

MCSE First Year EXAMINATION, 2019

Second Semester

Data Warehousing and Data Mining

Time : 3 hours

Full Marks : 100

Answer question no.1 and any four from the rest

All questions carry equal marks

1. Answer true or false stating reasons 2 x10 = 20
- (a) A temporal database typically stores sequences of ordered events, with or without concrete notion of time.
 - (b) NSGA differs from conventional GAs in the 'Selection' operation only.
 - (c) A clique decision problem is an NP-complete problem.
 - (d) A data warehouse is a department subset of a data mart.
 - (e) Customer reviews may contain sentiments about different aspects about a product or service that are mentioned in the reviews.
 - (f) Non-negative matrix factorization cannot be used for document clustering.
 - (g) Multi robot path planning problem consists of only one type of collision which is between a robot and an obstacle.
 - (h) In descriptive data mining, we perform inference on the current data in order to make predictions.
 - (i) Monitoring seismic waves for earthquake activities is a data mining problem.
 - (j) The optimal set of paths in a multi robot path planning problem must include the optimal paths for all the individual robots.
2. (a) Describe with a neat diagram the architecture and operation of a typical KDD process with an example. 14
- (b) Discuss briefly the limitations of a typical DBMS in handling data mining tasks. 6
3. (a) Describe briefly the significance of any three of the following advanced database systems in connection to data mining applications. 3 X 4

- i. Object-Relational Databases
- ii. Spatial databases
- iii. Text databases
- iv. World Wide Web.

(b) Describe how data mining can be useful in fraud detection and management. 8

4. (a) Narrate the significance of optimization techniques in solving problems of data mining. 4

(b) The following figure represents a simulated environment through which a robot can move around. The environment has been divided into a number of orderly numbered grids. The grids are rectangular in shape and equal sized. Shaded grids represent obstacles and the others represent spaces through which a robot can traverse freely. Now consider a multi robot path planning problem whose objective is to plan an optimal set of collision free paths where the paths are associated with each of the individual robots from a predefined start node (S) to target node (T) considering distance as the optimization criteria.

1	2	3	4	5	6
S					
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
					T

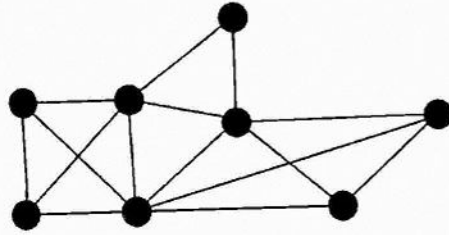
Describe a GAs based method to solve the aforesaid problem. 12

(c) What are the merits and demerits of your proposed GAs based method. 4

5. (a) State briefly the different types of community in a network. 4

(b) Describe a NSGA-II based algorithm for overlapping community detection problem. 12

- (c) Explain *Maximal k clique percolation* algorithm for finding overlapping communities on the following network considering the value of k as 3. 4



6. (a) What is aspect based opinion mining? What are the basic components of a typical aspect based opinion mining system? 2+3

- (b) Describe briefly a suitable technique for extracting certified buyers' opinion about different aspects of a product from an e-commerce website. Point out the merits and demerits of your proposed method. 12+3

7. Write short notes on any two of the following. 2 X 10

- (a) Pattern evaluation
- (b) Market basket analysis
- (c) Data Cubes