

# The Model of Energy Optimization

## 1 TRAINING COMPLETION TIME MODEL

For a specific training task (e.g., a particular DNN model), the time required to finish an training epoch on different devices is totally different. The training completion time is highly correlated with the following factors. First, since different devices can have different number of data objects to be trained, the training completion time can be determined by the amount of data objects  $d_i$ . Second, different clients can have totally different training platforms (e.g., CPU, GPU), thus the training completion time can also depend on different training platforms. Third, for a certain training platform, the hardware configuration can be different (e.g., number of cores, core frequency). Thus, we model the training completion time for one epoch as follows:

$$t_i = \frac{\theta * d_i}{c^\alpha f^\beta} \quad (1)$$

$\theta$  represents the coefficient which is different for different training platforms and devices.  $d_i$  represents the number of data objects required to be trained in this training epoch.  $c$  represents the number of cpu cores adopted during the training process.  $f$  represents the core frequency.  $\alpha$  and  $\beta$  are the corresponding coefficients that can be obtained through the system identification process. We can find that the more

data objects need to be trained, the longer time it requires to complete the training. The more number of cpu cores and high core frequency can help accelerate the training process.

## 2 DVFS POWER MODEL

During the training process, the power consumption (e.g., different devices, different core frequency) can be different. Thus, we model the training power from CPU as follows:

$$P_i = P_{BL, N_c} + \sum_i^{N_c} P_{core, U_i, f_i} \quad (2)$$

where  $N_c$  is the number of cores enabled,  $P_{BL, N_c}$  is the baseline CPU power with  $N_c$  enabled cores, and  $P_{core, U_i, f_i}$  is power increment of core  $i$  when it is working at frequency  $f_i$  with utilization  $U_i$ . Moreover, the single core power  $P_{core, U_i, f_i}$  can be modeled as follows:

$$P_{core, U_i, f_i} = \gamma * f_i * U_i + \gamma_{base} \quad (3)$$

in which,  $\gamma$  is the corresponding power coefficient. At last, the energy required for a specific training epoch on client  $i$  can be modeled as follows:

$$E_i = P_i * t_i \quad (4)$$