On Practical Implementation of Machine Learning for Microcontrollers

Mikhail Zolotukhin[1], Pyry Kotilainen[2], Riku Immonen[3] and Timo Hämäläinen[4]

[1]Faculty of Information Technology, University of Jyväskylä, Finland, e-mail: mizolotu@jyu.fi
[2]Faculty of Information Technology, University of Jyväskylä, Finland, e-mail: pyjopeko@jyu.fi
[3]Faculty of Information Technology, University of Jyväskylä, Finland, e-mail: rijuimmo@jyu.fi
[4]Faculty of Information Technology, University of Jyväskylä, Finland, e-mail: timoh@jyu.fi

Abstract

The recent progress in development of low-budget sensors and single-board computers has enabled deploying artificial intelligence (AI) on the edge which in turn allows service providers to deliver reduced latency, efficient bandwidth consumption, improved data security, increased privacy, and lower costs. Tiny machine learning (ML) models implemented on such smart devices can be used for object recognition [1] and time-series classification [2]. Furthermore, increasing computing and connectivity capabilities of modern microprocessors has made it possible to deploy tiny AI/ML models not only for inference but also for training [3]. In this study, we focus on implementation and evaluation of supervised and unsupervised machine learning models which can be deployed and trained on tiny smart devices for classification, regression and anomaly detection. We conduct our experiments in the context of three following use cases: wind speed estimation and anomalous vibration detection and classification. First, we study various traditional and neural network based ML methods which can be employed to solve the problems formulated under low computing and memory resource constraints. These include but are not limited to various deep learning architectures [4], decision tree ensembles [5] and stream clustering algorithms [6]. Next, we discuss techniques and tools which can be used to implement these algorithms on single-board devices. These for example involve automatic machine learning optimization to produce maximum performance from learning tools without human assistance [7]. Finally, several of the most efficient machine learning algorithms found are implemented on multiple modern microcontrollers and evaluated in terms of prediction error, classification accuracy, inference time and other metrics.

References

[1] Estrebou CA, Fleming M, Saavedra MD, Adra F, De Giusti AE. Lightweight Convolutional Neural Networks Framework for Really Small TinyML Devices. In: *SmartTech-IC*. Springer. 2022; pp. 3–16.

[2] Gupta S, Jain S, Roy B, Deb A. A TinyML Approach to Human Activity Recognition. In: *Journal of Physics: Conference Series*, vol. 2273. IOP Publishing. 2022; p. 012025.

[3] Ren H, Anicic D, Runkler TA. Tinyol: Tinyml with online-learning on microcontrollers. In: *IJCNN*. IEEE. 2021; pp. 1–8.

[4] Ruff L, Vandermeulen R, Goernitz N, Deecke L, Siddiqui SA, Binder A, Müller E, Kloft M. Deep one-class classification. In: *ICML*. PMLR. 2018; pp. 4393–4402.

[5] Friedman JH. Stochastic gradient boosting. *Computational statistics & data analysis*. 2002; 38(4):367–378.

[6] Silva JA, Faria ER, Barros RC, Hruschka ER, Carvalho ACd, Gama J. Data stream clustering: A survey. *ACM Computing Surveys (CSUR)*. 2013;46(1):1–31.

[7] Jin H, Song Q, Hu X. Auto-keras: An efficient neural architecture search system. In: *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 2019; pp. 1946–1956.