Guest Editorial: Automated Machine Learning

Hugo Jair Escalante, Quanming Yao, Wei-Wei Tu, Nelishia Pillay,
Rong Qu, *Senior Member, IEEE*, Yang Yu, and Neil Houlsby

1 Introduction

Nowadays the success of decision support systems crucially relies on human experts involved in all systems design stages. Commonly, humans take critical decisions that include: converting real world problems into machine learning problems, data collection, feature engineering, selecting or designing model architectures, tuning model hyper-parameters, evaluating model performance, deploying on-line systems, and so on. The complexity of these tasks, which is often beyond non-experts, together with the rapid growth of applications, have motivated a demand for off-the-shelf AI and machine learning methods that can be used easily and without expert knowledge. In the context of machine learning, the research area that targets progressive automation of machine learning is called AutoML (Automated Machine Learning) [1]. This broad definition comprises automation at different levels: from the adjustment of parameters or hyperparameters in fixed models, to the automated design of learning pipelines and deep learning architectures (see Liu *et al.* for a taxonomy of AutoML [2]).

AutoML solutions have been a dream for machine learning researchers for a long time [3]. Although this will remain a dream at least in the most general case (e.g., because of the NFL theorem), relaxed realizations of it have been successfully explored, see e.g., the surveys in [1], [4], [5], [6], [7], [8], [9], [10]. In fact, AutoML has been a very active research field in the last couple of years, spurred on by the organization of academic challenges[1] and workshops[2] on the subject. The increasing number of AutoML solutions available out there and the active and growing community working on this field show the time is ripe for a compilation with the latest advances in AutoML. With that goal in mind we have edited a special section of the *IEEE Transactions on Pattern Analysis and Machine Intelligence* (*TPAMI*) on AutoML.

The scope of the special section comprised all aspects of AutoML, including model selection and hyperparameter optimization, neural architecture search, feature engineering, hyperheuristics, transfer and meta-learning and related topics. A total of 43 high quality submissions were received and every paper was subject to the standard TPAMI review process. In the remainder of this note we briefly summarize the contributions of the articles included in the issue.

2 The AutoML Special Section

This special section is formed by 15 articles of outstanding quality that together comprise a snapshot of cutting edge AutoML research. Table 1 summarizes the main characteristics of the accepted papers.

Among the accepted papers, those approaching Neural Architecture Search (NAS) prevailed. This is not surprising given the predominance of deep learning based methodologies across most fields associated to data analysis. In terms of NAS, X. Zhang *et al.* introduced a pruning based methodology, where complete blocks are associated with scaling factors among connections, these are subject to a sparse regularization to remove uninformative connections. Optimization is done directly on the target network and dataset [11]. In [12], authors propose an approximator of architectures that is differentiable for NAS, the method aims to approximante a supernet by sampling child architectures, optimizing the relative weights and connections [12]. M. Zhang *et al.* also adopt a supernet based approach to NAS, but they present a method based on constrained continual learning optimization that implement *ad-hoc* loss functions that alleviate the catastrophic forgetting problem [13]. Zheng *et al.* introduce MIGO-NAS: a NAS method where the search space is formulated as a multivariate probabilistic distribution and it is optimized by a multivariate information-geometric optimization (MIGO) method [14]. The method is highly competitive and very efficient. Xu *et al.* [15] introduce a partially connected version of Differentiable Architecture Search (DARTS [26]). Channel connections are randomly removed trying to accelerate the search process and make the solution more robust to overfitting. Also, input features are downsampled aiming to eliminate spatial redundancy. Two approaches for transferring architectures in NAS are presented in [16] and [17]. Lu *et al.* optimize a supernet for NAS based on an archive

1. https://automl.chalearn.org/
2. https://www.automl.org/events/workshops/

- *Hugo Jair Escalante is with the Instituto Nacional de Astrofísica, Óptica y Eectrónica, Puebla 72840, Mexico. E-mail: hugojair@inaoep.mx.*
- *Quanming Yao is with the 4Paradigm Inc., Beijing 100085, China. E-mail: yaoquanming@4paradigm.com.*
- *Wei-Wei Tu is with the Science and Technology Department, 4Paradigm Inc., Beijing 100085, China. E-mail: tuweiwei@4paradigm.com.*
- *Nelishia Pillay is with the University of Pretoria, Pretoria 0002, South Africa. E-mail: npillay@cs.up.ac.za.*
- *Rong Qu is with the School of Computer Science, University of Nottingham, NG7 2RD Nottingham, U.K. E-mail: Rong.Qu@nottingham.ac.uk.*
- *Yang Yu is with the School of Artificial Intelligence, Nanjing University, Nanjing 210023, China. E-mail: yuy@nju.edu.cn.*
- *Neil Houlsby is with the Google Research, Brain Team, 8002 Zurich, Switzerland. E-mail: neilhoulsby@google.com.*

Digital Object Identifier no. 10.1109/TPAMI.2021.3077106

TABLE 1
Overview of Articles Included in the AutoML Special Section

Ref.	Problem approached	Model, optimizer, main contribution	Considered datasets, performance
[11]	Single shot NAS	Direct sparse optimization, regularized scaling factors associated to connections	CIFAR- 2.74% — ImageNet 25.4% (< 600M FLOPS)
[12]	NAS	Differentiable architecture approximation with ensemble gumbel softmax	CIFAR-10: 2.53% — 24.9% (< 600M FLOPS)
[13]	Single shot NAS	Constrained continual learning	CIFAR-10: 2.59% — CIFAR-100: 16.69% — ImageNet: 25.5% — NAS201 — (< 600M FLOPS)
[14]	NAS	Multivariate information-geometric optimization	CIFAR-10: 2.39% (1.5 GPU hrs)— ImageNet: 21.7 (96 GPU hrs)
[15]	NAS	Optimization of partially connected channels and spatial reduction of features.	CIFAR-10: 2.55% (0.07 GPU days)— ImageNet: 24.1% (2.8 GPU days)— MS.COCO: mAP: 28.8
[16]	NAS	Optimization of a supernet throughout an archive of subnets optimized with many objective evolutionary algorithms	CIFAR-10: 1.6%* — ImageNet: 22.5%*
[17]	NAS	Parameter mapping for architecture transfer	MS-COCO mAP: 24 (4.4M parameters)
[18]	NAS for facial PAD	Tailored search space for the PAD problem	State of the art performance on PAD
[19]	HPO - Network pruning	New additive tree-structured covariance function for HPO.	Synthetic functions, model compression/pruning
[20]	AutoML for multi-label classification	Variety of AutoML extensions for multi-label classification	Multi-label classification datasets
[21]	Prediction of run time performance	Supervised learning pipelines	Benchmark regression and classification datasets
[22]	AutoML for life long learning	AutoML frameworks are extended with mechanisms for dealing with concept drift	Benchmark datasets for learning under concept drift
[23]	AutoML for tabular data and NAS	Auto-PyTorch / variety of AutoML methodologies	Benchmark tabular data and NAS201 benchmark
[24]	AutoML for life long learning	Life long anchors for full pipeline construction	Benchmark data and AutoDL challenge data
[25]	AutoML for life long learning	Analysis of solutions to the AutoDL challenge	Benchmark data and AutoDL challenge data

NAS: Neural Architecture Search; HPO: Hyperparameter Optimization; PAD: Presentation Attack Detection. Performance is not comparable to that reported in the other references.

of subnets that are optimized with a many objective evolutionary algorithm [16]. Fang *et al.* introduce a parameter remapping technique for transferring (expanding) architectures pretrained on ImageNet to other tasks like object detection and pose estimation [17]. Last but not least, Yu *et al.* introduce a NAS methodology tailored to face spoofing detection that considers a search space ad-hoc for the task [18].

Articles that are part of the special section and do not deal with NAS approached quite diverse problems. Ma *et al.* propose an additive tree-structured covariance function in the context of Bayesian optimization for hyperparameter optimization (HPO). The proposed method improves the sample-efficiency of the associated Gaussian process when optimizing a black box function [19]. Wever *et al.* provide an overview of AutoML in the context of multi label classification. The authors reviewed relevant literature and conducted an experimental evaluation of AutoML methods adapted to the multi label classification setting [20]. Mohr *et al.* present a comprehensive study on predicting run time performance under different AutoML settings focusing on supervised learning pipelines [21]. Celik *et al.* present an experimental evaluation of different strategies to deal with evolving data using a variety of established AutoML frameworks. The study shows the relevance of incorporating this type of mechanisms when dealing with datasets affected by concept drift [22]. Zimmer *et al.* describe Auto-PyTorch, a framework for AutoML that allows the optimization of both architectures and hyperparameters of networks [23]. Auto-PyTorch implements cutting edge AutoML mechanisms and introduces a benchmark for studying multi-fidelity optimization. Experiments are reported mostly on tabular data, although authors also report results on the NAS-Bench 201. Two accepted articles were associated to the AutoDL challenge. On the one hand, Tang *et al.* proposed an AutoML method for full pipeline optimization that relies on life long anchors (prototypical solutions that have been explored in the past). Optimization is performed using an evolutionary algorithm. This solution achieved the best results in the AutoDL challenge [24]. On the other hand, Liu *et al.* describe the design of the AutoDL challenge and provide an extensive analysis of solutions [25].

3 Discussion

The special section is a compilation of outstanding contributions on AutoML. The compilation is representative of the trends and recent progress in the field. Based on the accepted papers and on the received submissions we have identified the following topics that will be decisive in the next few years:

- Architecture transfer in NAS: Meta-learning and transfer of architectures with NAS is a trending topic that will have a significant impact in the next few years.
- Benchmarking and reproducibility in NAS and AutoML in general: Although there are several benchmarks and platforms for the comparison of NAS methodologies (e.g., NAS201, NAS301), NAS contributions are still difficult to compare because the evaluation using such resources is still scarce.
- Task-specialized AutoML: While the main goal of NAS is to generalize across domains, NAS

methodologies tailored to specific domains (see e.g., [18]) could achieve better performance by incorporating domain knowledge.

- Explainability in AutoML: HPO, NAS and AutoML in general can be seen as black-box methods mostly, explainability and interpretability can make these methods more reliable for the final user, widespreading their usage.
- Interactive AutoML: Related with the preference of white/gray box models, interactive AutoML could lead to better solutions and the widespread usage of these methods in practice.
- AutoML in other ML processes: AutoML has been studied in the context of a wide variety of problems and domains. However, the focus so far has been on supervised learning and based on performance optimization, expanding AutoML to other tasks, and processes like feature engineering, clustering, semi-supervised learning etc. is a promising venue for research.

ACKNOWLEDGMENTS

The work of Hugo Jair Escalante was supported in part by INAOE and in part by CONACyT-Mexico under Grant CB-A1-S-26314.

REFERENCES

[1] F. Hutter, L. Kotthoff, and J. Vanschoren, Eds., *Automated Machine Learning - Methods, Systems, Challenges, (The Springer Series on Challenges in Machine Learning)*. Basel, Switzerland: Springer, 2019. [Online]. Available: https://doi.org/10.1007/978–3-030-05318-5

[2] Z. Liu, Z. Xu, J. J. Junior, M. Madadi, S. Escalera, and S. Rajaa, "Overview and unifying conceptualization of automated machine learning," in *Proc. Automat. Data Sci. Workshop*, 2019.

[3] J. R. Rice, "The algorithm selection problem," in "Computer Science Technical Reports. Paper 99," Dept. Comput. Sci., Purdue Univ., West Lafayette, IN, USA, Tech. Rep. 75–152, 1975

[4] Q. Yao, *et al.*, "Taking human out of learning applications: A survey on automated machine learning," 2018, *arXiv:1810.13306*.

[5] M.-A. Zöller and M. F. Huber, "Survey on automated machine learning," 2019, *arXiv:1904.12054*.

[6] X. He, K. Zhao, and X. Chu, "AutoML: A survey of the state-of-the-art," 2019, *arXiv:1908.00709*.

[7] L. Tuggener *et al.*, "Automated machine learning in practice: State of the art and recent results," *Swiss Conf. Data Sci.*, 2019, pp. 31–36. [Online]. Available: http://arxiv.org/abs/1907.08392

[8] T. Elsken, J. H. Metzen, and F. Hutter, "Neural architecture search: A survey," 2018, *arXiv:1808.05377*.

[9] N. Pillay, R. Qu, D. Srinivasan, B. Hammer, and K. Sorensen, "Automated design of machine learning and search algorithms [guest editorial]," *IEEE Comp. Intell. Mag.*, vol. 13, no. 2, pp. 16–17, May 2018. [Online]. Available: https://doi.org/10.1109/MCI.2018.2806988

[10] H. J. Escalante, . "Automated Machine Learning – A Brief Review at the End of the Early Years," 2021, *arXiv:2008.08516*.

[11] X. Zhang, Z. Huang, N. Wang, S. Xiang, and C. Pan, "You only search once: Single shot neural architecture search via direct sparse optimization," *IEEE Trans. Pattern Anal. Mach.Intell.*, vol. 43, no. 9, pp. 2891–2904, Sep. 2021.

[12] X. Zhang *et al.*, "DATA: Differentiable architecture approximation with distribution guided sampling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 9, pp. 2905–2920, Sep. 2021.

[13] M. Zhang *et al.*, "One-shot neural architecture search: Maximising diversity to overcome catastrophic forgetting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 9, pp. 2921–2935, Sep. 2021.

[14] X. Zheng *et al.*, "MIGO-NAS: Towards fast and generalizable neural architecture search," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 9, pp. 2936–2952, Sep. 2021.

[15] Y. Xu *et al.*, "Partially-connected neural architecture search for reduced computational redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 9, pp. 2953–2970, Sep. 2021.

[16] Z. Lu, G. Sreekumar, E. Goodman, W. Banzhaf, K. Deb, and V. N. Boddeti, "Neural architecture transfer," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 9, pp. 2971–2989, Sep. 2021.

[17] J. Fang *et al.*, "FNA++: Fast network adaptation via parameter remapping and architecture search," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 9, pp. 2990–3004, Sep. 2021.

[18] Z. Yu, J. Wan, Y. Qin, X. Li, S. Z. Li, and G. Zhao, "NAS-FAS: Static-dynamic central difference network search for face anti-spoofing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 9, pp. 3005–3023, Sep. 2021.

[19] X. Ma and M. B. Blaschko, "Additive tree-structured conditional parameter spaces in Bayesian optimization: A novel covariance function and a fast implementation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 9, pp. 3024–3036, Sep. 2021.

[20] M. Wever, A. Tornede, F. Mohr, and E. Hüllermeier, "AutoML for multi-label classification: Overview and empirical evaluation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 9, pp. 3037–3054, Sep. 2021.

[21] F. Mohr, M. Wever, A. Tornede, and E. Hüllermeier, "Predicting machine learning pipeline runtimes in the context of automated machine learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 9, pp. 3055–3066, Sep. 2021.

[22] B. Celik and J. Vanschoren, "Adaptation strategies for automated machine learning on evolving data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 9, pp. 3067–3078, Sep. 2021.

[23] L. Zimmer, M. Lindauer, and F. Hutter, "Auto-Pytorch: multi-fidelity metalearning for efficient and robust autoDL," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 9, pp. 3079–3090, Sep. 2021.

[24] L. Tangy *et al.*, "Evolving fully automated machine learning via life-long knowledge anchors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 9, pp. 3091–3107, Sep. 2021.

[25] Z. Liu *et al.*, "Winning solutions and post-challenge analyses of the chalearn autoDL challenge 2019, " *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 9, pp. 3108–3125, Sep. 2021.

[26] H. Liu, K. Simonyan, and Y. Yang, "Darts: Differentiable architecture search," 2019, *arXiv:1806.09055*

.

Hugo Jair Escalante has been a researcher scientist with Instituto Nacional de Astrofisica, Optica y Electronica, INAOE, Mexico, and the director of ChaLearn, a nonprofit dedicated to organizing challenges, since 2011. He has been involved in the organization of several challenges in computer vision and automatic machine learning. He was a coeditor of special issues in the *International Journal of Computer Vision*, the *IEEE Transactions on Pattern Analysis and Machine Intelligenc*, and the *Transactions on Automatic Control*. He was a competition chair and area chair of venues like NeurIPS, ICML, PAKDD, IJCNN, among others.

Quanming Yao received the PhD degree with the Department of Computer Science and Engineering, Hong Kong University of Science and Technology. He is currently a senior scientist at 4Paradigm. His research interests include machine learning, optimization, and automated machine learning. He has authored or coauthored more than 40 top-tier journals and conference papers, including International Conference on Machine Learning, Conference on Neural Information Processing Systems, the *Journal of Machine Learning Research*, and the *IEEE Transactions on Pattern Analysis and Machine Intelligence*. He was the recipient of the Forbes 30 Under 30, China, Young Scientist Awards, the Hong Kong Institution of Science, Wuwen Jun Prize for Excellence Youth of Artificial Intelligenc (issued by CAAI), and a winner of Google Fellowship (in machine learning).

Wei-Wei Tu received the master's degree from the Department of Computer Science, Nanjing University, China. He is currently the vice president, principal scientist, and head of the Science and Technology Department with 4Paradigm Inc., Beijing, China. He has authored or coauthored several papers with KDD, IJCAI, AAAI, ICLR, ICML, and NeurIPS. His research focuses on machine learning. He was a senior engineer with China's biggest search engine, Baidu. In Baidu, he built Baidu's first distributed GBDT training system running on hundreds of machines in 2012, deployed Baidu's first large scale deep learning based click-through rate prediction system in 2013, and co-designed Baidu's first distributed machine learning computation framework in 2014. In 4Paradigm Inc., he designed and developed the distributed machine learning computation framework, and led his team developed many large scale distributed machine learning algorithms supporting thousands of billions parameters and hundreds of billions instances. He is currently leading his team to build business AutoML systems at 4Paradigm Inc., one of the board members of ChaLearn, USA, data competition chair of PAKDD 2018 and PAKDD 2019, chair of AutoML workshop at KDD 2020 and PRICAI 2018, one of main organizers of NeurIPS2018 AutoML Challenge, main organizer of KDD Cup 2019 / 2020 AutoML Challenge, one of the main organizers of NeurIPS 2019 AutoDL competition, and advisor of NewInML session with NeurIPS 2019. He was the recipient of "Baidu Million Dollar Highest Prize" in 2015.

Nelishia Pillay is currently a professor at the University of Pretoria, South Africa. She is currently the multichoice joint-chair in machine learning and SARChI chair in artificial intelligence, chair of the IEEE Technical Committee on Intelligent Systems Applications, the IEEE Task Force on Hyper-Heuristics, and the IEEE Task Force on Automated Algorithm Design, Configuration and Selection. She has authored or coauthored in research areas in journals, national and international conference proceedings. Her research interests include hyper-heuristics, automated design of machine learning and search techniques, combinatorial optimization, genetic programming, genetic algorithms and deep learning. These are the focus areas of the NICOG (Nature-Inspired Computing Optimization) research group which she has established. She was with program committees for numerous national and international conferences and is a reviewer for various international journals.

Rong Qu (Senior Member, IEEE) is currently an associate professor at the School of Computer Science, University of Nottingham. Her current research interests include automated algorithm design for real-world optimization based on hybrid computational optimisation algorithms in operational research and artificial intelligence. She is currently an associated editor for the IEEE Computational Intelligence Magazine, the *IEEE Transactions on Evolutionary Computation*, the *Journal of Operational Research Society,* and the *PeerJ Computer Science*. She is a Senior IEEE Member since 2012 and the guest edited special issues on automated algorithm design and AI sheduling. She is currently the vice-chair of Evolutionary Computation Task Committee with the IEEE Computational Intelligence Society.

Yang Yu is currently a professor at the School of Artificial Intelligence, Nanjing University, China, and the founder and CEO of Polixir Technologies. His research interest includes machine learning, mainly on reinforcement learning and derivative-free optimization for learning. He was recognized as one of the AI's 10 to Watch by IEEE Intelligent Systems in 2018. He was invited to give an Early Career Spotlight Talk in IJCAI'18. He was the recipient of the CCF-IEEE CS Young Scientist Award in 2020 and the PAKDD Early Career Award in 2018.

Neil Houlsby received the PhD degree in Bayesian machine learning methods for active learning and matrix modeling from the University of Cambridge. He is currently a staff research scientist at Google Research, Brain Team, Zürich. His current research interests include AutoML, natural-language processing, and computer vision. He is currently a co-leader of a Computer Vision Team, Google Brain EMEA region, focussing on representation learning, transfer learning, and scalability.