**本资料为学长整理，判断与选择是标准答案，简答题是自己找的不保证全对**

**每年的范围可能不一样，请注意**

判断与选择

第一章

1. Computerized support is only used for organizational decisions that are responses to external pressures, not

for taking advantage of opportunities. F

2. The complexity of today's business environment creates many new challenges for organizations, such as

global competition, but creates few new opportunities in return. F

3. In addition to deploying business intelligence (BI) systems, companies may also perform other actions to

counter business pressures, such as improving customer service and entering business alliances. T

4. The overwhelming majority of competitive actions taken by businesses today feature computerized

information system support. T

5. PCs and, increasingly, mobile devices are the most common means of providing managers with information

to directly support decision making, instead of using IT staff intermediaries. T

6. In today's business environment, creativity, intuition, and interpersonal skills are effective substitutes for

analytical decision making. F

7. In a four-step process for decision making, managers construct a model of the problem before they evaluate

potential solutions. T

8. Due to the fact that business environments are now more complex than ever, trial-and-error is an effective

means of arriving at acceptable solutions. F

9. Group collaboration software has proved generally ineffective at improving decision-making. F

10. Due to the fact that organizations seek to store greater amounts of data than ever before, the cost per byte of

computer-based data storage devices is rapidly rising. F

11. Computerized information systems help decision makers overcome human cognitive limitations in

assembling and processing varied information. However, this is of little use in most analytical applications. F

12. In the Gorry and Scott-Morton framework of structured, semi-structured, and unstructured decisions,

computerized decision support can bring benefits to unstructured decisions. T

13. The term *decision support system* is a very specific term that implies the same tool, system, and development

approach to most developers. F

14. The access to data and ability to manipulate data (frequently including real-time data) are key elements of

business intelligence (BI) systems. T

15. One of the four components of BI systems, business performance management, is a collection of source data

in the data warehouse. F

16. Actionable intelligence is the primary goal of modern-day Business Intelligence (BI) systems vs. historical

reporting that characterized Management Information Systems (MIS). T

17. The use of dashboards and data visualizations is seldom effective in finding efficiencies in organizations, as

demonstrated by the Seattle Children's Hospital Case Study. F

18. The use of statistics in baseball by the Oakland Athletics, as described in the *Moneyball* case study, is an

example of the effectiveness of prescriptive analytics. T

19. Pushing programming out to distributed data is achieved solely by using the Hadoop Distributed File

System or HDFS. F

20. Volume, velocity, and variety of data characterize the Big Data paradigm. T21. In the Magpie Sensing case study, the automated collection of temperature and humidity data on shipped

goods helped with various types of analytics. Which of the following is an example of prescriptive analytics?

D

A) location of the shipment

B) real time reports of the shipment's temperature

C) warning of an open shipment seal

D) optimal temperature setting

22. In the Magpie Sensing case study, the automated collection of temperature and humidity data on shipped

goods helped with various types of analytics. Which of the following is an example of predictive analytics? A

A) warning of an open shipment seal

B) optimal temperature setting

C) real time reports of the shipment's temperature

D) location of the shipment

23. Which of the following is NOT an example that falls within the four major categories of business environment

factors for today's organizations? B

A) globalization

B) fewer government regulations

C) increased pool of customers

D) increased competition

24. Organizations counter the pressures they experience in their business environments in multiple ways. Which

of the following is NOT an effective way to counter these pressures? C

A) reactive actions

B) anticipative actions

C) retroactive actions

D) adaptive actions

25. Which of the following activities permeates nearly all managerial activity? B

A) planning

B) decision-making

C) directing

D) controlling

26. Why are analytical decision making skills now viewed as more important than interpersonal skills for an

organization's managers? D

A) because personable and friendly managers are always the least effective

B) because interpersonal skills are never important in organizations

C) because analytical-oriented managers tend to be flashier and less methodical

D) because analytical-oriented managers produce better results over time

27. Business environments and government requirements are becoming more complex. All of the following

actions to manage this complexity would be appropriate EXCEPT C

A) deploying more sophisticated tools and technique.

B) hiring more sophisticated and computer-savvy managers.

C) seeking new ways to avoid government compliance.

D) avoiding expensive trial and error to find out what works.

28. The deployment of large data warehouses with terabytes or even petabytes of data been crucial to the growth

of decision support. All the following explain why EXCEPT C

A) data warehouses have enabled the affordable collection of data for analytics.

B) data warehouses have assisted the collection of data for data mining.

C) data warehouses have enabled the collection of decision makers in one place.

D) data warehouses have assisted the collection of data from multiple sources.

29. Which of the following statements about cognitive limits of organizational decision makers is true? A

A) Cognitive limits affect both the recall and use of data by decision makers. B) Only top managers make decisions where cognitive limits are strained.

C) The most talented and effective managers do not have cognitive limitations.

D) All organizational decision-making requires data beyond human cognitive limits.

30. For the majority of organizations, evaluating the credit rating of a potential business partner is a(n) C

A) structured decision.

B) unstructured decision.

C) managerial control decision.

D) strategic decision.

31. For the majority of organizations, a daily accounts receivable transaction is a(n) C

A) strategic decision.

B) managerial control decision.

C) structured decision.

D) unstructured decision.

32. All of the following may be viewed as decision support systems EXCEPT D

A) an expert system to diagnose a medical condition.

B) a system that helps to manage the organization's supply chain management.

C) a knowledge management system to guide decision makers.

D) a retail sales system that processes customer sales transactions.

33. Business intelligence (BI) can be characterized as a transformation of A

A) data to information to decisions to actions.

B) Big Data to data to information to decisions.

C) data to processing to information to actions.

D) actions to decisions to feedback to information.

34. In answering the question "Which customers are most likely to click on my online ads and purchase my

goods?", you are most likely to use which of the following analytic applications? D

A) customer attrition

B) channel optimization

C) customer profitability

D) propensity to buy

35. In answering the question "Which customers are likely to be using fake credit cards?", you are most likely to

use which of the following analytic applications? C

A) customer segmentation

B) channel optimization

C) fraud detection

D) customer profitability

36. When Sabre developed their Enterprise Data Warehouse, they chose to use near-real-time updating of their

database. The main reason they did so was D

A) to be able to assess internal operations.

B) to aggregate performance metrics in an understandable way.

C) to provide a 360 degree view of the organization.

D) to provide up-to-date executive insights.

37. How are descriptive analytics methods different from the other two types? B

A) They answer "what to do?" queries, not "what-if?" queries.

B) They answer "what-is?" queries, not "what will be?" queries.

C) They answer "what-if?" queries, not "how many?" queries.

D) They answer "what will be?" queries, not "what to do?" queries.

38. Prescriptive BI capabilities are viewed as more powerful than predictive ones for all the following reasons

EXCEPT B

A) prescriptive models generally build on (with some overlap) predictive ones.

B) only prescriptive BI capabilities have monetary value to top-level managers.

C) understanding the likelihood of certain events often leaves unclear remedies.

D) prescriptive BI gives actual guidance as to actions. 39. Which of the following statements about Big Data is true? D

A) MapReduce is a storage filing system.

B) Data chunks are stored in different locations on one computer.

C) Hadoop is a type of processor used to process Big Data applications.

D) Pure Big Data systems do not involve fault tolerance.

40. Big Data often A

involves a form of distributed storage and processing using Hadoop and MapReduce. One reason for this is

A) the processing power needed for the centralized model would overload a single computer.

B) Big Data systems have to match the geographical spread of social media.

C) centralized storage creates too many vulnerabilities.

D) the "Big" in Big Data necessitates over 10,000 processing nodes.

第二章

1. When HP approaches problem-solving, the first step in solving business problems is building a model that

enables decision makers to develop a good understanding of the problem. F

2.In a decision making environment, continuous change always validates the assumptions of the decision

makers. F

3. The most important feature of management support systems is the computational efficiency involved in

making a decision. F

4. Web-based decision support systems can provide support to both individuals and groups that act in a

decision-making capacity. T

5. Single decision makers rarely face decisions with multiple objectives in organizations and so are not the focus

of data analytics tools. F

6. The design phase of decision making is where the decision maker examines reality and identifies and defines

the problem. F

7. Only after the failed implementation of a decision can the decision maker return a prior stage of decision

making. F

8. Web-based collaboration tools (e.g., GSS) can assist in multiple stages of decision making, not just the

intelligence phase. T

9. Uncovering the existence of a problem can be achieved through monitoring and analyzing of the

organization's productivity level. The derived measurements of productivity are based on real data. T

10. Qualitative elements of a problem cannot be incorporated into formal decision models, so one can only seek

to minimize their impact. F

11. Since the business environment involves considerable uncertainty, a manager cannot use modeling to

estimate the risks resulting from specific actions. F

12. A normative model examines all the possible alternatives in order to prove that the one selected is the best. T

13. Since a descriptive model checks the performance of the system for only a subset of all possible alternatives,

there is no guarantee that a selected alternative will be optimal. T

14. Generating alternatives manually is often necessary in the model-building process. The best option for the

decision makers is to generate as many of these alternatives as is conceivable. F

15. Generally speaking, people intuitively estimate risk quite accurately. F

16. A data warehouse can support the intelligence phase of decision making by continuously monitoring both

internal and external information, looking for early signs of problems and opportunities through a Web-based

enterprise information portal or dashboard. T

17. Business intelligence systems typically support solving a certain problem or evaluate an opportunity, while

decision support systems monitor situations and identify problems and/or opportunities, using analytic

methods. F

18. Artificial intelligence-based DSS fall into this category of document-driven DSS. F

19. The DSS component that includes the financial, statistical, management science, or other quantitative models

is called the model management subsystem. T

20. Knowledge-based management subsystems provide intelligence to augment the decision maker's own

intelligence. T

21. The HP Case illustrates that after analytics are chosen to solve a problem, building a new decision model from

scratch or purchasing one may not always be the best approach. Why is that? B

A) Analytic models work better when they are built from scratch or purchased.

B) A related tool requiring slight modification may already exist.

C) CIOs are more likely to allocate funds to new development.

D) Decision models should never be purchased, only developed in house.

22. Groupthink in a decision-making environment occurs when B

A) group members accept the same timeframe for problem solving without complaining.

B) group members all accept a course of action without thinking for themselves.

C) group members all use the same analytic tools without having a choice.

D) group members are all working together for the firm's success.

23. All of the following statements about decision style are true EXCEPT D

A) heuristic styles can also be democratic.

B) autocratic styles are authority-based.

C) decision styles may vary among lower-level managers.

D) decision styles are consistent among top managers.

24. A search for alternatives occurs in which phase of the decision making/action model? D

A) the intelligence phase

B) the implementation phase

C) the choice phase

D) the design phase

25. All of the following are benefits of using models for decision support EXCEPT B

A) you can find out probable outcomes of an action before actually taking it.

B) using well-designed models always guarantees you success in implementation.

C) it is easier to manipulate a model than a real system.

D) the cost of a model is usually much lower than manipulating the system in implementation.

26. In the design phase of decision making, selecting a principle of choice or criteria means that A

A) optimality is not the only criterion for acceptable solutions.

B) if an objective model is used with hard data, all decision makers will make the same choice.

C) risk acceptability is a subjective concept and plays little part in modeling.

D) using well-designed models guarantees you success in real life.

27. What form of decision theory assumes that decision makers are rational beings who always seek to strictly

maximize economic goals? C

A) satisficing decision theory

B) human optimal decision theory

C) normative decision theory

D) the theory of bounded rationality

28. When an Accounts Payable department improves their information system resulting in faster payments to

vendors, without the Accounts Receivable Department doing the same, leading to a cash flow crunch, what

can we say happened in decision-theoretic terms? C

A) cash flow problems

B) profit minimization

C) suboptimization

D) optimization

29. All of the following statements about risk in decision making are correct EXCEPT CA) all business decisions incorporate an element of risk.

B) decision makers frequently measure risk and uncertainty incorrectly.

C) most decision makers are pessimistic about decision outcomes.

D) methodologies are available for handling extreme uncertainty.

30. The Web can play a significant role in making large amounts of information available to decision makers.

Decision makers must be careful that this glut of information does not D

A) detract from the quality and speed of decision making.

B) take on the same credibility of internally-generated data.

C) increase their enthusiasm for data available on the web.

D) take on the same role as human intuition.

31. All of the following statements about the decision implementation phases are true EXCEPT D

A) ES and KMS can help in training and support for decision implementation.

B) implementation is every bit as important as the decision itself.

C) ERP, CRP, and BPM tools can all help track decision implementation.

D) employees need only the decisions from the CEO, not the rationale.

32. For DSS, why are semistructured or unstructured decisions the main focus of support? D

A) MIS staff prefer to work on solving unstructured and semistructured decisions.

B) There are many more unstructured and semistructured decisions than structured in organizations.

C) Unstructured and semistructured decisions are the easiest to solve.

D) They include human judgment, which is incorporated into DSS.

33. What class of DSS incorporates simulation and optimization? A

A) model-driven DSS

B) communications-driven/Group DSS

C) knowledge-driven DSS

D) data-driven DSS

34. When a DSS isD built, used successfully and integrated into the company's business processes, it was most

likely built for a(n)

A) one-off decision.

B) unimportant decision.

C) ambiguous decision.

D) recurrent decision.

35. The fact thatD many organizations share many similar problems means that in sourcing a DSS, it is often

wiser to acquire a(n)

A) offshored DSS.

B) consultant-developed DSS.

C) custom-made DSS.

D) ready-made DSS. 36. The

software

that manages the DSS database and enables relevant data to be accessed by DSS application programs is

called D

A) KWS.

B) CRM.

C) ERP.

D) DBMS.

37. The model management subsystem provides the system's analytical capabilities and appropriate software

management. Which of the following is NOT an element of the model management subsystem? D

A) MBMS

B) model base

C) model execution, integration, and command processor

D) DBMS

38. While Microsoft Excel can be an efficient tool for developing a DSS, compared to using a programming

language like C++, a shortcoming of Excel is A

A) errors can creep into formulas somewhat easily.

B) Excel is not widely understood compared to a language like C++.

C) it cannot be used effectively for small or medium sized problems.

D) it is not widely available for purchase.

39. What type of user interface has been recognized as an effective DSS GUI because it is familiar, user friendly,

and a gateway to almost all sources of necessary information and data? B

A) visual basic interfaces

B) Web browsers

C) ASP.net

D) mainframe interfaces

40. The user communicates with and commands the DSS through the user interface subsystem. Researchers

assert that some of the unique contributions of DSS are derived from B

A) the user being considered part of the system.

B) the intensive interaction between the computer and the decision maker.

C) some DSS user interfaces utilizing natural-language input (i.e., text in a human language).

D) the Web browser.

第三章

1. In the Isle of Capri case, the only capability added by the new software was increased processing speed of

processing reports. F

2. The "islands of data" problem in the 1980s describes the phenomenon of unconnected data being stored in

numerous locations within an organization. T

3. Subject oriented databases for data warehousing are organized by detailed subjects such as disk drives,

computers, and networks. F

4. Data warehouses are subsets of data marts. F

5. One way an operational data store differs from a data warehouse is the recency of their data. T

6. Organizations seldom devote a lot of effort to creating metadata because it is not important for the effective

use of data warehouses. F

7. Without middleware, different BI programs cannot easily connect to the data warehouse. T

8. Two-tier data warehouse/BI infrastructures offer organizations more flexibility but cost more than three-tier

ones. F

9. Moving the data into a data warehouse is usually the easiest part of its creation. F

10. The hub-and-spoke data warehouse model uses a centralized warehouse feeding dependent data marts. T

11. Because of performance and data quality issues, most experts agree that the federated architecture should

supplement data warehouses, not replace them.T

12. Bill Inmon advocates the data mart bus architecture whereas Ralph Kimball promotes the hub-and-spoke

architecture, a data mart bus architecture with conformed dimensions. F

13. The ETL process in data warehousing usually takes up a small portion of the time in a data-centric project. F

14. In the Starwood Hotels case, up-to-date data and faster reporting helped hotel managers better manage their

occupancy rates.T

15. Large companies, especially those with revenue upwards of $500 million consistently reap substantial cost

savings through the use of hosted data warehouses. F

16. OLTP systems are designed to handle ad hoc analysis and complex queries that deal with many data items. F

17. The data warehousing maturity model consists of six stages: prenatal, infant, child, teenager, adult, and sage.

T

18. A well-designed data warehouse means that user requirements do not have to change as business needs

change. F

19. Data warehouse administrators (DWAs) do not need strong business insight since they only handle the

technical aspect of the infrastructure. F

20. Because the recession has raised interest in low-cost open source software, it is now set to replace traditional

enterprise software. F

21. The "singleC version of the truth" embodied in a data warehouse such as Capri Casinos' means all of the following EXCEPT C

A) decision makers get to see the same results to queries.

B) decision makers have the same data available to support their decisions.

C) decision makers have unfettered access to all data in the warehouse.

D) decision makers get to use more dependable data for their decisions.

22. Operational or transaction databases are product oriented, handling transactions that update the database. In

contrast, data warehouses areA

A) subject-oriented and nonvolatile.

B) subject-oriented and volatile.

C) product-oriented and nonvolatile.

D) product-oriented and volatile.

23. Which kind of data warehouse is created separately from the enterprise data warehouse by a department and

not reliant on it for updates?D

A) sectional data mart

B) volatile data mart

C) public data mart

D) independent data mart

24. All of the following statements about metadata are true EXCEPT B

A) metadata gives context to reported data.

B) for most organizations, data warehouse metadata are an unnecessary expense.

C) metadata helps to describe the meaning and structure of data.

D) there may be ethical issues involved in the creation of metadata.

25. A Web client that connects to a Web server, which is in turn connected to a BI application server, is reflective

of a A

A) three tier architecture.

B) one tier architecture.

C) two tier architecture.

D) four tier architecture.

26. Which of the following BEST enables a data warehouse to handle complex queries and scale up to handle

many more requests? D

A) Microsoft Windows

B) a larger IT staff

C) use of the web by users as a front-end

D) parallel processing

27. Which data warehouse architecture uses metadata from existing data warehouses to create a hybrid logical

data warehouse comprised of data from the other warehouses? D

A) centralized data warehouse architecture

B) hub-and-spoke data warehouse architecture

C) independent data marts architecture

D) federated architecture

28. Which data warehouse architecture uses a normalized relational warehouse that feeds multiple data marts? B

A) federated architecture

B) hub-and-spoke data warehouse architecture

C) independent data marts architecture

D) centralized data warehouse architecture 29. Which approach to data warehouse integration focuses more on sharing process functionality than data

across systems? D

A) enterprise function integration

B) enterprise information integration

C) extraction, transformation, and load

D) enterprise application integration

30. In which stage of extraction, transformation, and load (ETL) into a data warehouse are data aggregated? C

A) extraction

B) load

C) transformation

D) cleanse

31. In which stage of extraction, transformation, and load (ETL) into a data warehouse are anomalies detected

and corrected? C

A) load

B) transformation

C) cleanse

D) extraction

32. Data warehouses provide direct and indirect benefits to using organizations. Which of the following is an

indirect benefit of data warehouses? D

A) extensive new analyses performed by users

B) simplified access to data

C) better and more timely information

D) improved customer service

33. All of the following are benefits of hosted data warehouses EXCEPT A

A) greater control of data.

B) frees up in-house systems.

C) better quality hardware.

D) smaller upfront investment. 34. When representing data in a data warehouse, using several dimension tables that are each connected only to a

fact table means you are using which warehouse structure? C

A) relational schema

B) dimensional schema

C) star schema

D) snowflake schema

35. When querying a dimensional database, a user went from summarized data to its underlying details. The

function that served this purpose is C

A) slice.

B) roll-up.

C) drill down.

D) dice.

3

6

.

36. Which of the following online analytical processing (OLAP) technologies does NOT require the

precomputation and storage of information? D

A) MOLAP

B) SQL

C) HOLAP

D) ROLAP

37. Active data warehousing can be used to support the highest level of decision making sophistication and

power. The major feature that enables this in relation to handling the data is B

A) nature of the data.

B) speed of data transfer.

C) country of (data) origin.

D) source of the data.

38. Which of the following statements is more descriptive of active data warehouses in contrast with traditional

data warehouses? A

A) large numbers of users, including operational staffs

B) restrictive reporting with daily and weekly data currency

C) detailed data available for strategic use only

D) strategic decisions whose impacts are hard to measure

39. How does the use of cloud computing affect the scalability of a data warehouse? A

A) Hardware resources are dynamically allocated as use increases.

B) Cloud vendors are mostly based overseas where the cost of labor is low.

C) Cloud computing has little effect on a data warehouse's scalability.

D) Cloud computing vendors bring as much hardware as needed to users' offices.

40. All of the following are true about in-database processing technology EXCEPT B

A) it pushes the algorithms to where the data is.

B) it is the same as in-memory storage technology.

C) it is often used for apps like credit card fraud detection and investment risk management.

D) it makes the response to queries much faster than conventional databases.

第四章

1. The WebFOCUS BI platform in the Travel and Transport case study decreased clients' reliance on the IT

function when seeking system reports. T

2. The dashboard for the WebFOCUS BI platform in the Travel and Transport case study required client side

software to operate. F

3. Data is the contextualization of information, that is, information set in context. F

4. The main difference between service level agreements and key performance indicators is the audience. T

5. The balanced scorecard is a type of report that is based solely on financial metrics. F

6. The data storage component of a business reporting system builds the various reports and hosts them for, or

disseminates them to users. It also provides notification, annotation, collaboration, and other services.F

7. In the FEMA case study, the BureauNet software was the primary reason behind the increased speed and

relevance of the reports FEMA employees received. T

8. Google Maps has set new standards for data visualization with its intuitive Web mapping software.T

9. There are basic chart types and specialized chart types. A Gantt chart is a specialized chart type. T

10. Visualization differs from traditional charts and graphs in complexity of data sets and use of multiple dimensions and

measures.T

11. When telling a story during a presentation, it is best to avoid describing hurdles that your character must

overcome, to avoid souring the mood.F

12. For best results when deploying visual analytics environments, focus only on power users and

management to get the best return on your investment.F

13. Information density is a key characteristic of performance dashboards. T

14. In the Dallas Cowboys case study, the focus was on using data analytics to decide which players would play

every week. F

15. One comparison typically made when data is presented in business intelligence systems is a comparison

against historical values. T

16. The best key performance indicators are derived independently from the company's strategic goals to enable

developers to "think outside of the box." F

17. The BPM development cycle is essentially a one-shot process where the requirement is to get it right the first

time. F

18. With key performance indicators, driver KPIs have a significant effect on outcome KPIs, but the reverse is not

necessarily true. T

19. With the balanced scorecard approach, the entire focus is on measuring and managing specific financial goals

based on the organization's strategy. F

20. A Six Sigma deployment can be deemed effective even if the number of defects are not reduced to 3.4 defects

per million. F

21. For those executives who do not have the time to go through lengthy reports, the best alternative is the D

A) last page of the report.

B) raw data that informed the report.

C) charts in the report.

D) executive summary.

22. All of the following are true about external reports between businesses and the government EXCEPT A

A) their primary focus is government.

B) they can be filed nationally or internationally.

C) they are standardized for the most part to reduce the regulatory burden.

D) they can include tax and compliance reporting.

23. Kaplan and Norton developed a report that presents an integrated view of success in the organization called

B

A) dashboard-type reports.

B) balanced scorecard-type reports.

C) metric management reports.

D) visual reports. 24. Which component of a reporting system contains steps detailing how recorded transactions are converted

into metrics, scorecards, and dashboards? D

A) assurance

B) extract, transform and load

C) data supply

D) business logic

25. Which of the following is LEAST related to data/information visualization? C

A) statistical graphics

B) information graphics

C) graphic artwork

D) scientific visualization

26. The Internet emerged as a new medium for visualization and brought all the following EXCEPT B

A) immersive environments for consuming data.

B) new forms of computation of business logic.

C) worldwide digital distribution of visualization.

D) new graphics displays through PC displays.

27. Which kind of chart is described as an enhanced variant of a scatter plot? B

A) heat map

B) bubble chart

C) pie chart

D) bullet

28. Which type of visualization tool can be very helpful when the intention is to show relative proportions of

dollars per department allocated by a university administration? C

A) heat map

B) bubble chart

C) pie chart

D) bullet

29. Which type of visualization tool can be very helpful when a data set contains location data? A

A) geographic map

B) tree map

C) bar chart

D) highlight table

30. Which type of question does visual analytics seeks to answer? D

A) What is happening today?

B) What happened yesterday?

C) When did it happen?

D) Why did it happen?

31. When you tell a story in a presentation, all of the following are true EXCEPT B

A) stories and their lessons should be easy to remember.

B) a well-told story should have no need for subsequent discussion.

C) a story should make sense and order out of a lot of background noise.

D) the outcome and reasons for it should be clear at the end of your story.

32. Benefits of the latest visual analytics tools, such as SAS Visual Analytics, include all of the following EXCEPT

C

A) there is less demand on IT departments for reports.

B) mobile platforms such as the iPhone are supported by these products.

C) they explore massive amounts of data in hours, not days.

D) it is easier to spot useful patterns and trends in the data.

33. What is the management feature of a dashboard? A

A) operational data that identify what actions to take to resolve a problem

B) summarized dimensional data to analyze the root cause of problems

C) graphical, abstracted data to monitor key performance metrics D) summarized dimensional data to monitor key performance metrics

34. What is the fundamental challenge of dashboard design? C

A) ensuring that the organization has access to the latest web browsers

B) ensuring that the organization has the appropriate hardware onsite to support it

C) ensuring that the required information is shown clearly on a single screen

D) ensuring that users across the organization have access to it

35. Contextual metadata for a dashboard includes all the following EXCEPT A

A) which operating system is running the dashboard server software.

B) whether any high-value transactions that would skew the overall trends were rejected as a part of the

loading process.

C) whether the dashboard is presenting "fresh" or "stale" information.

D) when the data warehouse was last refreshed.

36. Dashboards can be presented at all the following levels EXCEPT D

A) the static report level.

B) the visual dashboard level.

C) the self-service cube level.

D) the visual cube level.

37. Why is a performance management system superior to a performance measurement system? A

A) because measurement alone has little use without action

B) because performance management systems cost more

C) because performance measurement systems are only in their infancy

D) because measurement automatically leads to problem solution

38. Why is the customer perspective important in the balanced scorecard methodology? C

A) because customers should always be included in any design methodology

B) because companies need customer input into the design of the balanced scorecard

C) because dissatisfied customers will eventually hurt the bottom line

D) because customers understand best how the firm's internal processes should work

39. All of the following statements about balanced scorecards and dashboards are true EXCEPT B

A) scorecards are less preferred at operational and tactical levels.

B) scorecards are best for real-time tracking of a marketing campaign.

C) dashboards would be the preferred choice to monitor production quality.

D) scorecards are preferred for tracking the achievement of strategic goals.

40. What is Six Sigma? B

A) a methodology aimed at measuring the amount of variability in a business process

B) a methodology aimed at reducing the number of defects in a business process

C) a letter in the Greek alphabet that statisticians use to measure process variability

D) a methodology aimed at reducing the amount of variability in a business process

第五章

1. In the Cabela's case study, the SAS/Teradata solution enabled the direct marketer to better identify likely

customers and market to them based mostly on external data sources. F

2. The cost of data storage has plummeted recently, making data mining feasible for more firms. T

3. Data mining can be very useful in detecting patterns such as credit card fraud, but is of little help in

improving sales. F

4. The entire focus of the predictive analytics system in the Infinity P&C case was on detecting and handling

fraudulent claims for the company's benefit. F

5. If using a mining analogy, "knowledge mining" would be a more appropriate term than "data mining." T

6. Data mining requires specialized data analysts to ask ad hoc questions and obtain answers quickly from the

system. F

7. Ratio data is a type of categorical data. F

8. Interval data is a type of numerical data. T

9. In the Memphis Police Department case study, predictive analytics helped to identify the best schedule for

officers in order to pay the least overtime. F

10. In data mining, classification models help in prediction. T

11. Statistics and data mining both look for data sets that are as large as possible. F

12. Using data mining on data about imports and exports can help to detect tax avoidance and money

laundering.T

13. In the cancer research case study, data mining algorithms that predict cancer survivability with high

predictive power are good replacements for medical professionals. F

14. During classification in data mining, a false positive is an occurrence classified as true by the algorithm while

being false in reality. T

15. When training a data mining model, the testing dataset is always larger than the training dataset.F

16. When a problem has many attributes that impact the classification of different patterns, decision trees may be

a useful approach. T

17. In the 2degrees case study, the main effectiveness of the new analytics system was in dissuading potential

churners from leaving the company. T

18. Market basket analysis is a useful and entertaining way to explain data mining to a technologically less savvy

audience, but it has little business significance. F

19. The number of users of free/open source data mining software now exceeds that of users of commercial

software versions. T

20. Data that is collected, stored, and analyzed in data mining is often private and personal. There is no way to

maintain individuals' privacy other than being very careful about physical data security. F

21. In the Cabela's case study, what types of models helped the company understand the value of customers,

using a five-point scale? B

A) simulation and geographical models

B) clustering and association models

C) simulation and regression models

D) reporting and association models

22. Understanding customers better has helped Amazon and others become more successful. The understanding

comes primarily from C

A) collecting data about customers and transactions.

B) asking the customers what they want.

C) analyzing the vast data amounts routinely collected.

D) developing a philosophy that is data analytics-centric.

23. All of the following statements about data mining are true EXCEPT B

A) the valid aspect means that the discovered patterns should hold true on new data.

B) the process aspect means that data mining should be a one-step process to results. C) the novel aspect means that previously unknown patterns are discovered.

D) the potentially useful aspect means that results should lead to some business benefit.

24. What is the main reason parallel processing is sometimes used for data mining? D

A) because the hardware exists in most organizations and it is available to use

B) because any strategic application requires parallel processing

C) because the most of the algorithms used for data mining require it

D) because of the massive data amounts and search efforts involved

25. The data field "ethnic group" can be best described as D

A) ordinal data.

B) ratio data.

C) interval data.

D) nominal data.

26. The data field "salary" can be best described asA

A) ratio data.

B) nominal data.

C) ordinal data.

D) interval data.

27. Which broad area of data mining applications analyzes data, forming rules to distinguish between defined

classes? B

A) associations

B) classification

C) visualization

D) clustering

28. Which broad area of data mining applications partitions a collection of objects into natural groupings with

similar features? B

A) visualization

B) clustering

C) classification

D) associations

29. The data mining algorithm type used for classification somewhat resembling the biological neural networks

in the human brain is B

A) decision trees.

B) artificial neural networks.

C) association rule mining.

D) cluster analysis.

30. Identifying and preventing incorrect claim payments and fraudulent activities falls under which type of data

mining applications? B

A) retailing and logistics

B) insurance

C) computer hardware and software

D) customer relationship management

31. All of the following statements about data mining are true EXCEPT C

A) understanding the data, e.g., the relevant variables, is critical to success.

B) understanding the business goal is critical.

C) building the model takes the most time and effort.

D) data is typically preprocessed and/or cleaned before use.

32. Which data mining process/methodology is thought to be the most comprehensive, according to

kdnuggets.com rankings? A

A) CRISP-DM

B) SEMMA

C) KDD Process

D) proprietary organizational methodologies 33. Prediction problems where the variables have numeric values are most accurately defined as B

A) associations.

B) regressions.

C) computations.

D) classifications.

34. What does the robustness of a data mining method refer to? B

A) its ability to construct a prediction model efficiently given a large amount of data

B) its ability to overcome noisy data to make somewhat accurate predictions

C) its speed of computation and computational costs in using the mode

D) its ability to predict the outcome of a previously unknown data set accurately

35. What does the scalability of a data mining method refer to? A

A) its ability to construct a prediction model efficiently given a large amount of data

B) its ability to overcome noisy data to make somewhat accurate predictions

C) its ability to predict the outcome of a previously unknown data set accurately

D) its speed of computation and computational costs in using the mode

36. In estimating the accuracy of data mining (or other) classification models, the true positive rate is C

A) the ratio of correctly classified positives divided by the sum of correctly classified positives and

incorrectly classified negatives.

B) the ratio of correctly classified positives divided by the sum of correctly classified positives and

incorrectly classified positives.

C) the ratio of correctly classified positives divided by the total positive count.

D) the ratio of correctly classified negatives divided by the total negative count.

37. In data mining, finding an affinity of two products to be commonly together in a shopping cart is known as A

A) association rule mining.

B) decision trees.

C) cluster analysis.

D) artificial neural networks.

38. Third party providers of publicly available datasets protect the anonymity of the individuals in the data set

primarily by C

A) letting individuals in the data know their data is being accessed.

B) asking data users to use the data ethically.

C) removing identifiers such as names and social security numbers.

D) leaving in identifiers (e.g., name), but changing other variables.

39. In the Target case study, why did Target send a teen maternity ads? A

A) Target's analytic model suggested she was pregnant based on her buying habits.

B) Target's analytic model confused her with an older woman with a similar name.

C) Target was using a special promotion that targeted all teens in her geographical area.

D) Target was sending ads to all women in a particular neighborhood.

40. Which of the following is a data mining myth? A

A) Data mining requires a separate, dedicated database.

B) Newer Web-based tools enable managers of all educational levels to do data mining.

C) The current state-of-the-art is ready to go for almost any business.

D) Data mining is a multistep process that requires deliberate, proactive design and use.

第六章

1. In the opening vignette, the high accuracy of the models in predicting the outcomes of complex medical

procedures showed that data mining tools are ready to replace experts in the medical field. F

2. Though useful in business applications, neural networks are a rough, inexact model of how the brain works,

not a precise replica. T

3. The use of hidden layers and new topologies and algorithms renewed waning interest in neural networks. T

4. Compared to the human brain, artificial neural networks have many more neurons. F

5. In the mining industry case study, the input to the neural network is a verbal description of a hanging rock on

the mine wall. F

6. The network topology that allows only one-way links between layers, with no feedback linkage permitted, is

known as backpropagation. T

7. With a neural network, outputs are attributes of the problem while inputs are potential solutions to the

problem. F

8. The most complex problems solved by neural networks require one or more hidden layers for increased

accuracy. T

9. The task undertaken by a neural network does not affect the architecture of the neural network; in other

words, architectures are problem-independent. F

10. Prior to starting the development of a neural network, developers must carry out a requirements analysis. T

11. No matter the topology or architecture of a neural network, they all use the same algorithm to adjust weights

during training. F

12. Neural networks are called "black boxes" due to the lack of ability to explain their reasoning. T

13. Generally speaking, support vector machines are less accurate a prediction method than other approaches

such as decision trees and neural networks. F

14. Unlike other "black box" predictive models, support vector machines have a solid mathematical foundation in

statistics. T

15. In the student retention case study, support vector machines used in prediction had proportionally more true

positives than true negatives. T

16. Using support vector machines, you must normalize the data before you numericize it. F

17. The *k*-nearest neighbor algorithm is overly complex when compared to artificial neural networks and support

vector machines. F

18. The *k*-nearest neighbor algorithm appears well-suited to solving image recognition and categorization

problems. T

19. In the Coors case study, a neural network was used to more skillfully identify which beer flavors could be

predicted. T

20. In the Coors case study, genetic algorithms were of little use in solving the flavor prediction problem. F

21. In the opening vignette, predictive modeling is described as A

A) estimating the future using the past.

B) not yet accepted in the business world.

C) unable to handle complex predictive problems.

D) the least practiced branch of data mining.

22. In the opening vignette, which method was the best in both accuracy of predicted outcomes and sensitivity? B

A) CART

B) SVM

C) ANN

D) C5

23. Neural networks have been described as "biologically inspired." What does this mean? C

A) They are faithful to the entire process of computation in the human brain.

B) They have the power to undertake every task the human brain can. C) They crudely model the biological makeup of the human brain.

D) They were created to look identical to human brains.

24. Which element in an artificial neural network roughly corresponds to a synapse in a human brain? B

A) node

B) weight

C) output

D) input

25. Which element in an artificial neural network roughly corresponds to a dendrite in a human brain? C

A) output

B) node

C) input

D) weight

26. All the following statements about hidden layers in artificial neural networks are true EXCEPT C

A) more hidden layers increase required computation exponentially.

B) hidden layers are not direct inputs or outputs.

C) many top commercial ANNs forgo hidden layers completely.

D) more hidden layers include many more weights.

27. In developing an artificial neural network, all of the following are important reasons to pre-select the network

architecture and learning method EXCEPT A

A) most neural networks need special purpose hardware, which may be absent.

B) development personnel may be more experienced with certain architectures.

C) some neural network software may not be available in the organization.

D) some configurations have better success than others with specific problems.

28. Backpropagation learning algorithms for neural networks are D

A) used without a training set of data.

B) used without hidden layers for effectiveness.

C) the least popular algorithm due to their inaccuracy.

D) required to have error tolerance set in advance.

29. Why is sensitivity analysis frequently used for artificial neural networks? B

A) because it is generally informative, although it cannot help to identify cause-and-effect relationships

among variables

B) because some consequences of mistakes by the network might be fatal, so justification may matter

C) because it provides a complete description of the inner workings of the artificial neural network

D) because it is required by all major artificial neural networks

30. Support vector machines are a popular machine learning technique primarily because of C

A) their relative cost and relative ease of use.

B) their relative cost and superior predictive power.

C) their superior predictive power and their theoretical foundation.

D) their high effectiveness in the very few areas where they can be used.

31. In the student retention case study, which of the following variables was MOST important in determining

whether a student dropped out of college? C

A) marital status and hours enrolled

B) college and major

C) completed credit hours and hours enrolled

D) high school GPA and SAT high score math

32. In the student retention case study, of the four data mining methods used, which was the most accurate? C

A) ANN

B) LR

C) SVM

D) DT(C5)

33. When using support vector machines, in which stage do you transform the data? D

A) deploying the model B) experimentation

C) developing the model

D) preprocessing the data

34. When using support vector machines, in which stage do you select the kernel type (e.g., RBF, Sigmoid)?D

A) deploying the model

B) preprocessing the data

C) experimentation

D) developing the model

35. For how long do SVM models continue to be accurate and actionable? D

A) for as long as you choose to use them

B) for as long as the developers stay with the firm

C) for as long as management support continues to exist for the project

D) for as long as the behavior of the domain stays the same

36. All of the following are disadvantages/limitations of the SVM technique EXCEPT B

A) they have high algorithmic complexity and extensive memory requirements for complex tasks.

B) their accuracy is poor in many domains compared to neural networks.

C) model building involves complex and time-demanding calculations.

D) selection of the kernel type and kernel function parameters is difficult.

37. The *k*-nearest neighbor machine learning algorithm (*k*NN) is B

A) highly mathematical and computationally intensive.

B) regarded as a "lazy" learning method.

C) very complex in its inner workings.

D) a method that has little in common with regression.

38. Using the *k*-nearest neighbor machine learning algorithm for classification, larger values of *k* C

A) do not change the effect of noise on the classification.

B) increase the effect of noise on the classification.

C) reduce the effect of noise on the classification.

D) sharpen the distinction between classes.

39. What is a major drawback to the basic majority voting classification in kNN? D

A) Classes that are more clustered tend to dominate prediction.

B) Even the naive version of the algorithm is hard to implement.

C) It requires frequent human subjective input during computation.

D) Classes with more frequent examples tend to dominate prediction.

40. In the Coors case study, why was a genetic algorithm paired with neural networks in the prediction of beer

flavors? A

A) to complement the neural network by reducing the error term

B) to best model how the flavor of beer evolves as it ages

C) to enhance the neural network by pre-selecting output classes for the neural network

D) to replace the neural network in harder cases

第七章

1. In the chapter's opening vignette, IBM's computer named Watson outperformed human game champions on

the game show *Jeopardy!T*

2. Text analytics is the subset of text mining that handles information retrieval and extraction, plus data mining.

F

3. In text mining, inputs to the process include unstructured data such as Word documents, PDF files, text

excerpts, e-mail and XML files. T

4. During information extraction, entity recognition (the recognition of names of people and organizations)

takes place after relationship extraction. F

5. Categorization and clustering of documents during text mining differ only in the preselection of categories. T

6. Articles and auxiliary verbs are assigned little value in text mining and are usually filtered out. T

7. In the patent analysis case study, text mining of thousands of patents held by the firm and its competitors

helped improve competitive intelligence, but was of little use in identifying complementary products. F

8. The bag-of-words model is appropriate for spam detection but not for text analytics. T

9. Chinese, Japanese, and Thai have features that make them more difficult candidates for natural language

processing. T

10. Regional accents present challenges for natural language processing. T

11. In the Hong Kong government case study, reporting time was the main benefit of using SAS Business

Analytics to generate reports. T

12. Detecting lies from text transcripts of conversations is a future goal of text mining as current systems achieve

only 50% accuracy of detection. F

13. In the financial services firm case study, text analysis for associate-customer interactions were completely

automated and could detect whether they met the company's standards. T

14. In text mining, creating the term-document matrix includes all the terms that are included in all documents,

making for huge matrices only manageable on computers. F

15. In text mining, if an association between two concepts has 7% support, it means that 7% of the documents had

both concepts represented in the same document. T

16. In sentiment analysis, sentiment suggests a transient, temporary opinion reflective of one's feelings. F

17. Current use of sentiment analysis in voice of the customer applications allows companies to change their

products or services in real time in response to customer sentiment. T

18. In sentiment analysis, it is hard to classify some subjects such as news as good or bad, but easier to classify

others, e.g., movie reviews, in the same way. T

19. The linguistic approach to speech handles processes elements such as intensity, pitch and jitter from speech

recorded on audio. F

20. In the BBVA case study, text analytics was used to help the company defend and enhance its reputation in

social media. T

21. In the opening vignette, the architectural system that supported Watson used all the following elements

EXCEPT A

A) a core engine that could operate seamlessly in another domain without changes.

B) massive parallelism to enable simultaneous consideration of multiple hypotheses.

C) integration of shallow and deep knowledge.

D) an underlying confidence subsystem that ranks and integrates answers.

22. According to a study by Merrill Lynch and Gartner, what percentage of all corporate data is captured and

stored in some sort of unstructured form? B

A) 15%

B) 85%

C) 25%

D) 75% 23. Which of these applications will derive the LEAST benefit from text mining? A

A) sales transaction files

B) patent description files

C) customer comment files

D) patients' medical files

24. In text mining, stemming is the process of B

A) creating new branches or stems of recorded paragraphs.

B) reducing multiple words to their base or root.

C) categorizing a block of text in a sentence.

D) transforming the term-by-document matrix to a manageable size.

25. In text mining, tokenizing is the process of C

A) transforming the term-by-document matrix to a manageable size.

B) reducing multiple words to their base or root.

C) categorizing a block of text in a sentence.

D) creating new branches or stems of recorded paragraphs.

26. All of the following are challenges associated with natural language processing EXCEPT A

A) dividing up a text into individual words in English.

B) understanding the context in which something is said.

C) recognizing typographical or grammatical errors in texts.

D) distinguishing between words that have more than one meaning.

27. What application is MOST dependent on text analysis of transcribed sales call center notes and voice

conversations with customers? A

A) CRM

B) OLAP

C) finance

D) ERP

28. In text mining, which of the following methods is NOT used to reduce the size of a sparse matrix? D

A) eliminating rarely occurring terms

B) using singular value decomposition

C) using a domain expert

D) normalizing word frequencies

29. What data discovery process, whereby objects are categorized into predetermined groups, is used in text

mining? A

A) classification

B) trend analysis

C) association

D) clustering

30. In the research literature case study, the researchers analyzing academic papers extracted information from

which source? A

A) the paper abstract

B) the paper references

C) the main body of the paper

D) the paper keywords

31. Sentiment classification usually covers all the following issues EXCEPT D

A) range of polarity (e.g., star ratings for hotels and for restaurants).

B) classes of sentiment (e.g., positive versus negative).

C) range in strength of opinion.

D) biometric identification of the consumer expressing the sentiment.

32. In sentiment analysis, which of the following is an implicit opinion? C

A) The cruise we went on last summer was a disaster.

B) Our new mayor is great for the city.

C) The customer service I got for my TV was laughable. D) The hotel we stayed in was terrible.

33. In the Whirlpool case study, the company sought to better understand information coming from which

source? D

A) delivery information

B) customer transaction data

C) goods moving through the internal supply chain

D) customer e-mails

34. What do voice of the market (VOM) applications of sentiment analysis do? C

A) They examine the "market of ideas" in politics.

B) They examine employee sentiment in the organization.

C) They examine customer sentiment at the aggregate level.

D) They examine the stock market for trends.

35. How is objectivity handled in sentiment analysis? C

A) It is clarified with the customer who expressed it.

B) It is incorporated as a type of sentiment.

C) It is identified and removed as facts are not sentiment.

D) It is ignored because it does not appear in customer sentiment.

36. Identifying the target of an expressed sentiment is difficult for all the following reasons EXCEPT C

A) the review may not be directly connected to the target through the topic name.

B) sometimes there are multiple targets expressed in a sentiment.

C) strong sentiments may be generated by a computer, not a person.

D) blogs and articles with the sentiment may be general in nature.

37. In text analysis, what is a lexicon? A

A) a catalog of words, their synonyms, and their meanings

B) a catalog of customers, their words, and phrase

C) a catalog of customers, products, words, and phrase

D) a catalog of letters, words, phrases and sentences

38. What types of documents are BEST suited to semantic labeling and aggregation to determine sentiment

orientation? D

A) collections of documents

B) medium- to large-sized documents

C) large-sized documents

D) small- to medium-sized documents

39. Inputs to speech analytics include all of the following EXCEPT A

A) written transcripts of calls to service centers.

B) recorded conversations of customer call-ins.

C) videos of customer focus groups.

D) live customer interactions with service representatives.

40. In the Blue Cross Blue Shield case study, speech analytics were used to identify "confusion" calls by customers.

What was true about these calls? A

A) They were not documented by customer service reps for speech analytics.

B) They took less time than others as frustrated customers hung up.

C) They led customers to rely more on self-serve options.

D) They were difficult to identify using standard phrases like "I don't get it."

第八章

1. Participating in social media is so new that it is still optional for most companies in the United States. F

2. Web mining is exactly the same as Web analytics: the analysis of Web site usage data. F

3. Web crawlers or spiders collect information from Web pages in an automated or semi-automated way. Only

the text of Web pages is collected by crawlers. F

4. Generally, making a search engine more efficient makes it less effective. T

5. With the PageRank algorithm, a Web page with more incoming links will always rank higher than one with

fewer incoming links. F

6. The main purpose of frequent recrawling of some Web sites is to prevent search users from retrieving stale

search results. T

7. Search engine optimization (SEO) techniques play a minor role in a Web site's search ranking because only

well-written content matters. F

8. Clickstream analysis does not need users to enter their perceptions of the Web site or other feedback directly

to be useful in determining their preferences. T

9. Having more Web traffic coming from organic search than other types of search is the goal of most companies.

T

10. Since little can be done about visitor Web site abandonment rates, organizations have to focus their efforts on

increasing the number of new visitors. F

11. It is possible to use prescriptive tools for Web analytics to describe current Web site use comprehensively. F

12. Many Web analytics tools are free to download and use, including Google Web Analytics. T

13. Voice of customer (VOC) applications track and resolve business process and usability obstacles for a Web

site. F

14. Social network analysis can help companies divide their customers into market segments by analyzing their

interconnections. T

15. Decentralization, the need for specialized skills, and immediacy of output are all attributes of Web publishing

when compared to industrial publishing. F

16.

Consistent high quality, higher publishing frequency, and longer time lag are all attributes of industrial

publishing when compared to Web publishing. F

17. Web site visitors who critique and create content are more engaged than those who join networks and

spectate. T

18. Descriptive analytics for social media feature such items as your followers as well as the content in online

conversations that help you to identify themes and sentiments. F

19. Companies understand that when their product goes "viral," the content of the online conversations about

their product does not matter, only the volume of conversations. F

20. Social media analytics companies provide integrated support that is helpful to many parts of a business, not

only the Sales and Marketing functions. T

21. What does Web content mining involve? C

A) analyzing the universal resource locator in Web pages

B) analyzing the pattern of visits to a Web site

C) analyzing the unstructured content of Web pages

D) analyzing the PageRank and other metadata of a Web page

22. What does Web structure mining involve? B

A) analyzing the pattern of visits to a Web site

B) analyzing the universal resource locators in Web pages

C) analyzing the PageRank and other metadata of a Web page

D) analyzing the unstructured content of Web pages

23. In the extremist groups case study, what approach is used to discover the ideology and fund raising of

extremist groups through their Web sites? A

A) content analysis

B) physical visits to addresses on the site C) hyperlink analysis

D) e-mail responses to questions sent to the sites

24. Search engines do not search the entire Web every time a user makes a search request, for all the following

reasons EXCEPT C

A) it would take longer than the user could wait.

B) it is more efficient to use pre-stored search results.

C) most users are not interested in searching the entire Web.

D) the Web is too complex to be searched each time.

25. Breaking up a Web page into its components to identify worthy words/terms and indexing them using a set of

rules is called A

A) parsing the documents.

B) document analysis.

C) creating the term-by-document matrix.

D) preprocessing the documents.

26. PageRank for Webpages is useful to Web developers for which of the following reasons? B

A) They uniquely identify the Web page developer for greater accountability.

B) It gives developers insight into Web user behavior.

C) Developing many Web pages with low PageRank can help a Web site attract users.

D) It is used in citation analysis for scholarly papers.

27. Search engine optimization (SEO) is a means by which B

A) Web site developers can negotiate better deals for paid ads.

B) Web site developers can increase Web site search rankings.

C) Web site developers optimize the artistic features of their Web sites.

D) Web site developers index their Web sites for search engines.

28. In general, what is the best kind of Web traffic to a Web site? B

A) European Web traffic

B) organic Web traffic

C) paid Web traffic

D) bot-generated traffic

29. Clickstream analysis is most likely to be used for all the following types of applications EXCEPTB

A) predicting user behavior.

B) hiring new functional area managers.

C) designing cross-marketing strategies across products.

D) determining the lifetime value of clients.

30. What are the two main types of Web analytics? A

A) off-site and on-site Web analytics

B) data-based and subjective Web analytics

C) old-school and new-school Web analytics

D) Bing and Google Web analytics

31. Web site usability may be rated poor if B

A) users fail to click on all pages equally.

B) Web site visitors download few of your offered PDFs and videos.

C) the time spent on your Web site is long.

D) the average number of page views on your Web site is large.

32. Common sources of traffic to your Web site include all of the following EXCEPT C

A) direct links.

B) referral Web sites.

C) accidental visitors.

D) paid search from search engines.

33. Understanding which keywords your users enter to reach your Web site through a search engine can help

you understand B

A) the hardware your Web site is running on. B) how well visitors understand your products.

C) the type of Web browser being used by your Web site visitors.

D) most of your Web site visitors' wants and needs.

34. Which of the following statements about Web site conversion statistics is FALSE? D

A) The conversion rate is the number of people who take action divided by the number of visitors.

B) Analyzing exit rates can tell you why visitors left your Web site.

C) Web site visitors can be classed as either new or returning.

D) Visitors who begin a purchase on most Web sites must complete it.

35. A voice of customer (VOC) strategy involves all of the following EXCEPT B

A) capturing both unstructured Web data and enterprise data as a starting point.

B) connecting captured insights to unstructured data in order to take action.

C) taking actions related to your market, customers and services.

D) analyzing unstructured data with minimal effort on the user's part.

36. All of the following statements about social networks are true EXCEPT A

A) companies should invest equally to retain all members of a group.

B) it is possible to gain insights into how products go viral.

C) a group with all interconnected individuals is called a clique.

D) members of a group are affected by the behavior of others in the group.

37. What is one major way that Web-based social media is the same as publishing media? A

A) They can both reach a global audience.

B) They cost the same to publish.

C) They require the same skill and training to publish.

D) They have the same immediacy of updates.

38. What is one major way in which Web-based social media differs from traditional publishing media? C

A) Web-based media have a narrower range of quality.

B) They use different languages of publication.

C) They have different costs to own and operate.

D) Most Web-based media are operated by the government and large firms.

39. What does descriptive analytics for social media do? B

A) It examines the content of online conversations.

B) It helps identify your followers.

C) It identifies links between groups.

D) It identifies the biggest sources of influence online.

40. What does advanced analytics for social media do? D

A) It helps identify your followers.

B) It identifies the biggest sources of influence online.

C) It identifies links between groups.

D) It examines the content of online conversations.

第九章

1. Modeling can be viewed as a science in its entirety. F

2. In the Midwest ISO opening vignette, the solution provided by the model's output determined the best

output level to be produced by each power plant. T

3. If linear programming can be successfully applied a problem, the output is usually optimal. T

4. In the ExxonMobil case study, the approach taken was to find individual solutions to routing, transportation,

scheduling, and inventory management, and select the best solution for one of the variables. F

5. In order to be effective, analysts must use models to solve problems with no regard to the organizational

culture to find optimal results. F

6. In the Harrah's Cherokee Casino and Hotel case study, the revenue management system modified room

prices based on demand and offered the same price/availability to all customers at any one time. F

7. AHP can be used effectively for optimization with problems containing a small number of alternatives. T

8. The trend is towards developing and using Web tools and software to access and run modeling software. T

9. Using data cubes in OLAP systems opens the data up to analysis by more classes of models. F

10. Another name for result variables is independent variables. F

11. Taking a decision under risk is different from taking the decision under uncertainty. T

12. Spreadsheets are the second most popular tool for modeling. F

13. Linear programming seeks to optimally allocate resources among competing activities and is likely the best

known optimization model. T

14. When using Excel's Solver, we can have multiple constraints and multiple objective cells. F

15. Most managerial problems can be properly evaluated and solved using a single goal, such as profit

maximization. F

16. Sensitivity analysis seeks to assess the impact of changes in the input data and parameters on the proposed

solution. T

17. Goal seeking is roughly the opposite of "what-if" analysis. T

18. Using expected value (EV) with decision trees is totally appropriate for situations where one outcome could

lead to an immense loss for the company. F

19. In the U.S. HUD case study, the use of AHP brought standards and coherence to project selection, resulting in

a 10% decrease in project requests from 1999 levels. F

20. The analytic hierarchy process incorporates both qualitative and quantitative decision making criteria. T

21. Using modeling for decision support can currently achieve all of the following EXCEPT B

A) enable organizations to see likely results of their decisions.

B) replace strategy formulation at top levels of the organization.

C) enhance the decision making process.

D) reduce the costs of providing services to customers.

22. Environmental scanning is important for all of the following reasons EXCEPT B

A) organizational culture is important and affects the model use.

B) environments have greater impact on a model than the organization does.

C) environmental factors may have created the current problem.

D) it is critical to identify key corporate decision makers.

23. Today, it is critical for companies to consider D

A) how to package products in the right format.

B) how to sell products at the right price.

C) how to get products to the right customer.

D) all of the above

24. Models can be built with the help of human knowledge and expertise. Another source of help in building

these models is A

A) classification and clustering methods.

B) the customer.

C) business partners. D) customer service reps.

25. What is an influence diagram? D

A) a map of the environment around decision makers

B) a map of the environment around a model

C) a diagram showing the influence of decision makers

D) a graphical representation of a model

26. Spreadsheets are particularly useful for all of the following reasons EXCEPTB

A) they can be used to build static and dynamic models.

B) they easily import and manipulate massive databases.

C) they are able to import and export to many different file formats.

D) it is easy to manipulate data and see results instantly.

27. Linear programming belongs to a family of tools called B

A) decision tree models.

B) mathematical programming models.

C) qualitative models.

D) heuristic programming models.

28. Which of the following is NOT a component of a linear programming problem? A

A) internal metrics

B) decision variables

C) constraints

D) objective function

29. In an LP model, what does the fourth hidden component contain? D

A) product mix variables

B) financial and accounting variables

C) constraint and limit variables

D) slack and surplus variables

30. Managers in organizations typically have A

A) multiple goals that need to be simultaneously or jointly optimized.

B) single goals that cannot be optimized using linear and nonlinear programming.

C) single goals that can be optimized using linear and nonlinear programming.

D) a small number of goals that can be independently optimized using linear and nonlinear programming.

31. Sensitivity analysis is important in management support systems for all of the following reasons EXCEPT D

A) it permits the manager to input data to increase his/her confidence in the model.

B) it allows flexibility and adaptation to changing conditions.

C) it provides a better understanding of the model and the decision-making situation.

D) it improves the mathematical optimality of the generated solutions.

32. The question "What will total earnings be if we reduce our inventory stocking costs by 10%?" is a type of B

A) goal-seeking analysis.

B) what-if analysis.

C) sensitivity analysis.

D) utility modeling.

33. The question "What advertising budget is needed to increase market share by 7%?" is a type of A

A) goal-seeking analysis.

B) what-if analysis.

C) sensitivity analysis.

D) utility modeling.

34. The question "How many servers will be needed to reduce the waiting time of restaurant customers to less

than 9 minutes?" is a type of A

A) goal-seeking analysis.

B) what-if analysis.

C) sensitivity analysis.

D) utility modeling. 35. Decision trees are best suited to solve what types of problems? A

A) problems with a single goal

B) problems with a large number of alternatives

C) problems where probabilities are unknown

D) problems with a tabular representation

36. In handling uncertainty in decision modeling, the optimistic approach assumes B

A) the best possible outcome of most alternatives will occur.

B) the best possible outcome of each alternative will occur.

C) the best possible outcome of one alternative will occur.

D) the best possible outcome of some alternatives will occur.

37. In handling uncertainty in decision modeling, what does the pessimistic approach do? C

A) It assumes the worst possible outcome of one alternative will occur and then avoids it.

B) It assumes the worst possible outcome of some alternatives will occur and then selects the best of them.

C) It assumes the worst possible outcome of each alternative will occur and then selects the best of them.

D) It assumes the worst possible outcome of each alternative will occur and then selects the worst of them.

38. Which of the following statements about expected utility is true? B

A) In calculating utility, it assumes the decision will be made thousands of times, making the probabilities

more likely on average.

B) Used in decision making, it can bring huge risk to a small startup with limited resources.

C) It does not affect decisions made with expected values.

D) Used in decision making, it is an objective value, not subjective.

39. Which of the following statements about the analytic hierarchy process (AHP) is true? A

A) It can handle multiple criteria and goals.

B) It is really not a decision model at all.

C) It is based entirely on quantitative data.

D) It is an opaque "black box" in the same way as neural networks.

40. Which of the following statements about the end-of-chapter CARE International case study is true? D

A) CARE used a linear programming model for optimization.

B) CARE ran its own shipping operation with vehicles that needed route optimization.

C) CARE set out to exclusively use international suppliers with large capacity to better serve people

affected by disasters.

D) CARE's objective was to respond to natural disasters faster.

第十章

1. In the Fluor case study, redesigning the process of reviewing engineering changes had no discernible impact

on the bottom line. F

2. In the choice phase of problem solving, normative models involve selecting an optimal or best outcome. T

3. Analytical techniques for problem solving are best for unstructured rather than structured problems. F

4. Heuristic approaches are typically used to solve more complex problems. T

5. In the Chilean government case study, the government used complete enumeration to find the optimal

solution for deciding meal providers to schools. F

6. Genetic algorithms are heuristic methods that do not guarantee an optimal solution to a problem. T

7. A "what-if" model is most typically used for the most structured problems. F

8. In the Finnish Air Force case, the simulation had to take account of a finite number of possibilities relating to

task times, material handling delays, etc. T

9. The use of simulation models is desirable because they can usually be solved in one pass, without incurring

the time and cost of iterations. F

10. An advantage of simulation is that it allows model builders to solve problems with minimal interaction with

users or managers. F

11. Time compression in a simulation allows managers to test certain strategies with less risk. T

12. Simulation solutions cannot easily be transferred from one problem domain to another. T

13. Determining the duration of the simulation occurs before the model is validated and tested. F

14. Discrete events and agent-based models are usually used for middle or low levels of abstraction. T

15. In steady-state plant control design, time-independent simulation would be appropriate. T

16. Simulation does not usually allow decision makers to see how a solution to a complex problem evolves over

(compressed) time, nor can decision makers interact with the simulation. T

17. Visual interactive simulation (VIS) is a simulation method that lets decision makers see what the model is

doing and how it interacts with the decisions made, as they are made. T

18. In the RFID case study, the key variable tested by the simulation model was quicker availability of

information about the location of various parts in the supply chain. T

19. Visual interactive modeling (VIM) systems, especially those developed for the military and the video-game

industry, have "thinking" characters who can behave with a relatively high level of intelligence in their

interactions with users. T

20. In the Canadian pandemic case study, the macro-level simulation modeled aggregates of a population that

might experience a pandemic. F

21. How does blind search differ from optimization? C

A) Blind search cannot result in optimal solutions whereas optimization methods do.

B) Blind search represents a guided approach while optimization is unguided.

C) Blind search usually does not conclude in one step like some optimization methods.

D) Blind search is usually a more efficient problem solving approach than optimization.

22. In modeling, an optimal solution is understood to be C

A) a solution that can only be determined by an exhaustive enumeration and testing of alternatives.

B) a solution found in the least possible time and using the least possible computing resources.

C) a solution that is the best based on criteria defined in the design phase.

D) a solution that requires an algorithm for determination.

23. When is a complete enumeration of solutions used? B

A) when a solution that is "good enough" is fine and good heuristics are available

B) when there is enough time and computational power available

C) when the modeler requires a guided approach to problem solving

D) when there are an infinite number of solutions to be searched

24. All of the following are true about heuristics EXCEPT C

A) heuristics are used when the modeler requires a guided approach to problem solving.

B) heuristics are used when a solution that is "good enough" is sought. C) heuristics are used when there is abundant time and computational power.

D) heuristics are rules of good judgment.

25. Which approach is most suited to structured problems with little uncertainty? C

A) simulation

B) human intuition

C) optimization

D) genetic algorithms

26. Genetic algorithms belong to the family of methods in the A

A) artificial intelligence area.

B) optimization area.

C) complete enumeration family of methods.

D) non-computer based (human) solutions area.

27. All of the following are suitable problems for genetic algorithms EXCEPT D

A) dynamic process control.

B) pattern recognition with complex patterns.

C) simulation of biological models.

D) simple optimization with few variables.

28. Which approach is most suited to complex problems with significant uncertainty, a need for experimentation,

and time compression? A

A) simulation

B) optimization

C) human intuition

D) genetic algorithms

29. Which of the following is an advantage of simulation? A

A) It can incorporate significant real-life complexity.

B) It always results in optimal solutions.

C) Simulation software requires special skills.

D) It solves problems in one pass with no iterations.

30. In which stage of the simulation methodology do you determine the variables and gather data? B

A) defining the problem

B) constructing the simulation model

C) testing and validating the model

D) designing the experiment

31. In which stage of the simulation methodology do you determine how long to run the simulation? B

A) constructing the simulation model

B) designing the experiment

C) testing and validating the model

D) defining the problem

32. In which stage of the simulation methodology do you determine the system's boundaries and environment?

A

A) constructing the simulation model

B) defining the problem

C) testing and validating the model

D) designing the experiment

33. What BEST describes a simulation model with a limited number of variables, each with a finite number of

values? B

A) system dynamics simulation

B) discrete event simulation

C) continuous distribution simulation

D) Monte Carlo simulation

34. What BEST describes a simulation model in which it is not important to know exactly when a modeled event

occurred? BA) continuous distribution simulation

B) time-independent simulation

C) system dynamics simulation

D) discrete event simulation

35. The advantages of visual interactive simulation include all of the following EXCEPT B

A) improvements in training using the simulation.

B) reduced need for decision maker involvement.

C) the ability to see how a simulation works.

D) improved presentation of simulation results.

36. What can system dynamics modeling be used for? A

A) qualitative methods for analyzing a system

B) simulation models that test each subsystem in isolation

C) micro-level simulation models that examine individual values

D) studying system behavior at an instant in time

37. The EHR (electronic health record) system dynamics modeling example showed that C

A) increased electronic note-taking negatively affects compliance.

B) e-notes negatively affect radiology performance.

C) increased staff training results in increased electronic prescriptions.

D) adverse drug events help to decrease patient time.

38. In agent-based modeling, agents are C

A) the human workers or agents who use the system.

B) communication links between simulations.

C) autonomous rule-based decision making units.

D) the hardware platform used to conduct the simulation.

39. Agent-based modeling is best for all the following types of problem features EXCEPT B

A) complex interactions.

B) low uncertainty.

C) many interrelated factors.

D) irregular data.

40. What is the final stage of an agent-based modeling (ABM) methodology? C

A) identifying the agents and determining their behavior

B) determining agent-related data

C) validating agent behavior against reality

D) determining the suitability of ABM

Questions and Answer (2/10)

1. What is Business Intelligence Advanced Framework, describe the relationship among the different components. (Figure 1-4）



1. Briefly describe the steps of intelligence phase of decision making. （2.4节）

Problem(or Opportunity) Identification：The intelligence phase begins with the identification of organizational goals and objectives related to an issue of concern (e.g., inventory management, job selection, lack of or incorrect Web presence) and determination of whether they are being met.

Problem Classification：Problem classification is the conceptualization of a problem in an attempt to place it in a definable category, possibly leading to a standard solution approach.

Problem Decomposition：Many complex problems can be divided into subproblems. Solving the simpler subproblems may help in solving a complex problem

Problem Ownership：In the intelligence phase, it is important to establish problem ownership. A problem exists in an organization only if someone or some group takes on the responsibility of attacking it and if the organization has the ability to solve it.

问题（或机会）识别：情报阶段从确定与关注的问题相关的组织目标和目的开始（例如，库存管理、工作选择、缺乏或不正确的 Web 存在）并确定它们是否得到满足。

问题分类：问题分类是问题的概念化，试图将其置于可定义的类别中，可能导致标准的解决方案方法。

问题分解：许多复杂问题可以分为子问题。解决更简单的子问题可能有助于解决复杂的问题

问题所有权：在智能阶段，建立问题所有权很重要。只有当某人或某个团体承担起攻击它的责任并且组织有能力解决它时，组织中才会存在问题。

3. What are the characteristics of data warehousing? At least answering 5 characte ristics。(3.2.3)

Subject oriented

Integrated

Time-variant (time series)

Nonvolatile

Summarized

Not normalized

Metadata

Web based, relational/multi-dimensional

Client/server, real-time/right-time/active...

面向主题的

集成的

时变性（时间序列

非易失性

基于网络

关系的/多维的

客户端／服务器

实时

元数据

1. What are the components of a DSS. （2.11）

A DSS application can be composed of a data management subsystem, a model management subsystem, a user interface subsystem, and a knowledge-based management subsystem

DSS应用可以由数据管理子系统、模型管理子系统、用户界面子系统和基于知识的管理子系统组成

1. Briefly describe Simon’s four phases of decision making.

Simon (1977) said that this involves three major phases: intelligence, design, and choice. He later added a fourth phase, implementation.

The decision-making process starts with the intelligence phase; in this phase, the decision maker examines reality and identifies and defines the problem. Problem ownership is established as well. In the design phase, a model that represents the system is constructed. This is done by making assumptions that simplify reality and writing down the relationships among all the variables. The model is then validated, and criteria are determined in a principle of choice for evaluation of the alternative courses of action that are identified. Often, the process of model development identifies alternative solutions and vice versa.

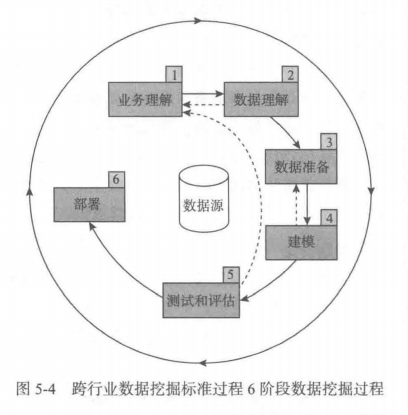
The choice phase includes selection of a proposed solution to the model (not necessarily to the problem it represents). This solution is tested to determine its viability. When the proposed solution seems reasonable, we are ready for the last phase: implementation of the decision (not necessarily of a system). Successful implementation results in solving the real problem. Failure leads to a return to an earlier phase of the process. In fact, we can return to an earlier phase during any of the latter three phases.

Simon（1977）说，这涉及三个主要阶段：情报、设计和选择。他后来增加了第四阶段，即实施。

决策过程从情报阶段开始；在这个阶段，决策者检查现实，识别并定义问题。问题所有权也已确立。在设计阶段，将构建一个表示系统的模型。这是通过做出简化现实的假设并写下所有变量之间的关系来实现的。然后对模型进行验证，并在选择的原则中确定标准，以评估所确定的替代行动方案。通常，模型开发过程会确定替代解决方案，反之亦然。

选择阶段包括选择模型的拟议解决方案（不一定是它所代表的问题）。对该解决方案进行测试以确定其可行性。当提出的解决方案看起来合理时，我们就为最后一个阶段做好了准备：实施决策（不一定是系统）。成功的实施会解决真正的问题。失败会导致返回到流程的早期阶段。事实上，在后三个阶段中的任何一个阶段，我们都可以回到更早的阶段。

1. Describe CRISP-DM process



数据挖掘的关键是知道要研究什么要解决这个问题，首先要深入了解新知识的管理需求，清楚地了解即将开展研究的商业目标

了解业务需求后，数据挖掘的主要任务是从大批可用数据库中识别相关的数据

数据准备，通常被称为数据预处理，它的目的是处理前面阶段识别的数据，为后续的数据挖掘分析做好准备

建模：在这一环节中，可以选择并应用不同的建模技术来处理已经准备好的数据集，解决特定的商业问题

第5步是对模型的精确性和推广性进行评估 步需要评估对所选用的模型在多大程度上满足业务目标，即是否需要开发建立和评估更多的模型 另外，如果时间及预算约束允许，还可以考虑在真实世界中测试通过数据挖掘建立的模型

完成建立并评估模型并不意味着数据挖掘项目的结束 即使模型的目的只是简单地探索

数据 也需要对探索获得的知识进行组织，并以终端用户能够理解和受益的方式展示

1. Using Hyperplane to explain what is SVM (6.5节)

支待向量机是通过标注训练集来实现输入和输出功能的有监督学习方法的一种 输入和

输出向量之间的函数既可以是一个分类函数（用来将个体分类到预定义类中），也可以是

个回归函数（用来估计所需输出结果的连续数值） 对于分类来说，通常会使用非线性核函

数来将输入数据转换（合理地代表高度复杂非线性关系）为对千输入数据来说是线性可分的

高维特征空间，然后构建最大间隔超平面来对训练数据中的输出类进行最优化划分

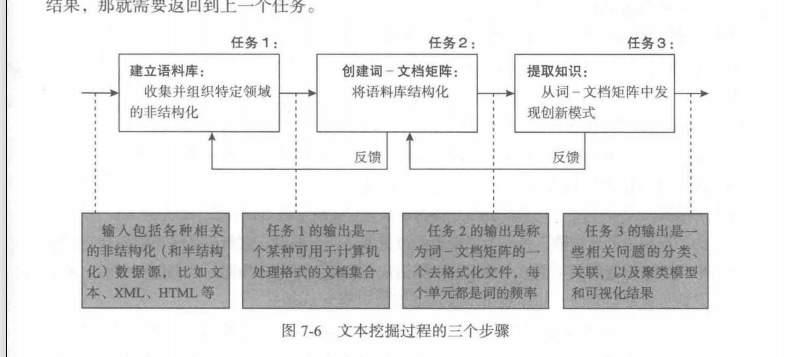
SVM 中使用的数据也许会多于两个维度（即两类） 在这种情况下，我们感兴趣的是使

n-I 维超平面来划分数据，其中 是维度数（即类别） 这是一个典型的线性分类器形式，

而我们感兴趣的是找到那个 n-1 维的超平面使得从超平面到最近数据点的距离最大 基本假设是这些平行超平面的间隔或距离越大 分类器的泛化能力（即 SVM 模型的预测能力）就越如果这些超平面是存在的，那么它们就能够用 次规划最优问题的方法来数学化地表达

这些超平面就是所谓的最大间隔超平面，而这个线性分类器就是所谓的最大间隔分类器

1. Describe Text mining processes.(three steps)



第一个任务的主要目的是收集所有与研究背景（兴趣领域）相关的文档，这包括了文本

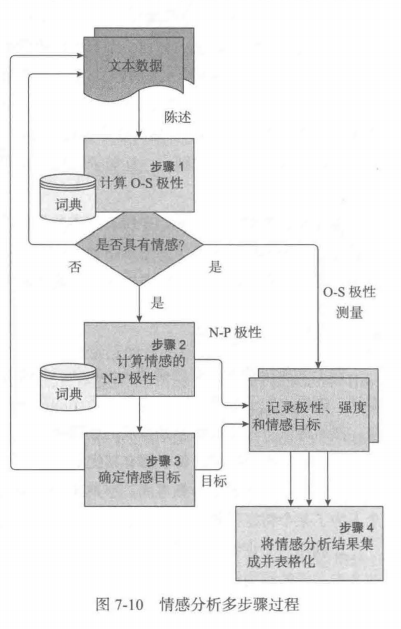
文档、 XML 文件、电子邮件、网页及便签 除了可用的文本数据外，语音也可以通过语音

识别算法来成为文本的一部分收集完后，文本文档就会被一个方式转化并组织，这样它们就能在进行计算机处理的时候都有相同的展现形式（比如 CII 文本文件）

第二任务中，我们使用经过组织的电子化文档（语料库）来创建 矩阵(TDM) TDM 中，行代表文档，列则代表词 词和文档的关系是由指数（比如，一种简单的关系可以是单词在各个文档中出现的次数）来表示的

任务3.利用结构化的 TDM 以及其他潜在的结构化数据元素，我们就可以将创新模式从特定问题的背景中提取出来

1. Describe Sentiment analysis process (7.9)



步骤1: 情感侦察在收集并准备 完文档后，敏感性分析的首要任务是客 观性侦察 这个步骤的目的是区分事实 和意见，即将文本分类为客观的或主观 这或许可以用 0-S 极性（客观—主 观极性，以范围为 的一个数来表 示）计算来表示 如果客观性值接近 I, 那么就没有意见可以挖掘（即这是 事实）

步骤 2: N-P 极性分类第二个重 要的任务是极性分类 给定 段包含意 见的文本，任务的目标是将意见分类为 两个相对的极性之一，或者确定其在两级之间的位置

步骤 3: 目标确认 这个步骤的目标是精确识别所表达的情感目标

步骤 4: 集成 旦文档中所有文本数据的情感都确认并计算好，它们就会在这步集成

并转化为一个单独的文档情感测量指标

10.What are the Web analytics Metrics？Just list name. (8.6.2)

网站可用性：他们如何使用我的网站？

• 流量来源：他们从哪里来？

• 访问者概要：我的访问者是什么样的？

• 转换数据：对千商务而言，这一切都意味着什么？

Concept, tool and methods: (5/20)

1. Data warehouse

A physical repository where relational data are specially organized to provide enterprise-wide, cleansed data in a standardized format

一个物理存储库，其中专门组织关系数据，以标准化格式提供企业范围的清洗后的数据

1. Visual analytics

而可视化分析 (visual analytics 的真正含义包括可视化和预测性分析两部分 信息可视化的目的是回答“发生了什么”和“正在发生什么”与商务智能（日常报表、计分卡、仪表盘）有密切联系 可视化分析主要回答“为什么会发生”和“将来可能发生什么”与业务分析（预测、分割 、关联分析）

1. Nominal data and Catalog data

标称数据 (nominal data) 是为对象分配的简单标签代码，是不可测量的

分类数据 (categorical data) 表示用于将变量划分为多个特定群体的类标签

1. OLAP vs OLTP

OLTP是联机事务处理传统的关系型数据库的主要应用，主要是基本的、日常的事务处理，记录即时的增、删、改、查，比如在银行存取一笔款，就是一个事务交易。OLAP即联机分析处理，是数据仓库的核心部心，支持复杂的分析操作，侧重决策支持，并且提供直观易懂的查询结果。典型的应用就是复杂的动态报表系统。

1. Dashboard

仪表板是许多企业用来帮助跟踪、分析和显示数据的工具，通常可用于深入了解组织、部门或特定流程的整体健康状况。

1. BPM

企业绩效管理 (business performance management, BPM) 是指企业用来评价 、监控和管理企业绩效的流程、方法、指标和技术，

1. Balence Scorecard

BSC即平衡计分卡（Balanced Score Card），是常见的绩效考核方式之一，平衡计分卡是从财务、客户、内部运营、学习与成长四个角度，将组织的战略落实为可操作的衡量指标和目标值的一种新型绩效管理体系。

1. Descriptive model

描述性模型 descripti ve model) 描述事物的本质或人们对事物的认知，这些模型通常

数学为基础

1. Business report (4.2.1)

业务报表是包含业务相关信息的书面文件。

1. ETL

数据提取、转换和加载

1. Data mining from technical viewpoint

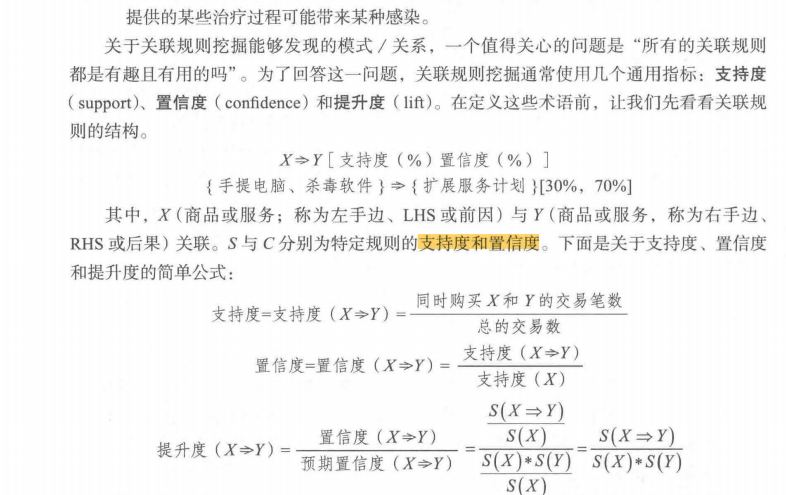
丛技术角度来讲，数据挖掘是综合运用统计学 数学和人工智能技术从大量数据中抽取

和识别有用信息及知识（或模式）的过程

1. k-fold cross validation

k折交叉验证 (k-fold crossvalidation), 也称为轮回评估，将整个数据集随机分成k个规模近似相等的互斥子集 分类模型经过k次训练和测试，每次在其中 k-1 个子集上训练，在另外1个子集上测试

1. Support and confidence



1. Text mining

文本挖掘 text mining, 又称文本数据挖掘，或者文本数据库中的知识发现）是从大量

的非结构化数据源中提取模式（有用的信息和知识）的半自动化过程

1. Web usage mining

网络使用挖掘（也称网络分析）是通过网络页面生成的数据访问和事务中提取出有用的

信息

1. Sentiment analysis

情感分析 (sentiment analysis) 是利用大量的文本数据源（网络记录形式的用户反馈）

来侦察用户对特定产品和服务的喜好与厌恶观点

1. SEO

搜索引擎优化 (SEO) 是一种可以影响电子商务网站可见性和搜索结果的策略性活动

一般来说，搜索结果页面排名越高，这个网站在搜索结果列表中出现得越频繁，这个网站会得到更多的访问者

1. CTR

点击率 (CTR): 点击一个产品广告的数量除以产品显示次数

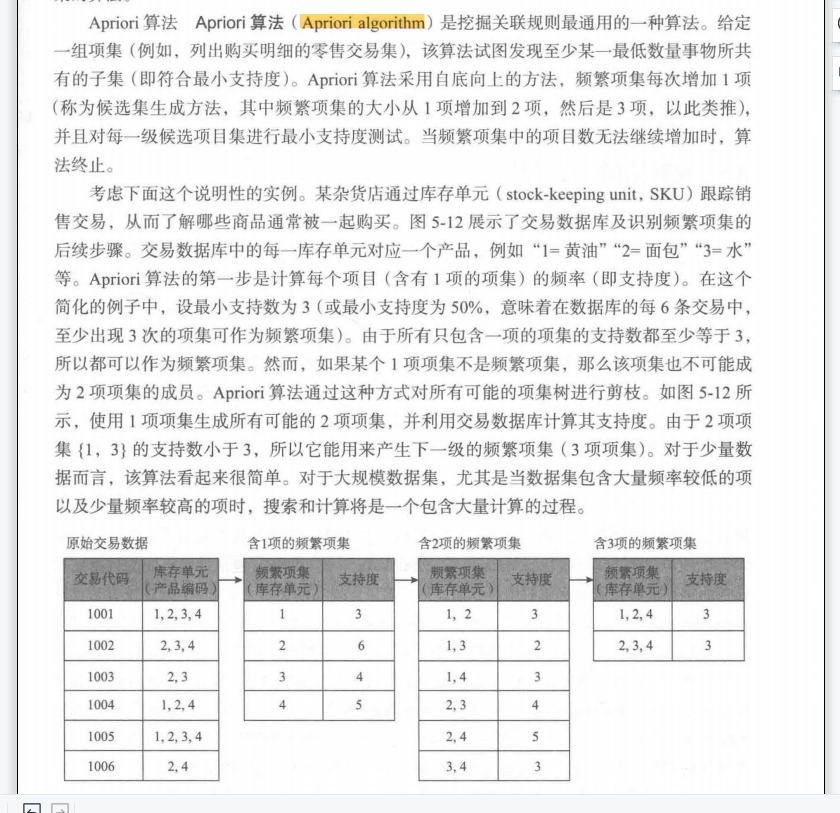
1. Web Crawler

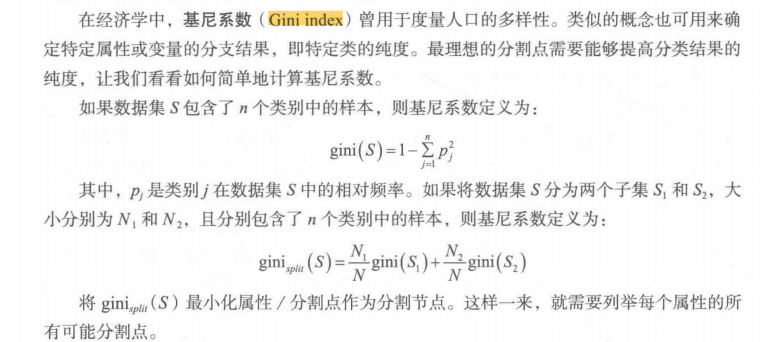
网络爬虫(Web crawler, 又称蜘蛛）是用千自动读取网络站点的内容

1. Tie strength (连接强度)

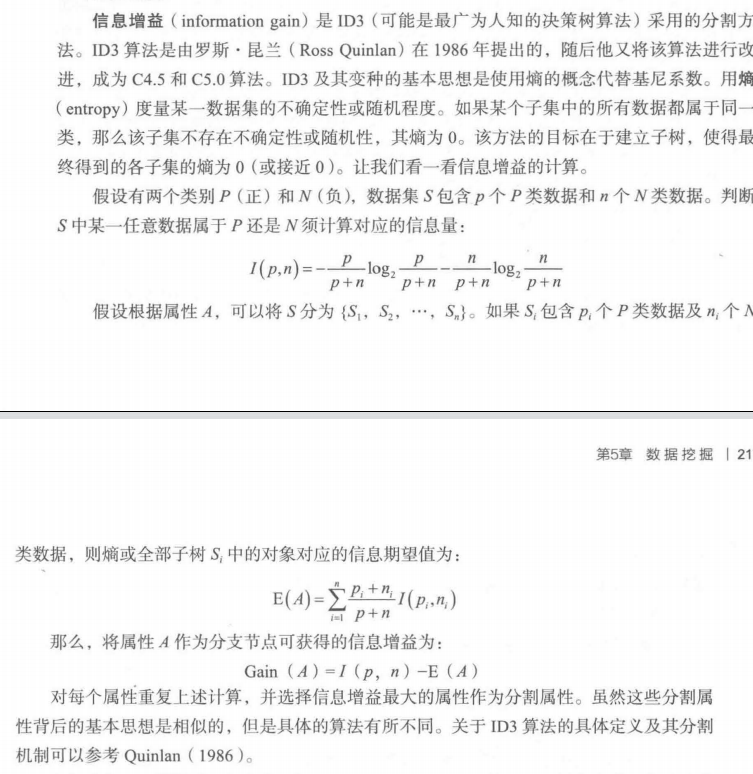
连接强度：定义为时间、情感强度、亲密关系 、互惠（即相互关系）的线性组合

计算：(3/6)

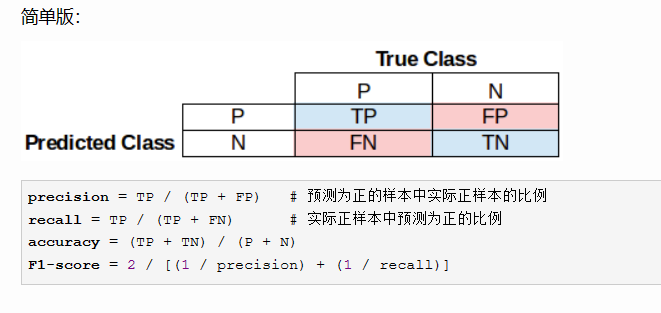
1. Apriori algorithm 计算支持度等于3的过程。
2. 
3. Gini index计算决策分割点。



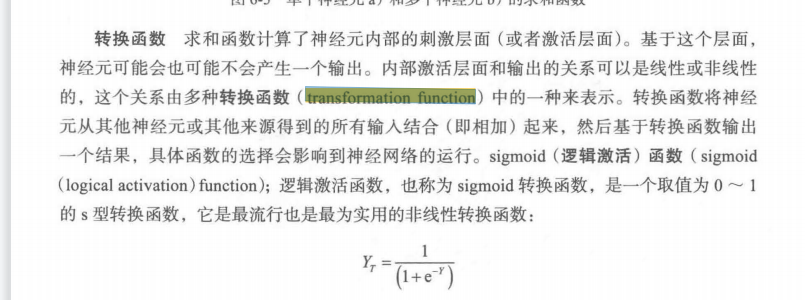
1. 基于最大熵的information gain计算决策分割点。



1. 计算Precision，accuracy， F1-Score。



5. 在神经网络中计算经过transformation function变换的输出



6. Minkowski distance，Manhattan distance曼哈顿距离，Euclidean distance欧几里得距离 and cosine similarity余弦相似度

