**Project Report**

**TITLE**

**Recurrent Neural Network Based Text Summarization Techniques By IBM Watson**

**Done by :**

**Nihal Aftab Patel**

**Manipal Academy of Higher Education - Dubai Campus**

**TABLE OF CONTENT**

|  |  |  |
| --- | --- | --- |
| Sr.No. | Topic | Page No. |
| 1 | Abstract | 3 |
| 2 | Introduction | 4 |
| 3 | Literature Review | 5 |
| 4 | Flow Chart | 6 |
| 5 | System Requirements | 7 |
| 6 | Experimental Results | 8 |
| 7 | Experimental Analysis | 10 |
| 8 | Advantages & Disadvantages of CNN | 11 |
| 9 | Future Scope | 12 |
| 10 | Conclusion | 13 |

Abstract

Every year over thousands of books, articles, newspapers, journals etc. are published worldwide. We live in a world where communication is the most important thing in our lives. We communicate my exchanging ideas and information to an individual or a group of persons. Over 84% of worlds population can speak and write. People read information that is widely available such as on internet, notice boards, media etc. Some people may have the habit of reading long paragraph or complex sentences whereas other people may not understand the paragraphs or sentences that are complex.

Text summarization came into picture for those who face understanding the paragraph or sentence. Text summarization summarizes the long paragraph into few sentences so that it can be understood by all.

A Recurrent Neural Network Based Text Summarization Techniques By IBM Watson is proposed to summarize the text.

Introduction

**Automatic summarization** is the process of shortening a set of data computationally, to create a subset (a [summary](https://en.wikipedia.org/wiki/Abstract_(summary))) that represents the most important or relevant information within the original content. For example, A three page of fire brigade information can be summarized into steps instructing the tenants or user of the building to action in case of accident.

Many technologies were developed for the text summarization. One of them is recurrent neural network which is an artificial neural network helpful in modelling sequence data. It is derived from feedforward networks.

Literature Review

Problem Statement: People seek knowledge and information by reading, listening and speaking. Text summarization converts large paragraphs into few sentences without losing the key themes of the text. There have been techniques developed on text summarization. One is Sentence Scoring based on Word Frequency, which uses weights to score the frequent words. Other technique is TextRank using Universal Sentence Embeddings.

Solution: The constant flow of data and the algorithms needed to train the data to make the machines takes decision has evolved over the years. Our solution to the text summarization is Recurrent Neural Network Based Text Summarization Techniques by IBM Watson. The system takes an input looks for the important key words of the theme and converts in to summarized text of few sentences.

Flowchart

Start

Import Libraries

End

Deploy on IBM Cloud

Prediction

Model Building

Data Preprocessing

System Requirements

Software:

Spyder (anaconda3)

Jupyter Notebook(anaconda3)

IBM Cloud

Anaconda Navigator

Hardware:

Processor: AMD Ryzen 3 4300U

OS: Windows 10

Storage: 1TB

**Experimental Results**

Output on Training the model on Local System

A picture containing text

Description automatically generated

Text

Description automatically generated

Deployment Output

Text, timeline

Description automatically generated

**Experimental analysis**

A few libraries were implemented for model building such as Tensorflow, Flask etc. And the train data was trained for 10 epochs and the model is implemented using recurrent neural network. The flask app is developed to link with the HTML file and is deployed on IBM cloud.

Advantages of RNN

* The principal advantage of RNN over ANN is that RNN can model a collection of records (i.e. time collection) so that each pattern can be assumed to be dependent on previous ones.
* Recurrent neural networks are even used with convolutional layers to extend the powerful pixel neighbourhood.

Disadvantages of RNN

* Gradient exploding and vanishing problems.
* Training an RNN is a completely tough task.
* It cannot system very lengthy sequences if the usage of Tanh or Relu as an activation feature.

Future Scope:

Text Summarization using Artificial Intelligence and Machine Learning can be used in various fields such as medical, aerospace and research field

Another suggestion is to add a grammar check feature which corrects the grammar mistakes in the input data before converting it into summarized text.

Conclusion:

The project developed to summarize the text using recurrent neural network and IBM Cloud. After 10 epochs the system can predict the text and produce the output. This project can be further be expanded to include enterprise system in various fields.