

Data Mining - Themes

Theoretical Foundations of Data Mining

The various theories for basis of data mining includes the following:

- **Data Reduction** - The basic idea of this theory is to reduce the data representation which trades accuracy for speed in response to the need to obtain quick approximate answers to queries on very large data bases. Some of the data reduction techniques are as follows:
 - Singular value Decomposition
 - Wavelets
 - Regression
 - Log-linear models
 - Histograms
 - Clustering
 - Sampling
 - Construction of Index Trees
- **Data Compression** - The basic idea of this theory is to compress the given data by encoding in terms of the following:
 - Bits
 - Association Rules
 - Decision Trees
 - Clusters
- **Pattern Discovery** - The basic idea of this theory is to discover patterns occurring in the database. Following are the areas that contributes to this theory:
 - Machine Learning
 - Neural Network
 - Association Mining
 - Sequential Pattern Matching
 - Clustering
- **Probability Theory** - This theory is based on statistical theory. The basic idea behind this theory is to discover joint probability distributions of random variables.
- **Probability Theory** - According to this theory data mining is finding the patterns that are interesting only to the extent that they can be used in the decision making process of some enterprise.

- **Microeconomic View** - As per the perception of this theory, the database schema consist of data and patterns that are stored in the database. Therefore according to this theory data mining is the task of performing induction on databases.
- **Inductive databases** - Apart from the database oriented techniques, there are statistical techniques also available for data analysis. These techniques can be applied to scientific data and data from economic & social sciences as well.

Statistical Data Mining

Some of the Statistical Data Mining Techniques are as follows:

- **Regression** - The regression methods are used to predict the value of response variable from one or more predictor variables where the variables are numeric. Following are the several forms of Regression:
 - Linear
 - Multiple
 - Weighted
 - Polynomial
 - Nonparametric
 - Robust
- **Generalized Linear Models** - Generalized Linear Model includes:
 - Logistic Regression
 - Poisson Regression

The model's generalization allow a categorical response variable to be related to set of predictor variables in manner similar to the modelling of numeric response variable using linear regression.

- **Analysis of Variance** - This technique analyzes:
 - Experimental data for two or more populations described by a numeric response variable.
 - One or more categorical variables (factors).
- **Mixed-effect Models** - These models are used for analyzing the grouped data. These models describe the relationship between a response variable and some covariates in data grouped according to one or more factors.
- **Factor Analysis** - Factor Analysis Method is used to predict a categorical response variable. This method assumes that independent variable follow a multivariate normal distribution.
- **Time Series Analysis** - Following are the methods for analyzing time-series data:
 - Autoregression Methods
 - Univariate ARIMA (AutoRegressive Integrated Moving Average) Modeling
 - Long-memory time-series modeling

Visual Data Mining

Visual Data Mining uses data and/or knowledge visualization techniques to discover implicit knowledge from the large data sets. The Visual Data Mining can be viewed as an integration of following disciplines:

- Data Visualization
- Data Mining

Visual Data Mining is closely related to the following:

- Computer Graphics
- Multimedia Systems
- Human Computer Interaction
- Pattern Recognition
- High performance computing

Generally data visualization and data mining can be integrated in the following ways:

- **Data Visualization** - The data in the databases or the data warehouses can be viewed in several visual forms that are listed below:
 - Boxplots
 - 3-D Cubes
 - Data distribution charts
 - Curves
 - Surfaces
 - Link graphs etc.
- **Data Mining result Visualization** - Data Mining Result Visualization is the presentation of the results of data mining in visual forms. These visual forms could be scatter plots and boxplots etc.
- **Data Mining Process Visualization** - Data Mining Process Visualization presents the several processes of data mining. This allows the users to see how the data are extracted. This also allow the users to see from which database or data warehouse data are cleaned, integrated, preprocessed, and mined.

Audio Data Mining

To indicate the patterns of data or the features of data mining results, Audio Data Mining makes use of audio signals. By transforming patterns into sound and musing instead of watching pictures, we can listen to pitches,tunes in order to identify anything interesting.

Data Mining and Collaborative Filtering

Today the consumer faced with large variety of goods and services while shopping. During live customer transactions, the Recommender System helps the consumer by making product recommendation. The Collaborative Filtering Approach is generally used for recommending products to customers. These recommendations are based on the opinions of other customers.

Source:

http://www.tutorialspoint.com/data_mining/dm_themes.htm