مجموعه جزوه های کاربردی در زمینه مدیریت تحول استراتژیک یکپارچه وجامع

تالیف :

دکتر ابراهیم معین نجف آبادی

فهرست مطالب

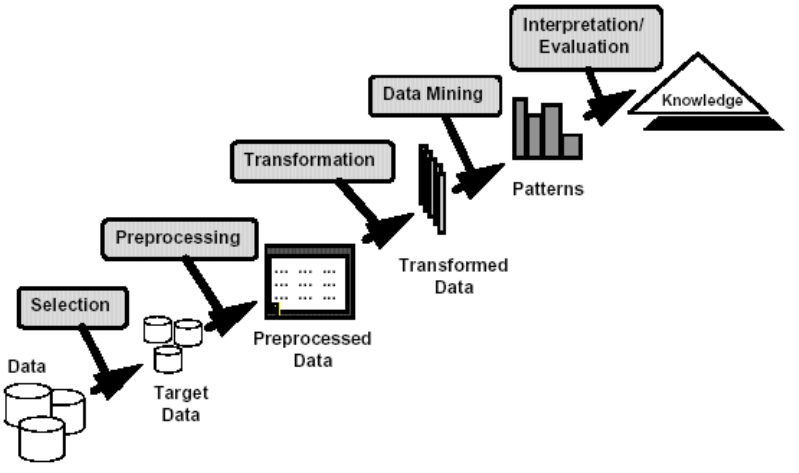
فهرست اشکال

فهرست جداول

۱- داده کاوی چیست ؟

داده کاوی[۱] فرایند تحلیل حجم زیادی از داده ها به منظور استخراج اطلاعات مفید ، واضح و با اهمیت ؛ الگوها وقوانین با معنی با استفاده از ابزارهای پیشرفتهٔ خودکار و نیمه خودکار است.

چنین داده هایی ممکن است در پایگاه های داده، انبارهٔ داده، و یا هر مخزنی از داده ها ذخیره شده باشند .هدف داده کاوی برای بسیاری از سازمان ها می تواند بهبود بازاریابی ،فروش یا عملیات مربوط به خدمات مشتری از طریق درک بهتری از مشتریان، شناخت الگوهای نا متعارف،پیش بینی آینده بر پایهٔ تجارب گذشته وبهبود روش های کاری فعلی باشد.



شکل شماره ۱: فرآیند داده کاوی

۲- ضرورت داده کاوی

درعصر اطلاعات ، سازمانهای کشور با مشکلاتی در ابعاد جهانی و داخلی به شرح زیر روبرو هستند:

۱-۲- مشکلات جهانی :

- ابهام در چشم انداز و مقصد نهایی
- هدفهای غیر قابل حصول
- مأموریت سازمانی ناپایدار وغیر موثر و ناکارا
- راهبردها وتاکتیکهای با نتایج ناخواسته
- چند وجهی بودن مشکلات و تصمیم گیریها

۱- Data mining

- ابهام درتشخیص و جذب فنآوری وابزارهای مناسب

۲-۲ – مشکلات داخلی

- ناتوانی درتحقق رسالت
- بازده سرمایه پایین
- کیفیت پایین محصولات و خدمات
- ظرفیتهای خالی وغیر مولد سرمایه های موجود
- کاهش مستمر قدرت جذب وحفظ نیروهای کیفی
- رشد منفی سرمایه گذاری واستهلاک تدریجی سرمایه ها
- محدودیت دردستیابی به تکنولوژیهای پایه ای، محوری، وپیشرو
- ناتوانی دربه انجام رساندن پروژه های بزرگ وپیچیده

۲-۳ – رهنمودهایی برای رفع مشکلات

- تحول یک ضرورت است.
- تحول یک موضوع ملی است.
- تحول نیازمند رهبری قدرتمند وسنت شکن است.
- تحول نیاز به تحلیل جامع و اندیشمندانه در اوضاع ودرک صحیح از مسایل دارد.
- تحول نیازمند تنظیم برنامه استراتژیک ویافتن راه حلهای بنیادی است.
- تحول نیازمند استفاده از فرایند استخراج دانش وبه کارگیری آن درتصمیم گیری است.

۲-٤ – اهداف پیشنهادی برای تحول

- پاسخگویی موثر به نیاز مشتری
- بهینه نمودن بازدهی سرمایه وبهره وری نیروی انسانی
- رضایت ،تعهد و بقای مستمر کارکنان
- ارتقاء مستمر استانداردها وکیفیت درابعاد مختلف
- انعطاف وقابلیت پاسخگویی سریع به تغییرات محیطی

۲-۵ – اقدامات اصلاحی برای تحقق اهداف پیشنهادی برای تحول

- آماده سازی بستر حرکت
  - سازمانی (اداری)
  - فرآیندی

- آشنایی با مفاهیم ونگرش منطبق با مدل اقتصادی مناسب
- ایجاد و بکارگیری ابزارها و روشهای تصمیم گیری،پیگیری و کنترل
- پویایی ساختاری
- تغییرات وسیع بعنوان شیوهٔ روزمرهٔ تجاری بر پایهٔ شرایط پویای محیطی
- نیاز به ابزارها وفنآوریهای نوین ، مدیریت وکاربرد اطلاعات و معرفت

٦-٢ – جایگاه داده کاوی بعنوان رهیافتی برای سازمانهای نوین

- تبدیل داده ها به دانش
- کمک به منطقی تر نمودن تصمیمات
- درک صحیح تر از ابعاد مختلف وپیچیدهٔ مسایل
- تبدیل داده های حجیم به نتایج مختصر
- واکنش به تغییرات سریع محیط : انعطاف پذیری
- ارزشیابی:

– راهبردها و تاکتیکها

– رویه ها

– نیروهای انسانی و تجهیزات و امکانات

–کارآیی تصمیم گیریها

- احترام به ارزشهای انسانی :

– رضایت مشتریان

– همکاران

- استخراج معرفت و عدم وابستگی به اطلاعات ذهنی کارکنان

۳- چارچوب داده کاوی

چارچوب داده کاوی شامل سه لایه است:

لایه *اول:* مربوط به فن آوریهای مورد استفاده درداده کاوی است . مثلاً، مدیریت پایگاه داده[۱]،یادگیری ماشین[۲]، آمار[۳]، بصری سازی[۴]، پردازش موازی[۵]،کمک در تصمیم گیری[۶]و انباره داده[۷].

لایه *دوم :* فنون انجام کار و ابزارهای داده کاوی قرار دارند.

لایه *سوم:* حوزه های مختلف داده کاوی قرار می گیرند .برای مثال، داده کاوی در پایگاه داده های توزیع شده[۸]،داده کاوی داده های چند رسانه ای[۹]،داده کاوی دروب[۱۰]، داده کاوی ابرداده ها[۱۱]، مساله امنیت داده ها و کنترل دسترسی.

٤- سیر تاریخی از جمع آوری داده تا داده کاوی

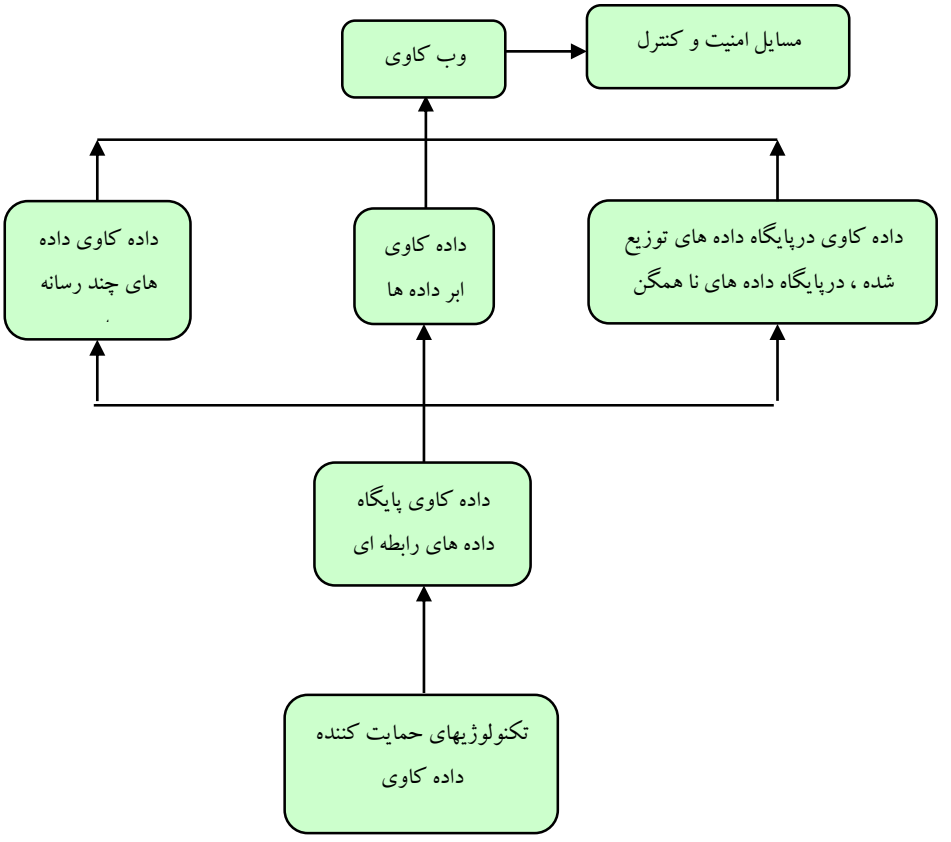
جدول شماره ۱: سیرتاریخی داده کاوی

Data Evolution	**Question**	***Activated Technology***
Data Collection ۱۹٦۰s	*How much is the total income of a person in last year?*	*Computer Tapes & Disks*
Data Access ۱۹۸۰s	*How Many units of a spesific product were soled durinig last month?*	*Relational Database ODBC,SQL*
Data warehouse DSS ۱۹۹۰s	*How Many units of a spesific product were soled durinig last month in comparison with other stores ?*	OLAP, DW
Data Mining Now	*What will happened in next month sell?*	*Perofessional algorithms multi processor computers*

٥- سیر تکامل داده کاوی

درشکل زیر سیر تکاملی داده کاوی مشاهده می شود:

۱-Database Management
۲-Machine learning
۳-Statistics
٤-Visualisation
٥-Parallel processing
٦-Decision support
۷-Data warehousing
۸-Mining Distributed database
۹-Mining Multimedia data
۱۰-Mining data on the world wide web
۱۱-Metadata aspects of Mining



شکل شماره۲ : سیرتکامل داده کاوی

٦- فنون مورد استفاده درداده کاوی

٦-١- سیستم های مدیریت پایگاه داده

سیستمهای مدیریت پایگاه داده، سیستمهایی هستند که وظیفهٔ سازماندهی ، ایجاد ساختار، ومدیریت داده ها را بعهده دارند. ازآنجا که داشتن داده های مناسب کلید بدست آوردن نتایج خوب و مورد انتظار است استفاده از سیستمهای مدیریت پایگاه داده حائز اهمیت می باشد .

از داده کاوی درهر نوع پایگاه داده می توان استفاده کرد .نمونه هایی از آن در زیرذکر شده است :

- پایگاه دادهٔ رابطه ای
- پایگاه دادهٔ شی گرا
- پایگاه دادهٔ چند رسانه ای

- پایگاه دادهٔ وب
- پایگاه دادهٔ توزیع شده
- انباره داده

۱-۱-٦- تاثیر سیستمهای پایگاه داده بر داده کاوی

ازآنجا که داده ، نقش اساسی درداده کاوی دارد ونحوهٔ نگهداری آن بر نحوهٔ بازیابی داده تاثیر می گذارد ،داده کاوی نیز تحت تأثیر داده قرار می گیرد .مراحل فرآیندی در طراحی پایگاه داده در رهیافت داده کاوی تاثیر می گذارد که شرح آن در زیر آمده است :

۱-۱-۱-٦- نحوهٔ مدل کردن داده ها[۱]

بسیاری از داده ها درحال حاضر در پایگاه داده های رابطه ای ذخیره شده اند و این در حالی است که حجم نسبتاً زیادی از داده ها هم در پایگاه داده های شی گرا [۲]،پایگاه داده های شی ـ رابطه ای[۳]،پایگاه داده های چند رسانه ای[۴]،ذخیره شده اند . از آنجا که اطلاعات کمی از داده کاوی پایگاه داده های شی گرا حاصل می شود ، ابتدا رابطهٔ بین اشیاء از چنین پایگاه داده هایی استخراج شده ودر یک پایگاه دادهٔ رابطه ای اعمال می شود.

۲-۱-۱-٦-معماری [۵]

می توان ابزارهای داده کاوی را به یک سیستم مدیریت پایگاه دادهٔ [۶]همه منظور اعمال کرد، یعنی یک سیستم مدیریت پایگاه داده تجاری ویک ابزار تجاری داده کاوی که با چنین سیستم هایی کار می کند خریداری کرده وبر آن اعمال نمود . با وجود مزایای چنین کاری ، مشکلاتی از قبیل پایین آمدن کارایی وجود دارد.

راه دیگر ، استفاده از موتورهای پایگاه داده ای است که ابزار داده کاوی را همراه خود دارند . درواقع ابزار کاوش در سیستم می باشند . به این ترتیب ،بسیاری از توابع و کارکردهای سیستم مدیریت پایگاه ،تحت تاثیر فنون داده کاوی قرار می گیرند.

---

[۱]-Data Modeling
[۲]-Object- oriented
[۳]-Object- relational
[۴]-Multimedia
[۵]-Architecture
[۶]-Database Management System

۳-۱-۱-٦- طراحی پایگاه داده[۱]

طراحی پایگاه داده نقش اصلی روی فرآیند داده کاوی دارد . درانباره داده روش های مختلفی (چون مدل های داده ای چند بعدی [۲]،مدل های پردازشی تحلیل بر خط [۳]و یا طرحهای[۴] مختلفی مثل طرح ستاره[۵]) برای مدل بندی ودر نهایت طراحی ارائه شده است . از آنجا که سازماندهی و ساختار داده در داده کاوی بسیار اساسی است ، انتخاب هر مدل وطراحی برداده کاوی تاثیر می گذارد.

٤-۱-۱-٦- مدیریت ونظارت پایگاه داده[٦]

مدیریت پایگاه داده تحت تاثیر داده کاوی قرار می گیرد . برای مثال اگر از سیستم مدیریت پایگاه داده استفاده کنیم که ابزار داده کاوی رادر خود دارد سوالاتی از قبیل زیر مطرح می شوند :

- چند وقت به چند وقت باید داده ها را تحت فرایند داده کاوی قرارداد؟
- آیا از داده کاوی می توان برای تحلیل داده های ممیزی شده[۷] استفاده کرد؟
- به روز رسانی داده ها در فواصل زمانی کوتاه چه تاثیری برداده کاوی خواهد داشت؟

۵-۱-۱-٦- بسیاری‌از توابع وکارکردهای پایگاه داده نیز تحت تاثیر داده کاوی قرار می گیرند:

- پردازش پرسش و پاسخ : معمولاً توسط زبانهای پردازش مثل زبان پرس و جوی ساخت یافته ۸ انجام می شود.
- مدیریت عملیات۹: از آنجا که داده کاوی معمولاً روی داده های مورد استفاده درتصمیم گیری صورت می گیرد ، تاثیر داده کاوی بر داده های عملیاتی کم است مگر در موارد خاص .
- ابر داده ها[۱۰]در مواردی که نمی توان " داده" را تحلیل کرد ، از کاوش در " ابر داده" می توان اطلاعات مفیدی استخراج کرد . مانند مواقعی که داده ها بدون ساختار هستند اما ابر داده دارای ساختار می باشد . از طرف دیگر " ابر داده " می تواند اطلاعات مفیدی دراختیار قرار دهد که در فرایند داده کاوی به ما کمک کند .

۱-Database Design
۲-Multidimentional Data
۳-On – line Analytical processing
٤-Schema
٥-Star- schema
٦-Database Administration
۷-Audit data
۸-SQL
۹-Transaction management
۱۰-Metadata

امنیت[۱]،صحت[۲]،کیفیت داده[۳]،میزان تحمل خطا[۴]: همه این موارد ، تحت تاثیر داده کاوی قرار می گیرند از داده کاوی می توان برای حفظ داده ها از تخریب ، با تشخیص تخریب و بهبود داده ها بهره گرفت . ازداده کاوی می توان برای حفظ سلامت داده ها در سیستمهای مختلف مثل سیستمهای کنترل ترافیک هوایی ،سیستمهای هسته ای وسیستمهای سلاحهای جنگی استفاده کرد.

۲-٦- انباره داده

با فنآوی انباره داده، داده های منابع مختلف بطریقی جمع آوری و نگهداری می شوند که براحتی درفرآیند تصمیم گیری مورد استفاده قرار گیرند . در اختیار داشتن یک انباره داده مناسب ، فرآیند داده کاوی را آسانتر می کند .

الف- مروری بر پیدایش انباره داده :

از آنجا که محیطهای قدیمی[۵]دارای مشکلات زیر هستند :

- سطوح عملیاتی[۶]مختلف
- واسط عملیاتی متفاوت
- نمایش داده ای مختلف
- اطلاعات ذخیره شده ناسازگار وناهمگن ودوباره ذخیره شده
- مجتمع نبودن اطلاعات درسیستمها
- نبود داده های تاریخی
- نبود دادهٔ خلاصه بندی شده[۷]

ب- نیاز مدیران :

- ازدید اطلاعاتی بتوانند تصویر واضح وکاملی نسبت به کار و تجارب خود داشته باشند
- داده های مجتمع درسطح عملیاتی داشته باشند
- داده های تاریخی ، مختصر وبا جزییات کافی داشته باشند

۱-Security
۲-Integrity
۳-Data quality
٤-Fault tolerance
٥-Legacy Environment
٦-Platform
۷-Summerize

بر این اساس، انباره داده سازوکاری برای بالا بردن سطح دسترسی به اطلاعات و تصمیم گیری بهتر است . صاحب نظری در تعریف انباره داده می گوید : یک انبارهٔ داده یک پایگاه دادهٔ موضوع گرا [۱]، منسجم[۲] ، متغیر با زمان[۳] و بدون تغییر حالت ناگهانی[۴] است. درواقع مجموعه ای است از داده ها که در فرآیند تصمیم گیری ، مدیریت رایاری می نماید.

پ- مزایای انباره داده:

- بهتر کردن کیفیت تصمیم گیری
- افزایش شانس موفقیت در رقابتهای تجاری
- برقراری ارتباط قویتر با مشتریان
- بهتر کردن سطح عملیات
- بهتر کردن رابطه با فروشندگان

ت- دلایل نیاز به انباره داده :

- وقتی که حجم داده ها برای تصمیم گیری خیلی زیاد است
- سرعت تغییر دربازار زیاد است
- اطلاعات درسراسر سازمان پخش شده اند
- هر واحدی با دید خودش به اطلاعات نگاه می کند در حالیکه نیاز به ایجاد دیدگاهی مشترک نسبت به اطلاعات می باشد .
- اطلاعات در قالبی تهیه می شود که کارکردن با آن مشکل است
- ارائه گزارش های ثابت و معین کاربردی ندارد

ث- تاثیر انباره داده در فرآیند داده‌کاوی

- دادهٔ منسجم : باداشتن داده ها بصورت منسجم امکان بررسی آنها به سرعت وبه آسانی برای داده‌کاو فراهم است .
- داده با جزئیات کامل : داده با جزئیات کامل ، کمک زیادی به داده کاو می کند تا بتواند آنرا در جزئی ترین شکل بررسی کرده و الگوهای مهم رادر پایین ترین سطح جزئیات کشف کند.

۱-Subject oriented
۲-integrated
۳-time variant
٤-non volatile

- دادۀ خلاصه بندی شده : وقتی داده ها بصورت خلاصه بندی شده موجود باشند از کار اضافی جهت تحلیل داده ها جلوگیری می شود. بطور معمول ، حجم کار جهت خلاصه بندی و تحلیل داده ها بسیار زیاد است .
- قطعات۱ مهم اطلاعات : قطعات مهم اطلاعات در دادۀ تاریخی مخفی می شوند .
- ابر داده : داده کاو نه تنها برای توصیف محتوای۲ اطلاعات از ابر داده ها استفاده می کند ،بلکه برای توصیف مفهوم۳ اطلاعات نیز می تواند از ابر داده ها بهره گیرد.

۳-٦- آمار

محققان روشهای آمار فنون خود را با فنون یادگیری ماشین [۴] مجتمع کرده اند تا روشهای آماری پیشرفته تری را برای داده کاوی تولید کنند. امروزه بسیاری از بسته های نرم افزاری که به طریقۀ تحلیل آماری عمل می کنند جزء ابزارهای داده کاوی هستند. بر این اساس آمار یکی از حوزه های اصلی مورد استفاده درداده کاوی است.

٤-٦- یادگیری ماشین

یادگیری ماشین بدین معنی است که ماشین از الگوهای انسانی ، قوانین مختلفی بیاموزد و سپس از این قوانین برای حل مسایل استفاده کند. درست است که اصول مورد استفاده در یادگیری ماشین و داده کاوی شبیه هم هستند ولی حجم داده ها در داده کاوی بسیار زیاد است .در نتیجه ،یکپارچگی مدیریت پایگاه داده و فنون یادگیری ماشین برای داده کاوی لازم است .

٥-٦- بصری سازی

محققین شاخۀ تصویرسازی کامپیوتری به داده کاوی با دید دیگری نگاه می کنند .هدف آنها داده کاوی محاوره ای است . برای این منظور از فنون بصری سازی استفاده می شود.

٦-٦ -کمک در تصمیم گیری

سیستمهای تصمیم یار مجموعه ای از ابزارها وفرآیندها هستند تا به مدیران در تصمیم گیری بهتر کمک کنند و راهنمای آنها در مدیریت باشند. بعنوان مثال می توان ابزارهای زمانبندی ملاقاتها ، سازماندهی رخدادها ، نمایش نموداری صفحات گسترده و ابزارهای ارزیابی کارایی را نام برد.

۷- روشهاي داده كاوي

۱- nuggets
۲-content
۳context-
٤-Machine learning

برای بررسی داده ها دو شیوه رایج است: از بالا به پایین[۱]واز پایین به بالا[۲] . در شیوه از بالا به پایین که آنرا آزمون فرضیه[۳]نیز می نامند، محقق بدنبال اثبات فرضیهٔ خود به بررسی داده های موجود می پردازد . اگر به موردی برخورد کند که فرضیه اش را تایید نکند ، فرضیه خود را اصلاح می کند . از بسیاری از روشهای آماری برای این منظور استفاده می شود. بطور کلی در آزمون فرضیه ، نظرات و مدلهایی ارائه می شود که با انجام ارزیابی ،اعتبار یا عدم اعتبار آنها تعیین می شود.

درشیوه پایین به بالا هیچ سوال یا فرضیه ای ارائه نمی شود . بلکه اجازه داده می شود که داده ها خود سخن بگویند . به این شیوه ،کشف دانش[۴]می گویند. این روش می تواند جهت دار[۵]یا بدون جهت[۶]باشد اگر ساختاری از پیش تعیین شده مورد نظر باشد آنرا جهت دار و در غیر اینصورت آنرا بدون جهت می نامند.

در داده کاوی ابتدا باید خروجی مورد انتظار را معین نمود سپس فن مورد نیاز برای تولید این خروجی را انتخاب کرد ودر نهایت روش انجام داده کاوی یعنی شیوهٔ بالا به پایین یا شیوهٔ پایین به بالا را اتخاذ نمود .

۸- فرآیند داده کاوی

مراحل زیر در فرایند داده کاوی طی می شوند:

- تعیین و تعریف مساله
- استفاده از فنون داده کاوی برای تبدیل داده ها به اطلاعات
- اقدام بر اساس اطلاعات بدست آمده
- ارزیابی نتایج حاصل شده

۹- وظایف داده کاوی

داده کاوی ، درحقیقت مجموعه ای از وظایف را تحت شرایط خاص خود به انجام می رساند .بسیاری از مسائل از قبیل مسائل اقتصادی ، تجاری ،و دهنی می تواند درمحدودهٔ شش وظیفهٔ زیر قالب بندی شود :

---

[۱]-Top- down
[۲]-Bottom-up
[۳]-Hypotesis testing
[۴]-Knowledge discovery
[۵]-Directed
[۶]-Undirected

- طبقه بندی۱
- تخمین۲
- پیش بینی ۳
- تحلیل سبد خرید یا وابستگی گروهی۴
- خوشه بندی۵
- توصیف۶

هیچ ابزار داده کاوی یا تکنیک خاصی موجود نیست که برای تمامی این وظایف تعریف شده باشد. در ابتدا هر کدام از این شش وظیفه را به اختصار توضیح می دهیم وسپس به بیان فنون مربوط که می تواند این وظایف را پوشش دهد خواهیم پرداخت.

۱۰- شرح وظایف داده کاوی

۱-۱۰- طبقه بندی *(Classifaction)*

مهمترین وظیفه داده کاوی طبقه بندی هایی است که به نظر می رسد وظیفه ای لازم الا جرا است . به منظور فهم بهترورارتباط راحت تر با دنیا ما همیشه در حال گروه بندی[۷]عناصر هستیم .

در طبقه بندی کردن اصولاً به بررسی خصوصیت یک شئ جدید و مر تبط کردن آن با یک مجموعه از طبقه بندی از پیش تعریف شده می پردازیم . اشیاء طبقه بندی شده بصورت مقداری در داخل پایگاه داده با پر کردن فیلدی با یک کد به نام کد طبقه [۸]معین می شوند. وظیفه طبقه بندی با تعریفی خوب از کلاس هاویک مجموعه آموزشی شامل مثالهایی از قبل تعریف شده می تواند مورد سنجش قرار بگیرد . مثالهایی از طبقه بندی رادر زیر می آوریم:

- طبقه بندی کردن اعتبارات بانکی ۹با اولویت های بالا، متوسط وپائین
- طبقه بندی کردن شماره تلفن هایی که به ماشین فاکس متصل می شوند.

۱- Classificatian
۲- Estimation
۳- Prediction
۴- Affinity Grouping
۵- Clustering
۶- Description
۷-Classifing and categorizing
۸-Class code
۹-Credits

در کلیه مثالهای بیان شده ، تعداد محدودی از کلاسهای تعریف شده موجود است و باید هر مقدار به کلاسی تخصیص داده شود. فنونی چون درخت تصمیم گیری[۱] می باشند . فن تحلیل پیوند[۲] در محدوده های خاصی مناسب این عمل می باشد .

۲-۱۰- تخمین[۳]

عمل طبقه بندی مربوط به نتایج گسسته[۴] است ،درحالی که عمل تخمین مربوط به نتایج پیوسته[۵]می باشد .با دادن مقادیری به عنوان ورودی ، از تخمین استفاده می کنیم تا یک متغیر ناشناس چون درآمد، میزان موجودی و ... را تخمین بزنیم.

مزیت بزرگ تخمین این است که می توان مقدارهایی را براساس میزان نمره شان مرتب کرد.به منظور مشخص شدن این اهمیت فرض کنید یک شرکت سازندۀ پوتین اسکی برای ارسال ۵۰۰/۰۰۰ کفش سرمایه گذاری کرده است . اگر از طبقه بندی استفاده شود و برای مثال ۱/۵۰۰/۰۰۰ نفر اسکی باز متقاضی کفش ، معین شود،یا می توان این روش را به کار گرفت و ۵۰۰/۰۰۰ نفر را بطور تصادفی از بین متقاضیان انتخاب کرد و یا از نمره اسکی هر رکورد استفاده شود ۵۰۰/۰۰۰ تایی را که نمره بیشتری دارند انتخاب نماییم .

مثالهایی ازتخمین را درزیر می آوریم:

- تخمین تعداد اولاد در خانواده
- تخمین کل درآمد یک خانواده
- تخمین ارزش یک مشتری۶

فن شبکه های عصبی[۷]،بسیار برای تخمین مناسب می باشد.

۳-۱۰- پیش بینی[۸]

پیش بینی همانند طبقه بندی و تخمین است با این تفاوت که رکورد براساس یک رفتار پیش بینی شده درآینده یا مقداری تخمینی در آینده ، طبقه بندی می شود. درعمل تنها کاری که میتواند دقت این طبقه

۱-Decision Trees
۲-Link analysis
۳-Estimation
۴-Discrete outcomes
۵-Contineously valued outcomes
۶-Life time value of a customer
۷-Neural networks
۸-Prediction

بندی را معین کند منتظر ماندن و ملاحظه نتایج درآینده می باشد . هر روشی که برای طبقه بندی و تخمین بکار می رود می تواند برای پیش بینی هم استفاده شود مشروط بر استفاده از مثالهایی که درآنها ارزش متغیر قابل پیش بینی در حال حاضر معلوم باشد و همچنین داده های جمع آوری شده برای مثال ها در دسترس باشد. این داده های قدیمی[۱]برای ساخت مدلی که رفتار فعلی قابل مشاهده را بیان میکنند بکار می روند . وقتی به این مدل ورودیهای جدید داده شود ، نتیجه در واقع پیش بینی رفتار درآینده خواهد بود.

فن تحلیل سبد خرید[۲] به منظور یافتن این که کدام اجناس در داخل سوپر مارکت ،بیشتر با هم خریداری می شوند بکار می رود و همچنین می تواند برای ساخت مدلی که معلوم کند چه خرید یا عمل آتی روی داده های فعلی انجام خواهد گرفت نیز استفاده شود.

مثالهایی از عمل پیش بینی را در زیر می آوریم :

- پیش بینی این که تا شش ماه آینده چه مشتریانی همچنان مشتری می مانند.
- پیش بینی این که کدام مشترک تلفنی درخواست مکالمات سه طرفه۳یا نامه صوتی۴ را خواهد داشت.
- فن تحلیل سبد خرید ، استدلال بر مبنای حافظه، درخت تصمیم و شبکه های عصبی همگی برای این نوع پیش بینی مناسب هستند . انتخاب فن مناسب ، بسته به طبیعت داده های ورودی ، نوع چیزی که نیاز به پیش بینی دارد و میزان پیش بینی ، احتمالاً تغییر می کند.

٤-١٠-تحلیل سبدخرید *(Market Basket Analysis or Affyning Grouping)*

عمل وابستگی گروهی در واقع تشخیص این است که چه چیزهای مرتبط با هم بهتر است کنار یکدیگر قرار بگیرند . یک فروشگاه زنجیره ای با استفاده از این روش می تواند به نظم دادن و گروه بندی اجناس در داخل قفسه های خود بپردازد بطوریکه بیشترین جذابیت را برای خریدار ومشتری داشته باشد . این عمل همچنین می تواند برای طراحی بسته بندی های جذاب محصولات کنار هم بکار رود .وابستگی های گروهی یک راه بسیار ساده برای رسیدن به قوانین در برگیرنده داده ها می باشد . در ادامه به تفصیل دربارهٔ این روش توضیح داده خواهد شد.

۱-Historical data
۲-Market Basket Analysis
۳-Three way calling
٤-Voice mail

۱۰-۵- خوشه بندی (Clustering)

خوشه بندی در واقع تقسیم بندی یک جمعیت ناهمگون[۱]به تعدادی از زیر مجموعه هایی که بیشتر همگون[۲]هستند می باشد که به آن خوشه [۳] اطلاق می شود.

وجه متمایز کننده خوشه بندی از طبقه بندی[۴] این است که خوشه بندی تکیه اش روی طبقات از قبل تعریف شده نمی باشد . در واقع در خوشه بندی طبقات از پیش تعریف شده نداریم بلکه مقادیر براساس شباهت ذاتی[۵]خودشان با یکدیگر دریک گروه قرار می گیرند .

۱۰-٦- توصیف (Description)

در بعضی مواقع ، هدف داده کاوی این است که بتوانیم براحتی از وقایع پیچیده ای که در داخل پایگاه داده داریم شرح و تفصیلی داشته باشیم ، بطوریکه میزان برداشت ما را افزایش دهد. یک توصیف خوب اصولاً یک تفسیر خوب را هم به همراه خواهد داشت. برخی از فنون داده کاوی چون تحلیل سبد خرید کاملاً جنبه توصیفی دارند.

۱۱- فنون داده کاوی

- -- تحلیل سبد خرید (Market basket analysis)
- -- استدلال براساس حافظه (Memory based reasoning (MBR))
- -- تشخیص خودکار خوشه ها (Automatic cluster detection)
- -- تحلیل پیوند ها (Link analysis)
- -- درخت تصمیم (Decision trees)
- -- شبکه های عصبی مصنوعی (Artificial neural networks)
- -- الگوریتم های ژنتیکی (Genetic Algorithms (GA))

در زیر به تفصیل، هر یک از این فنون توضیح داده خواهد شد .

۱۱-۱- تحلیل سبد خرید (Market basket analysis)

همانطور که قبلاً بیان شد ،تحلیل سبد خرید یک فرم از خوشه بندی است که برای شناخت گروهی از اجناس که در معاملات باهم اتفاق می افتند ، بکار برده می شود . برای مثال ، خرید توأم اجناس سوپر

۱-Heterogeneous population
۲-Homogeneous
۳- Cluster
۴- Classification
۵-Self - similarities

مارکت ها و نحوهٔ طبقه بندی قفسه ها را می توان نام برد که مثالی عملی برای این روش است . در واقع تنها اطلاعات اجناسی که با هم خریداری می شوند موجود است و سایر اطلاعات چون آمارگیری و تاریخچه مشتریان موجود نمی باشد . چرا که معاملات بدون نام مشتریان صورت می گیرد. . نتایج بدست آمده از این روش قابل پی گیری می باشد . اطلاعات بدست آمده می تواند برای مقاصد بسیاری استفاده شود . شامل مواردی از قبیل : ترتیب قرار گرفتن اجناس ، محدودیت قیمت گذاری محصولی خاص ، فروش کلی محصولات ، ارائه کوپن هایی که با جلب توجه مشتریان فروش بهتری برای محصولات خاصی داشته باشند. فرآیند اساسی این روش را می توان در موارد زیر خلاصه کرد:

- انتخاب کردن مجموعه ای درست از اجناس
- تولید قوانین با کشف نمودن تعداد در ماتریس رخداد توأم[۱]
- نتیجه گیری از محدودیت های عملی ایجاد شده توسط هزاران تا دهها هزار محصولی که در کنار هم قرار می گیرند تا حداکثر جذابیت برای مشتری را داشته باشد.

نقاط قوت این روش:

- نتایج حاصله کاملاً واضح و قابل فهم هستند
- این روش ،داده کاوی بدون جهت[۲]را پشتیبانی می کند
- این فن روی داده های با طول متغیر نیز کار می کند
- محاسبات استفاده شده در آن ساده و قابل فهم هستند

نقاط ضعف این روش:

- پیچیدگی این روش در مواقعی که اندازه مسئله بزرگ باشد بصورت نمائی افزایش می یابد و به کار محاسباتی خیلی زیاد نیاز دارد
- این روش پشتیبانی محدودی روی ویژگیهای[۳] داده ها دارد
- بسیار مشکل است که تعداد درست طبقات محصولات را بدست آوریم
- این روش محصولات اندکی را بصورت تخفیف داده شده معین می کند

---

[۱]-Co – occurance matrix
[۲]-Undirected data mining
[۳]-Attributes

موارد استفاده

روش تحلیل سبد خرید در مسائل داده کاوی بدون جهت قابل استفاده است که شامل تعریف درستی از محصولات صحیحی که با هم در یک طبقه قرار می گیرند می باشد . این مسائل در صنعت خرده فروشی بسیار به چشم می خورد . جایی که نقطه انجام معامله فروش [۱]، پایه و اساسی برای تحلیل و بررسی است . مسائلی مشابه در سایر صنایع نیز به چشم می خورد.

این روش همچنین می تواند روی برخی از مسائل داده کاوی جهت دار[۲]نیز اعمال شود. الگوریتم های اساسی این روش قابل تطبیق می باشد . برای مثال ، جهت یافت الگویی قابل فهم از فروش محصولی جدید می توان از این روش مدد جست.

مسائل سریهای زمانی[۳] محدودهٔ دیگری است که این روش جهت حل آنها قابل استفاده است . البته ذکر این نکته ضروری است که تبدیل نسبتاً ساده ای روی داده های سری زمانی باید صورت بگیرد تا قابل استفاده با این روش برای حل مسئله شوند.

۱۱-۲- استدلال براساس حافظه[٤](Memory Based Reasoning)

ما بر اساس تجارب گذشته مان قادر به تصمیم گیری هستیم . زمانی که شخص چهرهٔ آشنایی را درمیان جمعیت می بیند به دنبال شباهتی که این چهره با چهره آشنایی که از قبل می شناسد می گردد. اوّلین قدم دراین روش تشخیص موارد مشابه از روی تجارب کسب شده است . سپس از نتایج حاصله اطلاعاتی به مساله در مورد نظر داده می شود و این در واقع همان چیزی است که روش استدلال بر مبنای حافظه از آن استفاده می کند . بوسیلهٔ یک پایگاه داده از مقادیر شناخته شده ، این روش به دنبال همسایگان مشابه برای مقدار جدید می گردد و از این همسایگان جهت طبقه بندی و پیش بینی کمک می گیرد .

استدلال بر مبنای حافظه یک فن داده کاوی جهت دار می باشد که در واقع مدلی از چیزهای شناخته شده را جهت پیش بینی چیزهای ناشناس[۵]بکار می گیرد . این روش به دنبال نزدیکترین همسایه ها در فاصله شناخته شده می گردد و به ادغام ارزش آنها تا بدست آوردن طبقه بندی یا پیش بینی می پردازد.

---

[۱]-Point of sale transaction
[۲]-Directed data mining
[۳]-Time series problems
[٤]-MBR
[۵]- Unknown instances

یکی از محاسن مهم روش استدلال بر مبنای حافظه، قدرت آن است که میتواند روی هر منبع مجازی از داده ها عمل کند ،حتی بدون این که تغییری روی آن بدهد. دو عامل اساسی در این روش وجود دارد که یکی *تابع فاصله*[۱]، جهت بدست آوردن نزدیکترین همسایه ها و دیگری *تابع ادغام*[۲] است که عمل ادغام ارزش همسایه ها را برای بدست آوردن یک پیش بینی انجام می دهد . در مواردی ، این توابع می توانند بصورت جملات زبان پرس و جوی ساخت یافته روی پایگاه دادهٔ رابطه ای مستقیماً اجرا شده و برای اهداف داده کاوی بکار روند . البته استفاده از پایگاه داده های رابطه ای امروزه کارآیی چندان بالایی ندارد . در موارد دیگر این روش می تواند بر اساس انواع پیچیده تر داده ها چون متن ، تصویر و ... بکار گرفته شود که البته توابع فاصله و ادغام دراین دامنه ها باید معلوم باشد .

حسن دیگر این روش قدرت یادگیری دربارهٔ طبقه بندی های جدید فقط با معرفی مثالهای جدید به داخل پایگاه داده است . همانطور که بطور مختصر گفته شده این روش روی قالب و مقیاس داده های ورودی حساس نمی باشد و تنها وجود توابع فاصله و ادغام ضروری است . این توابع برای انواع استاندارد داده بصورت آماده موجود می باشد این روش همچنین روی داده های پیچیده ای چون موقعیت جغرافیایی و تصاویر و... نیز استفاده می شود که با سایر فنون داده کاوی کار با این نوع داده بسیار مشکل است.

نقاط قوت این روش:

- نتایج حاصله از این روش آماده و قابل فهم هستند
- قابل اعمال روی هر نوع داده اختیاری حتی داده های غیر رابطه ای نیز می باشد
- بطور مکفی روی تقریباً هر تعدادی از فلید کار میکند
- برای بدست آوردن یک مجموعه آموزشی۳، نیازمند تلاش اندکی است

نقاط ضعف این روش:

- از لحاظ محاسباتی ، انجام طبقه بندی و محاسبه ، پیش بینی هزینه بسیار بالایی دارد
- جهت مجموعه های آموزشی نیازمند فضای بسیار زیادی جهت ذخیره سازی می باشد
- چنانچه روی توابع فاصله و ادغام و تعداد همسایگان انتخاب داشته باشیم ، نتایج می تواند مستقل از هم بدست آید

۱- Distance function
۲- Combination function
۳- Trainig set

موارد استفاده :

استدلال بر مبنای حافظه یک فن داده کاوی جهت دار است که هم برای طبقه بندی و هم برای پیش بینی مناسب است . در مقایسه با سایر فنون زمانی که الگوی داده ها بصورت کاملاً محلی معلوم باشد بسیار خوب عمل می کند . بطور خلاصه این روش هنگامی قدرتمند است که از اطلاعات محلی برای طبقه بندی و پیش بینی استفاده می شود .

۱۱-۳- تشخیص خودکار خوشه ها (*Automatic Cluster detection*)

گاهی اوقات از ما خواسته می شود که به یک تصویر بسیار بزرگ نگاه کنیم اما این کار در موارد بسیاری جهت فهم دقیق این امر بسیار مبهم است . یک پایگاه داده وسیع ممکن است شامل تعداد بسیار زیادی متغیر باشد ، ابعاد بسیار بالایی داشته باشد و ساختار پیچیده ای دارا باشد که حتی بهترین فنون داده کاوی با بهترین میزان جهت دهی هم قادر به استخراج الگوهای با معنی از درون آنها نباشند . در بیشتر موارد مشکل این نیست که الگویی موجود نمی باشد بلکه مشکل این است که الگوها بسیار زیاد هستند .

وقتی انسان با سوالات پیچیده مواجه شد ، یاد گرفت که آنها را به سوالات کوچک تری بشکند بطوریکه هر قسمت راحت تربیان شود . برای مثال ، اگر از شما سوال شود که رنگ درختان جنگل را توصیف کنید جواب شما احتمالاً شامل ارتباط بین رنگ برگ انواع درختان جنگلی ، نوع فصل ، نوع زمین و ... می باشد . شما برای پاسخگویی ، اطلاعات کافی از جنگل دارید و صدها متغیری که مربوط به جنگل است مد نظر شماست ، در واقع ، فصل گونه درخت ،میزان اسیدیته خاک و ... چیزهایی هستند که برای تشکیل خوشه های درختان که رنگ مشابه دارند بکار می رود. اگر چه در بسیاری موارد این طور به نظر می رسد که در مجموعهٔ داده های با اختلالات بالا تعداد بیشتری خوشه موجود است ، اما ما هیچ ایده و خط مشی نداریم که چطور آنها را تعریف کنیم . اینجاست که فنونی چون تشخیص خودکار خوشه ها مطرح می شود . روش تشخیص خوشه در واقع ساخت مدل هایی است که مقادیر داده ها یی که مشابه هم هستند را پیدا کند. این مجموعه از شباهات ذاتی را خوشه می نامند . این روش یک روش داده کاوی بدون جهت است . چرا که در واقع هدف ، یافتن شباهت های شناخته نشده قبلی در داده ها می باشد . فنون بسیاری از قبیل روش های هندسی ، روش های آماری و شبکه های عصبی روشهای مناسبی برای شروع تحلیل خوشه بندی می باشد .

نقاط قوت روش:

- تشخیص اتو ماتیک خوشه در واقع یک فن اکتشاف دانش بدون جهت می باشد

- این روش روی داده های طبقه بندی شده ، عددی و متنی بخوبی کار می کند
- برای استفاده بسیار ساده می باشد

نقاط ضعف روش:

- مشکل است که میزان فاصله مناسب و وزن ها را به درستی تعیین کرد
- حساسیت این روش به پارامتر های ابتدایی زیاد است
- تفسیر خوشه های حاصل شده ، چندان ساده نمی باشد

موارد استفاده :

زمانی که با مجموعه بسیار بزرگ و پیچیده و با تعداد بسیار زیاد متغیر و ساختار های داخلی پیچیده از داده ها مواجهیم خوشه بندی ابزار بسیار منتسبی است . در ابتدای یک پروژه داده کاوی ، خوشه بندی تقریباً بهترین فن برای شروع است . بهر حال به ندرت به عنوان تنها ابزار شناخته می شود . زمانی که تشخیص خودکار خوشه ها در محدوده های فضای داده ای که شامل رکوردهای مشابه هستند به اتمام رسید ، سایر فنون داده کاوی شانس بهتری جهت کشف قوانین و الگوهای بین آنها خواهند داشت .

۱۱-٤- تحلیل پیوند (*Link analysis*)

تحلیل پیوند ، در پی گسترش مدل بر پایه الگوهای موجود ، ارتباط بین مقادیر را دنبال می کند این روش در واقع یک کاربرد از نظریه گراف می باشد که بر داده کاوی بنا شده است . امروزه ارتباط بین مشتریان از اهمیت خاصی برخوردار شده است و بیشتر مورد اهمیت قرار گرفته است زمانی که فروشندگان کالا بیشتر روی مشتریان حساب باز کرده اند . یک محدودهٔ کاربرد این روش ، موضوع ارتباطات است . هر تماس تلفنی ارتباط بین یک مشتری را با مشتری دیگر برقرار می کند این اطلاعات می تواند پایهٔ موفقیت تجاری شرکت ها قرار بگیرد . بعنوان یک ابزار داده کاوی تحلیل پیوند در پایگاه داده های رابطه ای با ناکار آمدی مواجه می شود . مهمترین محدوده ای که این روش کاربرد بالایی دارد محدوده قضایی است که ارتباط و رشته های ارتباطی بین جنایات را باهم مربوط می کند تا به یافتن عامل جنایت منتهی شود. در این رابطه چندین ابزار در اختیار می باشد که بیشتر روی بصری سازی ارتباطات کار می کنند تاروی تجزیه و تحلیل الگوها . بهر حال ، پرسش و پاسخ های زبان پرس و جوی ساخت یافته وی پایگاه داده های رابطه ای ، نیز می تواند پایه ای برای این روش باشد . تنها مورد مشکل ساز در این روش این است که هزینه پرسش و پاسخ های تحلیل ارتباطات اصولاً بالا می باشد چرا که در اینجا پیوند ها معادل جفت ها در مدل رابطه ای هستند . برای مقادیر کوچک داده فنآوری

مبتنی بر شی [۱]اغلب باعث کپسوله کردن ارتباطات در داخل پایگاه داده شده و راه مناسبی را جهت کار با آنها فراهم می کند.

مشکل دیگر این روش این است که کجا هابین عناصر پیوند برقرار گردد. در مورد شبکه تلفن و ارتباطات این امر چندان مشکل نیست و پیوند ها واضح هستند .در مواردی که این ارتباطات بطور اجباری به طور خودکار ایجاد شوند. سیستم تحلیل بسیار قوی که در اف بی آی استفاده می شود ارتباطات بین عناصر را از طریق تکنیکی مشابه تحلیل بر اساس حافظه بدست می آورد.

نقاط قوت این روش

- بر مبنای ارتباطات سرمایه گذاری می کند
- برای مقوله بصری سازی مناسب است
- ویژگی های درونی داده ها را ظاهر می سازد

نقاط ضعف این روش:

- برای بسیاری از انواع داده قابل استفاده نمی باشد
- تعداد اندکی از ابزارها چنین روشی را پشتیبانی می نماید
- اجرای این روش روی پایگاه های داده رابطه ای کارآیی بالایی ندارد

موارد استفاده :

تحلیل پیوند ها و ارتباطات ، ابزاری برای اکتشاف دانش می باشد که برای حل محدوده خاصی از مسائل کاربرد دارد . در موارد کاربرد ، قادر است الگوی بین داده ها را کشف کند که سایر فنون قادر به کشف آن نیستند . با وجود این قدرتی که داراست ، محدودیتهایی را نیز دارد از جمله این که روی بیشتر انواع داده قابل اعمال نمی باشد . بهر حال ، زمانی که نتایج بدست آمد ، به عنوان مشخصات برای سایر فنون از جمله شبکه های عصبی و درخت تصمیم گیری ، قابل استفاده هستند .

۱۱-۵- درخت تصمیم گیری[۲]

درختهای تصمیم گیری ابزار عمومی و مورد استفاده طبقه بندی و پیش بینی می باشند ، جذابیت روش هایی که مبنایشان بردرخت تصمیم گیری است نسبت به شبکه های عصبی این است که اینها قادرند قوانین رانشان بدهند. قوانین را میتوان به شکل زبان انگلیسی یا بهر نحو که قابل فهم باشد ملاحظه نمود.

---

[۱]-Object Oriented

[۲]-Decision trees

الگوریتم های متنوعی برای ساخت درخت تصمیم گیری موجود است .دو الگوریتم که بیشتر مورد استفاده قرار می گیرند عبارتند از :

CART[۱] و دیگر *CHAID*[۲] لازم به ذکر است که الگوریتم جدیدتری بنام **C٤,٥**در حال همگانی شدن است .

هر کسی که بازی بیست سوالی را تا به حال امتحان کرده باشد ، مشکلی در فهم کارکرد درخت تصمیم نخواهد داشت . درخت تصمیم نیز یک سری از سولات را نمایش می دهد. چنانچه سوالات به خوبی انتخاب شوند یک مسیر بسیار کوتاه جهت طبقه بندی مقادیر فعلی طی خواهد شد . مقادیر از ریشه وارد درخت می شود، در ریشه آزمون انجام می شود ویکی از نوه های فرزند به عنوان مسیر بعدی انتخاب می شود . تمام مقادیری که به یک برگ درخت میرسند، یکسان طبقه بندی می شوند . تنها یک راه یکسان از ریشه به هر برگ وجود دارد این مسیر درواقع بیانی از یک قانون برای طبقه بندی مقادیر می باشد.

بسیاری از برگهای متفاوت ممکن است طبقه بندی های مشابهی را بسازند ولی هر یک از آنها این طبقه بندی را به دلیل خاص خود تشکیل داده اند .برای مثال ، در درختی که رنگ میوه و سبزیجات را معین می کند ، برگهای سیب، گوجه فرنگی، و گیلاس همگی پیش بینی رنگ قرمز را برای میوه می کنند درحالی که به ندرت ممکن است گوجه فرنگی سبز یا گیلاس سیاه هم داشته باشیم.

نقاط قوت روش:

- قادر به ایجاد قوانین قابل فهم است
- درخت تصمیم قادر است طبقه بندی را بدون محاسبات زیاد انجام بدهد
- درخت تصمیم قادراست هم متغیر های پیوسته و هم متغیر های قابل دسته بندی را پوشش دهد
- درخت تصمیم قادراست یک برآورده واضحی از این که کدام فیلدها در طبقه بندی وپیش بینی مهم ترند، داشته باشد

نقاط ضعف روش:

- درخت تصمیم برای وظایف تخمینی مناسب نمی باشد از طرفی درخت تصمیم برای داده های سری زمانی مشکل ساز است مگر این که سعی بسیاری جهت ارائه داده به روش دیگر صورت گیرد .

---

[۱]-Classification And Registration Trees
[۲]-Chi-square Automatic interaction Detection

موارد استفاده :

درخت تصمیم در مواردی مناسب است که وظیفه داده کاوی دسته بندی مقادیر یا پیش بینی نتایج باشد. درخت تصمیم از طرفی بسیار سودمند خواهد بود چنانچه بدنبال بدست آوردن قوانین قابل فهم به زبان پرس و جوی ساخت یافته یا هر زبان طبیعی دیگر باشیم.

۱۱-٦- شبکه های عصبی مصنوعی[۱]

این روش در بسیاری از مسائل چون پیش بینی هزینه مالی جهت شرایط پزشکی خاص ، تشخیص دادن خوشه ها از مشتریان با اهمیت ، تشخیص نقل و انتقالات مالی کلاه برداران از طریق جعل کارت های اعتباری ، تشخیص پیش بینی نرخ خرابی موتورها و... کاربرد دارد.

کاری که روش شبکه عصبی می کند این است که پلی رابین تجارت بشری و ساختارهای صریح کامپیوتری بنا می کند تا این فاصله را بپوشاند ، و این کار را از طریق مدل سازی سلولهای عصبی انسان روی کامپیوتر های دیجیتال انجام می دهد. چنانچه بخوبی دامنه های آن تعریف شود قدرت آنها برای بدست آوردن ویادگیری از روی داده ها کاری مشابه قدرت ماناز طریق یادگیری از تجاربمان خواهد بود. این توان باعث شده که شبکه های عصبی در مقوله داده کاوی بصورت یک موضوع جالب توجه برای محققین درآید و نتایج امیدوار کننده ای رابرای فردایی نه چندان دور بهمراه آورد.

مشکلی که این فن دارد این است که نتایج گرفته شده از شبکه های عصبی به وزنی متکی است که علم شبکه نیز براین اساس است. این وزن باعث می شود نتوان گفت که چرا این راه حل معتبر است همانطور که محققین نمی توانند بگویند چرا یک تصمیم خاص بشری صحیح است. این وزنها براحتی قابل فهم نیستند و بطور گسترده فنون خبره و مصنوعی برای اکتشاف داخل شبکه های عصبی بکار گرفته می شوند تا تفسیری را از عملکرد آن داشته باشند . شبکه های عصبی در مورد مسائل مبهمی که همچون یک جعبه سیاه بوده که در داخل آن کارهای مرموزانه ای صورت می گیرد بسیار مورد استفاده دارد. جوابهای حاصله از این روش اصولاً هنگامی صحیح هستند که این نتایج در بسیاری از موارد ارزش به مراتب بالاتری تا شرح و تفسیر آنها داراست .

دراین روش یادگیری از طریق مجموعه های آموزشی انجام می شود و الگوهای درونی از آنها ایجاد شده وطبقه بندی وپیش بینی بر این اساس صورت می گیرد .این روش روی داده کاوی بدون جهت و پیش بینی های سری زمانی نیز کاربرد دارد. امروزه کاربردهای زیادی برای این روش درحال بررسی است و هر ماهه در کنفرانس های متعدد چالشهای جدید دراین مورد به بحث گذاشته می شود. مهمترین

[۱] Artificial neural networks

مزیت شبکه های عصبی ، گستردگی پوشش برای حل مسائل می باشد . ابزارهای متنوعی در بازار وجود دارد که شبکه عصبی را پوشش می دهد و در محیطهای متنوعی قابل استفاده است .از طرفی ، شبکه ها بدین خاطر جذابیت دارند که الگوهای داده ای مشابهی همانند نحوه فکر کردن انسان ایجاد می کنند که این خود در واقع پایه ابزار داده کاوی می باشد.

دو عیب عمده برای این روش گفته می شود اولاً که فهم مدل ایجاد شده چندان ساده نمی باشد ثانیاً نتایج حاصله شدیداً تحت تاثیر داده های ورودی است و نسبت به اینها حساس است به گونه ای که دسته های مختلفی از داده ها ، در نتایج تنوع ایجاد می نماید . بنابراین تشخیص نوع دادهٔ مناسب برای ورودی شبکه مهمترین بخش دراستفاده صحیح از این فن است.

نقاط قوت روش:

- دامنه وسیعی از مسائل متفاوت از این طریق قابل حل است
- حتی دردامنه های پیچیده نتایج خوبی را تولید می کند
- هر دوی متغیرهای پیوسته وقابل طبقه بندی[۱]را پوشش می دهد
- در بسته های نرم افزاری متعددی در بازار عرضه شده است

نقاط ضعف این روش :

- نیازمند گرفتن داده های ورودی در محدوده ۰و۱ می باشد
- شبکه ها نمی توانند شبکه حاصله را توضیح دهند
- ممکن است شبکه بطور نا بهنگام به سمت جواب نا مرغوب میل کند

موارد استفاده :

شبکه های عصبی انتخاب بسیار مناسبی برای مواردی که مسئله مورد نظر ، طبقه بندی و پیش بینی باشد و از طرف دیگر نتایج مدل برای فهم این که مدل چطور دارد کار میکند ، بسیار مهمتر باشد انتخاب مناسبی است . از آنجایی که این روش غیر شفاف و مبهم است ، استخراج قوانین از درون آنها امکان پذیر نیست . همچنین ، شبکه ها برای وظایف داده کاوی بدون جهت چون خوشه بندی نیز کاربرد دارد. دراین حالت شبکه ، خوشه هایی از مقادیر ورودی را که شبیه یکدیگر ند ایجاد می کند ولی نوع شباهت و این که چطور این ها شبیه هستند را مشخص نمی کند ، تنها جایی که این روش درست کار نمی کند وقتی است که تعداد بسیار متنوعی از ویژگیهای ورودی داریم واین خود منجر به این می شود که یافت الگو درداخل داده ها بسیار مشکل شود و در این موارد هرگز به جوابی خوب نخواهیم رسید.

---

[۱]-Categorical variables

این فن به خوبی با درخت تصمیم کار میکند . درخت تصمیم برای یافتن متغیرهای مهم بسیار مناسب است واین به نوبه خود می تواند برای آموزش یک شبکه مورد استفاده قرار گیرد .

۱۱-۷- الگوریتم های ژنتیکی

الگوریتم های ژنتیکی[۱] ، ساز و کار ژنتیک و انتخاب طبیعی را دارا می باشد که در جستجو برای یافتن بهترین مجموعه از پارامتر هایی که یک تابع پیش بینی نیاز دارد را به مدد می گیرد . این روش برای داده کاوی جهت دار مورد استفاده قرار می گیرد الگوریتم ژنتیکی به دلیل نیاز به شناخت کامل مدل ، مشابه علم آمار می باشد . دو روش تحلیل بر اساس حافظه و شبکه های عصبی مبنای قیاسی با فرآیندهای بیولوژیکی دارند . این فن این ایده کلی برای مسائل را دارد که راه حل به عنوان یک چیز منحصر به فرد[۲] بیان می شود و مسئله باید حداکثر همخوانی [۳]را با این راه حل فراهم کند . برای مثال کاربرد الگوریتم های ژنتیک می توانند در آموزش شبکه های عصبی مورد استفاده قرار گیرد . چیز خاص و منحصر به فرد می تواند مجموعهٔ وزنهای داخل شبکه باشد وچیزی که باید درآن جا بیفتد در واقع قدرت پیش بینی شبکه های عصبی بر اساس وزنهای موجود است مانند یک سیر تکاملی هر چه جفت شدن افراد جامعه بیشتر باشد باعث انتشار و تکاثر بیشتر آنها شده و نسلی برتر حاصل خواهد شد . اگر چه شانس ، نقش مهمی را در بقای هر موجودی دارد ، ولی در یک جمعیت زیادتر قانون تعداد بیشتر چیره شده و انتخاب های طبیعی مناسب جهت تکاثر منجر به ایجاد افرادی می شود که بیشترین میزان مناسبت را دارا هستند .

در سالهای اخیر الگوریتهای ژنتیکی در سه محدوده آزمایش شده و نتایج خوب و قابل ملاحظه ای نیز داشته اند . اینها عبارتند از : آموزش شبکه های عصبی ، ایجاد تابع ارزش دهی برای روش استدلال برمبنای حافظه و به عنوان موتورداخلی بهینه سازی در نرم افزار برنامه ریزی . البته این فنون درآموزش شبکه های عصبی کاربرد بسیار وسیع تری نسبت به سایرین دارد . بسیاری از بسته های شبکه های عصبی امروزه از این روش برای آموزش استفاده می کنند.

در دنیای داده کاوی و تحلیل داده استفاده از الگوریتم های ژنتیکی مانند سایر فنون گسترده نمی باشد . داده کاوی استفاده از این فنون عمدتاً در وظایفی چون طبقه بندی و پیش بینی صحه می گذارد تا بهینه سازی که در الگوریتمهای ژنتیکی موجود است اگر چه که به بسیاری از مسایل داده کاوی می توان بصورت مسائل بهینه سازی نیز نگاه کرد .

(GA) [۱]Genetic Algorithms
[۲]-Individual
[۳]-Fitness

الگوریتم های ژنتیکی به جای اینکه روی پارامترها یا متغیر های مسئله کار کند با قالب کد شده آنها سرو کار دارد که متداول ترین روش کد گذاری استفاده ازاعداد دو جمله ای می باشد . مفاهیم کورموزم ، جمعیت ، میزان برآن زندگی ، عملگرهای متقاطع و جهش از مقولاتی است که درالگوریتم های ژنتیکی موجود است .

نرم افزارهایی که اخیراً عرضه شده اند ،الگوریتم های ژنتیکی را نیز در بر میگیرند ، به هر حال الگوریتم های ژنتیکی مقوله گسترده ای است که هنوز تحقیقات وسیعی روی آن در حال انجام است . این طور برآورد کرده اند که الگوریتم های ژنتیکی سهم بسیار بسزایی در آینده داده کاوی و بهینه سازی روشهای آن خواهند داشت .

نقاط قوت این روش :

- نتایج تولید قابل توضیح هستند
- خیلی ساده می توان نتایج را بدست آورد
- قادر است دامنه بسیار وسیعی از داده ها را پوشش دهد
- قابل استفاده جهت بهینه سازی می باشد
- به خوبی با شبکه های عصبی قابل یکی شدن است

نقاط ضعف این روش :

- کد گذاری بسیاری از مسائل با این روش مشکل است
- تضمین برای بهینه سازی ندارد
- از لحاظ محاسباتی ، هزینه بسیار بالایی دارد
- در تعداد محدودی بسته نرم افزاری موجود است

۱۲- مختصری دربارهٔ پردازش تحلیلی بر خط

در چندین سال اخیر ابزارهای پردازش تحلیلی بر خط به عنوان ابزار پاسخگویی به سوالات مطروحه از پایگاه های داده بسیار بزرگ که یا داده درون یک انباره داده مرکزی [۱] یا انباره مجازی توزیعی [۲] و یا روی سیستم عامل بوده بکار می رود . این ابهام همچنین درباره پردازش تحلیلی بر خط بر جاست که برخی آنرا جایگزینی برای داده کاوی می دانند .پردازش تحلیلی بر خط یک ابزار سریع ، مناسب و قدرتمند است که برای گزارش گیری روی داده بکار می رود در حالی که ابزارهای داده کاوی روی

[۱]-Centralized data warehouse
[۲]-Virtual distributed warehouse

یافتن الگوهای درون داده ای است . پردازش تحلیلی بر خط و داده کاوی رهیافت هایی هستند که می توانند با یکدیگر کامل شده و هر یک سهم مهمی را در استخراج و بهره برداری از داده داشته باشند پردازش تحلیلی بر خط ابزاری برای ارائه است که به استفاده کننده امکان کشف اطلاعات را به صورت دستی [۱]می دهد . اساساً بسته به ذکاوت و قابلیت استفاده کننده دارد که تا چه حد این کار را به خوبی انجام دهد . اگر چه پردازش تحلیلی بر خط جزیی از داده کاوی محسوب نمی شود اما قسمتی از راه حلی است که با داده در محیط تجاری به کار برده می شود .داده کاوی می تواند از مزایای تحلیلی که از اجرای یک راه حل پردازش تحلیل بر خط ارائه می شود ، استفاده کند .

دنیای تجارت به طور مرتب در حال تهیه گزارشهای گوناگون در دهه های گذشته بوده و هست . قدیمی ترین روشی دستی که جهت تهیه گزارش استفاده می شده ، ابزارهای تولید گزارش از رایانه های بزرگ بوده است . پس از آنها نرم افزارها های تولید کننده پرس و جو های خارج از سیستم [۲] بود که به خاطر دسترسی شان به داده در دهۀ گذشته مورد استفاده گسترده ای داشتند . امروزه پردازش تحلیلی بر خط جایگزین آنها شده است که برای استفاده کننده نهایی ، دسترسی به داده ها را فراهم می کند . ابزاری سرویس دهنده – سرویس گیرنده [۳]می باشد که با وجود واسط کاربر گرافیکی قوی در آن ارائه داده ها به صورت نمودار سه بعدی وجود دارد . این نمودارهای سه بعدی یا در داخل پایگاه داده های رابطه ای با استفاده از طرح ستاره ، و یا داخل پایگاه داده های خاص چند بعدی [۴] برای بهینه سازی عملیات پردازش بر خط ، ذخیره می شوند . ابزارهای پرداز تحلیل بر خط پاسخگویی [۵] بسیار سریعی دارند که در حد ثانیه می باشد .پرس و جوهای روی پایگاه داده های رابطه ای چیزی در حد ساعت ها و یا روزها برای تحلیل اطلاعات مشابه نیاز دارند ، تا اطلاعات مشابهی را بدست آوردند ، به علاوه ابزارهای پردازش تحلیلی بر خط توابع تحلیل کاربردی را فراهم می کنند و این عمل کاری بسیار سخت و یا نشدنی در زبان پرس و جوی ساخت یافته می باشد . ایجاد این مکعب ها نیازمند تحلیل صحیح داده ها و همچنین استفاده از یک متخصص آشنا با داده و ابزار مورد نیاز می باشد . اگر چه که طراحی و بار کردن مکعب ها در ابتدای کار ممکن است روزها و هفته ها به طول بیانجامد ، لیکن در

---

[۱]-Manual
[۲]-Off–the–shell query
[۳]- Client-Server
[۴]-Special multi-Dimentional-Data bases
[۵]-Response time

نهایت دسترسی سریع و آموزنده برای استفاده کننده نهایی فراهم می کند که این امر بسیار مفید تر از نتایجی است که از ابزارهای تولید پرس و جو بدست می آید .

۱۳- رابطه پردازش تحلیلی بر خط و داده کاوی

داده کاوی روشی موفق در بهره برداری از داده ها در راستای اهداف پشتیبانی تصمیم [۱] می باشد . ملاحظه شد که موفقیت بیشتر این روش در گرو بهبود الگوریتم های تشخیص الگو ، نسبت به حالت فعلی می باشد . در راه حل های گسترده و بزرگ ،پردازش تحلیلی بر خط نقش واضح و مهمی را در دسترسی به داده ها و ملاحظه نتایج از جهات گوناگون فراهم می آورد . جایی که نیازمند به تصویر کشیدن بهترهستیم با استفاده کردن از پردازش تحلیلی بر خط روی همان داده ها با اطلاعاتی درباره خوشه ها که به عنوان بعد مکعب از آنها استفاده می شود می توانیم عامل تفکیک خوشه ها را تشخیص بدهیم . داده کاوی و پردازش تحلیلی بر خط مکمل یکدیگر هستند و می توان از پردازش تحلیلی بر خط در شبکه های عصبی استفاده نمود . داده کاوی با تعریف ابعاد مناسب برای مکعب ها و همچنین با تعیین چگونگی شکستن مقادیر یا ابعاد پیوسته ، کمک به ساخت مکعب های بهتری می کند ، از طرفی به واسطه خاصیت بصری سازی که دارا ست فهم بهتری از نتایج داده کاوی چون خوشه بندی و شبکه های عصبی را در پی خواهد داشت . در واقع قدرت یکدیگر را در زمینه بهره برداری از داده ها افزایش می دهند .

نقاط قوت پردازش تحلیلی بر خط

- یک ابزار بسیار قوی برای بصری سازی است
- زمان پاسخ بسیار سریع به صورت تعاملی است
- برای تحلیل سریهای زمانی مناسب است
- برای یافتن خوشه ها می تواند مورد استفاده قرار گیرد
- بسیاری از استفاده کنندگان ، ابزارهای موجود پردازش تحلیلی بر خط را ترجیح می دهند

نقاط ضعف

- شکل دادن مکعب ها می تواند کاری مشکل باشد
- روی متغیرهای پیوسته به خوبی کار نمی کند
- مکعب ها می توانند سریع خارج از رده محسوب شوند
- داده کاوی محسوب نمی شود

[۱]-Decision-Support Purposes

۱٤- ابزارهای داده کاوی

از ابزارهای داده کاوی می توان موارد زیر را نام برد :

۱-Microsoft OLAP analysis
۲-Cognos –PowerPlay
۳- Microstrategy
٤- DBMINER
٥- Macola
٦-Knowledge seeker
۷-DataEngine
۸- NeuNet Pro
۹- VisiRex

۱٥- خلاصه و نتیجه گیری

داده کاوی اکتشاف اطلاعات قابل پیش بینی مخفی شده در داخل پایگاه داده های عظیم است ، که فنآوری جدیدی بوده و از پتانسیل بسیار بالایی جهت کمک به شرکتها برخوردار است و این امکان را برایشان فراهم می کند که به اطلاعات بسیار مهمی که داخل انباره داده شان موجود است دسترسی پیدا کنند . ابزار های داده کاوی می توانند رفتارهای آتی سیستم ها را پیش بینی کنند و قدرت تصمیم گیری را بالا ببرند . قدرت تجزیه و تحلیل داده کاوی بسیار قوی تر از ابزارهای سیستم های پشتیبانی تصمیم است . این ابزارها می توانند سوالاتی را که اصولاً برای پاسخگویی بسیار وقت گیر بودند را در زمان اندکی پاسخ گو باشند . این ابزارها در پایگاه داده ها به دنبال الگوها می گردند ، تا اطلاعات قابل پیش بینی غیر قابل تشخیص متخصصین را استخراج کنند.

دراین زمینه تحقیقات بیشتری نیاز است که تا محیطی اکتشافی مبتنی بر ایجاد قیود[۱] حاصل شود . همچنین نیاز است که بدانیم چطور این اکتشاف چند بعدی و مبتنی بر قیود را به سایر اطلاعات چون طبقه بندی ، خوشه بندی و . . . اعمال کنیم . اینها هنوز مقولاتی است که در دستور کاری محققین وجود دارد و موضوع های مطرح تحقیقات آتی را تشکیل می دهد.

مطرح شدن زبانی برای پرس و جوی سطح بالا ، پردازش پرس و جوها و بهینه سازی پرس و جوها وظیفه ای اساسی در انقلاب و تکامل فنآوری پایگاه داده ایفا کرد و باعث شد سیستم های پایگاه داده با اجرای در حد بالا و قابل تطبیق شکل بگیرند .امروزه نگاه ما اینگونه است که نسل بعدی سیستم های داده کاوی بطور موفقیت آمیزی قابلیت های سنتی و قدیمی سیستم های مدیریت پایگاه داده ها را با

[۱]-Constraint - based

قابلیت های اکتشافی خود جایگزین خواهند کرد. اینگونه سیستم ها چارچوب بدون درزی ارائه خواهند داد که مدیریت سنتی پایگاه داده موقت و موردی را با تحلیل داده های دقیق و ابزارهای اکتشاف تلفیق خواهندکرد که به نظر می رسد وظیفه اساسی را اکتشاف چند بعدی بر مبنای قیود خواهد داشت.

قیود می تواند بعنوان یک نمونه کامل برای قیود سنتی مطرح باشد یا مربوط به سلسله مراتب بعد در پایگاه داده های چند بعدی باشد. قیود نه تنها چارچوب قوانین اکتشافات را قویتر می کنند و به استفاده کننده این امکان را میدهند که دقیقاً آن چیزی را که می خواهد مطرح کند بلکه باعث توسعه و بهینه سازی پرس و جوهای تحلیل و اکتشاف می شوند .

منابع وماخذ

الف-منابع فارسی

۱. معین نجف آبادی، ابراهیم، ضیغمی، شایا، و محرابی نژاد ، امیر .ً *مقدمات و بستر سازی استفاده از داده کاوی در سازمان نوین* ،سلسله سمینار داده کاوی ، شرکت پویندگان دانش نگار، بهار ۸۰

۲. معین نجف آبادی، ابراهیم.ً (فنون داده کاوی) کاربرد روشهای آماری در داده کاوی ً ،سلسله سمینار داده کاوی ، شرکت پویندگان دانش نگار ،بهار ۸۰

۳. آیت الله زاده شیرازی، محمد رضا، انبارۀ داده ً سلسله سمینار داده کاوی ، شرکت پویندگان دانش نگار ، بهار ۸۰

۴. محرابی نژاد، امیر،پردازش تحلیلی روی خط ((OLAP))ً،(مباحث ویژه در داده کاوی)ً، سلسله سمینار داده کاوی ، شرکت پویندگان دانش نگار ، بهار ۸۰

ب-منابع لاتین

٥. Bhavani Thuraisingham,” DATA MINING: Technologies, Techniques, Tools, and

٦. .Trends”, CRC Press LLC,١٩٩٩

٧. Michael J.A. Berry&Gordon S,Linoff.” DATAMINING Techniques”, John Wiley

٨. & Sons ,Inc, ١٩٩٧