

Ph.D/M.Tech (Computer Engineering), 2nd Semester Examination 2018
 Data Mining & Analytics
 Paper Code: MCEN – 201/MCEN-201(Ph.D)

Max. Marks: 60

Time: 3 Hours

Instruction to the candidates:

Attempt all questions by attempting any two parts from each question.
 Each part of the question carries 6 marks.

✓1.(a)	<p>Suppose we have the following two-dimensional data set:</p> <table border="1"> <thead> <tr> <th></th><th>A1</th><th>A2</th></tr> </thead> <tbody> <tr> <td>X1</td><td>1.5</td><td>1.7</td></tr> <tr> <td>X2</td><td>2.0</td><td>1.9</td></tr> <tr> <td>X3</td><td>1.6</td><td>1.8</td></tr> <tr> <td>X4</td><td>1.2</td><td>1.5</td></tr> <tr> <td>X5</td><td>1.5</td><td>1.0</td></tr> </tbody> </table> <p>Consider the data as two-dimensional data points.</p> <ul style="list-style-type: none"> I) Given a new data point, $x = (1:4;1:6)$ as a query, rank the database points based on similarity with the query using <ul style="list-style-type: none"> (i) Euclidean distance, (ii) Cosine similarity. II) Normalize the data set to make the norm of each data point equal to 1. Use Euclidean distance on the transformed data to rank the data points. 		A1	A2	X1	1.5	1.7	X2	2.0	1.9	X3	1.6	1.8	X4	1.2	1.5	X5	1.5	1.0	6
	A1	A2																		
X1	1.5	1.7																		
X2	2.0	1.9																		
X3	1.6	1.8																		
X4	1.2	1.5																		
X5	1.5	1.0																		
✓2(b)	<p>Explain the Principal Component Analysis (PCA) as applied in dimensionality reduction.</p>	6																		
✓3(c)	<p>Consider the double star graph given below with n nodes, where only nodes 1 and 2 are connected to all other vertices, and there are no other links. Answer the following questions (treating n as a variable).</p>	6																		

- (i) What is the degree distribution for this graph?
 (ii) What is the mean degree?
 (iii) What is the clustering coefficient for vertex 1 and vertex 3?
 (iv) What is the clustering coefficient $C(G)$ for the entire graph? What happens to the clustering coefficient as $n \rightarrow \infty$?
 (v) What is the betweenness value for node 2?

Q.a)

Given a simple transactional database.

6

Tid	Items
1	A,B,C,D
2	A,C,D,F
3	C,D,E,G,A
4	A,D,F,B
5	B,C,G
6	D,F,G
7	A,B,G
8	C,D,F,G

A, B, C, D

25 + 10
Total 35
8

Using the threshold support of 25% and confidence of 60%,

- i) Find all large itemsets in the database.
 ii) Find strong association rules for the database.
 iii) Analyze misleading associations for the rule set obtained above.

(b) (i)

We generally will be more interested in association rules with high confidence. However, often we will not be interested in association rules that have a confidence of 100%. Why? Then specifically explain why association rules with 98% or 99% confidence may be interesting (i.e., what might they indicate)?

6

(ii)

If the data set contains missing values, discuss the basic analyses and corresponding decisions you will take in the pre-processing phase of the data mining process.

Q.e)

You are given the transaction data shown in the Table below from a fast food restaurant. There are 9 distinct transactions (order : 1 – order : 9) and each transaction involves between 2 and 4 meal items. There are a total of 5 meal items that are involved in the transactions. For simplicity we assign the meal items short names (M1 – M5) rather than the full descriptive names.

6

Meal Item	List of item I.D.s	Meal Item	List of item I.D.s
Order : 1	M1, M2, M5	Order : 6	M2, M3
Order : 2	M2, M4	Order : 7	M1, M3
Order : 3	M2, M3	Order : 8	M1, M2, M3, M5
Order : 4	M1, M2, M4	Order : 9	M1, M2, M3
Order : 5	M1, M3	/	

The minimum support is $2/9$ and the minimum confidence is $7/9$. Note that you only need to achieve this level, not exceed it.

Apply the Apriori algorithm to the dataset of transactions and identify all frequent k itemsets. Show all of your work. You must show all candidates but can cross them off to show the ones that pass the minimum support threshold.

- 3.(a)(i) Use CHARM algorithm to find closed frequent itemset for the given dataset. Show the steps clearly. 6

Tid	Itemset
1	ABDE
2	BCE
3	ABDE
4	ABCE
5	ABCDE
6	BCD

- (ii) Is Maximal frequent itemsets are sufficient to determine all frequent itemsets with their support? 6

Given the following three sequences:

S1 : GAAGT

S2 : CAGAT

S3 : ACGT

Find all frequent subsequences with minsup = 2, but allowing at most a gap of 1 position between successive sequence elements. Explain clearly the steps on how the answer has been found out.

- (c) Suppose we have market basket data consisting of 100 transactions and 20 items. If the support for item A is 25%, the support for item B is 90% and the support for itemset {A,B} is 20%. Let the support and confidence threshold be 10% and 60% respectively.

6

- i) Compute the confidence of the association rule $\{A\} \rightarrow \{B\}$. Is the rule interesting according to the confidence measure?
- ii) Compute the interest measure for the association pattern $\{A, B\}$. What is the relationship between item A and item B in terms of interestingness?
- iii) Prove that if the confidence of the rule $\{A\} \rightarrow \{B\}$ is less than the support of $\{B\}$, then:
- $$\text{confidence } (\{\bar{A}\} \rightarrow b) > \text{confidence } (\{\bar{A}\} \rightarrow \{b\})$$
- $$c(\{\bar{A}\} \rightarrow \{B\}) > \text{support } (\{B\})$$

- 4.(a) The following dataset will be used to learn a decision tree for predicting whether a mushroom is edible or not based on its shape, color and odour.

6

Shape	Color	Odour	Edible
C	B	1	YES ✓
D	B	1	YES ✓
D	W	1	YES ✓
D	W	2	YES ✓
C	B	2	YES ✓
D	B	2	NO
D	G	2	NO
C	U	2	NO
C	B	3~	NO
C	W	3~	NO
D	W	3~	NO
C	B	2	NO
D	B	2	NO
C	W	2	YES ✓

BNGC

- i) What is the entropy $H(\text{Edible} | \text{Odour} = 1 \text{ or } \text{Odour} = 3)$?
 Which attribute would the C4.5 algorithm choose to use as the root of the tree?

- (b) Given a set of 5-D categorical samples

6

$$A = (1, 0, 1, 1, 0)$$

$$B = (1, 1, 0, 1, 0)$$

$$C = (0, 0, 1, 1, 0)$$

$$D = (0, 1, 1, 1, 0)$$

$$E = (1, 0, 1, 0, 1)$$

$$F = (0, 1, 1, 0, 0)$$

Suppose that the samples are distributed into two clusters:
 $C_1 = \{A, B, E\}$ and $C_2 = \{C, D, F\}$

Using K-nearest neighbor algorithm, find the classification for the following samples.

- i) $Y = (1, 1, 0, 1, 1)$ using $K = 1$
- ii) $Z = (0, 1, 0, 0, 0)$ using $K = 3$

What are ensemble methods? How bagging and boosting helps to enhance the accuracy of a classifier?

6

5. (a) Suppose we have the following data

A(2,0); B(1,2); C(2,2); D(3,2); E(2,3); F(3,3); G(2,4); H(3,4); I(4,8); J(3,5);

6

Identify the clusters by applying k-means algorithm with $k = 2$. Try using initial cluster centers as far as possible.

What is the principle of working of Expectation Maximization algorithm and how it is different from K-means?

6

Explain the working principle of any clustering algorithm that can find clusters of irregular shape. Why K-means and K-medoids can't find such clusters?

6

M.Tech./Ph.D.(Computer Engineering) II semester Examination 2018

Parallel Computing

Paper No: MCEN-202/MCEN202(Ph.D.)

Time:-03 Hours

Max. Marks:-60

Write your Roll No. on top immediately on receipt of question paper.

Attempt all questions, by attempting any two parts from each question.
Marks are written in the margin.

Q1(a) Consider a computer which can execute a program in two operational modes; regular mode versus enhanced mode with a probability distribution of $\{A, 1-A\}$ respectively. If A varies between a and b and $0 \leq a \leq 1$, derive an expression for the average speedup factor. (6)

Q1 (b)(i) Derive an expression for optimal number of stages in a pipeline.

? (ii) Distinguish between Superscalar Processor and VLIW processor (6)

Q1(c) What are Bernstein's conditions? Explain with help of suitable example. A sequential program consists of the following five statements S1 through S5. consider each statement as a separate process. Clearly identify input set I_i and output set O_i of each process. Restructure the program using Bernstein's conditions in order to achieve maximum parallelism between processes. If any pair of processes cannot be executed concurrently specify which of the Bernstein's conditions are not satisfied. (6)

S1: $A=B+C$ S2: $C=B*D$ S3: $S=0$ S4: For($I=0; I \leq 100; I++$) $S=S+X[I]$ S5: if($S > 1000$) $C=C^2$

Q2(a) Describe the following laws (6)

(i) Gustafson's law

(ii) Sun and Ni's law

Q2(b) Explain the various bounds on MAL. Consider the following reservation table for a

four stage pipeline with a clock cycle of 2ns. (6)

	1	2	3	4	5	6
S1	X					X
S2		X		X		
S3			X			
S4				X	X	

You are allowed to insert one non compute delay stage into this pipeline to make a latency of 1 permissible in the shortest greedy cycle. The purpose is to yield a new reservation table leading to an optimal latency equal to the lower bound. (6)

- ✓(i) find the modified reservation table
- ✓(ii) Draw state transition diagram for scheduling the pipeline.
- ✓(iii) find out all simple cycles and greedy cycle.
- ✓(iv) Prove that the MAL equals the lower bound
- ✓(v) What is the optimal throughput of this pipeline.

Q2(c) Discuss the architecture of Intel I7 processor. (6)

Q3(a) What is Tomasulo's algorithm? Explain its working with the help of suitable example and diagram. (6)

Q3(b) Describe Branch prediction techniques with help of suitable example and diagram. (6)

Q3(c) Explain the following Techniques. (6)

- (i) Software pipeline
- (ii) Simultaneous Multi threading (SMT) technique.
- (iii) Branch Predication technique.

Q4(a) Design and verify SIMD algorithm for $N \times N$ Matrix multiplication with computational complexity as $O(N^2)$. (6)

Q4(b) What is PRAM model? Design and verify PRAM Algorithm for EREW $N \times N$ matrix multiplication with computational complexity of $O(N)$. (6)

Q4(c) Parallelize Bucket sort. Compute its speed up with respect to its sequential algorithm. (6)

Q5(a) What is MPI? Write MPI program for computing $N!$ where N is an integer. (6)

Q5(b) How is open MP program written? Write an Open MP program to find maximum of N numbers. (6)

Q5(c) Explain how programs are written to exploit massive parallelism available in GPUs with the help of suitable programming example. (6)

M.Tech. Comp. Engg. IInd Sem Examination 2018
Soft Computing Techniques
Paper no MCEN 203

Time : Three hoursMax. Marks: 60

Note: Attempt all questions

Q1 a) What do you understand by Soft Computing? What are the various soft computing techniques? Briefly describe each using suitable examples. **6 Marks**

Q1 b) Differentiate between Hard Computing and Soft Computing in terms of problem solving. Also draw comparisons between Fuzzy logic and probability theory using examples. **6 Marks**

OR

Q1'a) Explain the architecture of a basic neural network, using a well labeled diagram, along with each of its internal components. **6 Marks**

Q1'b) What do you understand by Learning? Explain each of its types using suitable example. **6 Marks**

Q2'a) Briefly describe the various activation functions used in the design of neural networks. Also explain the McCulloch-Pitt's neuron model. Design an MP Neuron for AND problem. **6 Marks**

Q2'b) Explain the ADALINE architecture and its training process in details. **6 Marks**

OR

Q2' a) Design a perceptron network for the NOR function using binary inputs and bipolar targets. **6 Marks**

Q2' b) Explain the Error Backpropagation learning rule in details. Also derive the corresponding weight update rule for the same. **6 Marks**

Q3'a) Explain the various Heuristics used to improve the efficiency of the Backpropagation algorithm in details. **6 Marks**

Q3'b) Construct the Hetro associative network to store the vector $\{-1 \ 1 \ 1 \ -1, \ 1; \ 1 \ -1 \ -1 \ 1, \ 0\}$.
Also test your network using suitable inputs. **1x4** **6 Marks**

Q4a) Using suitable examples describe and verify the following properties in fuzzy sets: **6 Marks**

i) Commutativity

ii) Idempotency

iii) Identity

[P.T.O.]

Q4b) Explain the Mamdani FIS model using suitable block diagram. Also describe its points of difference from the Sugeno FIS model. **6 Marks**

Q5a) Describe the various Defuzzification methods using suitable diagram. Also consider the following two fuzzy sets A & B defined as **6 Marks**

$$\tilde{A} = \left\{ \frac{0.2}{x_1}, \frac{0.3}{x_2}, \frac{0.4}{x_3}, \frac{0.7}{x_4}, \frac{0.1}{x_5} \right\} \quad \tilde{B} = \left\{ \frac{0.4}{x_1}, \frac{0.5}{x_2}, \frac{0.6}{x_3}, \frac{0.8}{x_4}, \frac{0.9}{x_5} \right\}$$

Now generate the λ -cut sets using Zadeh notation of the followings

- i) $(\tilde{A} \cup \tilde{B})_{0.9}$
- ii) $(\tilde{B} \cap \tilde{A})_{0.5}$

Q5 b) Explain the Genetic algorithm in details. Explain how this technique can help you achieve the global minima.. Discuss some suitable examples where application of Genetic techniques will help yield better results. **6 Marks**

M.Tech. (Computer Engineering) 2nd Semester Examination, 2018

MOBILE COMPUTING

Paper No. MCEN-204

Time: Three Hours

Maximum Marks: 60

(Write your roll number on the top immediately on receipt of this question paper)

Note: Attempt any two parts from each question. All questions carry equal marks.

Q1. a. What is the concept of signal power in mobile computing? If 100 watt is applied to a unit gain antenna with 600MHz carrier frequency, find the received power in dBm at a free distance of 200m from the antenna.

b. Describe the frequency reuse concept in cellular communication system. Using the concept of cell geometry and co-ordinate system, drive the equation for the frequency reuse ratio.

A cellular system uses a frequency reuse factor $N=4$ ($i=0, j=2$). If the path loss exponent $\gamma=4$ and cell radius $R=5$ km, find the following quantities in decibels:

- The SIR for the system with no cell sectoring
- The SIR for the system when 120° sectoring is used
- The SIR for the system when 60° cell sectoring is used

Q1. c. What is GSM? Explain the architecture of GSM. What are the different types of logic channels? How these differ from physical channel?

Q2. a. What is Adaptive Multi Rate (AMR) in GSM? Explain the difference between Half rate and Full, rate AMR.

The power difference between the DPCCh and DPDCh of GSM is -3.0 dB for 12.2 kbps AMR speech. Calculate the coverage gain in link in dB by reducing the AMR bit rate from:

- 12.2 kbps to 7.95 kbps
- 12.2 kbps to 4.75 kbps

[P.T.O.]

b. Explain the architecture of CDMA with the help of block diagram. Also explain the key features of CDMA-2000. 6

c. How chip sequences are generated in CDMA? Explain direct sequence spread spectrum (DSSS) and frequency hopped spread spectrum (FHSS) in detail. 6

If the CDMA demux received the code (-1, -1, -3, +1), find the bits sent by the different stations.

Q3. a. Calculate the uplink cell load factor and number of voice users for WCDMA system using the following data: 6

Information rate $R_i = 12.2 \text{ kbps}$

Chip rate $R_c = 3.84 \text{ MCps}$

Required $E_b / N_t = 4 \text{ dB}$

Required interference margin = 3dB

Interference factor due to other cells = 0.5

Channel activity factor = 0.65

What is the pole capacity of the cell?

b. What are the different Mobile Adhoc network protocols? Explain Dynamic source routing (DSR) in detail with example. 6

c. Explain the operations of Mobile IP in detail. Why it is necessary to adapt different types of optimization strategies with regard to Mobile IP? 6

Q4. a. Explain the structure of wireless sensor network in detail. What are the different sensor deployment strategies used in wireless sensor network? 6

b. What do you mean by full coverage and full connectivity in WSN? Derive Voronoi Polygon formula for full coverage and full connectivity in Lattice Wireless sensor network. 6

c. A node X receives three beacons from nodes A, B, and C at (0, 0, 0), (2, 6, 0), and (3, 4, 0), respectively. From the received signal strength, it determines the distances to A, B, and C to be $\sqrt{26}$, $\sqrt{2}$, and $\sqrt{5}$ respectively. Find the coordinates of X. 6

Q5. a. What is an activity in android? Explain, how the lifecycle of an android activity works? 6

b. Explain the architecture of Android in detail. Also explain the concept of Android software stack. 6

c. What is mobile payment system? Explain its properties and security issues with suitable example. 6

M. Tech. (Computer Engineering) II Semester 2018

INTELLIGENT SYSTEMS

MM: 60

Time: 3 Hours*

Note: Attempt all questions

S. No.	Questions	Marks																											
1	<p>a) Calculate the probability of the ordering: Alarm(A), Burglary(B), Earthquake(E), JohnCalls(JC), MarryCalls(MC). i.e: $P[A, B, E, JC, MC]$.</p> <p>$P(B)$ 0.001</p> <p>$P(E)$ 0.002</p> <table border="1"> <tr><td>B</td><td>E</td><td>P(A)</td></tr> <tr><td>T</td><td>T</td><td>0.95</td></tr> <tr><td>T</td><td>F</td><td>0.95</td></tr> <tr><td>F</td><td>T</td><td>0.29</td></tr> <tr><td>F</td><td>F</td><td>0.01</td></tr> </table> <table border="1"> <tr><td>A</td><td>P(JC)</td></tr> <tr><td>T</td><td>0.90</td></tr> <tr><td>F</td><td>0.05</td></tr> </table> <table border="1"> <tr><td>A</td><td>P(MC)</td></tr> <tr><td>T</td><td>0.70</td></tr> <tr><td>F</td><td>0.01</td></tr> </table>	B	E	P(A)	T	T	0.95	T	F	0.95	F	T	0.29	F	F	0.01	A	P(JC)	T	0.90	F	0.05	A	P(MC)	T	0.70	F	0.01	6
B	E	P(A)																											
T	T	0.95																											
T	F	0.95																											
F	T	0.29																											
F	F	0.01																											
A	P(JC)																												
T	0.90																												
F	0.05																												
A	P(MC)																												
T	0.70																												
F	0.01																												
2	<p>b) Give an ADL solution of Spare Tire Problem.</p> <p>c) Write down the algorithm of partial Order Planning.</p>	6																											
	<p>a) Consider the following axioms, use resolution to conclude the final premises:</p> <ul style="list-style-type: none"> i. All hounds howl at night. ii. Anyonē who has any cats will not have any mice. iii. Light sleepers do not have anything which howls at night. iv. John has either a cat or a hound. v. (Conclusion) If John is a light sleeper, then John does not have any mice. <p>b) State the process(all the rules) of CNF conversion.</p> <p>c) Consider the following axioms:</p> <ul style="list-style-type: none"> i. Every child loves every candy. ii. Anyone who loves some candy is not a nutrition fanatic. iii. Anyone who eats any pumpkin is a nutrition fanatic. iv. Anyone who buys any pumpkin either carves it or eats it. v. John buys a pumpkin. vi. Lifesavers is a candy. vii. (Conclusion) If John is a child, then John carves some pumpkin. 	6																											
	Resolve through FOL principles-																												

✓ 3 a) Write down the DCG which accepts all of the following strings.
[the/a] [man/woman] [shoots] [the/a] [man/woman]

6

b) Write down the CFG recognizer with difference list which accepts
all of the following strings.
[the/a] [man/woman] [shoots] [the/a] [man/woman]

6

c) Draw the search tree for: len/2 predicate to check the length of a string.

6

a) What is a Red Cut and Green Cut. Explain with suitable examples.

6

b) Discuss the following predicates with suitable examples; findall/3,
setof/3, bagof/3

6

✓ 4 a) Draw the search tree for the following query ?- jealous(X,Y).

6

The knowledgebase is given below:

loves(vincent,mia).

V | M

loves(marsellus,mia).

jealous(A,B):-

 loves(A,C),

 loves(B,C).

✓ 5 Write short notes of any two of the following with their suitable applications.

6

a). Learning Decision Trees

6

b) Support Vector Machines

6

c) Naive Bayes Classification