MCA DEGREE IV SEMESTER EXAMINATION APRIL 2013

CAS 2402 DATA MINING

Time: 3 Hours Maximum Marks: 50

PART A
(Answer *ALL* questions)

(15 × 2 = 30)

I. (a) Define the KDD process.
(b) State some of data mining application areas.
(c) Explain sparse data, missing values and inaccurate values with examples.

II. (a) Give details on multidimensional hierarchies used in mining.
(b) State the OLAP rules.
(c) Explain techniques for data visualization.

III. (a) What is replicated subtree problem?
(b) List three important evaluation criteria for classification methods.
(c) Explain decision tree induction algorithm.

IV. (a) What are closed and maximal item sets? What are their relevance?
(b) Using probability define the concepts of support and confidence.
(c) Explain dynamic itemset counting algorithm.

V. (a) Describe the role of scaling and weighting in K-means algorithm.
(b) What is agglomerative method? Explain two situations suitable for this method.
(c) Explain how the quality of a clustering algorithm may be determined.

PART B

(5 × 4 = 20)

VI. A. Define data mining. Discuss data mining tasks.

OR

B. Outline the various measures of similarity and dissimilarity.

VII. A. Briefly describe summary statistics as ways to summarize the data.

OR

B. Describe the different multidimensional analysis techniques.

VIII. A. What is decision tree induction? Describe an algorithm for the same.

OR

B. What is lazy learner method? Discuss an algorithm for the same.

IX. A. Consider the following set of frequent itemsets {1, 2, 3}, {1, 2, 4}, {1, 2, 5}, {1, 3, 4}, {1, 3, 5}, {2, 3, 4}, {2, 3, 5}, {3, 4, 5}. Assume there are only five items. List all candidate 4-item sets using apriori algorithm.

OR

B. Illustrate the FP-Growth algorithm for frequent itemset generation.

X. A. Describe the basic agglomerative hierarchical clustering algorithm. How proximity between two clusters are defined?

OR

B. Outline DBSCAN algorithm. Illustrate it with an example.

***