Graph Mining

→ Graph mining is a process in which the mining techniques are used to find a patter or the relationship in the collection of graphs.

→ By mining the graph ~~sets~~ frequent substructures and the relationship can be identified

→ These will help in,

    (a) finding / clustering the graph sets

    (b) finding the relationship between the graph sets

    (c) characterizing the graphs.

→ By predicting these graph patterns any application model can be built in real world

→ To implement the graph mining, one must learn Frequent sub graph mining.

Frequent sub graph mining

Example: let us consider a graph h with

    an Edge set E(h) and

    vertex set V(h)

   → let us consider that

    the existence of subgraph isomorphism from h to h'

    such a way that h is a subgraph of h'.

→ Let us consider the labeled graph data set

$$F = H1, H2, H3 \ldots Hn$$

→ Let us consider the support as $S(h)$

→ find the support for each and every candidate

→ The two steps for frequent subgraph

step1: create frequent substructure candidates

step2: find the support for each and every candidate.

The two methods for frequent substructure mining are

1) The apriori based approach

2) The pattern growth approach.

1) The apprior based approach

→ This approach finds the graphs begin with the small sized graph.

→ This approach advances in bottom up way.

→ This algorithm is called apriorif graph.

2) Pattern growth approach

→ used for BFS (Breadth First search) and DFS (Depth First Search).

→ DFS are preferred due to less memory consumption.

Text mining

→ Text data mining is the process of extracting essential data from standard language text

→ All the data that we generate via text messages, documents, emails are written in common language text.

→ Text mining is useful to draw useful for patterns from such data

→ One of the primary reason for adapting the text mining is higher competition on the business market

→ Areas of Text mining.

1) Information extraction (IR)
2) Natural language processing (NLP)
3) data mining (DM)
4) Information Retrieval (IR).

} (explain each of them)

→ Text mining process
1) Text transformation
2) Text preprocessing
3) Feature selection
4) data mining
5) evaluate
6) Applications
7) Risk Management
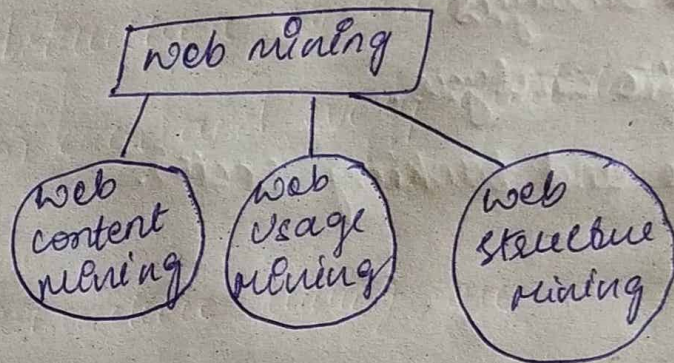8) customer care service
9) Business Intelligence
10) Social Media Analysis.

web mining

→ web mining is the process of automatically discovering and extracting the information from web documents such www.

→ world wide web has become a significant source of information for the popular business platforms as well.

→ Types of web mining

1) web content mining

→ web content mining is used to extract useful information knowledge from web page content.

→ each web page is considered as an individual document.

→ The primary task of content mining is data extraction where structured data is extracted from unstructured websites.

→ eg: user searches for a task on the search engine then the user will get a list of suggestions.

2) web usage mining

→ web ~~structure~~ usage mining used to extract the useful data information, knowledge from the weblog records.

→ It will assist in recognizing the user access patterns.

3) Web Structured Mining

→ used to find the link structure of the hyperlink.

→ It is used to find the data either links the web page or link network.

→ eg: It is useful for organization to regulate the network between two commercial sites

Data Visualization

→ data visualization is the process of graphically representing the quantitative information of the data, using graph charts and maps.

→ data visualization converts the large data set into small visualized data.

→ It provides a way to understand the outlier patterns etc.

History

→ The concept of data visualization was launched in 17th century, to understand the data from graphs, charts & maps.

→ In 1800's it was reinvented to pie chart.

→ After the decades the statistical graphics occurred.

Importance

→ It can identify the need of improvement & modification. area that

→ clarifies the each factor of customer behaviour.

→ helps to understand which products to place where.

→ it can predict the sales volume.

→ 1) graphs
2) chart
3) line chart
4) bar chart
5) pie chart

Data mining applications

1) Research

2) Education

3) Health sector

4) Transportation

5) market analysis

6) Scientific analysis.

7) finance and banking sector.