PREDICTING SIMILARITY IN SENTENCES THROUGH WORD EMBEDDING

A DISSERTATION

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE
OF

MASTER OF TECHNOLOGY

IN

INFORMATION SYSTEMS

Submitted By:

MAYANK KUMAR MITTAL

(2K17/ISY/10)

Under the supervision of
Dr. Anil Singh Parihar
(Associate Professor, Department of CSE**)**

DEPARTMENT OF INFORMATION TECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

2017 - 19

CANDIDATE'S DECLARATION

I, Mayank Kumar Mittal, Roll No. 2K17/ISY/10 student of M.Tech Information Systems, hereby declare that the project Dissertation titled "Predicting Similarity in Sentences through Word Embedding" which is submitted by me to the Department of Information Technology, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of the degree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any degree, Diploma Associateship, Fellowship or other similar title or recognition.

Place: Delhi   Mayank Kumar Mittal

Date:

CERTIFICATE

I hereby certify that the Project Dissertation titled “Predicting Similarity in Sentences through Word Embedding” which is submitted by Mayank Kumar Mittal, Roll No 2K17/ISY/10 Information Technology, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of the degree of Master of Technology, is a record of the project work carried out by the student under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Place: Delhi **Dr. Anil Singh Parihar**

Date: **SUPERVISOR**

ACKNOWLEDGMENT

I express my gratitude to my major project guide Dr. Anil Singh Parihar, Associate Professor, Department of CSE, Delhi Technological University, for the valuable support and guidance he provided in making this major project. It is my pleasure to record my sincere thanks to my respected guide for his constructive criticism and insight without which the project would not have shaped as it has.

I humbly extend my words of gratitude to other faculty members of this department for providing their valuable help and time whenever it was required.

Mayank Kumar Mittal

Roll No. 2K17/ISY/10

M.Tech (Information Systems)

E-mail: mkmittalofficial@gmail.com

ABSTRACT

In the field of natural language processing, learning the context from a given sentence is a very important and challenging task. Which is great source for predicting the intention of user, this prediction will help to detect the fake NEWS, for creating more interactive artificial intelligent bot that will interact better respond better act better, for giving the better recommendations such as recommending music for that purpose word embedding in used to bridge the gap between the computing machine and the real world. In this research we have examined the various existing models i.e, regression models like multilinear regression, support vector machine, random forest, match LSTM to detect the similar sentences. Also, we compared their results based on accuracy achieved. Moreover, we proposed new model based on convolution neural network warping with time distributed layer which outperform with respect to other models from 77.67% to 83.72%.

Table of Contents

List of Tables

List of Figures