（様式 5）

2021 年 12 月 16 日
Year Month Day

学位（博士）論文要旨

(Doctoral thesis abstract)

論文提出者 (Ph.D. candidate)	工学府博士後期課程 電子情報工学 専攻 (major) 2019 年度入学(Admission year) 学籍番号 19834311 氏名 蔡敬勇 (student ID No.) (Name)
主指導教員氏名 (Name of supervisor)	中條拓伯
論 文 題 目 (Title)	Log-or-Trig: Towards Efficient Learning in Deep Neural Networks.

Abstract

In this research, we analyze in depth the attributes of logarithmic quantization. Existing compression algorithms highly rely on retraining which requires heavy computational power. In such a situation, we propose a new logarithmic quantization algorithm to mitigate the deterioration on neural networks which contain layers of small size. Moreover, we present a weight-aware optimization objective for deep neural network quantization. Our method quantifies the importance of each weight by its magnitude. This method requires no sorting during quantization; therefore, the computation cost is very low and can be easily combined with existing quantization algorithms. We tested our method using existing state-of-the-art quantization methods. On Additive Powers-of-Two (APoT) quantization, our method can largely reduce the retraining time and on Piecewise Linear Quantization (PWLQ), our method improves the image classification accuracy.

In addition to the quantization algorithms, we also propose a new learning algorithm. Despite being heavily used in the training of deep neural networks, multipliers are resource-intensive and insufficient in many different scenarios. Previous discoveries have revealed the superiority when activation functions, such as the sigmoid, are calculated by shift-and-add operations, although they fail to remove multiplications in training altogether. Based on the aforementioned facts, we propose an innovative approach that can convert all multiplications in the forward and backward inferences of deep neural networks into shift-and-add operations. Because the model parameters and backpropagated errors of a large deep neural network model are typically clustered around zero, these values can be approximated by their sine values. Multiplications between the weights and error signals are transferred to multiplications of their sine values, which are replaceable with simpler operations with the help of the product to sum formula.

A rectified sine activation function is also proposed for further converting layer inputs into sine values. In this way, the original multiplication-intensive operations can be computed through simple add-and-shift operations. This trigonometric approximation method provides an efficient training and inference alternative for de- vices with insufficient hardware multipliers. Experimental results demonstrate that this method can obtain a performance close to that of classical training algorithms. The approach we propose sheds new light on future hardware customization research for machine learning.