Embedded ML in the Context of Human Perception and Emotion

Morris Blaustein
Purdue University
USA

1 INTRODUCTION

Machine Learning (ML) has enabled engineers to extract meaningful information from data in ways that were previously impossible. It has improved many aspects of society such as healthcare, finance, manufacturing, and transportation. Examples include: healthcare, where it can detect the early onset of cancer, finance, where it can be used for algorithmic trading and risk assessment, manufacturing, where it can improve supply chain efficiency, and transportation, where it is a pivotal aspect of autonomous vehicles. The papers summarized in Section 2 focus on embedded machine learning in the context of human perception and emotion. Wondering about the parallels between human intelligence, neuroscience, and artificial intelligence, I am able to glimpse into the future of innovation. Studying machine learning in the context of human perception and emotion helps ground my creative insights to something factual and researched.

ML algorithms are compute intensive, and chips optimized for parallel computations provide the best performance. They are often implemented with powerful desktop GPUs to achieve the best accuracy regardless of execution time [3]. However, advances in affordable microcontrollers have enabled engineers to implement ML algorithms closer to the source of data, which improves latency. Due to the tight resource constraints of embedded systems, engineers must evaluate the trade-off between accuracy and latency. For example, quantizing neural network weights can improve memory efficiency and execution time, but also decrease accuracy.

There are different types of ML algorithms that prove themselves optimal for certain applications. However, the success of an ML algorithm is ultimately less related to what particular algorithms are used and more about the domain understanding and problem statement [4]. Understanding the role of ML within a bigger ecosystem is vital for its progress. The research summarized in Section 2 each solved unique challenges that demonstrated powerful usage of embedded ML. The systems that were investigated were not only novel systems, but also contributed to the understanding of ML's role in society.

2 REVIEW

2.1 Evaluation of Classical ML Techniques towards Urban Sound Recognition on Embedded Systems

The purpose of this study was to determine whether ARM-based microprocessors can achieve similar performance to general-purpose CPUs when processing classical ML algorithms for Enviromental Sound Recognition (ESR) in Wireless Acoustic Sensor Networks (WASN). ESR can determine the source of sounds in an urban environment, which has many use cases such as surveillance and criminal investigation. In this study, audio data packets are statistically analyzed into a feature vector with 90 elements per one second frame, including mean, variance, and spectral rolloff. [2] The study used supervised ML algorithms that were expected to meet latency requirements of a WASN on an embedded system. Several audio data sets were combined into categories such as airplanes, animal sounds and sirens. Eighty percent of the data were used for training, and 20% were used for validation.

Scikit-learn was used for classifying and libROSA Python package was used for feature extraction. A cascaded classifier enabled the classifier chosen in the second stage to be dependent on the output of the first classifier stage. Thus, an optimal classifier could be selected and the execution could be pipelined. For example, if the first stage detected a construction sound, the second stage could use an classifier optimized to detect drilling and jackhammering.

The system was first evaluated on a Apple Macbook Pro (2.3GHz Intel Core i5) to optimize the classification stage and determine the fastest ML algorithm. An inverse relationship between number of sample categories and performance was discovered, as also concluded by Bountourakis et al. [1]. The results showed a trade-off between accuracy and execution time. The most time was consumed in the feature extraction phase. Computing the ML algorithms on a Raspberry Pi caused slightly lower accuracy and a 10-100 times slowdown compared to the Macbook. These results conclude that it is viable to perform ML for ESR with a Raspberry Pi.

2.2 On the Challenges of Embedded Real-Time Music Information Retrieval

This study examined the feasibility of using real time music information retrieval (rt-MIR) on a Raspberry Pi 4 to detect different types of guitar playing. Examples of Music Information Retrieval include beat-detection, onset detection, music transcription, and genre classification. Research on rt-MIR shows the potential for smart instruments that can detect music patterns and play along with musicians. The main challenges identified for rt-MIR are limitations of causal information, tradeoff between accuracy and latency, audio processing deadlines, real-time-safe programming rules, and embedded software and hardware limitations [5]. Offline MIR may process an entire signal at any given time, whereas rt-MIR is limited to causal information and can only process the current and past signal. To compromise with latency requirements, classifiers can be made shallower, narrower, and with reduced neuron weight precision. However, weight quantization reduced the accuracy of the study's results by 10-20% [5]. The latency requirement of this system was 30ms, because any longer would result in perceptible audio delays.

The study used Elk OS, a variant of Linux designed for low latency audio applications. According to Turchet and Fischione in [7], Elk OS is the only OS capable of meeting requirements of Internet of Music Things (IoMusT): hard real-time audio, high quality audio, and network functionality. A custom dataset was created using five guitarists, five guitars, and three intensities. The system consisted of onset detection, feature extraction, and classification. The Aubio library was used for the real-time onset detector and a C++ port of TimbreID library was used for the feature extraction. A synthetic tune was played along with the guitarist based on the rt-MIR analysis. The study tested the system with three tasks. Task A and Task B focused on pitch and percussion detection, which had high accuracy rates and low latency. However, Task C introduced more output possibilities with subtle differences which made the classification results inaccurate. This MIR system implemented on a Raspberry Pi showed important insight to understanding the capabilities of embedded ML for smart musical instruments.

2.3 Real-Time On-Chip Machine-Learning-Based Wearable Behind-The-Ear Electroencephalogram Device for Emotion Recognition

This was the first study to investigate a ML enabled wearable EEG device that is comfortable and portable [3]. It examined the viability of measuring brain waves with a behind-the-ear EEG sensor on a dedicated embedded system with a companion phone application. An embedded system was developed that performed data collection, preprocessing, and modeling which was connected to a smart phone app to display user's emotional state. The purpose of the system was to display the user's happy or sad state on the phone application. The majority of emotion detection systems currently use external devices to send data to dedicated powerful computers which perform the ML. However, this technique introduces latency which may be intolerable [6]. The embedded system was composed of a dual-core 32 bit STM32H747XI, Portenta H7 microcontroller, Murata H7 radio module, and a power management unit. This system allowed the EEG signal analysis to occur in real-time wherever the user was located, and not be reliant on an additional networked computer.

Three ML models were tested: one dimensional convolutional neural network (1D CNN), support vecctor machine (MLP), and multilayer perceptron (MLP). Results from the preliminary study they conducted with 7 females and 7 males indicated: 1D CNN mean accuracy was 94.87%, MLP mean accuracy was 91.13% and SVM mean accuracy was 85.19% [3]. Since 1D CNN performed best, it was chosen to be used in the real-time system. Tensorflow was used to build and train the models they used, and Tensorflow Lite was used to realize the models on the embedded real time device. A bandpass filter which passed frequencies between 1Hz-50Hz was applied to the EEG data before it was split into 5 bands: delta (1-4Hz), theta (4-8Hz), alpha (8-13Hz), beta (13-30Hz), and gamma (30-50Hz). The bandpass filter prevented unwanted noise from being significant in the spectral analysis. This allowed the power spectral density to be analyzed across the five frequency bands. The delta band showed highest power density and was surmised to contain the most relevant content for EEG emotion analysis [3]. The study mentioned that additional testing would help solidify the findings. It also noted that a more advanced band pass filter, such as blind source seperation (BSS), would help reduce noise from muscle movements.

3 DISCUSSION AND CONCLUSION

These research papers helped me understand the differences between ML and embedded ML. The trade-off between accuracy and latency is one of the main considerations for embedded ML. Understanding the domain and problem statement are also a crucial part of the ML design process. It is important to test ML models on high performance systems before placing them on resource constrained systems, because this will help discover which algorithm performs best per application.

Maximizing the potential of embedded ML requires the use of optimized hardware accelerators which were not used in these studies. Custom hardware such as FPGA has the capability of significantly improving latency due to parallel concurrency and custom data and control paths. Next generation phones and portable computers are including ML accelerators in their SoC's. ML algorithms on consumer CPU's have relatively longer latency and lower accuracy. However, an important distinction must be made in ML design that performance is only a portion of the puzzle. Understanding the problems embedded ML can solve is just as important. As an example, applying machine learning to physics requires not only a high performance computer and a machine learning engineer, but also a physicist who understands the domain.

Research studies such as these are important. For example, learning that analyzing EEG data can distinguish emotional states, we can infer that additional study in this field will uncover a deeper knowledge of the patterns between brain waves and emotions. The EEG study only analyzed the power spectral density of the five brain wave frequency bands, and not the information within the signal. Future research with embedded EEG systems will help provide more valuable real time emotion information to users.

Embedded ML is a powerful use of artificial intelligence that opens the door for real time data analysis at the data source. The future of research in this field will involve designing custom ML accelerators specifically for an application domain. Within the context of human perception and emotion, machine learning is able to analyze data that provides insight toward understanding our evolution as a species.

REFERENCES

[1] Vasileios Bountourakis, Lazaros Vrysis, and George Papanikolaou. Machine learning algorithms for environmental sound recognition: Towards soundscape semantics. In *Proceedings of the audio mostly 2015 on interaction with sound*, pages 1–7. 2015.

[2] Bruno da Silva, Axel W. Happi, An Braeken, and Abdellah Touhafi. Evaluation of classical machine learning techniques towards urban sound recognition on embedded systems. *Applied Sciences*, 9(18):3885, 2019.

[3] Ngoc-Dau Mai, Ha-Trung Nguyen, and Wan-Young Chung. Real-time on-chip machine-learning-based wearable behind-the-ear electroencephalogram device for emotion recognition. *IEEE Access*, 2023.

[4] Cynthia Rudin and Kiri L Wagstaff. Machine learning for science and society, 2014.

[5] Domenico Stefani, Luca Turchet, et al. On the challenges of embedded real-time music information retrieval. In *Proceedings of the International Conference on Digital Audio Effects (DAFx)*, volume 3, pages 177–184. MDPI (Multidisciplinary Digital Publishing Institute), 2022.

[6] Kirill Stytsenko, Evaldas Jablonskis, and Cosima Prahm. Evaluation of consumer eeg device emotiv epoc. In *MEi: CogSci Conference*, page 99, 2011.

[7] Luca Turchet and Carlo Fischione. Elk audio os: an open source operating system for the internet of musical things. *ACM Transactions on Internet of Things*, 2(2):1–18, 2021.