Guest editorial: special issue on adversarial learning in computational intelligence

Zixing Zhang, Dimitris N. Metaxas, Hung-Yi Lee, Björn W. Schuller

Angaben zur Veröffentlichung / Publication details:

Zhang, Zixing, Dimitris N. Metaxas, Hung-Yi Lee, and Björn W. Schuller. 2020. "Guest editorial: special issue on adversarial learning in computational intelligence." *IEEE Transactions on Emerging Topics in Computational Intelligence* 4 (4): 414–16. https://doi.org/10.1109/tetci.2020.3006295.

Guest Editorial Special Issue on Adversarial Learning in Computational Intelligence

I. Introduction

ADVERSARIAL learning has attracted tremendous attention in the community of machine learning over the past few years. It normally integrates two components that contest with each other in a two-player zero-sum game. Since its birth in 2014, adversarial learning has been widely applied to not only the generation of realistic images, but also many other research topics, such as data augmentation, domain adaptation, and adversarial attack, often leading to appealing performance. However, we have just witnessed the early rise of this technique, and still confront many challenges, for example, the mode collapse problem, and the interpretability of its results and failures. Computational Intelligence (CI) technologies are expected to provide efficient solutions to deal with the raised challenges. Moreover, most of the previous adversarial-learning studies are largely limited in addressing static images or feature vectors. It still remains largely an open question of how adversarial learning performs for other complex and temporally variational signals or modalities, such as speech and text.

This special issue aims to capture the most recent advances of adversarial learning from both theoretical and empirical perspectives. Moreover, it attempts to present its novel applications to other domains beyond image generation. Following a rigorous peer-review process, seven papers out of 20 received submissions have been accepted for inclusion in this special issue. The topics of these papers range from computer vision, natural language processing, and speech processing to cyber networks. Specifically, the special issue is organized as follows.

In the first paper, "Improving Adversarial Neural Machine Translation for Morphologically Rich Language", Mi, Xie and Zhang utilize generative adversarial networks (GANs) to improve the quality of translated language in neural machine translation (NMT). For NMT, when training the discriminator of a GAN, the conventional methods concatenate the word embeddings of two languages as input, while considering only one reference for the translated text. These procedures, however, are not reasonable in morphologically rich languages. To this end, the authors extend these methods by exploiting morphological word embedding as inputs of the discriminator of GAN, and using multiple reference translations instead of a single one. This method significantly improves the NMT performance on eight NMT tasks.

The next paper "Hardening Random Forest Cyber Detectors Against Adversarial Attacks" by Apruzzese, Andreolini, Colajanni and Marchetti, improves the model training strategy to enhance the model robustness in the context of a cyber attack. Conventionally, the machine learning-based cyber detectors are vulnerable to the targeted adversarial attacks, which partially owns to the rigid classification produced by hard class labels of training samples. To solve this issue, the authors introduce some degree of flexibility and uncertainty in the training process by using probability labels, which allows the algorithm to capture additional information between classes such as similarity and reduces the weakness of adversarial attack.

In the third paper "Static2Dynamic: Video Inference from a Deep Glimpse", Yeh, Liu, Chiu and Wang utilize GANs to improve the quality of synthesized videos in video inference. The video inference aims to infer a sequence from non-consecutive frames (images). For this purpose, the authors design a novel GAN structure, namely Stochastic and Recurrent Conditional GAN. Especially, the generator component contains an image encoder and a recurrent neural network (RNN) based temporal encoder to obtain the prior distribution of the given images. The outputs the encoders are then fed into another RNN-based temporal decoder and an image decoder to recover the original video sequence. The capability of this model is also shown in solving other video generation tasks, such as video interpolations and video predictions.

The fourth paper entitled "A System-Driven Taxonomy of Attacks and Defenses in Adversarial Machine Learning" by Sadeghi, Banerjee and Gupta, provides a comprehensive adversarial attack and defense survey, to help the researchers from the CI community select and design more robust machine learning model. This survey contributes to building a fine-grained system-driven taxonomy to specify adversarial system models in an unambiguous manner.

The paper "Unsupervised Representation Disentanglement using Cross Domain Features and Adversarial Learning in Variational Autoencoder based Voice Conversion", Huang, Luo, Hwang, Lo, Peng, Tsao and Wang exploit both the advantages of GAN and Domain adversarial training (DAT) for voice conversion (VC), where one of the major issues is how to disentangle the speech content and speaker-related information. In this paper, the authors present an extended cross-domain variational autoencoder VC framework, in which a GAN is used to approximate the distribution of real speech signals better, and DAT is applied to the latent code as an explicit constraint

to eliminate speaker-dependent factors and retain the speech content.

DAT has shown promising performance in mitigating the domain mismatch, whereas it cannot guarantee that the learned representation spaces from different domains are class-wise aligned. In the paper "Learning Class-aligned and Generalized Domain-invariant Representations for Speech Emotion Recognition", Xiao, Zhao and Li propose a class-aligned DAT, by adding a limited amount of annotated samples from different domains for supervised classification. This algorithm shows to be effective by an empirical evaluation in speech emotion recognition.

Last but not least, Chaturvedi and Garain in their paper "Attacking VQA Systems via Adversarial Background Noise" investigate adversarial attacks for visual question answering systems by only modifying the background pixels. It visualizes how an attention mechanism can be distracted by noise, and what is the difference of the disturbing noise for distinct models.

II. Conclusion

From this review, one can see that the special issue generally achieved its goal of capturing a snapshot of cutting-edge algorithms and applications of adversarial learning in a variety of domains. Particularly, the first and third articles focus on designing and implementing novel GAN structures and corresponding training strategies; the six article attempts to improve the classic DAT algorithms by taking limited supervised information into account; the fourth article coordinates GAN with DAT to utmost explore the advantage of adversarial learning; and the second, fourth, and seventh articles investigate or review the advanced adversarial attacks and defense strategies. Applications are well represented in contributions to this special issue, comprising the applications for NMT, cyber security, video inference, speech emotion recognition, voice conversion, and VQA. The data in most of these applications are in sequence, which is complex and dynamic than static images. Because of these high-quality contributions and their interesting findings, we foresee that this special issue will facilitate the future research and development work of adversarial learning in CI.

Acknowledgment

The guest editors are thankful to all the authors who submitted their excellent contributions to this special issue, and appreciate all the reviewers for dedicating their efforts in completing timely and constructive reviews. The guest editors especially thank the Editor-in-Chief, Prof Yew-Soon Ong, and members of the editorial team for their support during the editing process of this Special Issue.

Zixing Zhang, *Guest Editor*
Department of Computing
Imperial College London
London SW7 2AZ, UK
zixing.zhang@tum.de

Dimitris N. Metaxas, *Guest Editor*
Department of Computer Science
Rutgers The State University of New Jersey
NJ 08854-8019, USA
dnm@cs.rutgers.edu

Hung-Yi Lee, *Guest Editor*
Department of Electrical Engineering
National Taiwan University
hungyilee@ntu.edu.tw

Björn W. Schuller, *Guest Editor*
Chair of Embedded Intelligence for Health Care and Wellbeing
University of Augsburg
Augsburg 86159, Germany
schuller@ieee.org

Zixing Zhang received his master degree in physical electronics from the Beijing University of Posts and Telecommunications (BUPT), China, in 2010, and his PhD degree in computer engineering from the Technical University of Munich (TUM), Germany, in 2015. Currently, he is a Research Associate with the Department of Computing at Imperial College London (ICL), UK, since 2017. Before that, he was a Postdoctoral Researcher at the University of Passau, Germany, from 2015 to 2017. He has authored more than ninety publications in peer-reviewed books, journals, and conference proceedings to date. His research mainly focuses on deep learning technologies for speaker-centered state and health computing. He has organized special sessions, such as at the IEEE 7th Affective Computing and Intelligent Interaction (ACII) conference in 2017 and at the 43nd IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) in 2018. Moreover, he serves as a reviewer for numerous leading-in-their fields journals and conferences, and as a Program Committee Member and an Area Chair for many international conferences.

Dimitris N. Metaxas is a Distinguished Professor in the Computer Science Department at Rutgers University. He is director of the Center for Computational Biomedicine, Imaging and Modeling (CBIM) and the NSF I/UCR CARTA Center. He has also been a Tenured Faculty Member in the Computer and Information Science Department of the University of Pennsylvania. Prof. Metaxas received a Diploma with highest honors in Electrical Engineering and Computer Science from the National Technical University of Athens Greece, an M.Sc. in Computer Science from the University of Maryland, College Park, and a Ph.D. in Computer Science from the University of Toronto. His research emphasizes the development of novel Machine Learning methods, GANs, machine learning methods for representation and understanding of scenes, shape and motion, statistical object modeling and tracking, machine learning and sparse learning methods for segmentation and medical image reconstruction, deformable models and fluid simulations. Dr. Metaxas has published over 500 research articles in these areas and has graduated 57 PhD students. Dr. Metaxas has received several best paper awards, and holds 8 patents. He was awarded a Fulbright Fellowship in 1986, is a recipient of an NSF Research Initiation and Career awards, an ONR YIP, a Fellow of the MICCAI Society, a Fellow the American Institute of Medical and Biological Engineers and a Fellow of IEEE. He has been involved with the organization of several major conferences in vision and medical image analysis, including ICCV 2007, ICCV 2011, MICCAI 2008 and CVPR 2014.

Hung-yi Lee received the M.S. and Ph.D. degrees from National Taiwan University (NTU), Taipei, Taiwan, in 2010 and 2012, respectively. From September 2012 to August 2013, he was a Postdoctoral Fellow in the Research Center for Information Technology Innovation, Academia Sinica. From September 2013 to July 2014, he was a Visiting Scientist at the Spoken Language Systems Group of the MIT Computer Science and Artificial Intelligence Laboratory (CSAIL). He is currently an Associate Professor of the Department of Electrical Engineering of National Taiwan University, with a joint appointment at the Department of Computer Science & Information Engineering of the university. His research focuses on machine learning (especially deep learning), spoken language understanding and speech recognition. He owns a YouTube channel teaching deep learning (in Mandarin) with more than 4M views and 55k subscribers. He gave the tutorial on generative adversarial network (GAN) at ICASSP 2018, APSIPA 2018 and ISCSLP 2018.

Björn W. Schuller received his diploma in 1999, his doctoral degree in 2006, and his habilitation and Adjunct Teaching Professorship in the subject area of Signal Processing and Machine Intelligence in 2012, all in electrical engineering and information technology from TUM in Munich/Germany. He is Professor of Artificial Intelligence in the Department of Computing at Imperial College London/UK, where he heads GLAM, Full Professor Chair of Embedded Intelligence for Health Care and Wellbeing at the University of Augsburg/Germany, and CEO of audEERING. He was previously Full Professor Chair of Complex and Intelligent Systems at the University of Passau/Germany. Professor Schuller is President-emeritus of the Association for the Advancement of Affective Computing (AAAC), Fellow of the IEEE, and Senior Member of the ACM. He (co-)authored five books and more than 900 publications in peer-reviewed books, journals, and conference proceedings leading to more than overall 29 000 citations (h-index = 79). Schuller is former Editor in Chief of the IEEE TRANSACTIONS ON AFFECTIVE COMPUTING next to a multitude of further Associate and Guest Editor roles and functions as General, Program, and Area Chair.