TuePM1-3 **Pattern Recognition 1**
Chair: Tong Zhiqiang
Room: International Conference Hall III

13:20 - 13:40	Weighted Discriminant Analysis and Kernel Ridge Regression Metric Learning for Face Verification Siew-Chin Chong[1], Andrew Jin Teoh[2], Thian-Song Ong[1] [1]Multimedia University [2]Yonsei University
13:40 - 14:00	An Incremental One Class Learning Framework for Large Scale Data Qilin Deng[1] Yi Yang[1] Furao Shen[1] Chaomin Luo[2] Jinxi Zhao[1] [1]Department of Computer Science and Technology, Nanjing University [2]Department of Electrical and Computer Engineering, University of Detroit Mercy
14:00 - 14:20	Gesture Spotting by Using Vector Distance of Self-Organizing Map Yuta Ichikawa[1], Shuji Tashiro[1], Hidetaka Ito[1], Hiroomi Hikawa[1] [1]Kansai University
14:20 - 14:40	Cross-Database Facial Expression Recognition via Unsupervised Domain Adaptive Dictionary Learning Keyu Yan[1], Wenming Zheng[1], Zhen Cui[1], Yuan Zong[1] [1]Southeast University
14:40 - 15:00	Adaptive Multi-View Semi-Supervised Nonnegative Matrix Factorization Jing Wang[1], Xiao Wang[2], Feng Tian[1], Chang Hong Liu[1], Hongchuan Yu[1], Yanbei Liu[3] [1]Bournemouth University [2]Tsinghua University [3]Tianjin University

TuePM2-3 **Pattern Recognition 2**
Chair: Hiroomi Hikawa, Tong Zhiqiang
Room: International Conference Hall III

15:20 - 15:40	Robust Soft Semi-Supervised Discriminant Projection for Feature Learning Xiaoyu Wang[1], Zhao Zhang[1], Yan Zhang[1] [1]Soochow University
15:40 - 16:00	A Hybrid Pooling Method for Convolutional Neural Networks Tong Zhiqiang[1] Kazuyuki Aihara[1] Gouhei Tanaka[1] [1]The University of Tokyo
16:00 - 16:20	Multi-nation and multi-norm License plates detection in real traffic surveillance environment using Deep Learning Amira Naimi[1] Yousri Kessentini[1] Mohamed Hammami[1] [1]MIRACL
16:20 - 16:40	A study on cluster size sensitivity of fuzzy c-means algorithm variants Laszlo Szilagyi[1], Sandor Miklos Szilagyi[2], Calin Enachescu[2] [1]Sapientia University of Transylvania, Tirgu Mures, Romania [2]Petru Maior University
16:40 - 17:20	Invited talk Statistical mechanics of pre-training and fine tuning in deep learning Masayuki Ohzeki[1] [1]Graduate School of Informatics, Kyoto University.

Invited talk

Statistical mechanics of pre-training and fine tuning in deep learning

Masayuki Ohzeki[1]
[1]Graduate School of Informatics, Kyoto University

Abstract

Lack of analytical study on the architecture of deep learning hampers understanding of its origin of the outstanding performance. A recent study has formulated a theoretical basis for the relationship between the recursive manipulation of variational renormalization groups and the multi-layer neural network in deep learning [2]. Indeed, it is confirmed that the renormalization group indeed can mitigate the computational cost in the learning without any significant degradation [3]. The statistical mechanical approach is a rare successful approach to pave a way to the nature of the deep learning. We present a statistical-mechanical analysis on a part of architecture in the deep learning. We first elucidate some of the essential components of deep learning —pre-training by unsupervised learning and fine tuning by supervised learning. We formulate the extraction of features from the training data as a margin criterion in a high-dimensional feature-vector space. The self-organized classifier is then supplied with small amounts of labelled data, as in deep learning. For simplicity, we employ a simple single-layer perceptron model, rather than directly analyzing a multi-layer neural network. The surprising performance of the deep learning does not necessarily come from deep neural network, but rather stem from the potential of the neural network itself. We find a nontrivial phase transition that is dependent on the number of unlabelled data in the generalization error of the resultant classifier. The resultant phenomena exhibits the efficacy of the unsupervised learning in deep learning. Increasing the number of unlabelled data again leads to an improvement in the generalization error. A gradual increase in the number of labelled data allows us to escape from a metastable solution in multiple solutions. In this sense, fine tuning by supervised learning is necessary to achieve the lower-error state and mitigate the difficulties in reaching the desired solution. We should emphasize that the emergence of the metastable state does not come from the multi-layer neural networks, but from the combination of unsupervised and supervised learning. The analysis is performed by the replica method, which is a sophisticated tool in statistical mechanics. We validate our result in the manner of deep learning, using a simple iterative algorithm to learn the weight vector on the basis of belief propagation.

Reference:

[1] M. Ohzeki: Journal of the Physical Society of Japan 84, 034003 (2015)
[2] P. Mehta and D. J. Schwab: arXiv:1410.3831.
[3] K. Tanaka, S. Kataoka, M. Yasuda, and M. Ohzeki: Journal of the Physical Society of Japan 84, 045001 (2015)