Ex/PG/MTCT/T/122A/2018

M.TECH. COMPUTER SCIENCE FIRST YEAR SECOND SEMESTER - 2018

DATABASE TECHNOLOGY AND DATA MINING

Time : Three Hours. Full Marks : 100

*Answer Question No.***1,** *7 and any* **THREE** *from the rest*

1. (a) Distinguish with example between 3NF and BCNF
   (b) What do you mean by first quartile (Q1) and third quartile (Q3) of the data? Give a suitable example.
   (c) How do you measure the quality of a good cluster?
   (d) What is Outlier? Outlier are different from the noise data ?
   (e) Indicate the role of support and confidence in data mining. Give suitable example [4×5=20]

2. (a) Consider the following two sets of functional dependencies
   F= {A →C, AC →D, E →AD, E →H} and G = {A →CD, E →AH}. Check whether or not they are equivalent.
   (b) Consider the following relation for published books:
   BOOK (Book_title, Authorname, Book_type, Listprice, Author_affil, Publisher)
   Author_affil referes to the affiliation of the author. Suppose the following dependencies exist:
   Book_title → Publisher, Book_type
   Book_type → Listprice
   Author_name → Author-affil

   (i) What normal form is the relation in? Explain your answer.
   (ii) Apply normalization until you cannot decompose the relations further. State the reasons behind each decomposition. [7(3+10)]

3. (a) Show how you may specify the following relational algebra operations in both tuple and domain relational calculus. [6+(2+12)]

(i) PROJECT <A, B> (R(A, B, C)):
(ii) R(A, B, C) UNION S(A, B, C):
(iii) R(A, B, C) INTERSECT S(A, B, C):

(b) What is functional dependency? Prove or disproof of the following inference rules for functional dependencies.

(i) $\{X \rightarrow Z, Y \rightarrow Z\} \models \{X \rightarrow Y\}$
(ii) $\{X \rightarrow Y, XY \rightarrow Z\} \models \{X \rightarrow Z\}$
(iii) $\{XY \rightarrow Z, Y \rightarrow W\} \models \{XW \rightarrow Z\}$

4. (a) What is Similarity and Dissimilarity? Proximity? Data Matrix and Dissimilarity Matrix? Give suitable example Proximity Measure for Nominal Attributes and Proximity Measure for Binary Attributes?

(b) Cosine Similarity ?
Find the Cosine Similarity between the documents?(d1&d2, d1&d3, d1&d4, d2&d3, d2&d4, d3&d4) [8+(2+10)]

Document	*team*	*coach*	*hockey*	*baseball*	*soccer*	*penalty*	*score*	*win*	*loss*	*season*
Document1	5	0	3	0	2	0	0	2	0	0
Document2	3	0	2	0	1	1	0	1	0	1
Document3	0	7	0	2	1	0	0	3	0	0
Document4	0	1	0	0	1	2	2	0	3	0

5. (a) Discuss supervised learning vs. unsupervised learning. Elaborate the role(s) of Entropy, Information Gain and Gain Ratio in attribute selection

(b) Discuss the Apriori algorithm with example and its demerits. Compare Apriori algorithm with FP-growth algorithm in mining frequent pattern. [6+(7+7)]

6. (a) What is a cluster ?What is cluster analysis ? Explain with example(s) Clustering for Data Understanding and Applications

(b) Utility of Clustering? Indicate the Major Clustering Approaches? Discuss the advantages and disadvantages of partitioning based clustering approach. [(1+2+4)+(2+4+7)]

7. Any Four 4x5=20

(a) TRC vs. DRC
(b) Cardinality vs. Participation
(c) K-Means vs. K-Medoids
(d) Classification vs. Numeric Prediction
(e) Gain Ratio vs. Gini Index
(f) Dimensionality Reduction vs. Numerosity Reduction