MERCon 2020

Moratuwa Engineering Research Conference

WORKSHOP 1

6th International Multidisciplinary Engineering Research Conference

30th July 2020

University of Moratuwa, Sri Lanka

9am to 4pm

A HANDS-ON WORKSHOP IN WORD EMBEDDING:

From Word2Vec to StarSpace

Conducted by the National Language Processing Centre (NLPC) of University of Moratuwa
https://www.mrt.ac.lk/web/nlp

This workshop is focused on giving an introduction to the concept, implementation, and the uses of neural word embedding. Neural word embedding is the collective name for a set of language modeling and feature learning techniques based on neural networks, where words or phrases from the vocabulary are mapped to vectors of real numbers. Neural word embedding is gaining a huge momentum recently, where the use of the same has reported significant breakthroughs in Natural Language Processing (NLP) applications such as Machine Translation and Question Answering.

We have planned a full day workshop, which is broken down into 4 1.5-hour sessions. The first half (2 sessions) will introduce word embeddings and provide an introduction to the very first neural word embedding. Then we will critically review the power and limitations of neural word embedding. The second half will focus on some of the state-of-the-art developments in the recent years in word embedding and will provide hands-on experience with them. Specifically, it focuses on capturing order in context words using directional networks, scaling & rotation techniques and transfer learning techniques used by Star Space, Bert, and Wang2vec. We will be using https://colab.research.google.com cloud services ***so the participants only need to bring laptops with access to the internet.***

Word embedding is one of the most popular representation of document vocabulary. It is capable of capturing context of a word in a document, semantic and syntactic similarity, relation with other words, etc.

Workshop Objectives

- To introduce participants to the concept of neural word embedding and state-of the-art word embedding models
- To provide hands-on experience with platforms, libraries and fundamental code components relevant to word embeddings
- To develop an understanding on how neural word embedding is used in developing state-of-the-art NLP applications

Goals and expected outcomes

- Use of neural embedding in the end-user applications developed by the industry and research, which would lead to improved performance of the same
- Initiative to develop neural embeddings for local languages (Sinhala and Tamil) to strengthen local language computing efforts
- Awareness creation about the NLP Centre of University of Moratuwa as a knowledge hub for the latest technologies in local language computing

Session	Topic	Speaker
1	Introduction to Word Embedding	Dr. T. Uthayasanker
2	Word Embedding Applications	Dr. Surangika Ranathunga
3	Wordmebdding Rotation & Scaling	Mr. Yasas Senarath & Prof. Sanath Jayasena
4	State of the Art + Cloud Implementations of Word Embedding	Dr. T. Uthayasanker

Participation is **FREE** of charge

Register @:
http://mercon.mrt.ac.lk/workshops.html