

Peer Review

Review of: "Optimizing Edge AI: A Comprehensive Survey on Data, Model, and System Strategies"

Arda Yurdakul¹

1. Computer Engineering, Bogazici University, Turkey

The authors present a survey on Edge AI. The topic is timely, and the efforts of the authors are notable. The survey sets forth the proposed solutions from academia and industry for the problems related to Edge AI. However, the title may not fully reflect the content. For example, the manuscript contains application scenarios of Edge AI, but none of them are evaluated from the perspective of optimizing Edge AI.

The manuscript might be too long, as it aims to cover many topics related to Edge AI. In the beginning, long explanations for each example are given. As the evaluator moves further, she observes that there are only names or single sentences with no clear explanations. Towards the end, sections with a single paragraph exist. The authors may consider balancing the explanations throughout the manuscript and shortening the overall manuscript. They may also think of excluding some parts.

The manuscript can also be reorganized: 1) Repeated paragraphs or comments can be avoided. If a comment or a paragraph applies to more than one section, the authors may prefer merging them appropriately. 2) Some figures may be revised. For example, the intersection in Figure 4 does not reflect the manuscript. Figure 7 contains ambiguities: Do edge servers always collect data? To my knowledge, no. For example, model parameters are shared in federated learning or knowledge distillation. If the authors prefer to treat the end devices as only data sources, then the model compression part is redundant. Hence, there is a need for clarification. Model quantization in Figure 8 can be replaced with a more related drawing. 4) Some numbered/labeled sections or subsections contain only one paragraph. There is even one section where only one citation is given. It is common practice that sectioning is done when there is more than one paragraph or more citations to mention.

The manuscript can be strengthened by considering the following related and essential research topics: 1) Learning algorithms with their variants for Edge AI can be included as a subsection. For example, Federated Learning (FL) with heterogeneous devices is overlooked. Also, deep learning inference at the edge is important, as solutions like model partitioning exist in the literature. 2) Due to the resource-constrained edge devices, various data types and number formats have been proposed and are still being proposed as an optimization of Edge AI. The authors just name them, avoiding explaining what they are and why they are proposed. This part can also be included as a subsection. 3) If the authors want to keep the Security aspect in the manuscript, they need to watch out: Blockchain and AI are going hand in hand at the edge to enhance security and privacy. Even though there is a vast amount of research on this topic, the authors mention blockchain as a future solution in Edge AI security with no clear evaluation of the existing studies.

Declarations

Potential competing interests: No potential competing interests to declare.