School of Data Science
香港城市大學
City University of Hong Kong

SDSC SEMINAR

Making a Neural Network Smart Enough Not to Cross The Territory It Is Designed for: A Small Step

Date: 13 June 2024 (Thursday)

Time: 10:00am - 11:00am

Venue: Rm 2308, Li Dak Sum Yip Yio Chin Academic Building, City University of Hong Kong

ABSTRACT

Most of the "intelligent systems" lack "explainability", i.e., fail to explain why a decision has been made. Consequently, in application areas like medicine, defense, and judiciary domain experts are reluctant to use intelligent systems as the decision-making authority. One of the reasons for this is that often such systems make decisions going beyond the territories for which they were designed. For example, a neural network, deep or shallow, makes a decision, often with high confidence, even when a test data point is far from the training data used to design the system. Further, an MLP may be trained to distinguish between say, four kinds of childhood cancers, but if a test sample represents a normal patient or a colon cancer patient, the MLP will classify it into one of the four classes for which it was trained. So, as such, MLP (and many other intelligent decision-making systems) cannot deal with the "open world" nature of the problem. In this talk, we primarily focus on MLPs, but the idea is easily extendable to other learning systems. We want to address these problems by equipping neural networks not to make any judgments when they should not. We propose two approaches to get a practical solution to this problem. For this, we estimate the domain of the training data ("sampling window") using two schemes. We show an asymptotic optimal property of our estimate along with some other interesting results. Then to train the network we use some instances randomly drawn from outside the sampling window and label them as coming from a class called, "Don't know". In the first approach, for a c-class problem, we train c neural networks and fusion of them equips the network to refuse making decisions under appropriate conditions. As a byproduct, we obtain incremental learning ability of the system. In the second approach, we train a single network with c+1 classes. We devise an interesting mechanism so that the "Don't" class can be represented by a small number of instances. We illustrate our method with several data sets. Finally, we discuss the limitations of the proposed systems leading to the scope of further work.

Professor Nikhil R. Pal

GUEST SPEAKER'S PROFILE

Nikhil R. Pal was a Professor in the Electronics and Communication Sciences Unit and was the founding Head of the Center for Artificial Intelligence and Machine Learning of the Indian Statistical Institute. His current research interest includes brain science, computational intelligence, machine learning and data mining. He was the Editor-in-Chief of the IEEE Transactions on Fuzzy Systems for the period January 2005-December 2010. He has served/been serving on the editorial /advisory board/ steering committee of several journals including the International Journal of Approximate Reasoning, Applied Soft Computing, International Journal of Neural Systems, Fuzzy Sets and Systems, IEEE Transactions on Fuzzy Systems and the IEEE Transactions on Cybernetics. He is a Fellow of the West Bengal Academy of Science and Technology, Institution of Electronics and Tele Communication Engineers, National Academy of Sciences, India, Indian National Academy of Engineering, Indian National Science Academy, International Fuzzy Systems Association (IFSA), The World Academy of Sciences, and a Fellow of the IEEE, USA.

Enquiries: sdscgo@cityu.edu.hk

All are welcome