Invited Talk #2

VIETNAMESE NEURAL LANGUAGE MODEL FOR NLP TASKS WITH LIMITED RESOURCES

by Prof. Quan Thanh Tho, Vietnam National University Ho Chi Minh City, Vietnam

Abstract: A statistical language model is a probability distribution over sequences of words. Language modeling is used in various computing tasks such as speech recognition, machine translation, optical character and handwriting recognition and information retrieval and other applications. Whereas n-gram is considered as a traditional language model, neural language model has been emerging recently as a means to approximate the probability of a sentence using neural networks and word embeddings. An advantage of a neural language model is that it can be further applied to other NLP tasks where the training datasets may be limited. In this talk, we realize this idea by introducing the usage of a Vietnamese neural model language trained from a large corpus of social media data. When further applying this neural model language with other NLP tasks including entity recognition, spam detection and topic modeling with relatively small training datasets; we witness improved performance achieved, as compared to other existing approaches using deep learning with typical word embedding techniques.

Bio: Dr. Quan Thanh Tho is an Associate Professor in the Faculty of Computer Science and Engineering, Ho Chi Minh City University of Technology (HCMUT), Vietnam. He received his B.Eng. degree in Information Technology from HCMUT in 1998 and received Ph.D degree in 2006 from Nanyang Technological University, Singapore. His current research interests include formal methods, program analysis/verification, the Semantic Web, machine learning/data mining and intelligent systems. Currently, he heads the Department of Software Engineering of the Faculty. He is also serving as the Chair of Computer Science Program (undergraduate level).