Unit 5

Data mining

Data mining

- Data mining is simply filtering through large amounts of raw data for useful information that gives businesses a competitive edge. This information is made up of meaningful patterns and trends that are already in the data but were previously unseen.

برتری رقابتی

روندها

- The most popular tool used when mining is Artificial Intelligence (AI). Intelligent guesses, learning by example and using deductive reasoning. Some of the more popular AI methods used in data mining include neural networks, clustering and decision trees.

استدلال قیاسی

Data mining

- Neural networks look at the rules of using data, which are based on the connections found or on a sample set of data. As a result, the software continually analyses value and compares it to the other factors and it compares these factors repeatedly until it finds patterns emerging. These patterns are known as rules. The software then looks for other patterns based on these rules or sends out an alarm when a trigger value is hit.

Data mining

- Clustering divides data into groups based on similar features or limited data ranges. Clusters are used when data is not labeled in a way that is favorable to mining. For instance, an insurance company that wants to find instance of fraud wouldn’t have its records labeled as fraudulent or not fraudulent. But after analyzing patterns within clusters, the mining software can start to figure out the rules that point to which claims are likely to be false.

کلاهبرداری

کشف کرد

Data mining

- Decision trees, like clusters, separate the data into subsets and then analyze the subsets to divide then into further subsets and so on (for a few more levels). The final subsets are then small enough that the mining process can find interesting patterns and relationships within the data.

Data mining

- Once the data to be mined is identified, it should be cleaned. Cleansing data frees it from duplicate information and erroneous data. Next, the data should be stored in a uniform format within relevant categories or fields. Mining tools can work with all types of data storage, from large data warehouses to smaller desktop databases to flat files. Data warehouses and data when the process is complete, the mining software generates a report. An analyst goes over the report to see if further work needs to be done, such as refining parameters, using other data analysis tools to examine the data, or even scrapping the data if it's unusable. If no further work is required, the report proceeds to the decision makers for appropriate action.

دسته های مرتبط
چک کردن
بهبود
دور ریختن
تقدیم می شود

Data mining

دادگاه عالی

- The power of data mining is being used for many purposes, such as analyzing Supreme Court decisions, discovering patterns in health care, pulling stories about competitors from newswires, resolving bottlenecks in production processes and analyzing sequences in the human genetic makeup. There really is no limit to the type of business or area of study where data mining can be beneficial.

جمع آوری اطلاعات راجع به رقبا از رسانه های خبری

Exercises 1

- **1. Mark the following statements as True or false (T/F):**

a. Data mining is a process of analyzing know patterns in data.
b. Artificial intelligence is commonly used in data mining.
c. In data mining, patterns found while analyzing data are used for further analyzing the data.
d. Data mining is used to detect false insurance.
e. Data mining is only useful for a limited range of problems.

Exercises 2

2. Find the answers to these questions in the text.

1. What tool is often used in data mining?
2. What AI method is used for the following processes?
   a. Separate data into subsets and then analyze the subsets to divide them into further subsets for a number of levels.
   b. Continually analyze and compare data until patterns emerge.
   c. Divide data into groups based on similar features or limited data ranges.

Exercises 2

2. Find the answers to these questions in the text.

3. What term is used for the patterns found by neural networks?
4. When are clusters used in data mining?
5. What types of data storage can be used in data mining?
6. What can an analyst do to improve the data mining results?
7. Name some of the ways in which data mining is currently used.

Exercises 3

- **3. Match the terms in table A with the statements in Table B:**

Table A	Table B
a. Data mining	i. Storage method of archiving large amounts of data to make it easy to access
b. AI	ii. Data free from duplicate and erroneous information
c. Cleansed data	iii. A process of filtering through large amounts of raw data for useful information
d. Data warehouse	iv. A computing tool that tries to operate in a way similar to the human brain

Exercises 4

- **4. Complete the following description of the data mining process using words from the text:**

Large amounts of data stored in data……are often used or data

……..The data is first……..to remove……..information and errors.

The…….is then analyzed using a tool such as……An analysis report is then analyzed by an……who decides if the…….need to be refined, other data…….tools need to be used, or if the results need to be discarded because they are……. . The analyst passes the final results to the……makers who decide on the…….action.