Code: 9A05706

1

B.Tech III Year II Semester (R09) Regular & Supplementary Examinations, April/May 2013

DATA WAREHOUSING & DATA MINING

(Information Technology)

Time: 3 hours Max. Marks: 70

Answer any FIVE questions
All questions carry equal marks

*****

1. (a) What is data mining? Explain its role in knowledge discovery process.
   (b) Discuss concept hierarchy generation for categorical data with examples.

2. (a) Give the three-tier data warehouse architecture. Explain it.
   (b) Explain BUC algorithm for the computation of sparse or iceberg queries.

3. What is a frequent item set? How to find frequent item sets for a transactional database? Explain any one approach with illustrations.

4. (a) Discuss rule quality measures.
   (b) What is the significance of learning rate in back propagation algorithm?
   (c) How to measure the accuracy of a classifier? Explain.

5. (a) Discuss the typical requirements of clustering in data mining.
   (b) Describe deviation-based outlier detection.

6. (a) Explain Viterbi algorithm.
   (b) Discuss mining alternative substructure patterns in graph mining.

7. Describe various types of text databases. What is meant by text mining? Which data mining functionalities are applicable to text databases?

8. (a) How to choose a data mining system? Discuss.
   (b) Discuss ubiquitous and invisible data mining.

*****

Code: 9A05706

2

B.Tech III Year II Semester (R09) Regular & Supplementary Examinations, April/May 2013

DATA WAREHOUSING & DATA MINING

(Information Technology)

Time: 3 hours Max. Marks: 70

Answer any FIVE questions
All questions carry equal marks

*****

1. (a) Discuss briefly various data mining functionalities.
   (b) Explain dimensionality reduction as a preprocessing activity.

2. (a) Define data warehouse. Differentiate between data warehouse and database system.
   (b) Explain mining class comparisons with AOI.

3. (a) Discuss ECLAT algorithm to find frequent patterns.
   (b) Explain association rule clustering system with examples.

4. (a) Why information gain is considered as attribute selection measure? Illustrate with an example.
   (b) How to derive rules from a decision tree?
   (c) Discuss ensemble methods to increase the accuracy of a classifier.

5. (a) Discuss interval-scaled variables and their standardization.
   (b) Discuss the categorization of major clustering methods.
   (c) Describe a typical dimension-reduction sub space clustering methods.

6. (a) Explain Baum – Welch algorithm.
   (b) What is a social network? Discuss its characteristics.

7. (a) How to construct a spatial data cube? Discuss the types of measures in a spatial data cube.
   (b) Describe similarity search in multimedia data.
   (c) Explain locality preserving indexing.

8. Discuss data mining for biological data analysis.

*****

Code: 9A05706

3

B.Tech III Year II Semester (R09) Regular & Supplementary Examinations, April/May 2013

DATA WAREHOUSING & DATA MINING

(Information Technology)

Time: 3 hours Max. Marks: 70

Answer any FIVE questions
All questions carry equal marks

*****

1. What is the need for preprocessing the data? Explain briefly various forms of data preprocessing.

2. (a) Discuss multidimensional data model and explain various schemes for multidimensional data model.
   (b) Explain indexing OLAP data.
   (c) Describe types of OLAP servers.

3. (a) Find frequent itemsets for the following table using FP-Growth algorithm. Assume relevant thresholds.

T_{id}	List of item ids
T_1	I_1, I_3, I_5
T_2	I_2, I_4, I_1
T_3	I_1, I_2, I_3, I_4
T_4	$I_5, I_3, I_2,$
T_5	$I_1, I_2, I_5,$
T_6	$I_3, I_4, I_5,$

   (b) Discuss constraint based mining.

4. (a) State Baye's theorem. Explain how it can be adopted for classification.
   (b) Describe case-based reasoning as a lazy learner.
   (c) List the measures for classifier's accuracy.

5. (a) Discuss hierarchical methods for clustering. List their merits and demerits.
   (b) Explain statistical based outlier detection.

6. Discuss the characteristics of social networks and the tasks challenges in mining social networks

7. (a) Write about probabilistic latent semantic indexing method.
   (b) Explain HITS algorithm.
   (c) What is meant by an object cube?

8. (a) Discuss data mining for financial data analysis in brief.
   (b) Write a note on statistical data mining.

*****

Code: 9A05706

4

B.Tech III Year II Semester (R09) Regular & Supplementary Examinations, April/May 2013

DATA WAREHOUSING & DATA MINING

(Information Technology)

Time: 3 hours Max. Marks: 70

Answer any FIVE questions
All questions carry equal marks

*****

1. (a) List and explain data mining task primitives.
   (b) How to measure the central tendency of data?
   (c) Describe data cleaning process.

2. (a) Explain OLAP operations in the multidimensional data model.
   (b) Discuss star-cubing algorithm.

3. (a) Consider the following table to find frequent item sets using vertical data format. Support threshold 30%

T_{id}	List of items
T_{01}	milk, biscuits, surf powder, teabags
T_{02}	teabags, sugar, soap
T_{03}	milk, sugar, bread, soap
T_{04}	bread, teabags, biscuits
T_{05}	chocolates, milk, biscuits
T_{06}	milk, teabags, bread
T_{07}	bread, biscuits, chocolates
T_{08}	milk, surf powder, bread

   (b) How to mine multilevel association rules? Discuss

4. (a) Explain classification by association rule analysis.
   (b) How does a Bayesian belief network learn?
   (c) What is the necessity of tree pruning in decision tree induction?

5. (a) Discuss chameleon algorithm for clustering.
   (b) Describe model-based clustering methods briefly.

6. (a) What is multirelational data mining?
   (b) Discuss mining customers' networks for viral marketing.
   (c) Describe Hoeffding Tree algorithm.

7. (a) What is a multimedia database? Explain multidimensional analysis of multimedia data.
   (b) Discuss the basic measures for text retrieval.
   (c) Describe DOM structure of a web page.

8. Explain the social impacts of data mining in detail.

*****