

ADAPT Undergraduate Internship Programme 2017

PROJECT DESCRIPTION

Institution/Team:	Theme B, DCU	
Project Title:	Investigation on different sentence embedding methods	
Suitable for students who are studying in the following areas:	Computer science, mathematics, machine learning	
Skills needed:	<p>Strong programming skills, be familiar with at least one of the following programming languages: Python/Perl/Java/C++</p> <p>Must be familiar with Linux based operating systems</p> <p>Excellent problem solving and communication skills</p> <p>Knowledge of machine learning is desirable</p>	
Project Description:	<p>Sentence embedding/vector is a distributed semantic representation of sentence or paragraph levels which is very important for current popular deep learning-based natural language processing (NLP) tasks, such as machine translation, information retrieval, question/answering, sentiment analysis.</p> <p>In this project, we will first investigate different methods of modelling sentence/paragraph embedding, such as simple words centroid, deep LSTM-RNN (Long-short term memory – Recurrent Neural Networks), CNN (Convolutional Neural Networks) etc. Then we will compare different methods on different NLP tasks as mentioned above.</p>	
The Role of the student & benefits gained from participation in this project:¹	<p>The student is expected to have the following duties:</p> <ul style="list-style-type: none"> • Word embedding: learning how to use neural networks to train distributed word representations given the data as the basis of sentence embedding. • From word embedding to sentence embedding: investigate some mainstream methods in relation to sentence/paragraph embedding • Comparison experiments: examine performance of different methods on different NLP tasks and conclude the disadvantages and advantages of each method. • <p>Through this project, the student will have an understanding of deep learning, word embedding and sentence embedding etc.</p>	
Who will be working with you?	<p>The student will be working directly with Dr. Jinhua Du, a research fellow at ADAPT. The student will also receive further guidance and support from Prof. Andy Way.</p>	
Short description of the group:	<p>We are a big group in ADAPT working on machine translation, natural language processing and machine learning etc. led by Prof. Andy Way.</p>	
Recommended Reading Material:	<ul style="list-style-type: none"> • Tomas Mikolov et al., Efficient Estimation of Word Representations in vector space • Quoc V. Le and Tomas Mikolov. Distributed Representations of Sentences and Documents. CoRR, abs/1405.4053, 2014. • Hamid Palangi et al. Deep Sentence Embedding Using the Long Short Term Memory Network: Analysis and Application to Information Retrieval. CoRR, abs/1502.06922, 2015 	
Other information:		
For further details on this project please contact:	Name: Phone: E-Mail: Website:	Jinhua Du 01 700 6714 Jinhua.du@adaptcentre.ie

¹ This is an initial description of the role of the student and it is liable to change following discussions between the investigators and the student.