Experiment No.1
Study various applications of NLP and Formulate the Problem Statement for Mini Project based on chosen real world NLP applications
Date of Performance:
Date of Submission:

Aim: Study various applications of NLP and Formulate the Problem Statement for Mini Project based on chosen real world NLP applications.

Objective: Understand the different applications of NLP and their techniques by reading and critiquing IEEE/ACM/Springer papers.

Theory:

1. **Machine Translation**

Machine translation is a process of converting the text from one language to the other automatically without or minimal human intervention.

2. **Text Summarization**

Condensing a lengthy text into a manageable length while maintaining the essential informational components and the meaning of the content is known as summarization. Since manually summarising material requires a lot of time and is generally difficult, automating the process is becoming more and more popular, which is a major driving force behind academic research.

Text summarization has significant uses in a variety of NLP-related activities, including text classification, question answering, summarising legal texts, summarising news, and creating headlines. Additionally, these systems can incorporate the creation of summaries as a middle step, which aids in shortening the text.

The quantity of text data from many sources has multiplied in the big data era. This substantial body of writing is a priceless repository of data and expertise that must be skillfully condensed in order to be of any use. A thorough investigation of NLP for automatic text summarization has been necessitated by the increase in the availability of documents. Automatic text summarising is the process of creating a succinct, fluid summary without the assistance of a human while maintaining the original text's meaning.

3. Sentiment Analysis

Sentiment analysis, often known as opinion mining, is a technique used in natural language processing (NLP) to determine the emotional undertone of a document. This is a common method used by organisations to identify and group ideas regarding a certain good, service, or concept. Text is mined for sentiment and subjective information using data mining, machine learning, and artificial intelligence (AI).

Opinion mining can extract the subject, opinion holder, and polarity (or the degree of positivity and negative) from text in addition to identifying sentiment. Additionally, other scopes, including document, paragraph, sentence, and sub-sentence levels, can be used for sentiment analysis.

Businesses must comprehend people's emotions since consumers can now communicate their views and feelings more freely than ever before. Brands are able to listen carefully to their customers and customise their products and services to match their demands by automatically evaluating customer input, from survey replies to social media chats.

4. Information Retrieval

A software programme that deals with the organisation, storage, retrieval, and evaluation of information from document repositories, particularly textual information, is known as information retrieval (IR). The system helps users locate the data they need, but it does not clearly return the questions' answers. It provides information about the presence and placement of papers that may contain the necessary data. Relevant documents are those that meet the needs of the user. Only relevant documents will be pulled up by the ideal IR system.

5. Question Answering System (QAS)

Building systems that automatically respond to questions presented by humans in natural language is the focus of the computer science topic of question answering (QA), which falls under the umbrella of information retrieval and natural language processing (NLP).

CSE-DS
Tanvi Patil (45)
Aditya Joshi (24)

Sentiment Analysis Using BERT

Abstract:

Sentiment analysis, also known as opinion mining, is a natural language processing (NLP) task that involves determining the sentiment or emotional tone expressed in a piece of text. This analysis plays a crucial role in understanding public opinion and feedback across various domains, including social media, customer reviews, and news articles. In this study, we explore the application of Bidirectional Encoder Representations from Transformers (BERT), a state-of-the-art NLP model, for sentiment analysis. BERT is renowned for its ability to capture context and semantics effectively, making it well-suited for sentiment analysis tasks. We utilize pre-trained BERT models and fine-tuning techniques to classify text into different sentiment categories, such as positive, negative, or neutral. This research aims to provide an in-depth analysis of BERT's performance in sentiment analysis and its potential applications in understanding and interpreting human emotions.

Methodology:

1. **Data Collection:** Gather a dataset containing text samples, such as social media posts, reviews, or news articles, labeled with sentiment categories (e.g., positive, negative, neutral). Datasets can be obtained from various sources, including research repositories and commercial sources.
2. **Data Preprocessing:** Prepare the text data by cleaning and tokenizing it. Remove any noise, such as special characters, punctuation, and HTML tags. Perform text normalization and handle any missing or null values in the dataset.
3. **Data Splitting**: Split the dataset into training, validation, and testing sets. Common splits include 70% for training, 15% for validation, and 15% for testing.
4. **BERT Model Selection:** Choose a pre-trained BERT model, such as BERT-base or BERT-large, as the backbone for sentiment analysis. These models have been pre-trained on a massive amount of text data and are readily available in NLP libraries.

5. **Fine-Tuning:** Fine-tune the selected BERT model on the training data for the specific sentiment analysis task. This process involves modifying the model's parameters to adapt it to the sentiment classification objective. Fine-tuning typically involves training for a specific number of epochs with appropriate hyperparameters.
6. **Model Evaluation:** Use the fine-tuned BERT model to predict sentiment labels for the test data. Evaluate the model's performance using various metrics, including accuracy, precision, recall, and F1-score.

Tools and Libraries:

1. **Transformers:** A popular Python library that provides pre-trained models for BERT and other transformer-based NLP architectures.
2. **PyTorch or TensorFlow**: Deep learning frameworks used for building and fine-tuning BERT-based models.
3. **Pandas:** A Python library for data manipulation and analysis to handle the dataset effectively.
4. **Matplotlib and Seaborn**: Data visualization libraries for creating informative plots and visualizing model performance.
5. **Scikit-Learn**: A machine learning library for evaluating model performance with various metrics.

Sentiment Analysis with BERT: BERT (Bidirectional Encoder Representations from Transformers) is a transformer-based model that excels in understanding context and semantics in natural language. It is pre-trained on massive text corpora and has achieved state-of-the-art results in a wide range of NLP tasks. For sentiment analysis, BERT can capture the nuances and intricacies of sentiment in text by considering the entire context and word relationships.

In conclusion, sentiment analysis using BERT is a powerful approach that leverages the capabilities of modern NLP models to extract valuable insights from text data, enabling businesses and researchers to better understand public sentiment and user feedback in various domains. This research explores the application of BERT in sentiment analysis and demonstrates its effectiveness in classifying text into sentiment categories.