Total Pages 12 (Twelve)

SYMBIOSIS INTERNATIONAL (DEEMED UNIVERSITY)

(Established under Section 3 of the UGC Act 1956)

Re-accredited by NAAC with 'A++' Grade | Awarded Category - I by UGC

Verified all entries & found correct
Jr. Supervisor's Signature & Date
22/10/24

No. 075785
SIU-24 E

Centre: SCMHRD

Seat No. (in figures) 2 0 4 0 0 5

Seat No. (in words) Two Zero Four Zero Zero Five

Examination : Day & Date : 22nd Oct '24, Tuesday.

Programme : MMM - HR Semester : III

Course : Natural Language Processing

Main Ans. Script 1 + No. of Supplementary Answer Scripts 1 =Total 2

Q.No.																Total	Signature of Examiner
Max. Marks																	
Marks Obtained																	

INSTRUCTIONS

1. Mention your details only in the space provided for in the main answer script & the supplement. If any other details (including seat number, name, contact details, etc.) are written anywhere else in the answer script and or supplement it will be treated as adoption of unfair means and the performance will be treated as null and void for the entire examination.
2. Write answer in legible hand. Answers written in an illegible and undecipherable hand are liable to be marked as zero.
3. An act of copying or of impersonation at an examination is punishable under the Maharashtra Prevention of Malpractices at University, Board and other specified examinations Act 1982.
4. Candidates should write answers in BLUE/BLACK ink only. Use of Pencil and other colors are permitted only in case of diagrams, graphs etc. Answer Scripts written with pencil or ink of other colors will not be evaluated.
5. Write on both sides of paper.

BEGIN WRITING HERE :

1. Natural Language Processing :- (NLP)

NLP enables the system or machine to interact with the human, to understand the human language and try to replicate the language, so it can help in various application in real world.

The two real world application are as follows :-

(a) Sentiment Analysis.

How the NLP helps in understanding the essence or sentiment of the market or public, by processing the text or document

an an input and giving out the sentiment which can be positive or negative.

eg: Understanding the sentiment of public or customer towards a new launch of product.

Benefits:-

→ This helps in understanding the performance of the product in the market.

→ This will help the business in strategizing the actions to be taken in future.

Challenges:-

→ NLP may find difficulty in capturing the sentiment clearly due to complex feedback or text.

→ The processing of the text or document given can take time and also can be computationally intensive.

(5) Speech Recognition.

→ NLP can analyse the text or any speech. eg

For instance, the analysis of Federal Bank President speech so that it can help in prediction of impact of the speech on the stock market.

Benefits:-

→ Helps in strategizing the investment and trading position as per the speech given.

→ It saves time on the real impact of the speech on taken time.

Challenges:-

→ If the speech in different language, then NLP can find difficulty in converting it into local language.

→ Phrases or sign used in speech, is difficult to understand by NLP.

2. The four commercial use of NLP are as follows:-

(a) NLP is used in understanding the sentiment of the customers about the new launch of product.
This helps in depicting the performance of the market campaign and also to understand the real time feedback.
Accordingly, the organization can strategize the action to be taken in future.

(b) Analyzing the Ancient Symbols or Pictures:-

NLP has great contribution in archaeology field. The old script

written on the sculptures are difficult to analyse.

Here the machine can use suitable approach to analyse and get the maximum information from the sculptures script.

(c) Machine Translation:-

NLP has been greatly leveraged in translating the speech from one language to another.
For instance, when there is an interaction of people from different cultures or background then they can communicate to each other in real time, in their very own mother tongue.

(d) Chatbots

NLP is being leveraged by the organization to built the chatbots. This chatbots are responsible to answer the query of the customer in real time.

This requires machine to understand the human language, break down into text and analyse the question so that it can properly give the feedback in human language.
This saves a lot of time and money for the organization in form of accuracy and manpower optimization

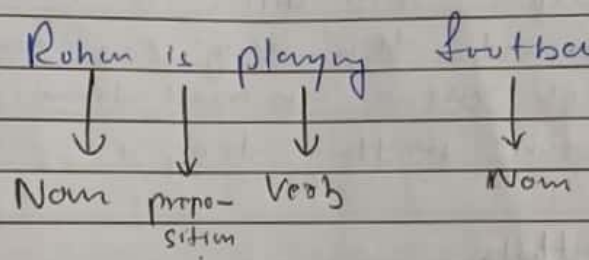
3. POS Tagging :-

Parts of Speech tagging is the process of analysing the text as per grammar rule.

Significance :-

This helps in broadly down the sentence to understand the parts of speech of the words belong to.
This bifurcate the words into Noun, Pronoun, Adjective, Adverbs, Preposition, Conjuction and interjuction.

eg.



-> This breakdown of the text, helps in understandy the meany and the action happening.

-> If the POS tagg is not done, then machine can confuse with the action, object subject relationship.

-> POS tagging who helps in for the subsequent steps of NLP.
Once the speech and grammar understandy is clear for the machine, it can now relate or analyse the relationship between the different sentences and also relate it with the real world.

SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU

4. N-gram Models:-

N-gram are the sequence of words in the given text.

For example.

The article is good.

Bigram → "The article", "is good"

Unigram → 'The' 'article', 'is', 'good'

Language Modelly:-

In lagage modelly, we try to predict the next word or sentence as per the given text.

→ Now, lagage modelly, works on the principle of probabilistic modelly, which try to find the probability of the next word to occur in the text.

→ Here, N-Gram Model helps in a way that the sequence of word present determines the probability of next word or words. This is based on Markov or Hidden Markov Model.

For instance.

Playy is good for health.

In this, after seeing the word like 'is' and 'for' which do not contribute much.

'Playy' and 'food' determines the probability of

the next word to be 'health'.

Limitation :-

→ There are words whose probability can be zero. And incorporating this word can disturb the overall analysis. So, we use various smoothing technique to tackle this problem. The methods like Laplace, Additive, Katz Methods, etc are used.

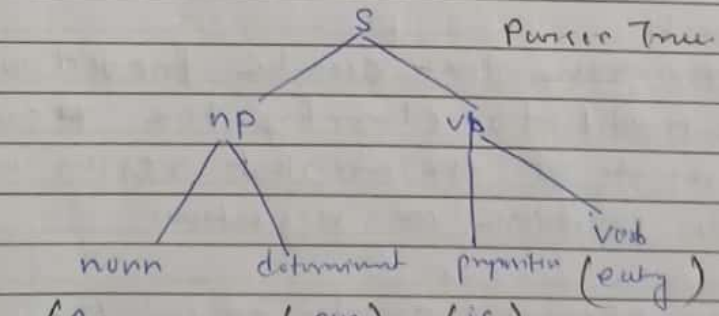
→ In N-gram model, the sequence of words can be such which has least contribution in analysing the sentiment of the document or text.
So, this create issue in building the accuracy of the model.

5.	Constituency Parsing	Dependency Parsing
(1)	In this, the text or document is segmented into constituents or group of words as per grammar rules.	In this process, the aim is to find the relationship between the words as per the grammar rule.
(2)	Parse tree formed here is top-down approach.	Parse tree formed by bottom to up approach.

Consisteny Pasing.	Dependy Pasing
(3) This is not so faster in general.	This is faster in comparison to the consisteny pasing.
(4) Used in application such as text preprocessing, speech recognition, Spam detection, etc.	The application where it used are heavy such as used in machine translation, question-answer, etc.

(5) For instance,
Rohan is eating an apple.

In consisteny pasing:-



From above we can see how the top down approach parse tree is formed.
Whereas in dependey pasing, this is not formed in this manner.

SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU

6. Naive Bayes Classifier:-

The works on the principle of bayes's theorem. It is basically a mathematical model which works on the probability theory.

$$P(A/B) = \frac{P(B/A) \times P(A)}{P(B)}$$ → Bayes's Theorm

→ In text classification, the Naive Bayes Classifier helps in classify the overall sentiment of the text into categorical outcomes.

→ By instance, NBC is used in classifying the document or text into sentiments like positive and negative.

→ Suppose we have a collection of feedback in a raw format from the customers who have used our new product launch.
Now the raw text can be feeded in the model to classify the overall sentiment of the market region wise into positive or negative.

→ Another application can be analysing the speech of the finance minister on the "Budget Day" to understand and categorize the sentiment of the budget into → (1) Bullish
(2) Bearish
(3) Neutral

SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU

7. Text Pre-processing:-

It is one of the crucial step in the NLP Pipeline.
Before the text Preprocessing, the data is collected and stored properly. Then the following steps are taken to pre-process the text before it is exposed to feature extraction and modelling:-

So, Text Preprocessing is all about breaking down the text and converting it into smallest units.

(1) Tokenization:- It is process of breaking down the text into smaller units
It can be done by word tokenization or sentence tokenization.

~~Two Crucial Steps In~~

(a) Word tokenization:- Dividing the sentences into words.

(b) Sentence tokenization:- Dividing the sentences or paragraphs into individual sentences.

(c) Stemming:- In this the ~~sentence~~ word is broken down into its smallest unit by removing the prefixes or suffix.

eg. Running → Runn
Playing → Play

i.e removing 'ing', 'ly', 'es', etc.

(d) Lemmatization :-

In this the word is broken down into its root word.

for example,

Running → run

Ran → run

Studying → study.

Note:- Lemmatization uses corpus or dictionary to get the root word and it is more accurate than stemming. But stemming is faster.

(2) POS tagging :- This segments the words on pos grammar rule.

for example:- Rohan is playing football

Noun verb Noun.

preposition

(3) Text Segmentation

(4) Dependency Parsing

Importance of Text Pre-processing :-

(1) Text preprocessing is helpful in a perspective that it makes easier for the machine to convert the human language into right form so that it can use the text.

(2) It gives the machine the understanding of feature for modelling.

Total Pages 8 (Eight)

SYMBIOSIS INTERNATIONAL (DEEMED UNIVERSITY)

(Established under Section 3 of the UGC Act 1956)

Re-accredited by NAAC with 'A' grade (3.58/4) | Awarded Category - I by UGC

Verified all entries & found correct
Jr. Supervisor's Signature & Date
29/10/21

Supplement No. SIU-20 A
58220

Centre: SCMHRD

Seat No. (in figures) 2 0 4 0 0 8

Seat No. (in words) Two Zero Four Zero Zero Eight

Programme : MBA - BA Semester : III

Course : Natural Language Processing

Supplementary Answer Script No. I / II / III

BEGIN WRITING HERE :

(3) It enhance the accuracy of the model going forward.

(4) Feature extraction becomes easy if the preprocessing is done better.

P.T.O.

8. NLP (Natural Lyge Brocessy)

The purpose is to make machine understand the human lagyge, so that it can be used for various application.

There are two basic components of NLP:-

(a) NLU (Natural lagyge Understandy.)

→ This converts the input given in human lagyage into machine digit form.

→ In this the text given as input, are assigned digits and thus pattern formation is used by the machine for task assigned.

SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU

(b) NLG (Natural Language Generation)

-> This converts the machine interpretation into human language. The generated output is received by the user.

For instance,

→ The Use of Chatbots for Question - Answer.

Chatbots works on these two components discussed above.
When the customer put his/her query to the bot, the NLU translate into machine language and NLG generate the final output or the answer to the customer.

9. Language Modelling:-

Language modelly is used to predict the next word or text or sentence as per the given document or text.

The different methods of language modelly are as follow:-

(1) Statistical Models:-

This model works on the principle of probabilistic model. The next word or sentence is predicted on the bases of the previous state of the sentence.

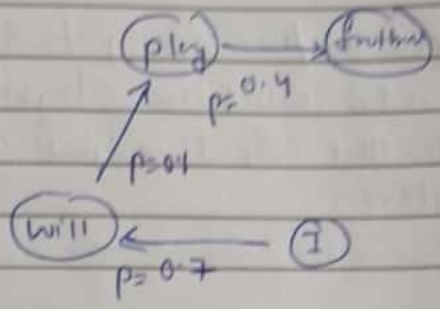
$$P(w_{n+1}) \simeq P(w_1 w_2 w_3 w_4 w \ldots w_n) \simeq P(w_n)$$

-> As per Markov Model, the future state of the word depends on the current state.

-> In this the probability is generated for the next word to happen.

For instance,

I will play football.



-> There is a transmission probability of going one state to another.

-> Models like N-gram Model, Naive Bayes Classifier are the examples of the Statistical Models.

(2) Neural Model:-

-> This is advance version of language model. This doesn't consist of any labeling of the input text.
The machine develop the understandy of the text and try to figure out the relationship.

-> Recurrent Neural Networks and Convolutional Neural Networks are the examples of this type.

-> This is fastest and are fraded with enormous amount of input.
-> It uses corpus to train itself and understand language so that it can predict.

-> Chatbots and Answer-Question application are few of the examples of this type of model.

10. Text Pre-processing:-

This is used in cutting the input data by tokenization, pos tagging, etc in order to pre-process before it is fed into the model.

Challenges in pre processing Chinese or Arabic can be:-

(1) The language texture is symbolic in nature. For instance, the symbol used in Chinese depicts the whole word.
Also, in arabic, the texture is quite symbolic and logothermic in nature.

This makes difficult for the machine to apply the methods like tokenization, pos tagging.

(2) The texture looks very similar to each other. The words are very similar in look wise.
This makes challenge for the machine to

different to how each other.

(3) Arabic is also challenging in sense that it goes backward in reading.

Suggestion to Overcome:-

(a) Use unilingual corpus, to understand the meaning and analysing the text in Chinese and Arabic

(b) It is also important that the input data is properly collected and stored. Any discrepancy in the first step, can create more challenge for the machine to pre process.

(c) Use dependency parsing, it is more suitable in this case.

Total Pages 12 (Twelve)

SYMBIOSIS INTERNATIONAL (DEEMED UNIVERSITY)

Established under Section 3 of the UGC Act 1956

Re-accredited by NAAC with 'A++' Grade | Awarded Category - I by UGC

Verified all entries & found correct
Jr. Supervisor's Signature & Date
22/10/24

No. SIU-24 E 075639

Centre: SCMHRD

Seat No. (in figures) 2 0 4 0 0 6

Seat No. (in words) Two Zero Four Zero Zero Six

Examination : Day & Date : Tuesday, 22/10/24

Programme : MBA-BA Semester : III

Course : Natural Language Processing

Main Ans. Script 1 + No. of Supplementary Answer Scripts 1 =Total 2

Q.No.																	Total	Signature of Examiner
Max. Marks																		
Marks Obtained																		

INSTRUCTIONS

1. Mention your details only in the space provided for in the main answer script & the supplement. If any other details (including seat number, name, contact details, etc.) are written anywhere else in the answer script and or supplement it will be treated as adoption of unfair means and the performance will be treated as null and void for the entire examination.
2. Write answer in legible hand. Answers written in an illegible and undecipherable hand are liable to be marked as zero.
3. An act of copying or of impersonation at an examination is punishable under the Maharashtra Prevention of Malpractices at University, Board and other specified examinations Act 1982.
4. Candidates should write answers in BLUE/BLACK ink only. Use of Pencil and other colors are permitted only in case of diagrams, graphs etc. Answer Scripts written with pencil or ink of other colors will not be evaluated.
5. Write on both sides of paper.

BEGIN WRITING HERE :

Q ①

NLP :

Natural Language Processing is a part of Artificial Intelligence that was developed to develop communication between humans and computers using plain human language.

When a computer passes Turing test, it can be deemed fit for NLP task Eg : Eliza.

(i) Chatbots :

Earlier chatbot had pre-defined responses inside. Now, chatbots are able to understand word by word of the text sent as input and provide relevant response as output.

(iii) Speech to Text Conversion :

This is used to convert the speech [input] into a text, store it in a document.

Q ②

CHATBOT

Reasons

* Accurate infor

* No extra lu

* Chatbot ma
the con

* Eg : On
user
detected

CHATBOT	SPEECH TO TEXT
BENEFITS	
* Accurate information response	* Elimination of typing long documents.
* No extra/unwanted response	* Saves a lot of time.
CHALLENGES	
* Chatbot may not understand the correct context.	* Understanding the text input as it may have semantic errors.
* Eg: Orders, sarcasm by user cannot be properly detected.	* Eg: The dairy looks good. When translating, it can type dairy as diary → Notebook; ↳ Milk product.

Q ②

NLP has a lot of commercial uses:

(i) Document Retrieval:

When using packages like "QuePy", queries can be used to retrieve documents from a database through sentence matching, regular expressions.

When there are thousands of documents and the naming conventions involve numbers, finding the right document can be a hassle.

In such cases, NLP can be used.

(ii) Spell Check:

When the document contains typos, one need to go through every word to check the spelling, which takes a lot of time. This can be eliminated through NLP.

Eg: The technical compitency requires a degree

CORRECTED: The technical competency requires a degree

Eg: The water born disease is contagious.

CORRECTED: The water borne disease is contagious

(iii) Grammar Check:

NLP understands the language through various corpus. So, it can identify the grammatical errors.

Applications such as Grammarly are developed on this principle.

Eg: I see a bird fly.

CORRECT: I am seeing a bird fly (or) I saw a bird fly.

Eg: I would have ate the cake tomorrow.

CORRECT: I would have eaten the cake by tomorrow.

(iv) Text Summarization:

Companies have a lot of requirements, SOP's to follow. Text Summarization allows to give a summary of the input text.

This helps the person in saving time such as HR's, legal team to save time in reading everything.

Applications such as Merlin are built on this and are used for a web-page summary.

Q(3) POS Tagging: Parts of Speech Tagging is a crucial process in the steps of NLP.

It helps to identify each of the tokenised words by its POS form.

Eg Noun, Verb, Adverb, Adjective, etc

Steps in POS Tagging :

(i) Sentence Tokenisation, Word Tokenisation -
It breaks down paragraph into sentences and sentences into words.

(ii) Importing language models such as Spacy, NLTK for pre-processing.

(iii) Stemming, Lemmatization to reduce the words to lemma (or) root form with meanings.

(iv) Apply POS tagging on the words.

(v) Result Analysis to understand the results.

A simple break up of words give no understanding.

POS tagging provides a semantic network for the words.

With POS tagging, we can understand the link between the words as it helps in defining the structure.

This step proves crucial to Named Entity Recognition and even Consistency Parsing.

Eg :- I love cakes.

Word Tokenised Result = ["I", "love", "cakes"] → Have no meaning

Eg: I → love → cakes
Subject Verb Object

Eg: I / have / a flight / tomorrow
Subject Verb Object Complement

Q ⑨ N-grams are continuous n-words that help in prediction of next word, in Language modelling.

N-grams canbe of
* Unigram - One word
* Bigram - Two words
* Trigram - Three words
* N-gram - Many words

When there are multiple words in an - N-gram, the complexity increases.

I saw a cat on a wall.

N-gram = [I saw a cat on a ___]
Answers can be sofa, wall, table etc.

In this the answers are endless, the model will be confused.
To reduce this, we can use tri-gram, bi-gram

Bigram = [I, saw] [saw, a], [a, cat], [cat on], [on, a], [a, ___]
This reduces the complexity.

LIMITATIONS:
(i) Data sparcity, the model will not have enough corpus to predict the outcome. This brings the probability to 0.
(ii) Out of corpus words add ambiguity
N-gram works on Statistical approach, if probability is 0, the model becomes useless.

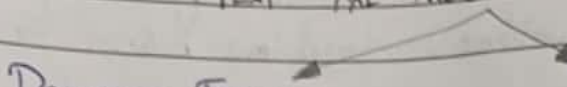
(iii) To overcome this, Smoothing techniques such as Laplace Smoothing, Additive Smoothing, Backoff, Interpolation Omit 2x Gram, Smoothing are used.

Q ⑤ CONSISTENCY PARSI
(i)
Consistency Parsing to identify the words in a
(ii) It is a Top-
(iii) It provides the words like sub
(iv) Best suited f Natural Langua
(v) Suitable for ta POS tagging, N
(vi) It is more f analysis, uses Grammar.
(vii) It provides a output.
(viii) Good for la Semantics. Thi

Q (5)

	CONSISTENCY PARSING	DEPENDENCY PARSING
(i)	DEFINITION	
	Consistency Parsing is used to identify the constituent words in a sentence	Dependency Parsing is used to identify the dependency relations between the words
(ii)	It is a Top-Down Approach	It is a bottom-up approach
(iii)	It provides the type of words like subject, verb etc.	It identifies the relationship such as subject-verb agreement.
(iv)	Best suited for NLU Natural Language Understanding (complex)	Best suited for NLG Natural Language Generation complex task
(v)	Suitable for tasks, like POS tagging, NER, Stemming	Suitable for Text to Speech, Machine Translation etc.
(vi)	It is more for semantic analysis, uses Free Free Grammar.	It is used for less semantic tasks, has dependency based grammar
(vii)	It provides a parse tree as output.	It provides a dependency tree as output.
(viii)	Good for language with higher semantics, like Japanese, Thai.	Not able to pick up semantics as suitable for English, French.

Q (7) Text-Preprocessing is the process of getting a text as input, perform various operations to break down and understand the text.



DOCUMENT TRIAGE:

Document triage is a part of NLU. Natural Language Understanding.

Here, a document is received by the system. It scans and converts the document into a set of digitalized words, which are machine readable.

TEXT SUMMARISATION:

It is a part of NLG. Natural Language Generation. This involves various steps

* Sentence Tokenisation: Breaking the paragraph into various sentences.
* Word Tokenisation: Breaking the sentences into words.
* Stemming: Converting words to root form which may (or) may not have meaning.

Eg:	Happiness	Intelligence	Better
Stemmed:	Happi	Intelligen	Bett

* Lemmatization: Converts words into Meaningfull words

Eg:	Happiness	Intelligence	Better
Lemma:	Happy	Intelligent	Good

* Identify Stop Words: Remove mostly repeated fillers like "a", "an", "the" etc.

Words that can be removed and still form an meaningful sentence.

* PosTagging : Identify Parts of Speech of each word

* NER : Named Entity Recognition

This tags each word into Name, Organisation, Date, Place etc.

Ian went to HSBC Bank Head Office in New Orleans.

Organisation : HSBC Bank
Name : Ian
Location : New Orleans

IMPORTANCE:

It is very important to pre-process the text. Otherwise any model run on the random text will yield many errors.

Or, the model will not be able to predict.

Q 8) NLP has two main components NLU and NLG

i) Natural Language Understanding :

NLU is a way to understand text fed from various metadata that is input to the system.

It converts the human language into a machine understandable form.

NLU is very hard to process than NLG. This is because, NLU is dependent on corpus of words, its quality. It will also take a lot of time to train with the corpus to get accurate results.

SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU SIU

Natural Language Generation:

NLG converts the machine format back into human understandable form.

When the machine process the input and about to deliver the result, the NLG helps to convert it.

NLG involves Tokenisation, Stemming, NER, stopwords analysis, etc.

Once these analysis is done, data is finally chunked into various words for the user to understand.

Q(9) LANGUAGE MODELLING:

Language modelling is a way of getting tokenised words/lexicons as input and to check if they are are syntactically correct to be accepted as the corpus.

i) STATISTICAL MODELLING:

Statistical Modelling uses probabilistic models to build the corpus, predict the next word.

If a $P=W$, the probability of occurrence of a word,

$P(P=W) \mid W_1, W_2, \ldots W_{n-1})$

Markovs and Hidden Markovs Models are under this category.

format back

input and
NLG helps

ing, NER,

is finally
used to.

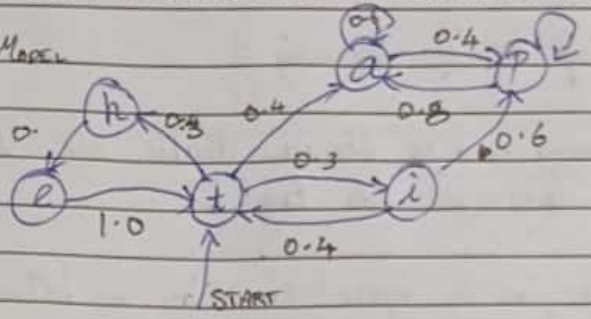
ing tokenised
ey are are
corpus.

lls to

a word,

this

MARKOV MODEL



The start is identified and the probability of the next word is identified based on the current state / word.

HIDDEN MARKOV :

Hidden markov model is used in speech to text
It consists of hidden state which transcend to an output state

NEURAL ~~N~~ LANGUAGE MODELLING :

It makes uses of Neural networks to predict the words such as

CNN - Convolution Neural Networks
RNN - Recurrent Neural Networks.

It also uses the concept of word embeddings

Word embedding is used to convert a word / logograph / pictogram into a lower order function to easily interpret.

LOGOGRAM		Word Embedding
Happy Smiley	=	[1, 0, 0, 0]
Sad "	=	[0, 1, 0, 0]
Neutral "	=	[0, 0, 1, 0]
Frustrated "	=	[0, 0, 0, 1]

Q (10) Challenges of Text Pre-Processing :

Languages such as Chinese, Arabic are very different from languages like English.

English has a "space" character in between the words, this information can be generalised to provide as input for word tokenization.

Languages like Chinese, Arabic are complex and do not have such space between them.

Hence it is diffult to analyse the language Syntaxically, Semantically. This makes tokenisation difficult.

RECOMMENDATIONS

(i) Hence more corpus, more understanding of these language is required.

(ii) This also takes a longer time for getting robust model.

(iii) Competency Parsing has the ability to understand the intricacies and complexities as it tries to understand the words by itself using context Free Grammars.

(iv) Advent of Internet era provides more data which can be incorporated.

(v) Global API's for each language can be used for training if available.

(vi) Go for Neural Network modelling as it does not require Feature Importance and it can identify on its own.

SYMBIOSIS INTERNATIONAL (DEEMED UNIVERSITY)

(Established under Section 3 of the UGC Act 1956)

Re-accredited by NAAC with 'A' grade (3.58/4) | Awarded Category - I by UGC

Verified all entries & found correct
Jr. Supervisor's Signature & Date

Supplement No. SIU-19 A
04430

Centre: SCMHRP

Seat No. (in figures) 2 0 4 0 0 6

Seat No. (in words) Two Zero Four Zero Zero Six

Programme : MBA-BA Semester : III

Course : Natural Language Processing

Supplementary Answer Script No. (I) II / III

BEGIN WRITING HERE :

Q 6 Naive Bayes Classifier of Text Classification is based on Naive Bayes Theorem.

Assumptions :

(i) No Feature Importance : Does not provide any importance to one particular feature.

(ii) Features of Equal Importance : All the features are treated equally

(iii) Numerical Variables : Numerical variables are normally distributed

(iv) Categorical Variables : Categorical variables all follow Multi-nominal classification.

(v) No Missing Data : There should not be any missing data

Naive Bayes Classifier is used for Categorical classification such as Text Classification.

Eg: A cat is white in color

If P = Cat The probability of the next word occurring

(P = cat / is) : This can be found.

The word with the highest probability that comes next can be found using this algorithm.