University of Glasgow

School *of* Computing Science

Essence: Data Engineering & Distributed Computing

Machine Learning Models Re-usability in Edge Computing Environments

Xenia Skotti

Kostas Kolomvatsos

Christos Anagnostopoulos

WORLD CHANGING GLASGOW

Table of Contents

What is Edge Computing?

- Paradigm where the nodes at the edge of the network both produce and consume data
- More efficient to process the data at the edge
- Applied in a variety of contexts such as smart homes and smart cities.

University of Glasgow

School *of* Computing Science
Essence: Pervasive &
Distributed Computing

How to analyse information?

Machine Learning Methods

What is the problem?

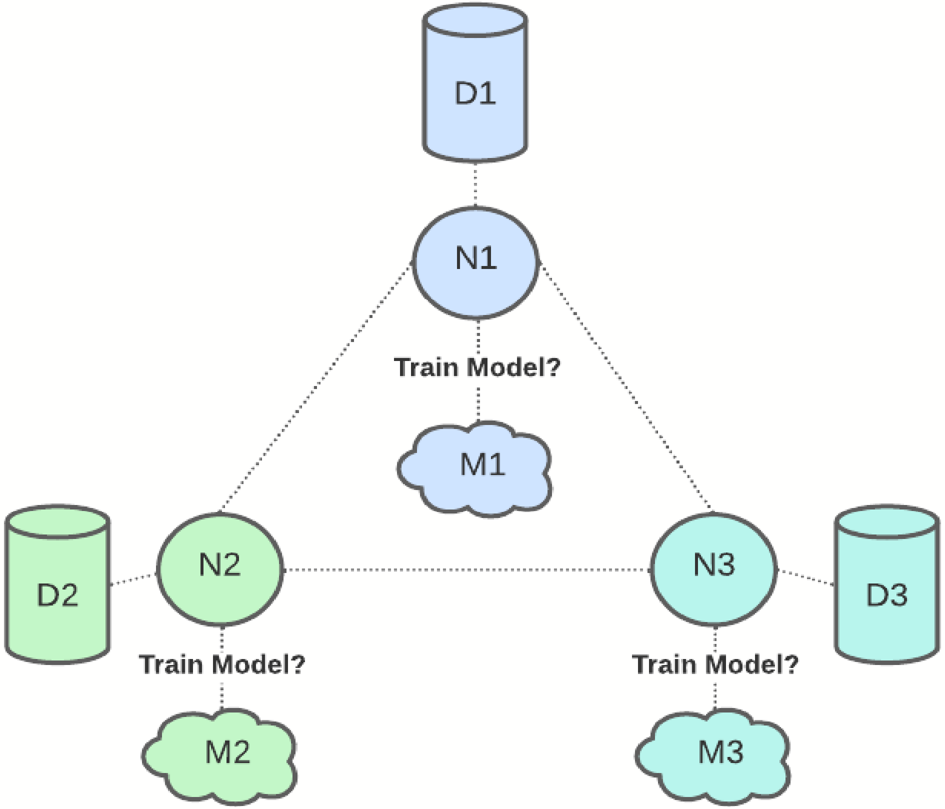
- computation
- energy consumption
- storage

What now?

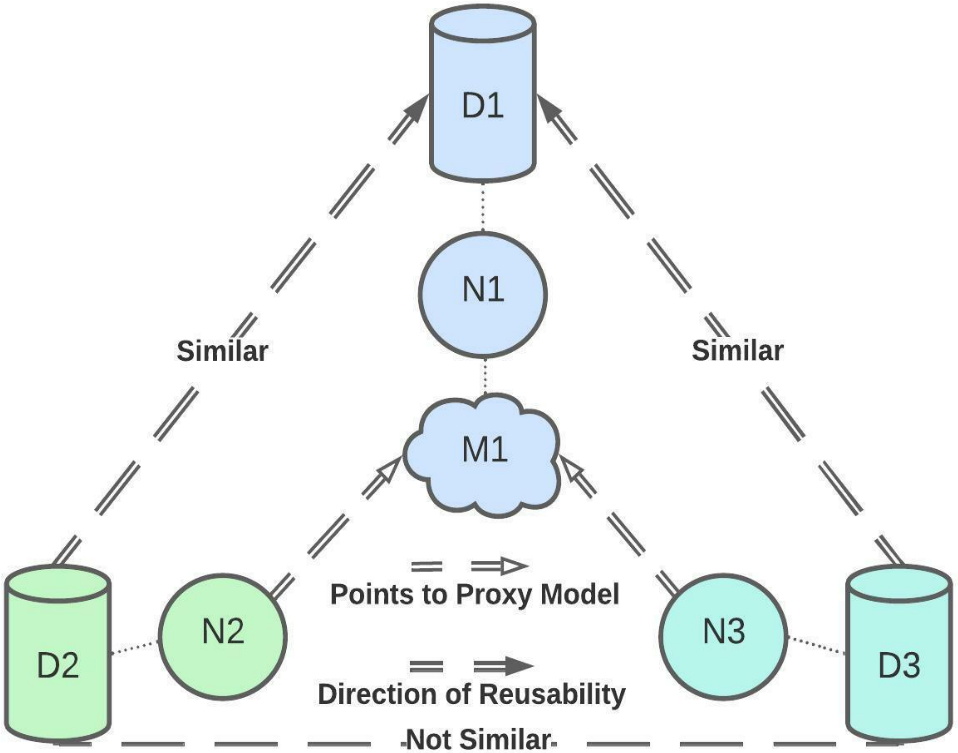
Reduce the number of models needed to be trained in the first place by reusing existing models.

University of Glasgow

Aim



Framework Requirements



Pair Similarity Detector

The similarity detector is based on the Maximum Mean Discrepancy (MMD).

MMD quantifies the mean discrepancy of two data distributions in a kernel space in order to determine if two samples are drawn from different distributions.

The detector compares the MMD value of the pair against a threshold we've dubbed the Average Similarity MMD (ASMMD) to distinguish between similar and non-similar pairs.

Direction of Reusability Detector

The direction of model reusability detector is based on the inlier data space overlap.

A predictor for inlier space overlap is the probability of correctly predicting the non-native inliers of a model.

We used the One-class Support Vector Machines (OCSVM) to determine which points are inliers.

Given two nodes and their corresponding OCSVM models, we can use each OCSVM model to predict the other node's inliers and then find the probability of detecting them, hence their overlap.

Experimental Setup

Datasets:

- GNFUV Unmanned Surface Vehicles Sensor Data Set (GNFUV)
- UCI Bank Marketing Dataset (BM)

ML Models:

- Support Vector Regression (SVR)
- Logistic Regression (LR)

We used Grid Search to optimise the parameters of the models.

We experimented with a linear and non-linear kernel for regression, standardised and non-standardised version of the regression data and a balanced and non-balanced version of the classification data.

Model Reusability Metrics

Speedup: the gain in performance by avoiding to train some of the models

Precision: the number of times the framework is correct. Assessed for each component individually and combined.

Results: Precision

Regression:

The framework performs better on the original data across all three levels of precision with 0.77 combined precision at threshold 0.8 and a drop of 15% for standardised ones.

The threshold simply states how close to the true model performance is the proxy model performance.

The linear kernel is better suited for the original data, while the opposite is true for standardised data even though this difference is not large.

The rbf kernel models have higher performance on their native datasets compared to linear ones, but nevertheless have higher discrepancy. On average linear models provide better results.

Results: Precision

Classification:

The combined precision is extremely high regardless of whether we distinguish between the configurations or not, if we are not strict.

If we are strict this performance drops at 0.55 on average and this is a direct reflection of the OCSVM precision.

Considering how good the performance is overall, the real combined precision of the framework is the one given by the non-strict measure.

Non-strict and strict refers to whether we allow a 0.05 margin of error on the direction of reusability detector.

Results: Speedup

Overall, the speedup of the framework for the particular datasets used for regression and classification are 26% and 29% to 41% respectively .

The GNFUV dataset has four nodes and on average there is one good pair for reusability regardless of the data configuration hence one node's model is not trained.

The nodes of the BM dataset were derived from two clusters so ideally we would only train two models. Nevertheless the results are lower than this average case due to the fact we use samples of the dataset hence the true reusability differs from sample to sample.

Hence, for both the classification and regression case we can argue that the framework is effective in identifying the true number of similar pairs.

Conclusions

- Presented a novel online model reuse framework in edge computing.
- The framework considers all possible pairs of nodes in the network and infers which are good reusability pairs using MMD as well as which of the two nodes' model can be used as a replacement model for the other per pair.
- The node model that is chosen to be reused in each pair is the one with the highest inlier data space overlap.

Conclusions

- Proposed an algorithm which given the results of the framework so far, it can maximise the number of nodes which use reused models along with a list of potential replacement models.

- Experiments in the context of both regression and classification indicate the framework achieves good precision.

Limitations & Future Work

- Extend the evaluation of the framework to check it's compatibility with more domain models and data configurations.
- Amend the framework to preserve user privacy.
- Consider more than one outlier detection model.
- Define decision making algorithms which can produce either the performance optimal or partially optimal solution.

University of Glasgow

School *of* Computing Science

Essence: Data Engineering & Distributed Computing

Thank you!

Questions?

Email me at: xeniaskotti@gmail.com

http://www.dcs.gla.ac.uk/essence/