Reg. No. : ☐☐☐☐☐☐☐☐☐☐☐

Question Paper Code: 31983

B.E. / B.Tech. DEGREE EXAMINATION, MAY 2016

Elective

Information Technology

01UIT921 - DATA MINING CONCEPTS AND TECHNIQUES

(Regulation 2013)

Duration: Three hours   Maximum: 100 Marks

Answer ALL Questions

PART A - (10 x 2 = 20 Marks)

1. List the major features of a data warehouse system.

2. Enumerate three classes of schemas that are popularly used for modeling data warehouses.

3. Define data mining.

4. The mean and the SD values for the attribute income are $54,000 and $16,000 respectively. Transform the attribute income $70,000 using Z-Score normalization.

5. What are the two interestingness measures related to association rule mining?

6. Summarize the challenges faced by the Frequent Pattern Mining.

7. Given two objects represented by the tuples (22, 1, 42, 10) and (20, 0, 36, 8):
   (i) Compute the Euclidean distance between the two objects
   (ii) Compute the Manhattan distance between the two objects

8. What is an outlier?

9. What do you meant by HITS?

10. Define spatial data mining.

PART - B (5 x 16 = 80 Marks)

11. (a) (i) Compare and contrast OLAP with OLTP. (8)

(ii) Illustrate with suitable diagram the various data models. (8)

Or

(b) (i) Suppose that a data warehouse consists of the three dimensions time, doctor, and patient, and the two measures count and charge, where charge is the fee that a doctor charges a patient for a visit.

(1) Enumerate three classes of schemas that are popularly used for modeling data warehouses

(2) Select the best suitable schema diagram for the above data warehouse

(3) Starting with the base cuboid [day, doctor, patient], what specific OLAP operations should be performed in order to list the total fee collected by each doctor in 2004? (12)

(ii) How is indexing performed in a data warehouse? (4)

12. (a) (i) Describe the various data mining functionalities in detail. (10)

(ii) Justify how the evolution of database technology led to data mining. (6)

Or

(b) Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70.

(i) What is the mean of the data? What is the median?

(ii) What is the mode of the data? Comment on the data's modality (i.e., bimodal, trimodal, etc.).

(iii) What is the midrange of the data?

(iv) Can you find (roughly) the first quartile (Q1) and the third quartile (Q3) of the data? (16)

2

31983

13. (a) A database has five transactions. Let min sup = 60% and min con f = 80%.

TID	items bought
T1	{ M, O, N, K, E, Y}
T2	{ D, O, N, K, E, Y }
T3	{ M, A, K, E }
T4	{ M, U, C, K, Y }
T5	{ C, O, O, K, I ,E }

    (i) Find all frequent item sets using Apriori algorithm

    (ii) List all strong association rules with support and confidence   (16)

Or

(b) Briefly outline the major steps involved in decision tree classification with an example.   (16)

14. (a) With suitable illustration point-out the methodology, pros and cons of any three types of clustering algorithms.   (16)

Or

(b) Briefly describe the different approaches behind statistical-based outlier detection, distanced-based outlier detection, density-based local outlier detection, and deviation-based outlier detection.   (16)

15. (a) What do you meant by Web document classification and analyzes how to use Web linkage information to improve the quality of classifying documents.   (16)

Or

(b) Explain spatial data mining by considering the highway traffic scenario.   (16)

—————————

31983