase می توانند با این سیستم ترکیب شوند علاوه بر این ما می توانیم از الگوریتم های تصفیه توزیع شده مانند درخت جنگل پیاده سازی شده درApache mahout جهت افزایش کارایی و مقیاس پذیری استفاده کنیم.[1]

منابع

1. Boyd, D.; Crawford, K. (2012). "Critical Questions for Big Data". Information, Communication & Society. 15 (5): 662–679.
2. Brijiain R Patel, Mr.Kushik K Rana (2014). A Survey on Decision Tree Algorithm for classification.
3. De Mauro, Andrea; Greco, Marco; Grimaldi, Michele (2016). "A Formal definition of Big Data based on its essential Features". Library Review. 65: 122–135
4. Drink deRoos, Paul C. Zikopoulos, Bruce Brown, Rafael Coss, Roman B. Melnyk – Hadoop For Dummies, John Wiley & Sons, Inc. Hoboken, New Jersey,2014
5. Hilbert, Martin; López, Priscila (2011). "The World's Technological Capacity to Store, Communicate, and Compute Information". Science. 332 (6025): 60–65.
6. Ibrahim; Targio Hashem, Abaker; Yaqoob, Ibrar; Badrul Anuar, Nor; Mokhtar, Salimah; Gani, Abdullah; Ullah Khan, Samee (2015). "big data" on cloud computing: Review and open research issues". Information Systems. 47: 98–115
7. Lee, Jay; Wu, F.; Zhao, W.; Ghaffari, M.; Liao, L (January 2013). "Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications". Mechanical Systems and Signal Processing. 42 (1).J
8. Magoulas, Roger; Lorica, Ben (February 2009). "Introduction to Big Data". Release 2.0. Sebastopol CA: O'Reilly Media (11).
9. Mayer-Schönberger, V., & Cukier, K. (2013). Big data: a revolution that will transform how we live, work and think. London: John Murray.
10. O'Donoghue, John; Herbert, John (1 October 2012). "Data Management Within mHealth Environments: Patient Sensors, Mobile Devices, and Databases". Journal of Data and Information Quality. 4 (1): 5:1–5:20.
11. Recent advances delivered by Mobile Cloud Computing and Internet of Things for Big Data applications: a survey". International Journal of Network Management. 11 March 2016. Retrieved 14 September 2016.
12. Reichman, O.J.; Jones, M.B.; Schildhauer, M.P. (2011). "Challenges and Opportunities of Open Data in Ecology". Science. 331 (6018): 703–5
13. Snijders, C.; Matzat, U.; Reips, U.-D. (2012). "'Big Data': Big gaps of knowledge in the field of Internet". International Journal of Internet Science. 7: 1–5.
14. Wei Dai and Wei Ji. (2014). A MapReduce Implementation of C4.5 Decision Tree Algorithm. International Journal of Database Theory and Application.

[1]built in
Scalable vector Graphic
Real time

مطالعات مدیریت و کارآفرینی

دوره ۳، شماره ۱/۲، بهار ۱۳۹۶، صفحات ۲۵۷- ۲۴۹

Analyzing Customer Behavior Using Big Data analysis

Alireza Jabari Roodi

Payame Noor University of Mashhad, Faculty of Business Administration

Abstract

Although many systems have been implemented for customer behavior analysis, there is still a lot of potential for better advancement in this field. Big Data is one of the fastest-growing technology trends that can significantly change the ways used by business organizations which use customer behaviors to analyze and transform them into valuable insights. Even decision trees can be used effectively for data analysis. At the end of this paper, a Map Reduce implementation using the well-known statistical classification of decision tree algorithm c4.5 has been presented. Apart from this system, it tends to implement visual representation of customer data using data driven documents (d3.js), which allows for the creation of customized graphics.

Keywords: Big Data analysis, C4.5 algorithm, D3.js, visual representation of data, Hadoop, Map Reduce