

Workshop#5

Recent Advances in Interpreting and Understanding DNNs

Dr. Erfan Loweimi, Dr. Samira Loveymi

Deep neural networks (DNNs) are among the major breakthroughs in Machine Learning and Artificial Intelligence. They play a central role in building state-of-the-art pattern recognition systems and have led to remarkable performance improvement in a wide variety of tasks and applications. However, despite outstanding progress from engineering perspective in developing reliable large-scale data and information processing systems, the understanding about their deep structure has remained shallow. This has triggered an expanding body of work aiming at deciphering the DNNs behaviour as black boxes.

This tutorial reviews four lines of research which have shed some light on such black boxes and have been insightful in interpreting and understanding the DNNs:

- We start by looking into the information bottleneck technique which employs some concepts from Information Theory to explain the learning process evolution in DNNs.
- The second part is dedicated to a fast-growing research direction which addresses the following questions: why despite being highly overparameterized, and contrary to the conventional wisdom in machine learning, DNNs generalize well?
- In the final part, we look at the training dynamics of the DNNs: using a set of experiments in a controlled condition, we demonstrate how during the training process their behaviour evolves towards finding an optimal solution.

Dr. Erfan Loweimi is a research associate with the Department of Engineering, King's College London (KCL), working on the EPSRC-funded Speech Wave Project. From 2018 to 2021 he was a postdoctoral researcher in the Centre for Speech Technology Research (CSTR), University of Edinburgh. He received his B.Sc., M.Sc. and Ph.D. degrees from the Shahid Chamran University of Ahvaz (2007), Amirkabir University of Technology (2010) and University of Sheffield (2018), respectively. He is author/co-author of more than 35 peer-reviewed papers and his research interests revolve around acoustic modelling from raw signal representations, end-to-end speech processing, model-based robust ASR and phase-based speech signal processing.



Dr. Samira Loveymi is an adjunct lecturer in the Computer Engineering Department, Shahid Chamran University of Ahvaz, Ahvaz, Iran. She holds a Ph.D. degree in Artificial Intelligence from Bu-Ali Sina University, Hamedan, Iran (2020). She received her B.Sc. in Software Engineering and M.Sc. in Artificial Intelligence from the Shahid Chamran University of Ahvaz, in 2007 and 2010, respectively. Her current research interests include machine vision, deep learning, and visual question answering.



Skyroom Link: <https://www.skyroom.online/ch/mvip2022/workshop>

Time and Date (In Iran):

5PM-7PM {22 February 2022 (3th of Esfand 1400)}