School of Engineering and Technology, DYPU, Ambi

A.Y : Fourth Year, B.Tech **Semester :7th** **Branch: CE, IT and AI & DS**

Course : Elective- IV (Natural Language Processing) **Course Code: UCEPE 704.3, UITPE 704.3 and UDSPE 704.3**

Question Bank

Unit I Introduction to Natural Language Processing

Q Select correct option and re-write sentence.

1. What is the primary motivation behind NLP?
a) Process spoken languages b) Enable computers to understand/generate human language c) Create new programming languages d) Translate binary to text

2. Which is NOT an application of NLP?
a) Machine Translation b) Sentiment Analysis c) Compiler Design d) NER

3. In NLP, "ambiguity" refers to:
a) Clarity of sentences b) Multiple interpretations c) Predefined grammar rules d) Storage issues

4. Correct order of language processing stages?
a) Morphology→Syntax→Semantics→Pragmatics b) Syntax→Semantics→Pragmatics→Morphology c) Semantics→Morphology→Syntax→Pragmatics d) Pragmatics→Syntax→Morphology→Semantics

5. Stage focusing on meaning of words/sentences:
a) Syntax b) Morphology c) Semantics d) Pragmatics

6. Pragmatics in NLP deals with:
a) Sentence structure b) Word meaning c) Context/intended meaning d) Word formation

7. Which is a major challenge in NLP?
a) Ambiguity b) Large vocabulary c) Slang language d) All of the above

8. Data sparsity in NLP mainly affects:
a) Large datasets b) Language modeling c) Speech recognition d) Syntax parsing

9. Goal of a language model is to:
a) Generate random words b) Assign probabilities to word sequences c) Translate code d) Summarize text

10. Trigram model considers:
a) 1 previous word b) 2 previous words c) 3 previous words d) 4 previous words

11. Perplexity evaluates:
a) Syntax parsers b) Language models c) Chunking models d) Translation systems

12. PoS tagging assigns:
a) Word frequency b) POS category to each word c) Semantic meaning d) Chunk boundaries

13. NER identifies:
a) Verbs/adverbs b) Person, organization, location c) Sentence boundaries d) POS categories

14. Chunking identifies:
a) Word dependencies b) Noun/verb phrases c) Morphological changes d) Sentence semantics

15. HMM assumes:
a) Future depends on entire past b) Future depends only on current state c) States fully observable d) No probability needed

16. In HMM, hidden part is:
a) Observed words b) Transition probabilities c) Underlying states (tags) d) Training data

17. Algorithm for parameter estimation in HMM:
a) Forward b) Viterbi c) Baum-Welch (EM) d) Gradient descent

18. Algorithm for decoding best tag sequence in HMM:
a) Forward b) Backward c) Viterbi d) Naive Bayes

19. Observed sequence in PoS tagging with HMM is:
a) Words b) POS tags c) Hidden states d) Grammar rules

20. Main advantage of HMM for tagging:
a) Handles ambiguity probabilistically b) Needs no training data c) Ignores context d) Always 100% accurate

Unit 2 Parsing and Context-free Grammars

Q Select correct option and re-write sentence.

1) Parsing in NLP refers to: a) Splitting sentences b) Analyzing grammatical structure c) Translating code d) Tokenization

2) Context-Free Grammar (CFG) is defined by: a) Terminals, non-terminals, start symbol, production rules b) Only terminals and rules c) States and transitions d) Words and meanings

3) Which is an example of a CFG production rule? a) NP → Det N b) NP = N + V c) NP == VP d) NP in VP

4) Syntax of English is captured using: a) Regular expressions b) CFGs c) Automata d) Tokens

5) Ambiguity in CFGs arises when: a) No parse exists b) Multiple parse trees for same sentence c) Wrong grammar rules d) Only one rule applies

6) Example of structural ambiguity: a) “old men and women” b) “fast car” c) “red apple” d) “the cat sleeps”

7) Probabilistic CFGs (PCFG) assign: a) Cost b) Frequency c) Probabilities to rules d) Tags to words

8) Main use of PCFGs: a) Resolve ambiguity b) Remove terminals c) Define tokens d) Build automata

9) CKY algorithm is based on: a) Bottom-up dynamic programming b) Top-down parsing c) BFS search d) Heuristic grammar

10) CKY works only if grammar is in: a) Normal English b) Chomsky Normal Form c) Backus-Naur Form d) Regular Form

11) Time complexity of CKY algorithm: a) O(n) b) $O(n^2)$ c) $O(n^3)$ d) O(n log n)

12) Weakness of PCFGs: a) No probabilistic outputs b) Ignores lexical information c) Too fast d) Handles context too well

13) Lexicalization of a treebank means: a) Removing rules b) Adding words (heads) to grammar rules c) Deleting terminals d) Expanding sentences

14) Lexicalized PCFG improves parsing by: a) Ignoring probabilities b) Using headwords in rules c) Removing ambiguity d) Stopping recursion

15) In Lexicalized PCFG, headword refers to: a) The main word of a phrase b) Random terminal c) Non-terminal only d) Probability value

16) Parameter estimation in lexicalized PCFG often uses: a) Maximum likelihood b) Random choice c) Viterbi approximation d) Manual assignment

17) The main limitation of CFG in English syntax: a) Cannot handle recursion b) Cannot handle agreement/long dependencies c) No rules allowed d) Too probabilistic

18) Ambiguity example in parsing: a) “Visiting relatives can be annoying” b) “Dogs bark loudly” c) “She eats food” d) “Sun rises daily”

19) A CFG without probabilities differs from PCFG because: a) It allows ambiguity b) It cannot parse c) It doesn’t rank parse trees d) It is faster

20) CKY algorithm is mainly used for: a) Lexical tagging b) Probabilistic parsing c) Syntax highlighting d) Tokenization

Unit 3: Feedforward Neural Networks

1. What is a log-linear model? Recall its mathematical formulation and explain its applications in NLP tasks such as classification or tagging.

2. Discuss the importance of learned representations (e.g., word embeddings) in NLP. How do they improve over handcrafted linguistic features?

3. Explain with a diagram how a single-layer feedforward network can be used for text classification in NLP.

4. Describe how multi-layer feedforward neural networks are applied in NLP tasks. Give an example where depth improves performance.

5. Explain how distributed word representations are incorporated into neural networks for tasks like sentiment analysis, PoS tagging, or machine translation.

Unit 4: Computational Graphs, and Backpropagation

Q match the pair and write correct answer in front of column A

Column A	Column B
A1. Computation graph	B1. Outputs values between 0 and 1
A2. Regularization	B2. Technique to prevent overfitting
A3. Sigmoid function	B3. Activation with faster convergence, avoids saturation

Column A	Column B
A4. Multi-layer neural network	B4. Optimization method used to minimize loss
A5. Activation function	B5. Algorithm to compute gradients using chain rule
A6. Dropout	B6. Helps avoid vanishing or exploding gradients
A7. Gradient descent	B7. When model performs well on training but poorly on unseen data
A8. Chain rule	B8. Adds penalty proportional to squared weight values
A9. Overfitting	B9. Introduces non-linearity into the network
A10. ReLU function	B10. Limited to linear separability
A11. Single-layer network	B11. Information flows only in one direction (input → output)
A12. Backpropagation	B12. Represents mathematical operations as nodes and edges
A13. Weight initialization	B13. Randomly ignores neurons during training for regularization
A14. Feedforward network	B14. Uses multiple hidden layers for better feature learning
A15. L2 regularization	B15. Fundamental rule for computing derivatives in backpropagation

1) Explain the structure of a multi-layer neural network. Discuss the roles of input, hidden, and output layers, and illustrate how neurons in different layers are connected. Include an example of a simple network with two hidden layers.
2) Describe the purpose of activation functions in neural networks. Explain at least three commonly used activation functions, their mathematical expressions, and the impact on the network's ability to learn non-linear patterns.
3) Explain the chain rule of calculus. Show how it is used to compute the gradient of the loss function with respect to the weights in a multi-layer neural network. Provide a simple numerical example.
4) Define a computation graph and explain its significance in training neural networks. Draw a computation graph for a simple feedforward network with one hidden layer and demonstrate how the forward pass is represented.
5) Describe the backpropagation algorithm in detail. Explain how it propagates the error backward through the network, calculates gradients, and updates weights. Include the mathematical formulation for weight updates.
6) Explain the concept of overfitting in neural networks. Discuss at least two regularization techniques (such as L1/L2 regularization and dropout) and how they improve the generalization of the network.
7) Discuss the importance of weight initialization in multi-layer neural networks. Explain common initialization strategies (such as random, Xavier, and He initialization) and their impact on training performance and convergence.

Unit 5: Recurrent Networks, and LSTMs, for NLP

1) Explain the architecture and working of a Simple Recurrent Neural Network (RNN). How does it capture sequential dependencies in NLP tasks?

2) What are the exploding and vanishing gradient problems in RNNs? Discuss their impact on training deep neural networks and possible solutions.

3) Describe the architecture and functionality of LSTM and GRU networks. How do they overcome the limitations of simple RNNs?

4) Discuss the role of Convolutional Neural Networks (CNNs) in NLP. Compare their advantages and limitations with RNNs for text processing tasks.

5) What is the attention mechanism in neural networks? Explain its importance and applications in modern NLP systems

Unit 6: Neural Machine Translation

1) What is the machine translation problem in NLP? Explain its challenges and importance in real-world applications.

2) Describe the architecture of the sequence-to-sequence (Seq2Seq) model. How does it handle variable-length input and output sequences?

3) Explain how LSTMs are used for encoding and decoding sequences in Seq2Seq models. Why are LSTMs preferred over simple RNNs in this context?

4) Discuss the role of the attention mechanism in sequence decoding. How does it improve the performance of Seq2Seq models in machine translation?

5) Compare machine translation models with and without attention. What advantages does attention provide in terms of handling long sentences?

Q Select correct option and re-write sentence.

1. Which of the following correctly describes a multi-layer neural network?
A) A network with only input and output layers
B) A network with one or more hidden layers between input and output layers
C) A network that cannot learn non-linear functions
D) A network with no weights

2. Which activation function is commonly used to introduce non-linearity in hidden layers?
A) Linear function
B) Step function
C) ReLU (Rectified Linear Unit)
D) Identity function

3. In neural networks, the chain rule is used to:
A) Initialize weights
B) Compute gradients during backpropagation
C) Normalize input data
D) Decide the number of hidden layers

4. What is a computation graph?
A) A diagram showing data flow and operations in a network
B) A graph of loss versus epochs
C) A plot of activation functions
D) A histogram of weights

5. Backpropagation in neural networks is used to:
A) Generate random weights
B) Propagate the error backward and update weights
C) Choose the best activation function
D) Reduce the number of hidden layers

6. Which of the following is a regularization technique to reduce overfitting?
A) Gradient descent
B) Dropout
C) ReLU activation
D) Batch normalization

7. Xavier and He initialization methods are used for:
A) Selecting activation functions

B) Weight initialization
C) Computing loss
D) Regularization