Cartel: A System for Collaborative Transfer Learning at the Edge

Harshit Daga* | Patrick K. Nicholson* | Ada Gavrilovska* | Diego Lugones*

*Georgia Institute of Technology, +Nokia Bell Labs
Multi-access Edge Computing (MEC)

- Compute & Storage closer to the end user
- Provides ultra-low latency
Machine Learning @ Edge

- There is tremendous growth of data generated at the edge from end-user devices and IoT.

- We explore machine learning in the context of MEC:
  - Results are only needed locally
  - Latency is critical
  - Data volume must be reduced
Existing Solution

Centralized System

Problems

- **Data movement** is time consuming and uses a lot of backhaul network bandwidth.

- **Distributed ML** across geo-distributed data can **slow down** the execution up to 53X\(^\text{[1]}\).

- **Regulatory constraints** (GDPR)

---

An Alternative Approach

**Isolated System**

- Train machine learning models independently at each edge, in isolation from other edge nodes.
- The isolated model performance gets heavily impacted in scenarios where there is a need to adapt to changing workload.
Motivation

Can we achieve a balance between centralized and isolated system?

Leverage the resource-constrained edge nodes to train customized (smaller) machine learning models in a manner that reduces training time and backhaul data transfer while keeping the performance closer to a centralized system?

Opportunity

• Each edge node has its own attributes / characteristics → a full generic model trained on broad variety of data may not be required at an edge node.
Solution Overview

Cartel: A System for Collaborative Transfer Learning at the Edge

- Cartel maintains small customized models at each edge node.
- When there is change in the environment or variations in workload patterns, Cartel provides a jump start to adapt to these changes by transferring knowledge from other edge(s) where similar patterns have been observed.
Key Challenges

C1: When to request for model transfer?
C2: Which node (logical neighbor) to contact?
C3: How to transfer knowledge to the target edge node?
Solution Design

**Raw data v/s Metadata**

- Do not share raw data between any edge nodes or with the cloud.
- Use Metadata
  - Statistics about the network
  - Software configuration
  - Active user distribution by segments
  - Estimates of class priors (probability of certain classes), etc.

Cartel maintains and aggregates metadata locally and in the metadata server (MdS).
C1: When to request for model transfer?

**Drift Detection**

- Determine when to send a request to collaborate with edge nodes for a model transfer.
- In our prototype we use a threshold-based drift detection mechanism.
C2: Which neighbors to contact?

Logical Neighbor

• Find the neighbor that has similar class priors to the target node.

• We call them as “logical neighbors” as they can be from anywhere in the network.

• In our prototype class priors are undergoing some shift, the empirical distributions from the target node is compared with those from the other nodes at the MdS to determine which subset of edge nodes are logical neighbors of the target node.
C3: How to transfer knowledge to the target?

Knowledge Transfer

• Two steps process
  1. Partitioning
  2. Merging

Help Me (SOS)
Solution Overview

Edge Node

Collaborative Component

Existing ML Library*

Data
Solution Overview

Edge Node

Collaborative

Register
Accuracy Trend
Distribution Drift
Transfer

Learning

ML Model

Predict
Train
Partition
Merge

Data
Evaluation

Goals

• How effectively system adapts to the change in workload?

• How effective is Cartel in reducing data transfer costs, while providing lightweight and accurate models?

• What are the costs in the mechanisms of Cartel and the design choices?

• How does Cartel perform in a real-world scenario?

Methodology

• Workload

  ![Graph](image)

  - Introduction Workload
  - Fluctuation Workload

• Machine Learning Model – ORF & OSVM

• Datasets used - MNIST & CICIDS2017
Evaluation

Goals

• How effectively does the system adapt to the change in workload?
• How effective is Cartel in reducing data transfer costs, while providing lightweight and accurate models?
• What are the costs in the mechanisms of Cartel and the design choices?
• How does Cartel perform in a real-world scenario?

Methodology

• Workload

  ![Introduction Workload Graph](image1)

  ![Fluctuation Workload Graph](image2)

• Machine Learning Model – ORF & OSVM
• Datasets used - MNIST & CICIDS2017
Evaluation

Adaptability to Change in the Workload

- Online Random Forest (ORF)
- Centralized
- Isolated
- Cartel

Introduction Workload

- Number of Requests
- Time
- Introduction Workload

Model Error %

- 30-
- 10
- 0

Batch ID

- 20
- 40
- 60
- 80
- 100

Online Random Forest (ORF)
Evaluation

Adaptability to Change in the Workload

Fluctuation Workload

Time

Class 1  Class 2

Online Support Vector Machine (OSVM)

Batch ID

20  40  60  80  100
Evaluation

Adaptability to Change in the Workload

• When changes in the environment or variations in workload patterns require the model to adapt, Cartel provides a jump start by transferring knowledge from other edge(s) where similar patterns have been observed.

• Cartel adapts to the workload changes up to 8x faster than isolated system while achieving similar predictive performance compared to a centralized system.
Evaluation

Data Transfer Cost

• Data/Communication cost includes the transfer of raw data or metadata updates.

• Model transfer cost captures the amount of data transferred during model updates to the edge (periodically in case of centralized system or partial model request from a logical neighbor in Cartel).

• Cartel reduces the total data transfer cost up to 1500x when compared to a centralized system.
Summary

- We introduce Cartel, a system for sharing customized machine learning models between edge nodes.

- Benefits of Cartel include:
  - Adapts quickly to changes in workload (up to 8x faster compared to an isolated system).
  - Reduces total data transfer costs significantly (1500x ↓ compared to a centralized system).
  - Enables use of smaller models (3x ↓) at an edge node leading to faster training (5.7x ↓) when compared to a centralized system.
Cartel: A System for Collaborative Transfer Learning at the Edge

Harshit Daga* | Patrick K. Nicholson* | Ada Gavrilovska* | Diego Lugones+

*Georgia Institute of Technology, +Nokia Bell Labs