Skysong Innovations
Advancing the Arizona State University Knowledge Enterprise

1475 N. Scottsdale Road, Suite 200
Scottsdale, AZ 85287-3538
Phone: 480 884 1996 Fax: 480 884 1984

Case ID:M22-093P
Published: 1/4/2023

Inventors

skytech_inventors does not have any rows

Contact

Team

Memory Efficient, Multi-Domain On-Device Machine Learning

-One practical limitation of deep neural network (DNN) is its high degree of specialization to a single task or domain (e.g., one visual domain). This motivates the development of algorithms that can adapt DNN model to multiple domains sequentially while still performing well on past domains. This is known as multi-domain learning.

Conventional multi-domain learning algorithms only focus on improving accuracy with minimal parameter update while ignoring high computing and memory costs during training. This makes it difficult to deploy multi-domain learning into widely used resource-limited edge devices, like mobile phones, IoT devices, edge devices, embedded systems, etc. For example, an IoT device collects massive amounts of new data crossing various domains/tasks in daily operations. Conventionally, to process the new data, learning/training is performed on cloud servers and then the learned DNN model is transferred back to the IoT device for inference only. This approach (i.e., learning-on-cloud and inference-on-device) is inefficient due to communication cost between cloud and device as well as data privacy concerns (e.g., sensitive information being shared back and forth with cloud). Thus, there is a need for on-device multi-domain learning.

Researchers at Arizona State University (ASU) have developed a memory-efficient, on-device multi-domain machine learning method. This method mitigates activation memory buffering for reducing memory usage during training and serves as a dynamic gating mechanism to reduce computation cost for fast inference.

Related publication: DA3: Dynamic Additive Attention Adaption for Memory-Efficient On-Device Multi-Domain Leaning

Potential Applications:

- For multi-domain on-device learning on devices including:
  - Cell phones
  - Medical devices
  - Wearables
  - Sensors
  - Other IoT devices

Benefits and Advantages:

- Significant reduction in training costs (i.e., time and memory)
  - Tested ASU's machine learning method on an edge GPU, results show a reduction in on-device training memory consumption (by 19-37x) and a reduction in training time (by 2x) in comparison to baseline methods (e.g., fine-tuning, series and parallel residual adapters, Piggyback, etc.)

- Largely eliminates the storage of intermediate activation feature map to greatly reduce overall memory usage