

Practical- 3

Aim- Introduction to open source NLP tools like NLTK, GATE or UIMA.

Theory -

Natural language processing (NLP), combines computational linguistics - rule based modeling of human language - with statistical and machine learning models to enable computers & digital devices to recognize, understand & generate text & speech.

NLP tools -

- i) NLTK : Natural language toolkit, is a popular library for working with human language data in python .
 - It has a variety of tools that can break down sentences into words (tokenization), group similar words together (stemming & lemmatization), understand the grammatical structure (parsing) and even analyse the meaning (semantic reasoning) to some extent .
 - It also provides access to pre-existing collection of text data (corpora) & resources like WordNet, which helps understand relationship b/w words .
 - NLTK Ps a friendly environment to learn the fundamentals of NLP and experiment with building your own language processing applications .

2] GATE : General architecture for text engineering, is another valuable open source NLP tool.

- While NLTK with its Python focus and beginner friendliness, GATE offers a different approach. Imagine GATE as a workbench for building NLP pipelines.
- It provides a graphical user interface & a modular design allowing you to chain together various processing components like tokenization, named entity recognition, & sentiment analysis. This modularity makes GATE powerful for complex NLP tasks.
- GATE requires some knowledge of Java.
- If you need a more customizable platform for building complex NLP workflows, GATE is a powerful option.

3] UIMA : unstructured information management architecture, is another open source framework for NLP tasks, similar to GATE.

- UIMA focuses on building complex NLP workflows.
- It provides a platform for integrating various NLP components, allowing you to create pipelines for tasks like info extraction, sentiment analysis and summarization.
- Both UIMA & GATE offers a modular design, enabling you to assemble different NLP components into customized workflows. This makes them ideal for tackling complex tasks that require multiple processing steps.
- Unlike GATE's graphical user interface, UIMA requires more technical expertise and programming knowledge, often in Java. This makes it less beginner friendly.

compared to NLTK's python focus.

- UIMA is known for its scalability & flexibility. It can handle large volumes of text data & integrates well with enterprise systems.

Conclusion - Thus, we have studied about the open source NLP tools like NLTK, GATE and UIMA in detail.

Practical-4

Aim - Perform text pre processing on the chosen dataset.

Theory -

- Text preprocessing is the foundation of working with textual data for computers.
- It is like building a bridge between the raw messy world of human language & the structured world that computer understand.
- The process involve cleaning & transforming text data to prepare it for analysis. This can involve several steps -
 - 1) Breaking down the text - sentences are broken into words and punctuation may be removed.
 - 2) Standardization - Text is converted to lowercase to avoid confusion.
 - 3) Reducing redundancy - Common words like the & a might be removed.
 - 4) Understanding the core - Techniques like stemming & lemmatization can group different forms of a word together.
- By pre processing text, we make it easier for computers to analyze the input into complete tasks like sentiment analysis etc.
- Whenever the data is gathered from different sources it is collected in raw format which is not feasible for the analysis. It is called the data pre-processing.
- By cleaning, transformation & integrating your data, Preprocessing leads to more reliable & insightful results.

The steps used for data preprocessing usually fall into two categories -

- 1) selecting data objects and attributes for the analysis, creating / changing the attributes.
 - 2) The following data preprocessing techniques are implemented here : handling of missing data, splitting the dataset into training & testing datasets, feature scaling.
- Data preprocessing tackles the issue to ensure the data is high quality & formatted correctly for the specific task, at hand.

Practical - 2

Aim - Write a program to develop robot traversal problem using perceptron model & multi category model.

Theory -

i) Perceptron Model -

- It is the simplest artificial neural network architecture.
- It was introduced by Frank Rosenblatt in 1957s.
- It is the simplest type of ~~back~~ feedforward neural network, consisting of a single layer of input nodes that are fully connected to a layer of output nodes.
- It can learn the linearly separable patterns. It uses slightly different types of artificial neurons known as threshold logic units.

• Types of Perceptron -

- 1) single layer
- 2) multi layer .

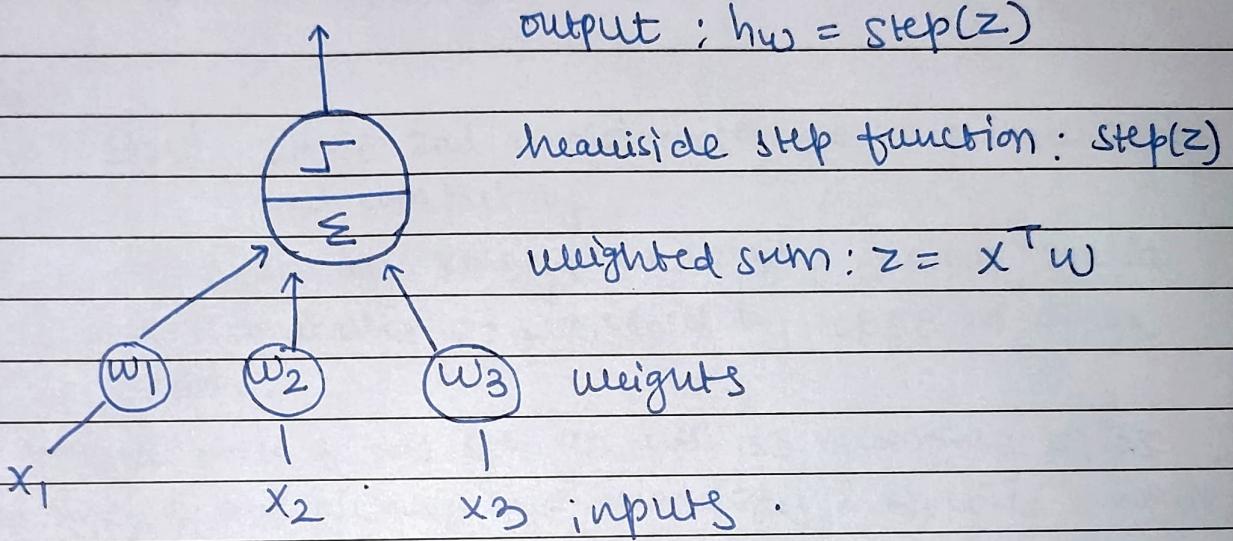
• Basic components -

- 1) Input feature
- 2) weights
- 3) summation function
- 4) Activation function
- 5) Output .

6) Bias

7) Learning algorithm .

- A perceptron has a single layer of TLU connected to all inputs.



- limitations :-

- 1) limited to linearly separable problems
- 2) convergence issues with non-separable data
- 3) requires labeled data
- 4) sensitivity to input scaling
- 5) lack of hidden layers .

Conclusion - Thus, we have studied how to use the perceptron model & multi category model .

Practical - 8

Aim - submission of case study (Research paper) on Topic "Improving emergency medical services : Using AI & image preprocessing to track & detect ambulances .

Theory -

WCONF 2024 : IEEE 2nd world conference on communication and computing .

- It is organized by Kalinga university , Raipur , India , and is technically co sponsored by IEEE madhya pradesh section .
- It target state of the art as well as emerging topics pertaining to engineering and technology & effective strategies for implementation an intelligent system .
- The objective of this world conference is to provide opportunities for the researchers , academics , industry persons and students to interact and exchange ideas , experience and expertise in the current trend & strategies .

The paper " Improving emergency medical services : Using AI & image preprocessing to track & detect ambulance " is published in this WCONF 2024 , with the paper ID - 1436 & was submitted on 20th April 2024 .

Expt. No. 8

Date _____

Page No. 15

Conclusion - thus, we have studied the case study research paper in the practical & successfully submitted it in the WCONF 2024.

Teacher's Signature _____