Code: 9A05706

R09

B.Tech IV Year II Semester (R09) Advanced Supplementary Examinations, July 2013

DATA WAREHOUSING & DATA MINING

(Electronics and Computer Engineering)

Time: 3 hours Max. Marks: 70

Answer any FIVE questions

All questions carry equal marks

*****

1 (a) List and describe the five primitives for specifying a data mining task.

(b) In real world data, tuples with missing values for some attributes are a common occurrence. Describe various methods for handling this problem.

(c) Define each of the following data mining functionalities:
(i) Characterization. (ii) Discrimination. (iii) Classification.

2 (a) Briefly compare the following concepts. You may use an example to explain your points:
(i) Snowflake schema, fact constellation starnet query model.
(ii) Data cleaning, data transformation, refresh.
(iii) Enterprise warehouse, data mart, virtual warehouse.

(b) What are the differences between the three main types of data warehouse usage: information processing, analytical processing and data mining? Discuss the motivation behind OLAP mining.

3 (a) Find association rules with 50% support and 75% confidence for the transactions given below.

Tid	Items
1	I_1, I_2, I_3, I_4
2	I_1, I_2, I_4
3	I_1, I_5, I_6
4	I_1, I_4, I_5
5	I_2, I_4, I_5

Generate frequent item sets using candidate generation method.

(b) What is multilevel association mining? Illustrate with examples how multilevel association rules can be mined.

4 (a) Discuss the major steps of decision free classification.

(b) Describe the K-nearest neighbor classifier.

Contd. in Page 2

Code: 9A05706

R09

5 (a) Suppose that the data mining task is to cluster the following 8 points into 3 clusters:

A_1 (2, 10), A_2 (2,5), A_3 (8,4)
B_1 (5, 8), B_2 (7, 5), B_3 (6, 4)
C_1 (1, 2), C_2 (4, 9)

Distance function is Euclidean distance. Suppose initially we assign A_1, B_1 and C_1 as the center of each cluster respectively. Use the K-means algorithm to show only (i) The 3 cluster centers after the first round execution. (ii) The final 3 clusters.

(b) How does DBSCAN find clusters? Explain.

6 (a) What is data stream mining? Discuss the stream OLAP and stream data cubes.

(b) Outline an efficient method that may find strong correlation rules in a large multi relational database.

7 (a) Discuss the basic measures for text retrieval.

(b) Discuss the different categories of association that can be mined in multimedia data.

8 (a) Write a short notes on data mining system products and research prototypes.

(b) Discuss the social impacts of data mining.

*****