

An FPGA-Based Implementation of Factorization Machine for FMQA Black-Box Optimization

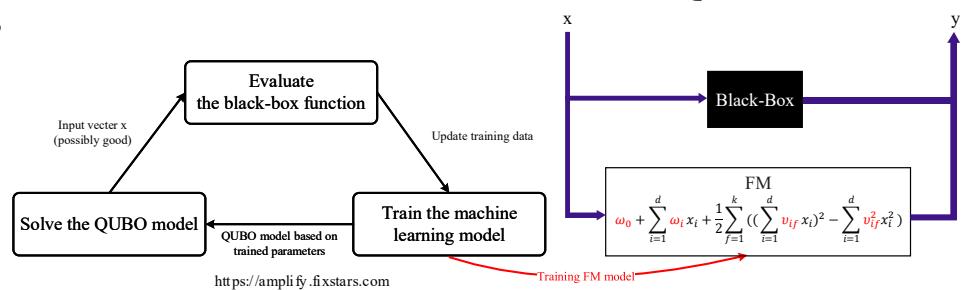
Junjie Xie, Hasitha Muthumala Waidyasooriya and Masanori Hariyama, Tohoku University, Japan
 Tadashi Kadowaki, National Institute of Advanced Industrial Science and Technology, Ibaraki, Japan
 DENSO CORPORATION, Tokyo, Japan

Background and Objectives

Many real-world optimization problems involve objective functions that are unknown or too complex to model analytically. Black-box optimization addresses such cases by iteratively evaluating input–output relationships without requiring explicit knowledge of the objective function. Factorization-Machine-based Quantum Annealing (FMQA) combines a Factorization Machine (FM) model with an annealing solver to optimize black-box functions. The FM approximates the objective as a quadratic function of binary variables and is iteratively trained with new samples, while quantum or simulated quantum annealing (SQA) searches for optimal configurations based on the trained model.

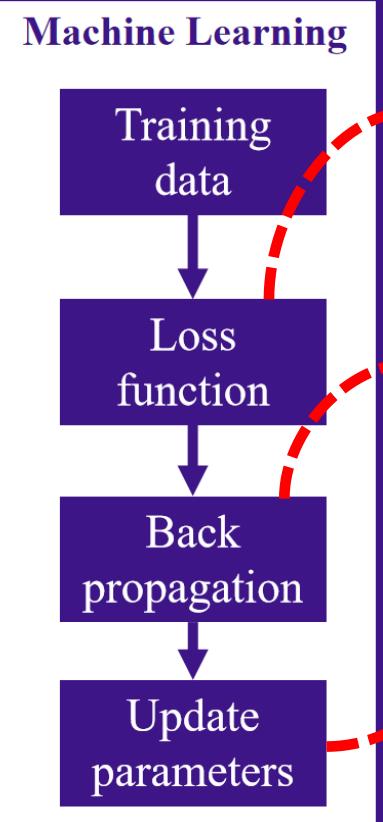
Although efficient FPGA-based SQA accelerators already exist, FM model training remains the bottleneck. In this work, we propose an FPGA-based FM training accelerator that achieves 3.7–7 times speedup over CPU implementation. This work also establishes the foundation for a complete FPGA-based black-box optimization accelerator by integrating the FM training and SQA components within a unified system.

Process flow of FMQA



Methods

Process flow of FM model training



$$\hat{y} = \omega_0 + \sum_{i=1}^d \omega_i x_i + \frac{1}{2} \sum_{f=1}^k \left(\left(\sum_{i=1}^d v_{if} x_i \right)^2 - \sum_{i=1}^d v_{if}^2 x_i^2 \right)$$

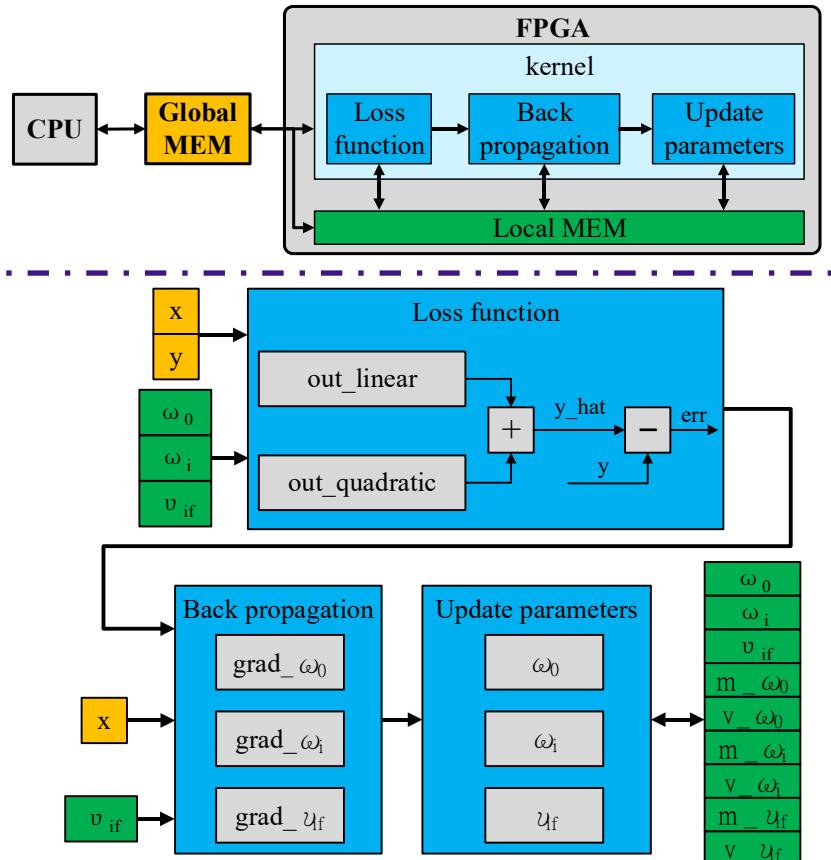
Loss function (MSE): $\frac{1}{2} (\hat{y} - y)^2$

$$\begin{aligned} \text{Gradient} &\leftarrow \frac{\partial L}{\partial \theta} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial \theta} = (\hat{y} - y) \cdot \frac{\partial \hat{y}}{\partial \theta} \\ \text{grad}_{\omega_0} &\leftarrow \frac{\partial L}{\partial \omega_0} = (\hat{y} - y) \cdot \frac{\partial \hat{y}}{\partial \omega_0} = (\hat{y} - y) \\ \text{grad}_{\omega_i} &\leftarrow \frac{\partial L}{\partial \omega_i} = (\hat{y} - y) \cdot \frac{\partial \hat{y}}{\partial \omega_i} = (\hat{y} - y) \cdot x_i \\ \text{grad}_{v_{if}} &\leftarrow \frac{\partial L}{\partial v_{if}} = (\hat{y} - y) \cdot \left(x_i \cdot \sum_{j=1}^d v_{jf} x_j - v_{if} x_i^2 \right) \end{aligned}$$

$$\begin{aligned} \mathbf{m}_t &\leftarrow \beta_1 \cdot \mathbf{m}_{t-1} + (1 - \beta_1) \cdot \mathbf{g}_t \\ \mathbf{v}_t &\leftarrow \beta_2 \cdot \mathbf{v}_{t-1} + (1 - \beta_2) \cdot \mathbf{g}_t^2 \\ \hat{\mathbf{m}}_t &\leftarrow \frac{\mathbf{m}_t}{1 - \beta_1} \\ \hat{\mathbf{v}}_t &\leftarrow \frac{\mathbf{v}_t}{1 - \beta_2} \\ \theta_t &\leftarrow \theta_{t-1} - \alpha \cdot \frac{\hat{\mathbf{m}}_t}{\sqrt{\hat{\mathbf{v}}_t} + \epsilon} \end{aligned}$$

Adam

FPGA architecture of the FM model training



Result

CPU vs. FPGA training time

(learning rate, feature size)	CPU training time (ms)	FPGA		Speedup
		Training time (ms)	Frequency (MHz)	
(0.001, 8)	21346	4335	332	5
(0.001, 16)	55447	7675	318	7
(0.1, 32)	106319	28561	317	3.7
(0.1, 64)	210856	48819	314	4.3

The experimental results demonstrate that implementing the FM training on an FPGA significantly enhances performance compared to a software-only implementation.

CPU: Intel Xeon 4316
 FPGA: Intel Agilex 1A-840F

Future works

- Further optimizations, such as performance can be further enhanced by increasing parallelism.
- Integrating FM training with SQA accelerators, fully FPGA-based optimizer that delivers both high performance and low power.