Design and Optimization of Energy-Efficient Cascaded Classifiers on Silicon

Aosen Wang and Wenyao Xu

Cascaded classifier is a prevalence approach to deploy binary primitives in multi-classification problems in machine leaning. Many work have achieved excellent performance in accuracy and time-efficiency. However, there are few attempts exploring the optimization strategy of the cascaded classifiers towards energy-efficiency. In this paper, we design and optimize the energy-efficient cascaded classifier on hardware. By training the binary classifier following conventional method, we model the energy consumption of the testing phase. Moreover, we formulate a closed form to jointly optimize the energy and accuracy of cascading scheme for the multi-classification problem. Experimental results demonstrate our framework can build energy-efficient cascaded classifier and maintain the accuracy.

Introduction: Cascaded classifier [1] is widely applied for multi-classification problem in machine learning and data mining. The state-of-the-art training techniques guarantee the high accuracy of binary classification. Facing with the multi-class classification task, the cascading strategy is employed to organize the binary primitives. Due to the large size of basic binary classifiers, the cascaded classifier has a huge design space. However, most endeavours are paid on algorithmic optimization to improve the accuracy of cascaded classifier. There is no sophisticated technique to explore the cascaded classifier to minimize the energy consumption on silicon. In this paper, we design and optimize the energy-efficient cascaded classifier. We model the RTL-level energy consumption of the binary classifiers. We integrate this energy model into the framework of the cascading, jointly optimized with accuracy.

Motivational Example: We illustrate an example to compare the accuracy and energy consumption for different cascading methods in Fig. 1. There are two cascaded classifiers for 4-class classification problem of Electroencephalography (EEG) signals using supporting vector machine (SVM) algorithm. The left method is using one-to-one strategy, separating one category for each binary classifier. The right case first divides all the categories into two equals, and then iterate this procedure until indivisibility is satisfied. Although the two classifiers hold the same accuracy, 83.3%, the left cascading has about 20% energy savings than the right case. Therefore, we can observe that the energy consumption of the cascaded classifier on silicon may have drastically reduction with the help of cascading strategy selection [2], while keeping the accuracy.

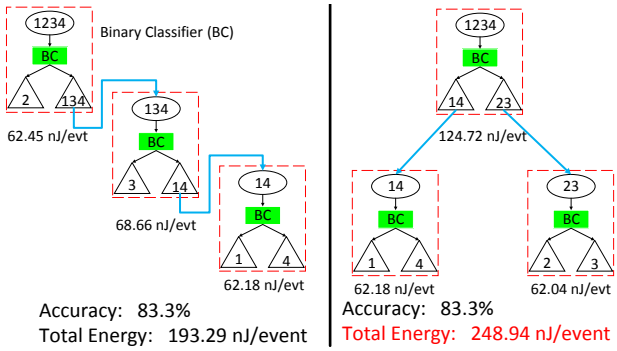

Fig. 1 *An example of different cascaded classifiers with the sample accuracy but different energy consumptions.*

Energy-Efficient Cascaded Classifier: We compare the conventional design flow and our energy-ware approach in Fig. 2. The traditional design flow follows the solid black arrow, covering data preparation, binary classifier training, cascading, on-silicon implementation and eventually obtain the classifier on silicon. Our method is indicated by the dashed red arrow. Different from the traditional solution, we add an on-silicon simulation after the binary classifier training. Based on the simulation, we build a power library to model the energy consumption in the testing phase of the binary classifiers. This RTL-level model also influences the cascading strategy. We formulate a closed form to find the optimal case, which will be discussed in details later. So the new cascading strategy and power library form the energy-aware cascading. Eventually, the classifier on silicon is also energy-efficient yet accurate.

Our energy model is based on the on-silicon simulation. After binary classifier training, we obtain parameters and calculation procedures of all the primitives in testing phase. We use Synopsys Design Suit to accomplish

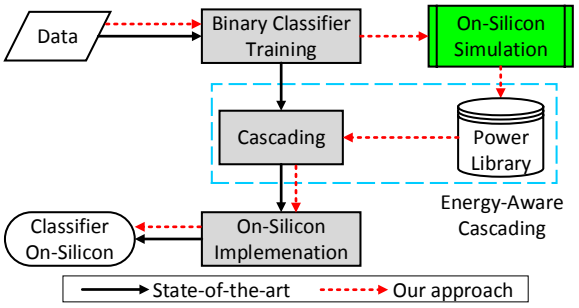

Fig. 2 *The block diagrams of the conventional flow and our energy-aware procedure.*

all the binary primitive simulation. We take TSMC 130nm standard cell library for Synopsys and implement testing phase by verilog in Verilog Compile Simulator. We carry out logic simulation and record the logic toggles, the SAIF file. Then we adopt Design Compiler to synthesize the simulation and report the energy consumption by the Power Compiler.

For each binary classifier, we apply proper binary classifier to divide the category set with divisibility. If we use a super dummy node to connect all the classifier whose categories before classification are most, this classifier tree will contain all the possible cascaded classifier schemes for the multi-classification problem. Our purpose is to find the optimal cascading to trade off the accuracy and energy consumption. We define Ad, the accuracy density, as the criterion to evaluate the trade-off between the accuracy and energy consumption: $Ad = \frac{accuracy}{energy}$. It is an indicator of the accuracy gain by consuming one unit energy for an event. Therefore, our objective is to maximize the total Ad for the cascaded classifier. We divide the binary classifier tree into several levels according to the category number dealt with by each binary primitive. Then, the problem to maximize the total Ad value can be reformulated as the multi-stage decision problem. We define $dp[i][j]$ as the max sum of Ad at the j-th node on the i-th level. We apply a dynamic programming algorithm on this binary classifier tree to obtain the optimal cascading layout, with its recursive formula as follows:

$$dp[i][j] = Ad(i,j) + \underbrace{\max dp[l_1][k_1]}_{Left\ Subset} + \underbrace{\max dp[l_2][k_2]}_{Right\ Subset}, \quad (1)$$

Experiments: In the experiment, we choose 120 continuous 128-sample EEG segments from Physionet [3], 60 for training and 60 for testing. To achieve better energy-efficiency, we label these EEG segments by their Pareto's curve. We take SVM algorithm to train binary classifier. We compare our energy-aware cascading (EAC) with accuracy-aware cascading (AAC) and one-to-rest strategy (OTR) [2] in Table 1.

Table 1: The comparison of accuracy and energy consumption

Method	Accuracy (%)	Energy (nJ/evt)	Energy Saving (%)
EAC	**85.00**	**187.88**	–
AAC	86.67	224.73	**16.39**
OTR	83.33	341.68	**45.02**

We can observe that compared to accuracy-aware cascading, our result saves energy by 16.39%, while only compromising 1.6% accuracy. For the one-to-rest strategy, it needs six binary classifiers, so that the total energy consumption is larger than the one-to-one strategy. Our result holds 45% energy savings in this case.

Conclusion: We investigated a design and optimization for the cascaded classifier to improve the energy efficiency on hardware. We illustrated our framework, presented the RTL-level energy model and integrated it into a closed form formulation. Experiments on EEG demonstrated our energy-efficient cascaded classifier design.

A. Wang and W. Xu (*Department of Computer Science and Engineering, University at Buffalo, the State University of New York, NY, 14260 USA*)

E-mail: wenyaoxu@buffalo.edu

References

1 G. Heitz, S. Gould, A. Saxena, and D. Koller, "Cascaded classification models: Combining models for holistic scene understanding," in *Advances in Neural Information Processing Systems*, 2009, pp. 641–648.

2 P. Lingras and C. Butz, "Rough set based 1-v-1 and 1-vr approaches to support vector machine multi-classification," *Information Sciences*, vol. 177, no. 18, pp. 3782–3798, 2007.

3 G. B. Moody, R. G. Mark, and A. L. Goldberger, "Physionet: a web-based resource for the study of physiologic signals," *IEEE Eng Med Biol Mag*, vol. 20, no. 3, pp. 70–75, 2001.