ASSIGNMENT 1 (B)

Name : Jyoti Sharma

Class : BE-IT

Roll No : 60

Subject : AI

Batch : I3

DOP	DOC	Marks	Sign

Assignment 1 Part B

Q1. Explain PEAS descriptors for wumpus world.

→ PEAS descriptors

1. Performance measure

- +100 for grabbing the gold and ~~combi~~ coming back to the starting position,
- -200 if the player (agent) is killed.
- -1 per action,
- -10 for using the arrow.

2. Environment

- Empty Rooms.
- Room with wumpus.
- Rooms neighbouring to wumpus which are smelly.
- Rooms with bottomless pits.
- Rooms neighbouring to bottomless pits which are breezy.
- Room with gold which is glittery.
- Arrows to shoot the wumpus.

3. Sensors (assuming a robotic agent)

- Camera to get the view
- Odour sensor to smell the stench
- Audio sensor to listen to the scream and bump

4. Effectors (assuming a robotic agent)

- Motor to move left, right
- Robot arm to grab the gold
- Robot mechanism to shoot the arrow

The WUMPUS world agent has following characteristics :

1. Fully Observable
2. Deterministic
3. Episodic
4. Static
5. Discrete
6. Single agent

Q2. Explain various elements of Cognitive System.

→ Cognitive Computing is a new type of computing with the goal of more accurate models of how the human brain/mind senses, reasons, and responds to stimulus. Generally, the term cognitive computing is used to refer to new hardware and/or software that mimic the functioning of the human brain thereby improving human decision-making. Cognitive computing applications link data analysis and adaptive page displays i.e. Adaptive User Interfaces, to adjust content for a particular type of audience.

- Following are some of the features of cognitive systems :

1. Interactive: They may interact easily with users so that those users can define their needs comfortably. They may also interact with other processors, devices, and cloud services, as well as with people.

2. Adaptive: They may be engineered to feed on dynamic data in real time. They may learn as information changes and as goals and requirements evolve. They may resolve ambiguity and tolerate unpredictability.

3. Contextual: They may understand, identify, and exact contextual elements such as meaning, syntax, time, location, appropriate domain, regulations, user's profile, process, task and goal. They may draw on multiple sources of information, including both structured or unstructured digital information, as well as sensory inputs like visual, ~~ges~~ gestural, auditory or sensor-provided.

4. Iterative and Stateful: They may aid in defining a problem by asking questions or finding additional source of input if a problem statement is ambiguous or incomplete. They may "remember" previous interactions in a process and return information that is suitable for the specific application at that point in time.

Q3. Write note on Language Model.

→ The goal of a language model is to compute a probability of a token (e.g. a sentence or a sequence of words) and are useful in many different Natural Language Processing applications.

→ Language Model actually a grammar of a language as it gives the probability of word that will follow.

→ For example, they have been used in Twitter Bots for 'robot' accounts to form their own sentences.

Language Model Definition:

- In case of probabilistic language modeling the probability of a sentence of words is calculated:

$$P(W) = P(w_1, w_2, w_3, \ldots\ldots w_n)$$

- It can also be used to find the probability of the next word in the sentence:

$$P(w_5 \mid w_1, w_2, w_3, w_4)$$

- A model that computes either of these is called a Language Model.

- There are various Language models in available in practice. Following are few of them:

1. Methods using the Markov assumption:

- A process which is stochastic in nature, is said to have the markov property if the conditional probability distribution of future states of the process depends only upon the present state, not on the sequence of events that happened in the past.

- In other words, the probability of the next word can be estimated given only the previous K no. of words.

For example, if K : 1 :

P(transparent | its water is so) ≈ P(transparent | so)

or if K = 2 :

P(transparent | its water is so) ≈ P(transparent)

- Following is the general equation for the Markov Assumption, k = i :

$$P(w_i \mid w_1\ w_2 \ldots\ldots w_{i-1}) \approx P(w_i \mid w_{i-k} \ldots\ldots w_{i-1})$$

2. N-gram Models:

- From the Markov Assumption, we can formally define N-gram models where k = n-1 as the following:

$$P(w_i \mid w_1 w_2 \ldots\ldots w_{i-1}) \approx P(w_i \mid w_{i-(n-1)} \ldots w_{i-1})$$

- The simplest versions of this are defined as the Unigram Model (K = 1) and the Bigram model (K = 2)

3. Unigram Model (K=1):

$$P(w_1 w_2 \ldots w_n) \approx \prod_i P(w_i)$$

4. Bigram Model (K=2):

$$P(w_i | w_1 w_2 \ldots w_{i-1}) \approx P(w_i | w_{i-1})$$

- These equations can be extended to compute trigrams, 4-grams, 5-grams, etc. This is an insufficient model of language because sentences often have long distance dependencies. For example, the subject of a sentence may be at the start whilst our next word to be predicted occurs more than 10 words later.

- Following is the Maximum Likelihood Estimate model to Estimating Bigram Probabilities:

$$(w_i | w_{i-1}) = \frac{Count(w_{i-1} \ldots w_i)}{Count(w_{i-1})}$$

Example

Given a corpus with the following three sentences, let's find out the probability that "I" starts the sentence. Here "<s" and "/s>" denote the start and end of the sentence.

<s I am Sam /s>
<s Sam I am /s>
<s I do not like green eggs and ham /s>

- Therefore, we have:

$$P(I \mid <S) = \frac{Count(<S\,I)}{Count(<S)} = \frac{2}{3}$$

In other words, of the three times the sentence started in our corpus, "I" appeared as the first word in two sentences.

- Language modeling is one of the most important parts of modern Natural Language Processing. There are many sorts of applications for language modeling, like: Spell Correction, speech Recognition, Machine Translation, Question Answering, Summarization, Sentiment analysis etc. All these tasks require use of language model. Language model is supposed to represent the text to a form understandable from the machine point of view.

- Moreover, language modeling must also consider the correlated ordering of tokens. As every language is based on some grammar, where order has a lot of influence on the meaning of a text.

Q4. Write a note on Machine Translation.

→ **Machine Translation**

- Machine Translation is the classic test of language understanding. It consists of both language analysis and language generation. Many machine translation systems have huge commercial use. Following are few of the examples:

1. Google Translate goes through 100 billion words per day.
2. eBay uses Machine Translation techniques to enable cross-border trade and connect buyers and sellers around the world.
3. Facebook uses machine translation to translate text in posts and comments automatically, in order to break language barriers and allow people around the world to communicate with each other.
4. Systran became the first software provider to launch a Neural Machine Translation engine in more than 30 languages back in 2016.
5. Microsoft brings AI-powered translation to end users and developers on Android, iOS, and Amazon Fire, whether or not they have access to the internet

- In a traditional Machine Translation system, parallel corpus a collection of texts is used each of which, is translated into one or more other languages than the original. For example, given the source language e.g. French and the target language e.g. English, multiple statistical models need to be build, including a probabilistic formulation using the Bayesian rule, a translation model P(f|e) trained on the parallel corpus, and a language model p(e) trained on the English only corpus.

- It is obvious that, this approach skips hundreds of important details, requires a lot of human feature engineering, consists of many different and independent machine learning problems, and overall is a very complex system.

A. Neural Machine Translation (NMT):

- The process is modelled through one big artificial neural n/w, known as a Recurrent neural network (RNN) which, is a stateful neural n/w. It has connections b/w passes and connections through time. Neurons are fed information not just from the previous layer but also from themselves from the previous pass. This means that the order in which we feed the input and train the n/w matters: feeding it "Donald" and then "Trump" may yield different results compared to feeding it "Trump" & then "Donald".

- Standard Neural Machine Translation is an end-to-end neural n/w where the source sentence is encoded by a RNN called encoder, and the target words are predicted using another RNN known as decoder. The RNN Encoder reads a source sentence one symbol at a time, and then summarizes the entire source sentence in its last hidden state. Following are the features of NMT:

1. End-to-End training: All parameters in NMT are simultaneously optimized to minimize a loss of function on the network's output.

2. Distributed representations: NMT has a better exploitation of word and phrase similarities. Hence, it forms a robust translator.

3. Better exploration of context: NMT can use a much bigger context for both source and partial target text in order to translate more accurately.

4. More fluent text generation: Deep learning text generation is of much higher quality than the parallel corpus way.

B. Long Short-Term Memory (LSTM):

- LSTM works as a solution to the vanishing gradient problem by introducing gates and an explicitly defined memory cell. Each neuron has a memory cell and three gates: input, output and forget. The function of these gates is to safeguard the information by stopping or allowing the flow of it.

1. The input gate determines how much of the information from the previous layer gets stored in the cell.

2. The output layer takes the job on the other end and determines how much of the next layer gets to know about the state of this cell.

3. The forget gate seems like an odd inclusion at first but sometimes it's good to forget: if it's learning a book and a new chapter begins, it may be necessary for the new to forget some characters from this section.

- LSTMs are able to learn complex sequences, such as writing like Shakespeare or composing primitive music. Note that each of these gates has a weight to a cell in the previous neuron, so they typically require more resources to run.

C. Gated Recurrent Units (GRU):

- They are a slight variation on LSTMs and are extensions of Neural Machine Translation. They have one less gate and are wired slightly different. GRU has an update gate instead of an input, output and a forget gate. This update gate determines how much information to be kept from the last state and how much information to forget from the previous layer.

- The reset gate functions much like the forget gate of an LSTM, but its located slightly different. They dont have an output gate.

- There have been further improvements in neural machine translation systems over the past few years. Below are the most notable developments:

- Sequence to Sequence, learning with Neural Networks proved the effectiveness of LSTM for Neural Machine Translation. It presents a general end-to-end approach to sequence learning that makes minimal assumptions on the sequence structure. The method uses a multilayered LSTM to map the input sequence to a vector of a fixed dimensionality, and then another deep LSTM to decode the target sequence from the vector.
- NMT by Jointly Learning to Align and Translate introduced the attention mechanism in NLP. Acknowledging that the use of a fixed-length

vector is a bottleneck in improving the performance of NMT, the ~~reshe~~ researchers propose to extend this by allowing a model to automatically (soft)-search for parts of a source sentence that are relevant to predicting to a target word.

- Convolutional over Recurrent Encoder for NMT arguments the standard RNN encoder in NMT with additional convolutional layers in order to capture wider context in the encoder output.
- Google built its own NMT system, called Google's NMT, which addresses many issues in accuracy and ease of deployment. The model consists of a deep LSTM network with 8 encoder and 8 decoder layers using residual connections as well as attention connections from the decoder n/w to the encoder.
- Instead of using Recurrent Neural networks Facebook AI Researchers uses Convolutional neural networks for Sequence-to-Sequence learning tasks in NMT.

Q5. Explain following terms :

a. Phonology

It is the study of organizing sounds systematically in an NLP (natural language processing) system.

b. Morphology

It is the study of construction of words from primitive meaningful units.

c. Lexical Analysis

Lexicon is words and phrases in language. Lexical Analysis deals with recognition and identification of structure of sentences. It divides paragraphs in sentences, phrases & words.

d. Syntatic Analysis

In this, sentences are parsed as noun, verbs, adjective and other parts of sentences. In this phase grammer of sentence is analyzed in order to get relationship among different words in sentences.
e.g. 'Mango eats me' will get rejected by the analyzer.

e. Word sense disambiguation

While using words that have more than one meaning, we have to select meaning which makes most sense in context. For e.g, we are typically given list of words associated word senses e.g, from dictionary or from an online Resource such as word net.