Chapter 8
NLP for Chatbot Application:
Tools and Techniques Used for Chatbot Application, NLP Techniques for Chatbot, Implementation

Shyamala Devi Nithyanandam
VISTAS, India

Sharmila Kasinathan
VISTAS, India

Devi Radhakrishnan
VISTAS, India

Jebathangam Jebapandian
VISTAS, India

ABSTRACT

The chatbot is one of the increasing number applications in the era of conversational series. It is a virtual application that can efficiently interact with any human being using the deep natural language processing skills. In NLP, for chatbot application, the various techniques needed for chatbot using NLTK tool are explained and implemented. The process of converting the text to numerical value is called text embedding. In NLTK tool, various text embedding tools are available such as TF-IDF vectorization and bag of words. Deep NLP is an efficient way to implement the chatbot. Thus the chatbot is implemented with sequence-to-sequence networks.

INTRODUCTION

The Chatbot's are one of the applications which are growing in the era of conversational series. It is a virtual application which can efficiently interact with any human being using the Deep Natural language processing skills. The reality of Chatbots is the integration of machine learning technique where the data is trained to build a relatable model. The proficiency with chatbots is its ability to understand the queries to provide quick and relevant response to the users. Its instantaneous adroitness pertaining to messaging framework augments the efficiency of service multifariously. The construction of a chatbot application can be easily implemented due to its autonomist nature that accelerates quick responses. The natural language processing is categorised into classical and deep learning models, where the former is a set of tools and techniques which enable the computer machine to learn and understand the natural linguistic language such as text and voice of the humans. The tools and techniques that are used are machine translation, text summarization, information retrieval and multilingual cross language information retrieval etc. but there is certain limitation in the classical natural language processing system, where it will not train huge amount of online social media data for chatbot application. Thus, the classical natural language processing system is taking a backseat, with more migrative utilization towards the Deep Natural language processing system. Deep Neural network which has multiple hidden layers aids in training the deep expressive data and renders good result. The deep NLP is the combination of deep learning and Natural language processing system where the former comprises of if-else rules, Audio frequency component analysis (Speech recognition), Bag of words of model (Classification) and Convolution neural network for text recognition (text classification).

The deep NLP holds an end-to-end deep learning model, and applies the deep neural network architecture with various deep learning algorithms for classifying the text-based inputs from the neural network. Some of the stratifications of these algorithms are logistic regression, linear regression, Naïve Bayes, random forest, support vector machine and passive aggressive classifier.

The deep NLP is categorised into different word embedding techniques for converting the words or text in to Numerical vector The word or text which is generated by the human being can be converted to a numerical vector space, to be subsequently fed into the deep neural network architecture. Once the completion of text vectorization is done, the weighted data is applied to deep neural network. Some of the models used in this process are Bag of words, binary encoding, TF-IDF vectorization.

The various deep neural network frameworks are elaborately detailed through the recurrent neural network architecture, sequence to sequence architecture, convolution neural network architecture, gated recurrent neural network and contextual LSTM architecture. Recurrent Neural Network (RNN) is a family of neural networks,that generates the output of the previous layer to be passed as input to the current layer. Sequence to sequence model is a derivation of the RNN. Convolution neural network is a most efficient model to recognize the image of the text, and gated neural network allows the network to find the increment of layers. The Contextual LSTM is used to learn the context of text and to understand the sematic of the text entailed.

All the above content which gives an explanation to implement the Chatbot application hold lesser reference to the data pre-processing techniques for developing the chatbot application. Thus, some of the techniques that requires further exploration for adequate cognizance to effectively delve into the chatbot data pre-processing are tokenization, lemmatization, stemming and stop word removal. The sequence-to-sequence model is a dialog and machine translation system which comprises of two recurrent neural

Related Content

News Classification to Notify About Traffic Incidents in a Mexican City

Alejandro Requejo Flores, Alejandro Ruiz, Abraham Lópezand Raul Porras (2021). *Handbook of Research on Natural Language Processing and Smart Service Systems (pp. 227-244).*

www.irma-international.org/chapter/news-classification-to-notify-about-traffic-incidents-in-a-mexican-city/263104

Light Weight Structure Texture Feature Analysis for Character Recognition Using Progressive Stochastic Learning Algorithm

S. Rubin Bose, Raj Singh, Yashodaye Joshi, Ayush Marar, R. Reginand S. Suman Rajest (2024). *Advanced Applications of Generative AI and Natural Language Processing Models (pp. 144-158).*

www.irma-international.org/chapter/light-weight-structure-texture-feature-analysis-for-character-recognition-using-progressive-stochastic-learning-algorithm/335837

Mathematical Information Retrieval Trends and Techniques

Pankaj Dadure, Partha Pakrayand Sivaji Bandyopadhyay (2021). *Deep Natural Language Processing and AI Applications for Industry 5.0 (pp. 74-92).*

www.irma-international.org/chapter/mathematical-information-retrieval-trends-and-techniques/284204

A Critical Review of the Current State of Natural Language Processing in Mexico and Chile

César Aguilarand Olga Acosta (2021). *Natural Language Processing for Global and Local Business (pp. 365-389).*

www.irma-international.org/chapter/a-critical-review-of-the-current-state-of-natural-language-processing-in-mexico-and-chile/259797

Geospatial Situation Analysis for the Prediction of Possible Cases of Suicide Using EBK: A Case Study in the Mexican State of Aguascalientes

Carlos Manuel Ramirez López, Martín Montes Rivera, Alberto Ochoa, Julio César Ponce Gallegosand José Eder Guzmán Mendoza (2021). *Handbook of Research on Natural Language Processing and Smart Service Systems (pp. 327-346).*

www.irma-international.org/chapter/geospatial-situation-analysis-for-the-prediction-of-possible-cases-of-suicide-using-ebk/263109