

UNIT - 3

Multidimensional Analysis

* Multidimensional analysis :-

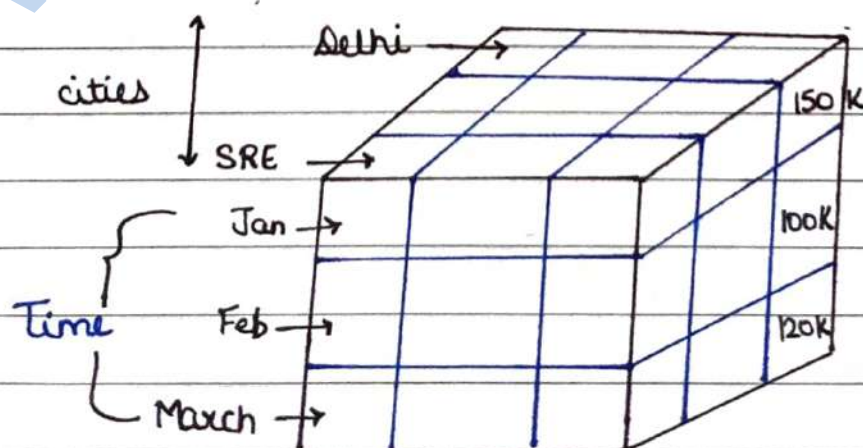
In many application, data contain structured information that is multidimensional and multi-level in nature such as- e-commerce, tele communication, retail, stock, scientific data etc.

In last decade, we have been facing research afford on data warehousing to get better view of multidimensional data. This strategy of searching pattern in multidimension databases is called data cubing.

The basic of multidimensional analysis is effective and efficient competition of agregating function

* Multidimensional data cubing :-

A multidimensional data cubing contain information for analytical purpose. It is a data structure that allows fast analysis of data according to multidimensional that define a problem. In cubing every pha face hold uniquely identity and each identity are linked together in business problem.



Computer Security
chain (K)

* Visualisation in multidimensional analysis :-

Visualisation is an area related to data mining. It is used to visualise knowledge and data. It can be used on its own (usually for description and summary task).

It can be used in combination with other data mining techniques. Visualisation can be done by following techniques -

- 1- Visualisation for decision trees
- 2- Cluster visualisation
- 3- Visualisation of association rules
- 4- Sub-group visualisation

24

* Knowledge Discovery :-

It is define as the process of identifying valid, valuable, potentially and ultimately understandable model/pattern in data.

Data mining is a key step in knowledge discovery process, performed by using data mining technique for extracting model or pattern from the data.

(Knowledge Discovery in Database (KDD)) -

● Process of Knowledge Discovery - There are five steps in knowledge discovery process -

- 1- Selection
- 2- Pre-processing
- 3- Transformation
- 4- Data mining
- 5- Interpretation evaluation

* Knowledge discovery and detection :-

In this sub-section deals with discovering the knowledge that the process apply over the organisation. Once knowledge is created it can ^{be} reuse or shared. it must be properly recognised and categorized.

Knowledge discovery and detection following with sub-section -

- 1- Explicit knowledge - This is largely process of sorting data through documents and other reports which apply many patterns like intelligence gathering, data mining, text mining etc.
- 2- Tacit knowledge - The role of this method to management of data understanding of what their company's expert actually knows. It is a process of observation and awareness any organisation terms and condition.
- 3- Embedded knowledge - This kind of knowledge trapped inside organisational routine, process, product etc.

• Knowledge discovery sub-process :-

There are two sub-process used in knowledge-

1. Combination - It enabling the discovery of new explicit knowledge.
2. Socialisation - It enabling the discovery of new tacit knowledge.

• Technologies used in Knowledge discovery :-

1. Database
2. World Wide Web
3. Sorting
4. Adding
5. Combining

* Data mining :-

It is a process of discovering knowledge, meaningful new correlation, pattern and trends. In data, By shifting through large amount of data using pattern recognition as well as statistical and mathematical techniques is called data mining.

Data mining provides tools and techniques for producing useful knowledge from information.

• Data mining applications -

- 1- In business for increasing marketing sale.
- 2- Fraud reduction
- 3- To improve product and process
- 4- For text mining
- 5- For finding new best solution for difficult problem.

* Data Sources :-

- 1- Relational database
- 2- Data warehouse
- 3- WWW
- 4- Audio
- 5- Printed material

• Data mining tasks :-

1. Predictive — It perform on current data.
2. Descriptive — It characterised general properties of data.
3. Pattern — It identify interactive method.

• Issues of data mining —

1. Data mining methodology
2. User interaction
3. Performance (accuracy / robustness)
4. Interestingness
5. Heterogeneous database

★ Clustering for database in data mining :-

This concept is define as
Set of data points, each having a set of attributes
and similarity measure among them.

- (a) Data points in one cluster are more similar to another cluster.
- (b) Data points in seperate cluster are less similar to one another.

★ Predictive data mining -

- 1- Often referred to as regression.
- 2- Data are object characterized with attributes.
- 3- Given object described with attribute numeric value.

• Predictive data mining techniques -

1- Bayesian Methods

(a) Simple Bayesian methods

(b) Main Bayesian methods

(b.a) Naive Bayesian classifier

(b.b) Semi Naive - Bayesian classifier

(b.c) Gaussian - Bayesian classifier

2- Decision tree framing

(a) Create nodes

(b) Selecting labels and classes of each nodes

3- Classification of rule learning

(a) Unordered set of rules

(b) Ordered set of rules

4- Classifier evaluation

(a) Accuracy and error

(b) N-fold cross validation

• Genetic algorithm in data mining :-

Genetic algorithm is a search heuristic that mimics the process of natural evolution. This is routinely used to generate useful solution for optimization and search problem which generate solution techniques inspired by natural evolution, inheritance, mutation, selection and cross over etc.

* Link analysis :-

Link analysis focus on analysis of relationship among nodes through visualisation methods (network chart, association matrix). There are three purpose of link analysis -

1. Find match in data for known-pattern of interact.
2. Find anomalies where known patterns are violated.
3. Discover new pattern of interaction like social media analysis.

* Decision-tree :-

Decision tree is one of the most used techniques in data mining because its model is easy to understand for users. In decision tree, the root of decision tree is the simple question or condition that have multiple answer. Each answer is lead with set of questions or condition that help us determine the data to find best answer or solution.

* Neural Network :-

These network are an approach to computing that involve developing mathematical structure with ability to learn. Neural network have broad applicability to real world business problem and have been already successfully applied in many industry areas.

Neural networks are the best at identifying patterns or trends in data. They are well-suited for prediction or future forecasting. Need including -

1. Sale forecasting
2. Industrial process control
3. Customer research
4. Data validation
5. Risk management
6. Target marketing

★ Memory Based Reasoning (MBR) :-

MBR is identifying similar like experience. It is applying the information from these cases to the problem to hand. MBR finds neighbours similar to new record and use neighbours for classification and prediction.

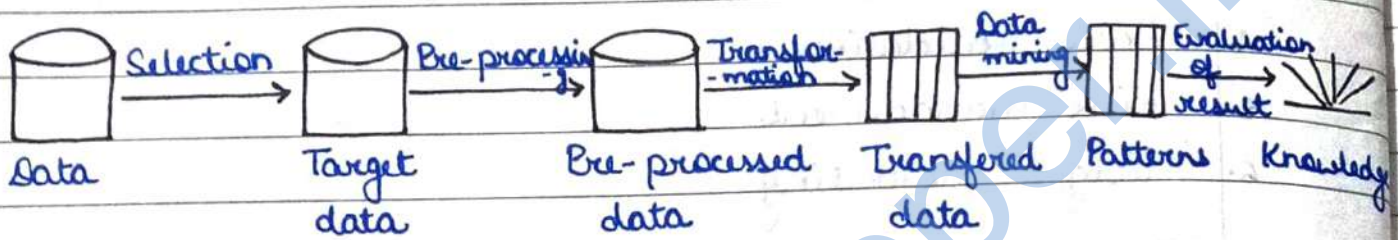
MBR mainly perform two operations -

1. Distance function (assign the distance between any two words).
2. Combination function (combine the result from the neighbour to arrive at an answer).

● Application areas of MBR -

1. Fraud detection
2. Customer Response Prediction
3. Classifying Response
4. Appropriate set of historical record
5. Help in medical treatment

Ques - How you relate data mining techniques to knowledge discovery?



Ques - Difference between data mining and data warehouse.

- 1- Collecting - data collection, data extraction through tools
- 2- Organising - The collected data need to be organised.
- 3- Summarizing - The lengthy info is presented in tabular or graphical format.
- 4- Analyzing - The info is analyzed in order to find the relationships, redundancies and patterns.
- 5- Synthesizing - At this point, information becomes knowledge.
- 6- Decision making - The knowledge is used for decision making.