Friday
19-06-2020

DWDM Assignment-5

UNIT-5

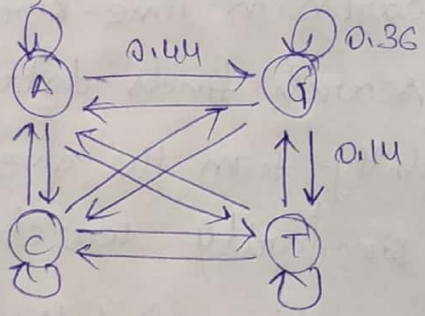
N.V.S.k.kalyani
174N1A0584
3rd B.tech CSE-C

2 Marks

1A) Spatial Data Mining:-

It is the process of discovering interesting & previously unknown, but potentially useful patterns from large spatial databases. The complexity of spatial data and intrinsic spatial relationships limits the usefulness of conventional data mining techniques for extracting spatial patterns.

3A) Markov chain:- It is a model that generates sequences in which the probability of a symbol depends only on the previous symbol.



4A) Time-series database:- It consists of sequences of values or events obtained over repeated measurements of time.

Ex:- Economic and sales forecasting utility studies and the observation of natural phenomena

2A) Hoeffding Tree:- This algorithm is a decision tree learning method for stream data classification. It was initially used to track web clickstream and construct model to predict which web hosts and web sites a user is likely to access.

17UMIDS84

5A) Stream Data :- Stream data are data that are massive, temporally ordered fast changing and potential infinize.

Ex:- Telecommunication data, transaction data from the retail and data from electronic power grids.

10 Marks

1A) world wide web mining :-

The world wide web server as a hug, widely distributed global information service center for news, advertisements consumer information, financial management, education, e-commerce, and many other information services.

* The complex web-~~research~~ to be too pages it for gather than the of any traditional text document collection.
* The web-seams to be too ~~fudges~~ for effective data ware housing and data mining.
* The web is a highly dynamic information source
* The web server a broad diversity of user communi-ities.
* only a small portion at the information on the web is truly relevant or useful.

There are several issues related to web mining and are as follows :

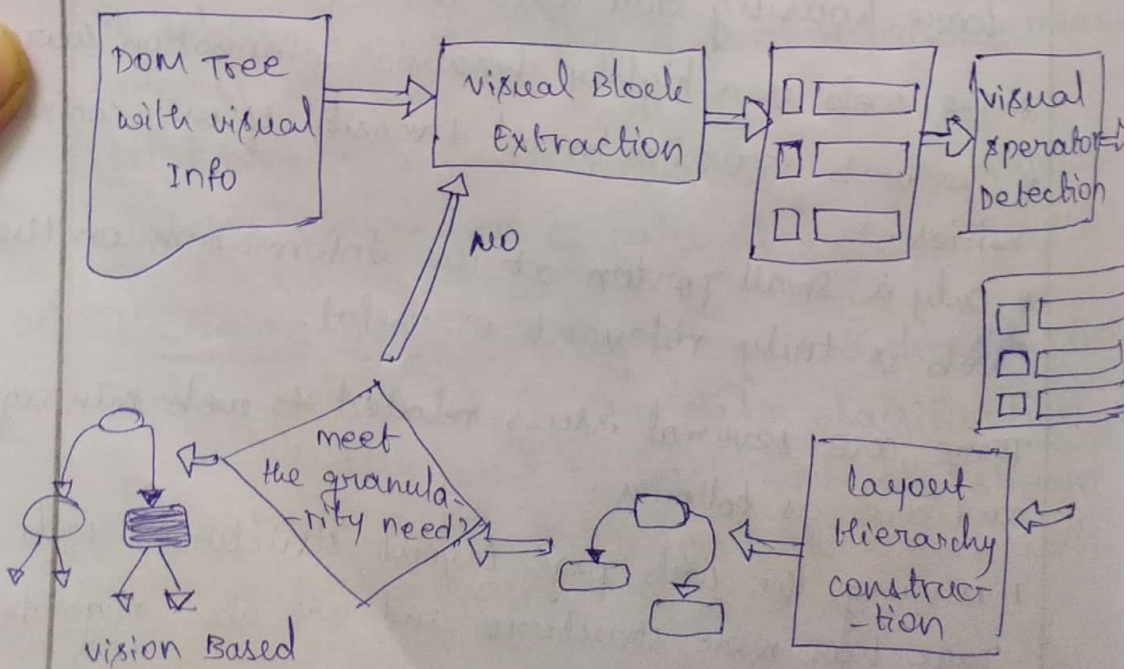
1. Mining the web page layout structure : web page has more structure and are also regarded

as Semi-structured data. The basic structure a web page is its DOM structure. The structure of a web page is a tree structure. where every Html tag in the page corresponds to a node in the DOM tree. The web page can be segmented by some predefined structure tags.

useful tags include <p> <table> , <h1> <h2> etc.,

```
<tr>
<td> <img src = "t1.jpg"> </td> <td> <img src = t2.jpg">
<td> <img src = "t3.jpg"> </td> <td> <img src = "t4.jpg"> </td>
</tr>
<tr>
<td> Timber wolf </td>, <td> Giraffes </td>
<td> Elephant sunrise </td> <td> fox </td>
</tr>
```

An Algorithm called vision-based page segmentation (VIPS) aims to extract the semantic



2. Mining the web's link structure to identify Authoriative web page; To retreving pages that are relevant and hoping a high quality which can also of hyperlinks pointing from one page to another web linkage structure has some unique features as

* Not every hyperlink represents endorsement
* The authority and hubs weights are updated based on the following Equations

$$a_p = \Sigma (q \text{ such that } q \to p)\, h_q$$

$$h_p = \Sigma (q \text{ such that } q \leftrightarrow p)\, a_q$$

3. Mining multimedia data on the web: A huge amount of multimedia data are available on the web in different forms: These Includes video audio, Images, pictures and graphs.

4. Autoniatic classifications of web document; In this each document is assigned a class label from a set of predestion topic categories, based on set of ex of documents.

1b) Spatial data mining:-

. objects of
types:- points
lines
polygons etc., Spatial data mining is the process of discovering Interesting, useful, non-trivial patterns from large spatial datasets.

→ Extracting interesting and useful patterns from spatial datasets is more difficult than extracting the corresponding patterns from traditional numeric and categorical data due to the complexity of spatial data types, spatial relationships, and spatial auto correlation.

Ex:- crime hots spots for planning police patrol routes.

Characteristics of spatial Data Mining:

* Rich data datatypes.
* Implicit spatial relationship among variable.
* Auto correlation.
* Large no.of patterns, large dataset sizes.
* patterns usually have to be defined in the spatial attributes subspace and not in the complete attribute space.
* patterns exist at different levels of granularity
* Spatial patterns, e.g. spatial clusters can have arbitrary shapes.
* Longitude and latitude are the glue that links different data collections together.

Spatial mining task:-

I. clustering: Help to find outlier detection which is useful to find suspicious knowledge.

Ex:- Group crime location.

II. Discriminate Rules: It mainly describes about the difference b/w two parts of database.

Eg:- comparison of price range of house in different graphic regions.

iii. characteristic rule: A spatial characteristic rule is general description of spatial data.

iv. classification: It defines whether a spatial entity belong to a particular class or many classes will be classified.

Eg: Remove sensor images based in spectrum data.

22) i multimedia database:- A system stores & manages a large collection of multimedia data. Such as studio, video, images, graphics, speech, text doc and hypertext data contain text, text markups, and linkages, multimedia data mining methods includes:

* Similarity search in multimedia data.
* multidimensional analysis.
* classification and prediction analysis and
* mining associations in multimedia data.

1. Similarity search in multimedia data: For Similarity searching in multimedia a data, consider two main families of multi indexing and retrieval system.

2. Description-Based retrieval system: Build indicates and perform object retreival system on image descri-ptions. such as keywords, captions, size.

2. Multidimensional Analysis of multimedia dat
3. classification and prediction Analysis
4. Mining Association in multimedia data.
5. Audio and video data mining.

ii) Time series database:-

optimized for collecting, storing, retrieving processing of time series data.

Compare this to:

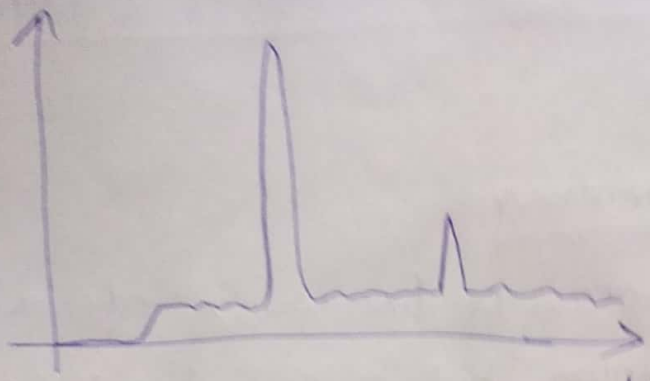
- Document databases optimized for storing doc
- search databases optimized for full-text sear
- Traditional relational Database optimized for the tabular storage of related data in row and columns.

unique Time-series:

- Time-series workload
- Lifecycle managment
- Summarization
- large range scans of many records

Time series data is a sequence of data points typically consisting of successive measurements made from the same source over a time interva

plot the points on a graph and one of your axes would always be time.



Trend analysis consists of the four major compo-nents or movements for characterizing time-series data.

* Trend or long-term movements
* cyclic movements or cyclic variations
* Seasonal movements or seasonal variations
* Irregular or random movements.

2. Similarity search in Time-series Analysis:

Similarities search finds data sequence that differ only slightly from the given query sequence. Two types of similarity searches as follows:

* Subsequence matching find the sequence in that subsequences that are similar to a given query sequence x.
* whole sequence matching finds a set of sequence in S that are similar to each other.