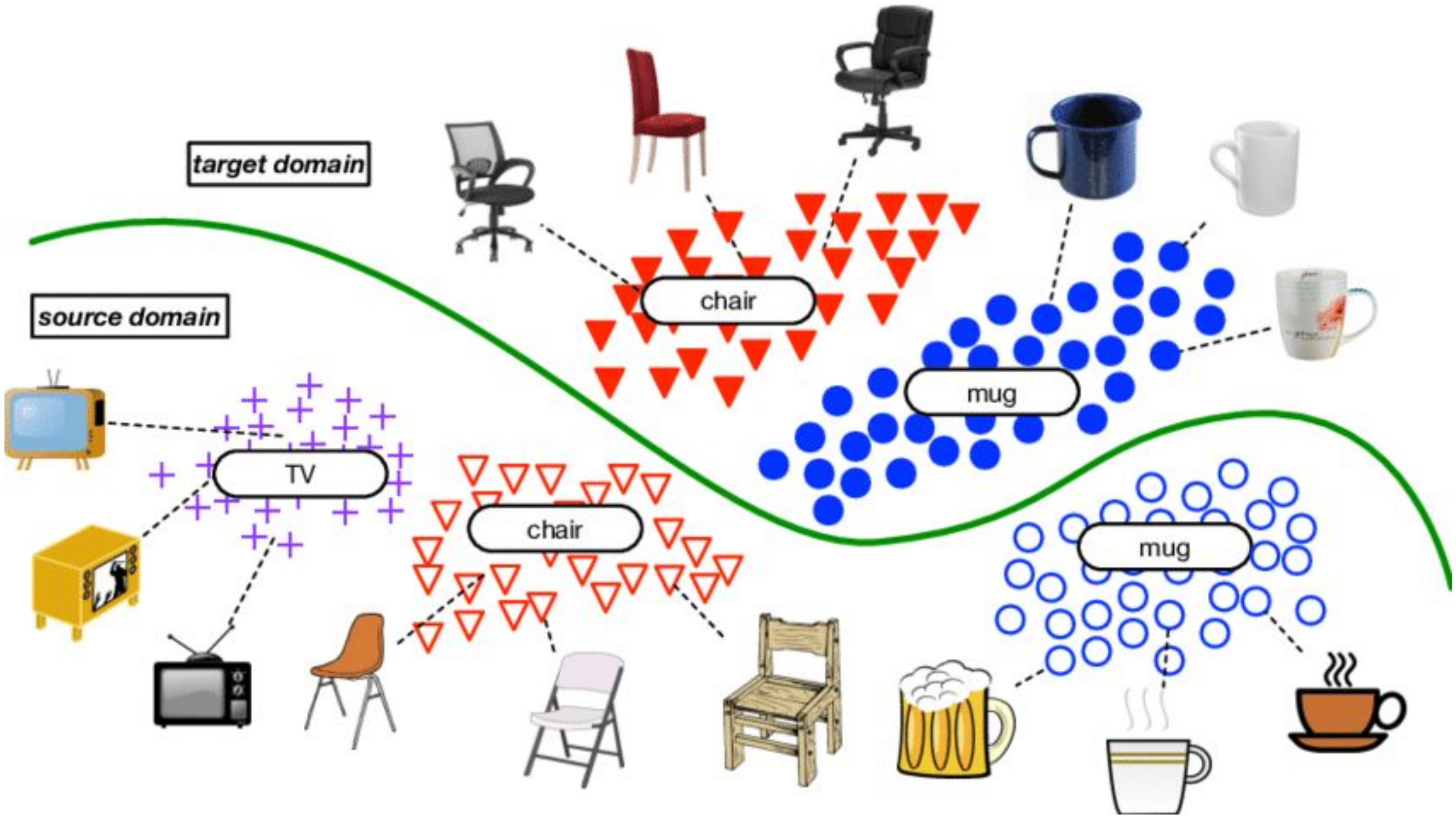
Domain Adaptation for Predictive Maintenance

Sepideh Pashami

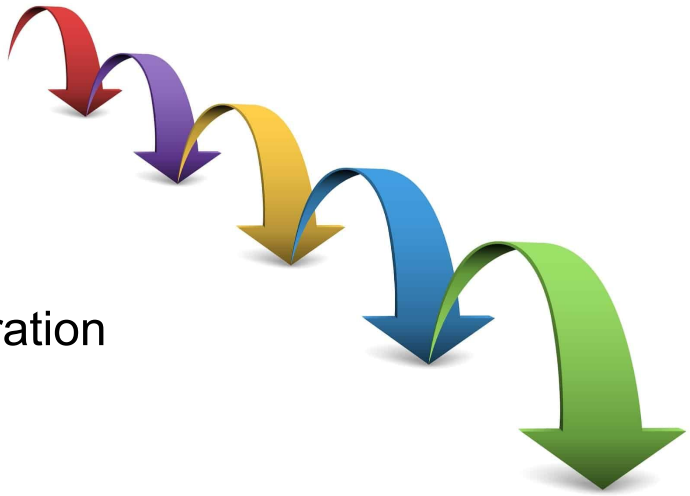
sepideh.pashami@{ri.se, hh.se}

June 29th, 2023


target domain
source domain
chair
mug
TV
chair
mug

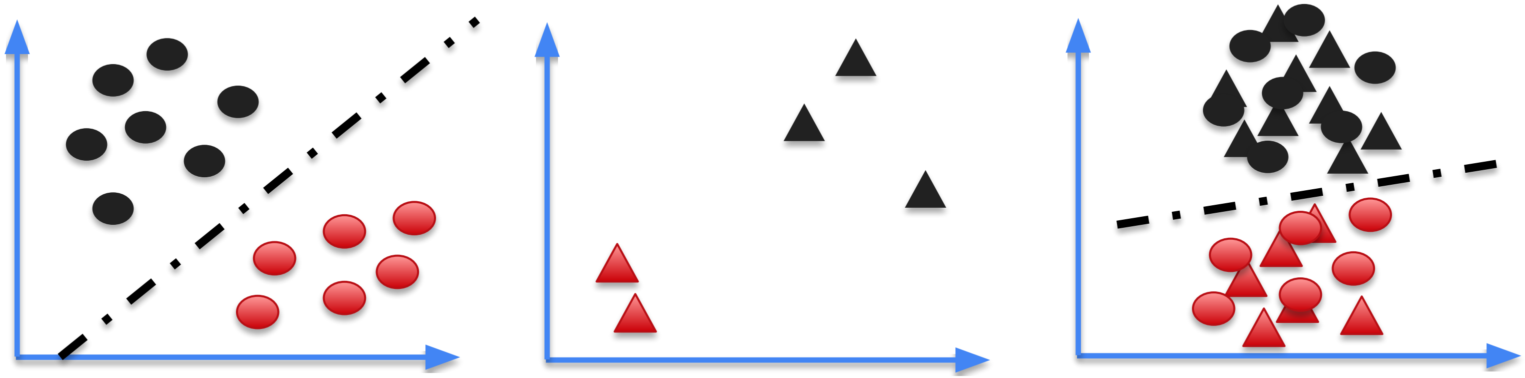
Why Transfer?

- To avoid building a model from scratch
  - Time consuming to train a model
- To reuse existing knowledge
  - Not to develop a model for each conditions
- To deal with a new system or a new configuration
  - Avoid collecting a lot of new data
- To preserve privacy
  - Share models rather than the data



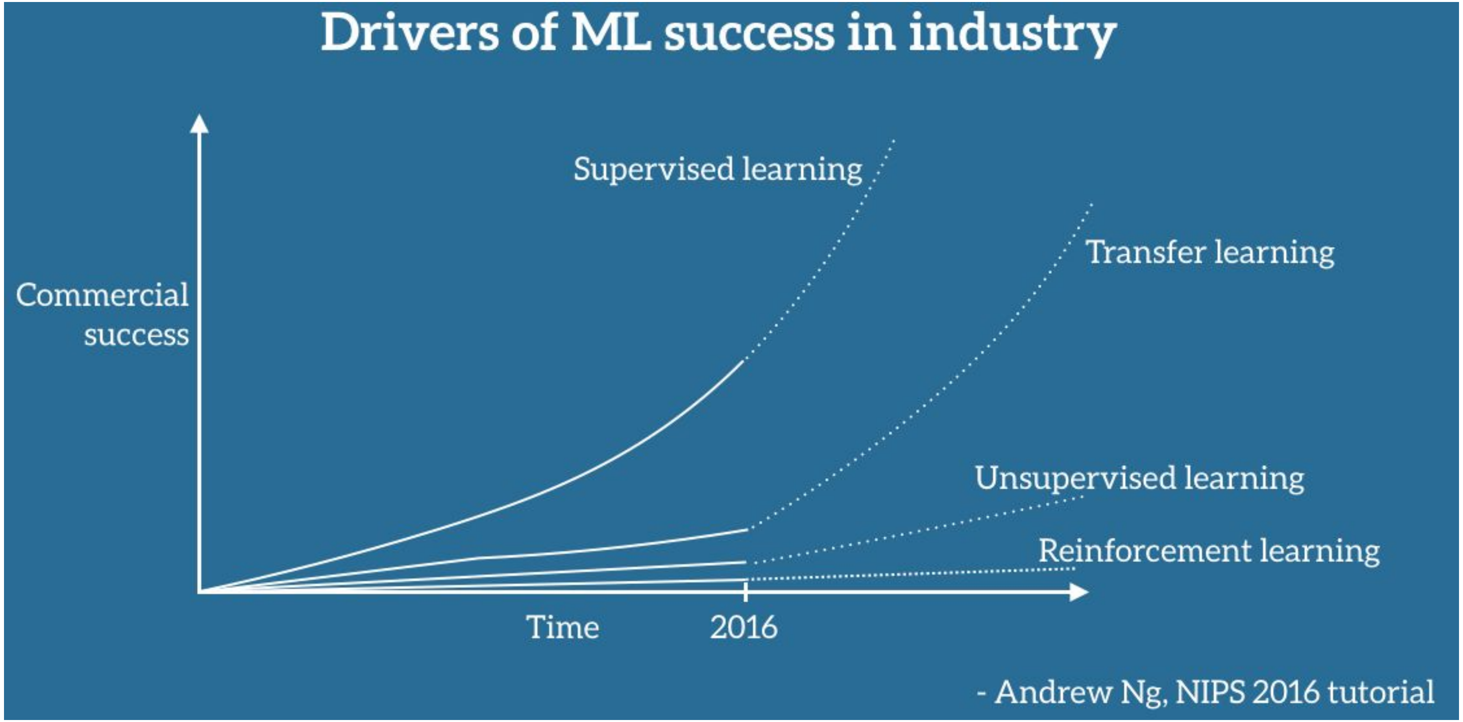
Transfer Learning

The ability of a system to recognise and apply knowledge and skills learned in **previous** domains/tasks to **novel** domains/tasks, which **share some commonality**.



Source domain | Target domain | Transfer Learning

Is transfer learning considers as ML's next frontier?



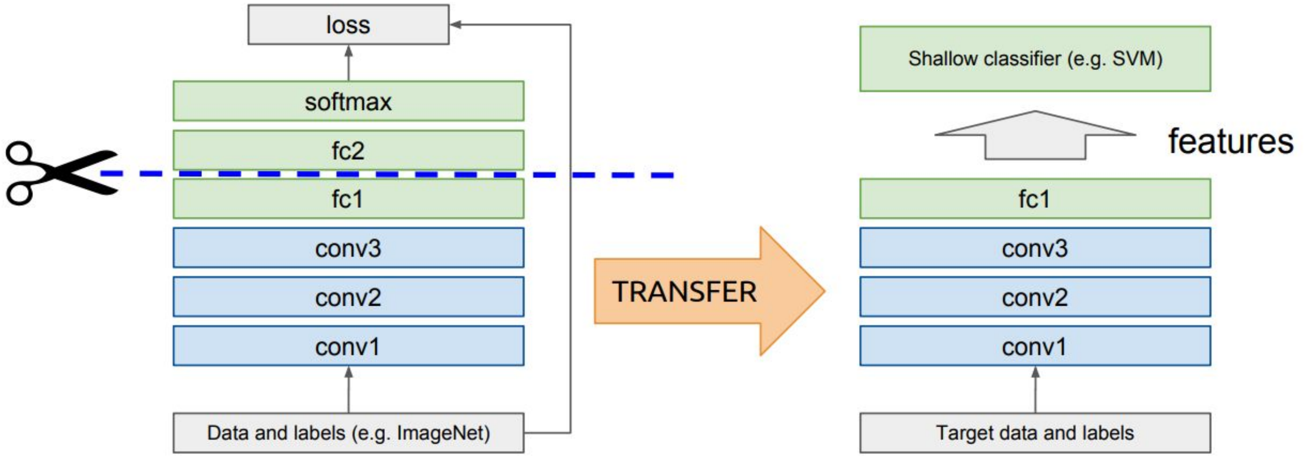
https://ruder.io/transfer-learning/index.html

Domain adaptation

“Off-the-shelf” strategy

Idea: use outputs of one or more layers of a network trained on a different task as generic feature detectors. Train a new shallow model on these features.

Assume: $\mathbf{D_S = D_T}$



Fine-tuning: supervised domain adaptation

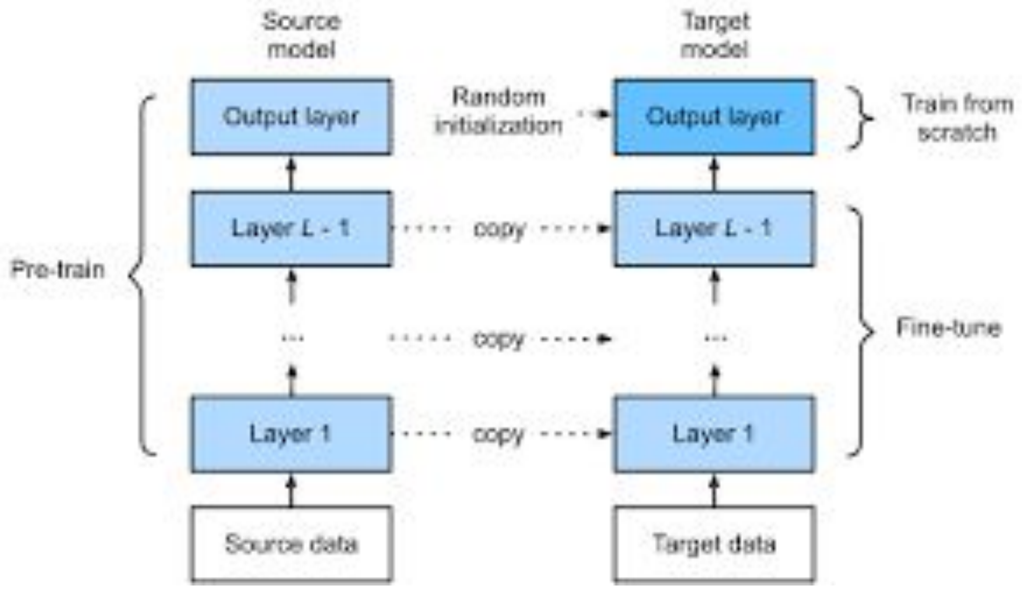
Pretrain a network which it is easy to get labels or select a pretrained network

- E.g. ImageNet classification
- Pseudo classes from augmented data
- Slow feature learning, ego-motion

Replace the last layer with supervised objective for target domain

Fine-tune network with labels for target domain

Aligns D_S with D_T



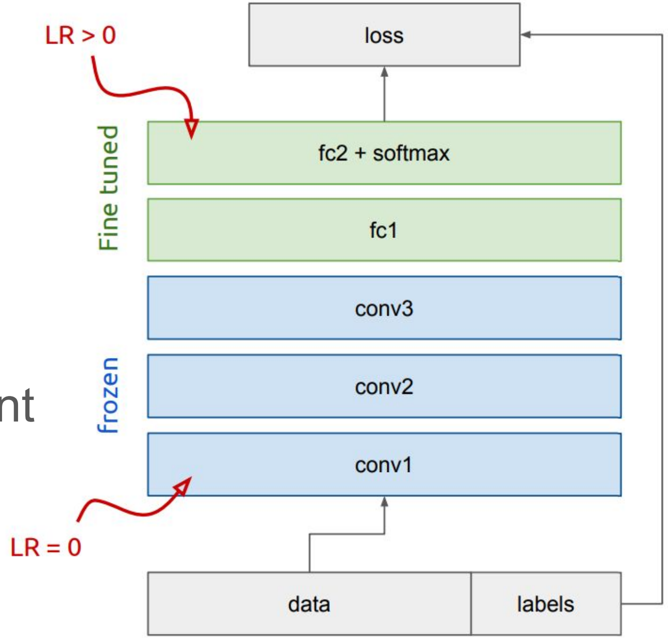
Freeze or fine-tune?

Bottom n layers can be frozen or fine tuned.

- **Frozen:** not updated during backprop
- **Fine-tuned:** updated during backprop

Which to do depends on target task:

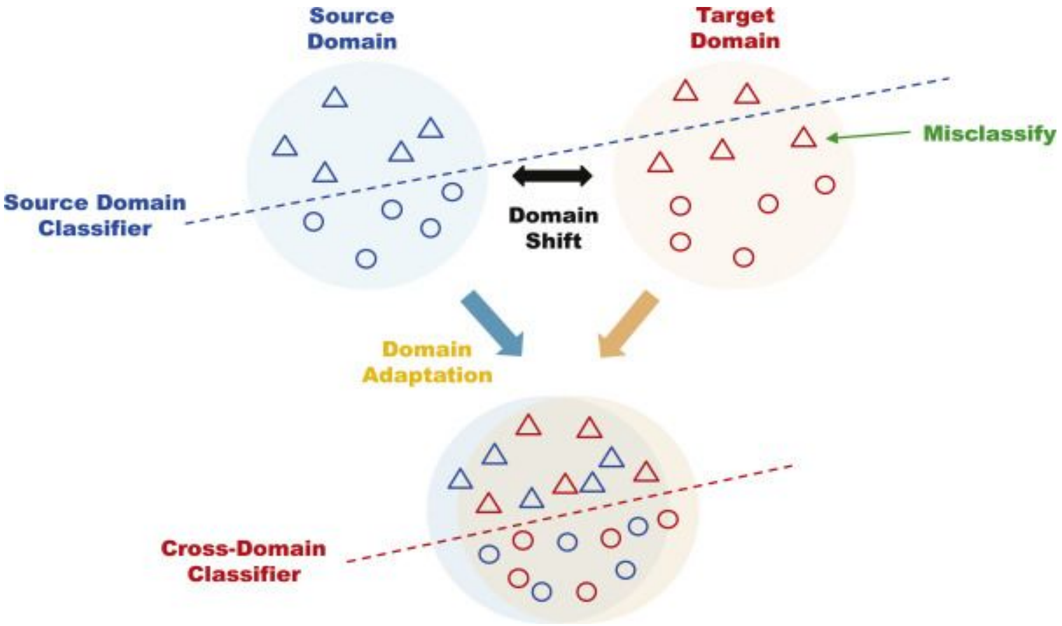
- **Freeze:** target task labels are scarce, and we want to avoid overfitting
- **Fine-tune:** target task labels are more plentiful In general



Learning with domain adaptation

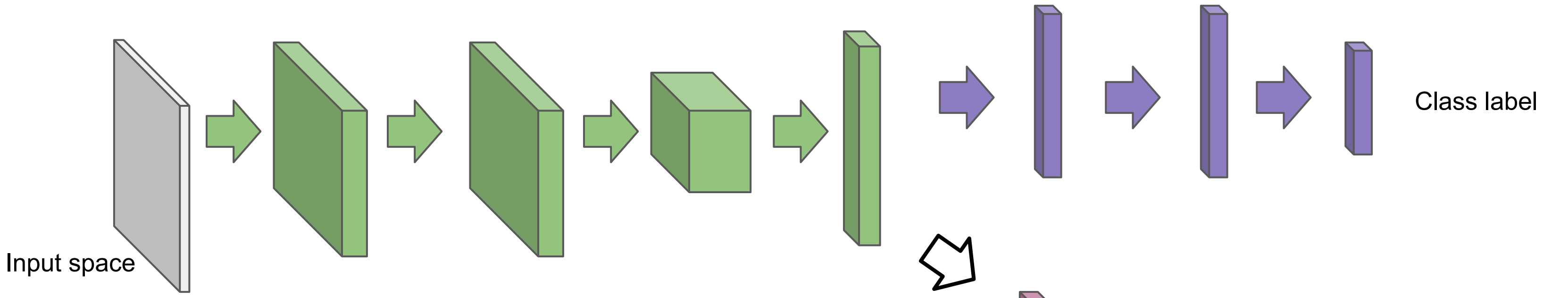
When data at source and target domains come from similar but different distributions

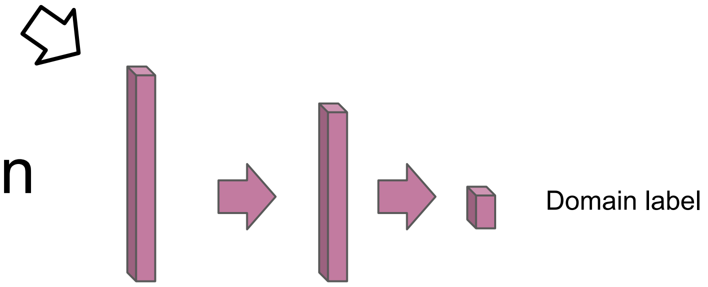
Create a new feature space that **cannot discriminate between the source and target domains** and **classify correctly in both domains**



https://arxiv.org/pdf/2010.03978.pdf

Domain Adversarial Neural Network (DANN)





Build a network for domain adaptation

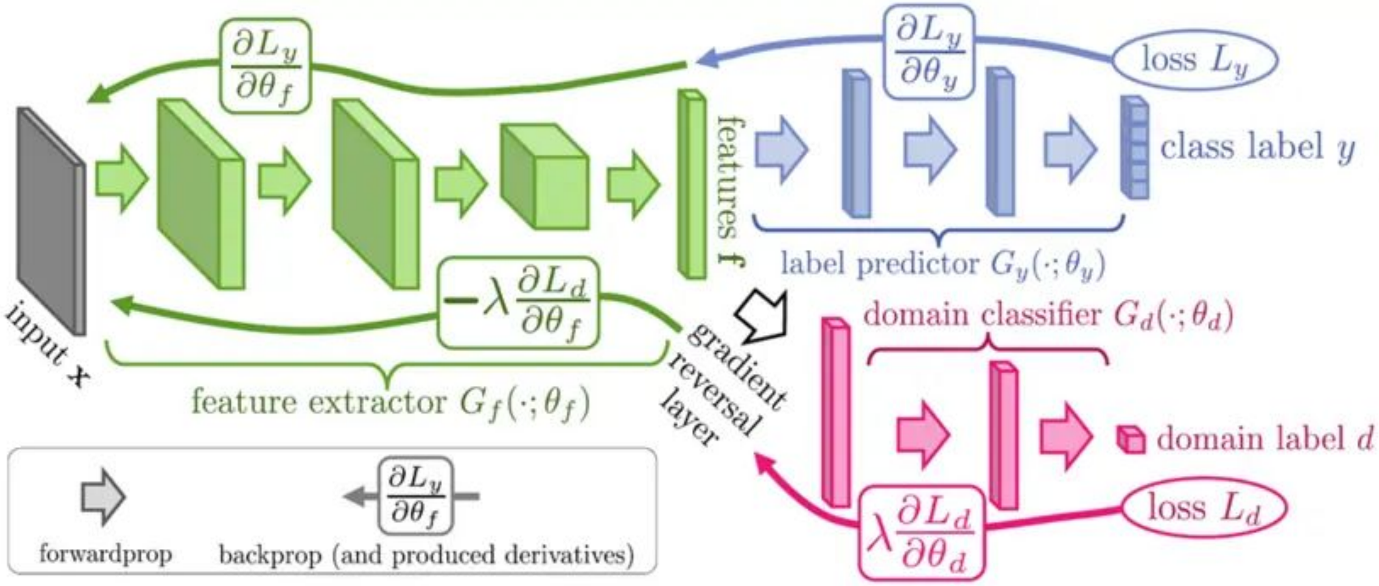
- Train a feature extractor and class predictor on **source data**
- Train a feature extractor and domain adaptor on **source and target data**
- Use a feature extractor and class predictor at the **test time**

DANN, https://arxiv.org/pdf/1505.07818.pdf

Backpropagation during training

Design the loss function in a way that we have

- Good class predictor
- Bad domain classifier



Minimising domain shift during backprop.

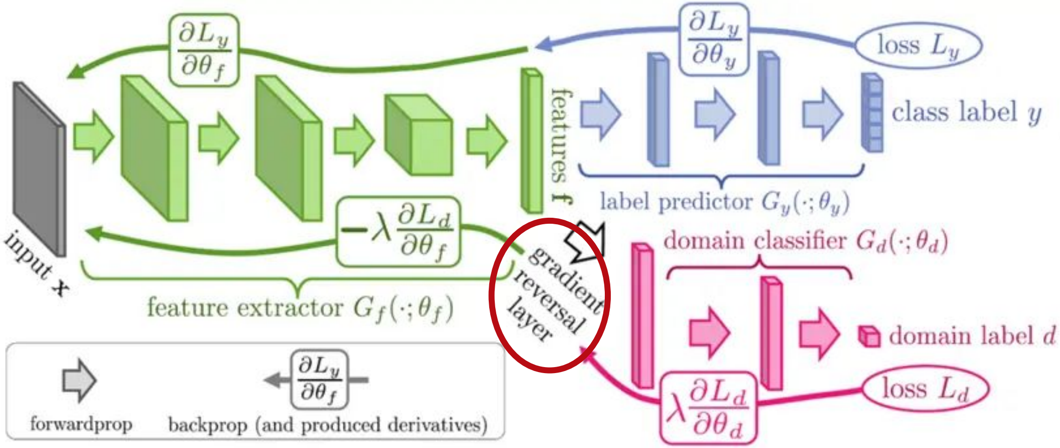
gradient reversal layer

- leaves the input unchanged during forward propagation
- reverses the gradient by multiplying it by a negative scalar during the backpropagation

$$\theta_f \longleftarrow \theta_f - \mu\left(\frac{\partial \mathcal{L}_y^i}{\partial \theta_f} - \lambda \frac{\partial \mathcal{L}_d^i}{\partial \theta_f}\right),$$

$$\theta_y \longleftarrow \theta_y - \mu \frac{\partial \mathcal{L}_y^i}{\partial \theta_y},$$

$$\theta_d \longleftarrow \theta_d - \mu\lambda \frac{\partial \mathcal{L}_d^i}{\partial \theta_d},$$

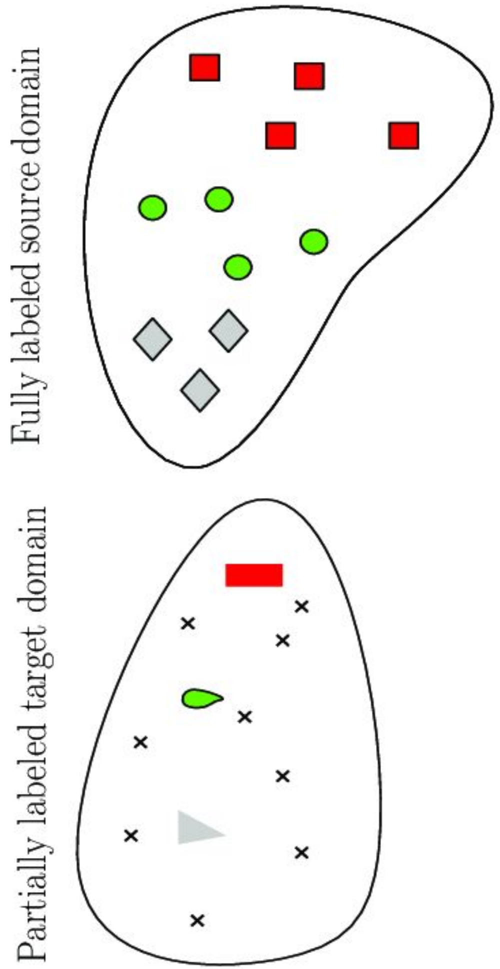


Semi-supervised domain adaptation

When some labels are available in the target domain, then we can use these when doing domain adaptation.

Simultaneously optimizing different criteria:

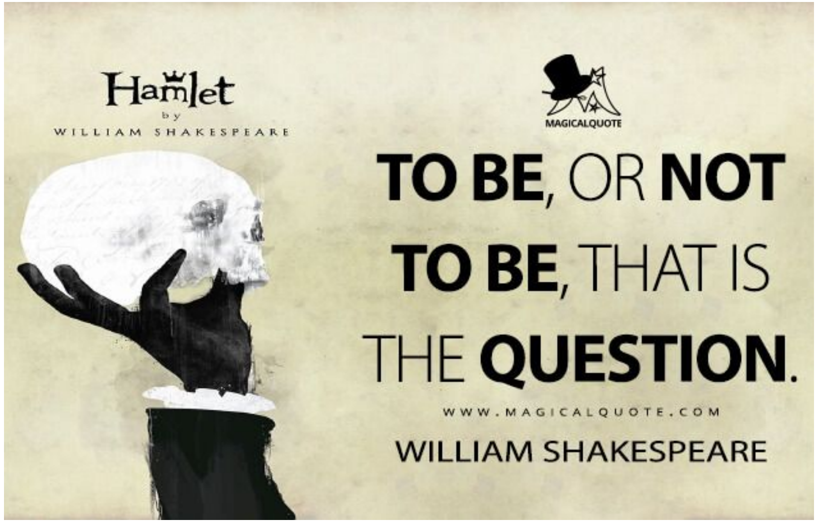
- Classification accuracy on both source and target datasets
- Goodness of mapping different domains



To transfer or not to transfer

Use transfer learning when

- Source and target domains are similar enough!
- Target data set is relatively small
- Source model has been trained on a quite big data set
- Source domain contain diverse set of the data, and not from a domain-specific task



VOLVO
Volvo Buses
Volvo Connected Solutions

EVE - Extending life of Vehicles within Electromobility era

HALMSTAD UNIVERSITY

VINNOVA

2019 - 2023

Goals

- Predicting lifetime of each individual battery
- Suggesting actions to extend battery lifetime
- Monitoring of other components in electric driveline,
  e.g. battery contactor
- Estimating energy consumption and proposing services for different behaviour groups

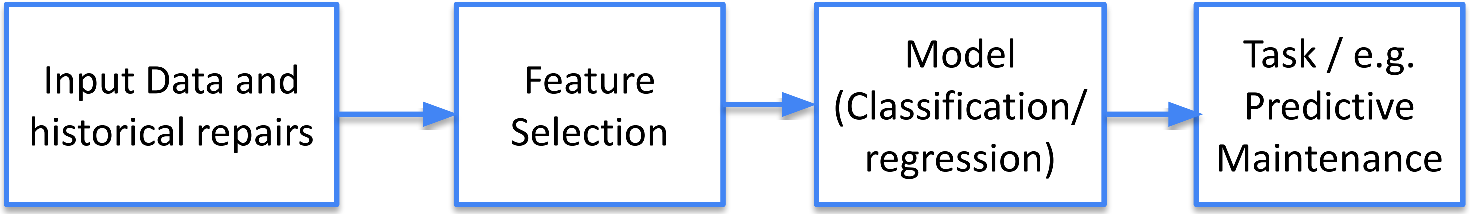
Monitoring drive batteries

Challenges

Real data is complex

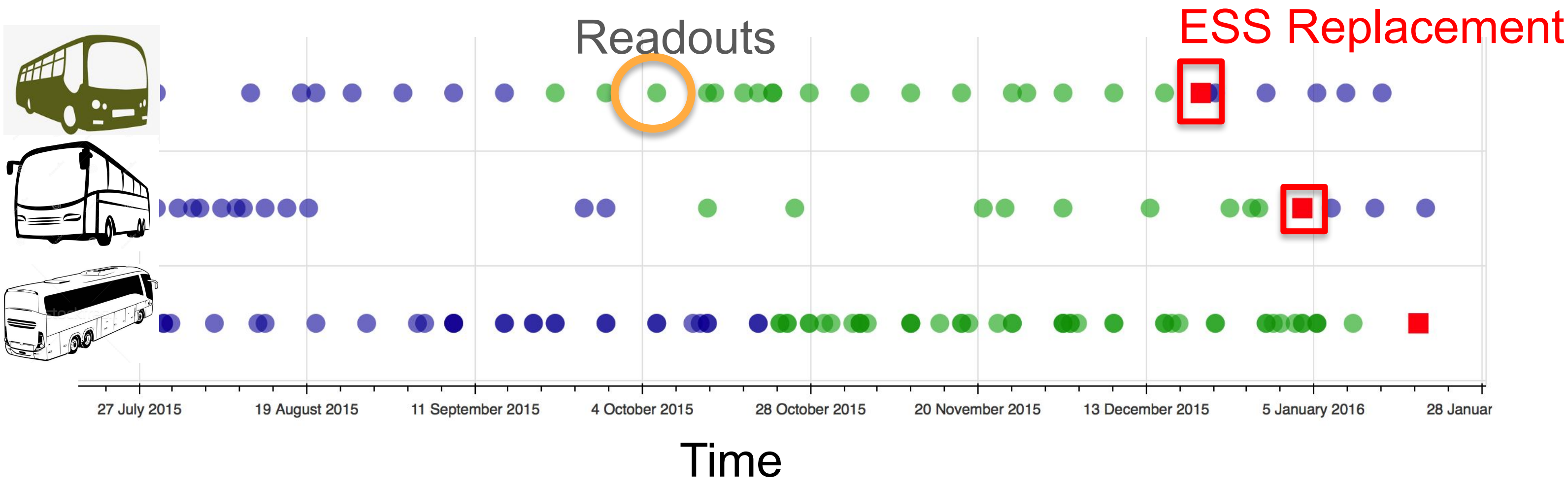
- Relevant information is seldomly directly measured
  - Labels (repair vs actual fault)
  - Sensor measurements (limited number of sensors)
- Data is heavily imbalanced
- Normal operation is difficult to characterise
  - Concept drift due to usage & external conditions
- Missing data
- Low data frequency

Typical Approach - Classical supervised machine learning



Analysing the current replacement strategy

- Can we create a model that predicts the current strategy?
- How can we improve the current strategy?



Results of ESS replacement model

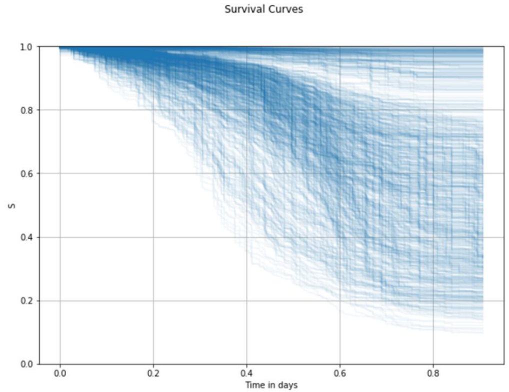
- A ML model to estimate if a bus is approaching an ESS replacement due date, based on operating parameters.
- Trained with 360 buses data and tested on 90 buses.

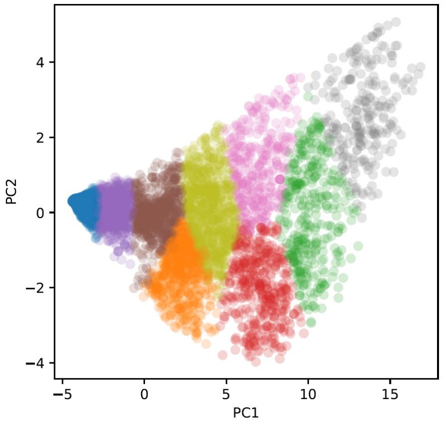
All - 180 Days	Predicted: **No**	Predicted: **Yes**
Actual: **No**	7,441	1,427
Actual: **Yes**	1,041	1,834

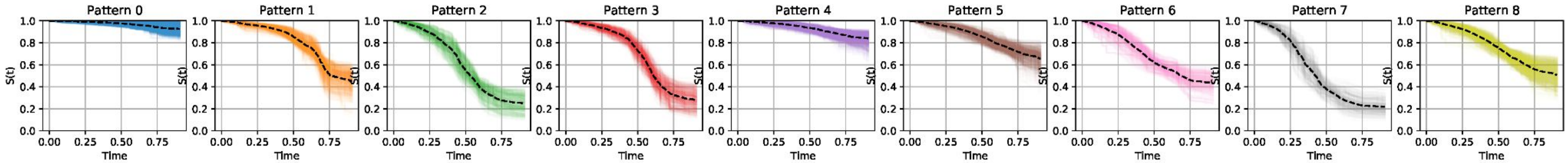
B5LH – 180 Days	Predicted: **No**	Predicted: **Yes**
Actual: **No**	3,516	933
Actual: **Yes**	654	2,540

Survival of batteries for hybrid buses

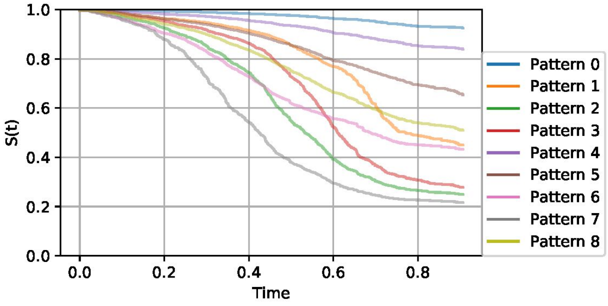
Different survival patterns



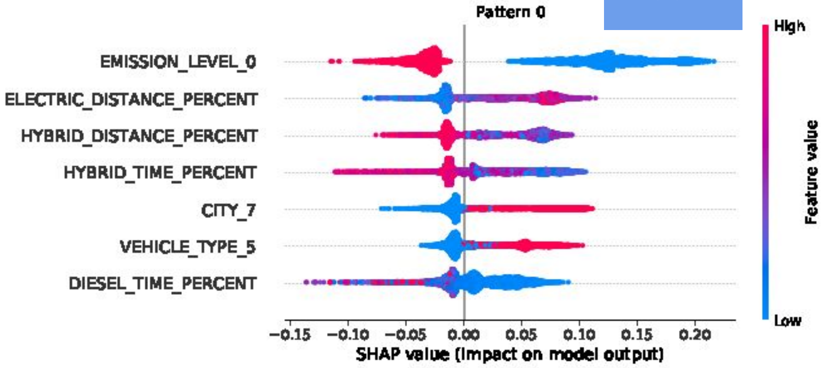




Explanation for pattern



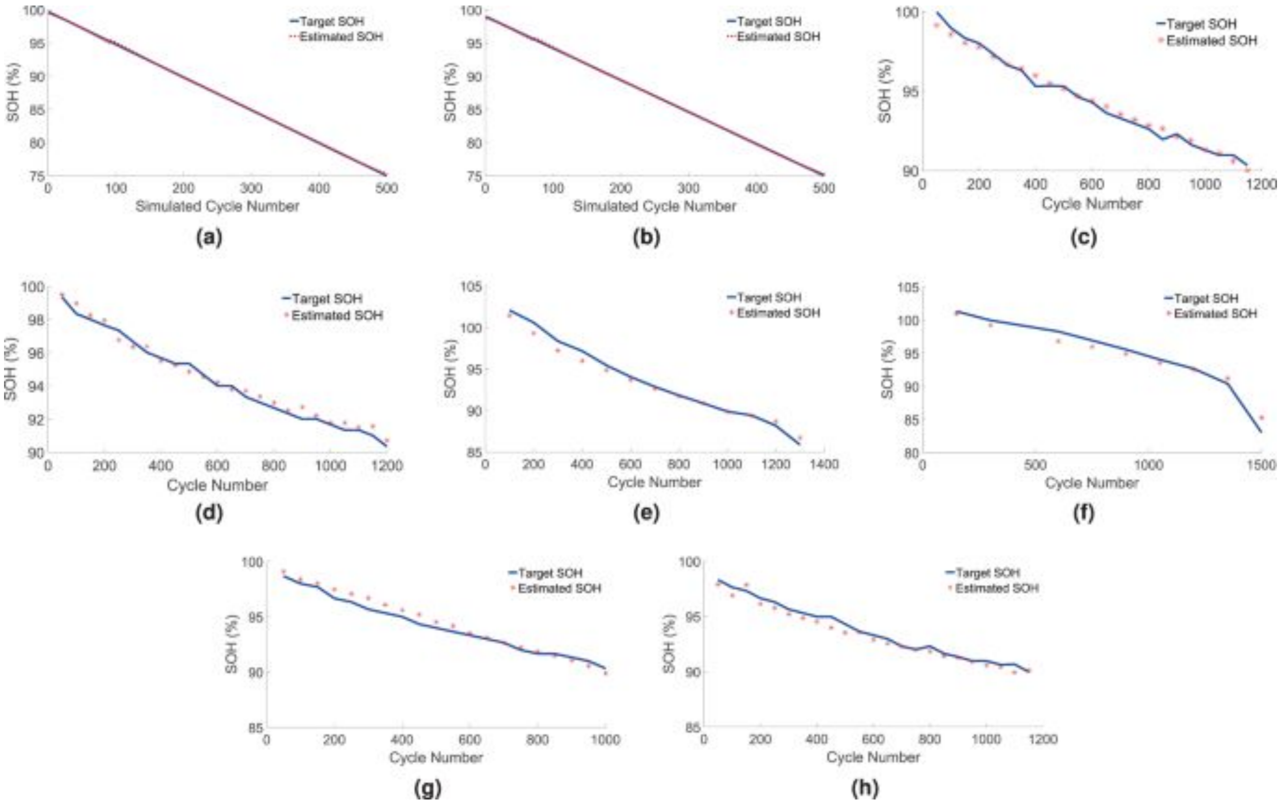
Best



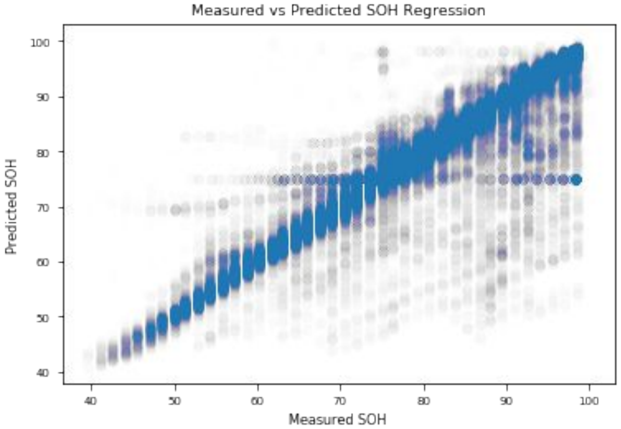
Best Pattern:

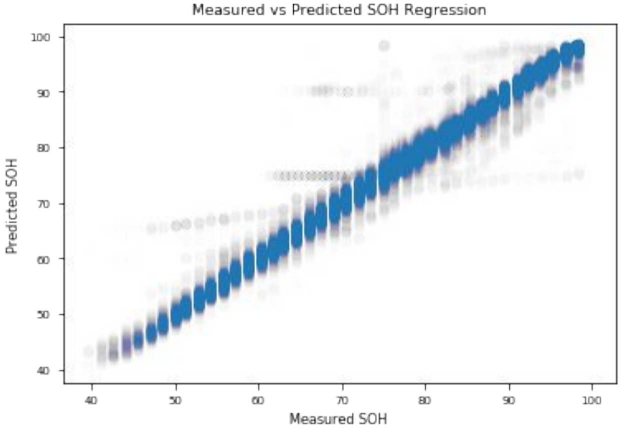
- Not Emission Level 0
- High usage of the electric mode and low usage of hybrid mode (healthy battery)

Estimating and predicting the State of Health (SOH) for batteries for hybrid buses



Estimating and predicting the State of Health (SOH) for batteries for hybrid buses





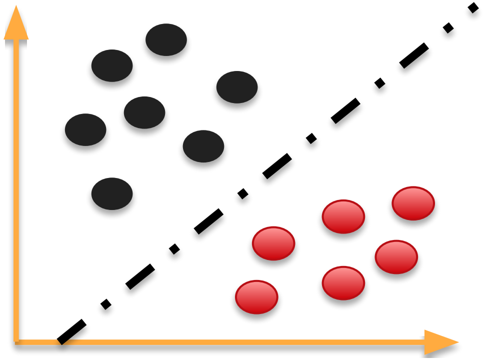
Metric	All (3212 Buses)	Monotonic-decreasing Function (2049 Buses)
MAE	2.60	1.04
R^2	0.81	0.98
Correlation	0.90	0.99

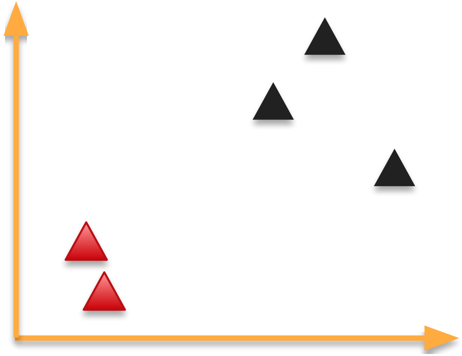
M. G. Altarabichi, Y. Fan, S. Pashami, S. Nowaczyk, and T. Rögnvaldsson. **Predicting State of Health and End of Life for Batteries in Hybrid Energy Buses**. ESREL-PSAM 2020 Conference, Italy, 1 - 6 November 2020.

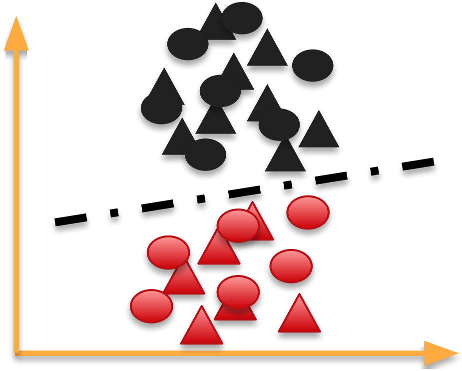
There are more challenges ...

- How to deal with a new system or a new configuration?
  - Can we avoid collecting a lot of new data?
  - E.g. Hybrid vs. full-electric, single vs. double deck
- How to deal with a different usage pattern?
  - E.g. deployment in a new country, long haul vs delivery vehicles
- How about diversity in operating conditions?
  - Do we need a new model for every possible operating condition?

Domain adaptation







Source domain

Target domain

Domain adaptation

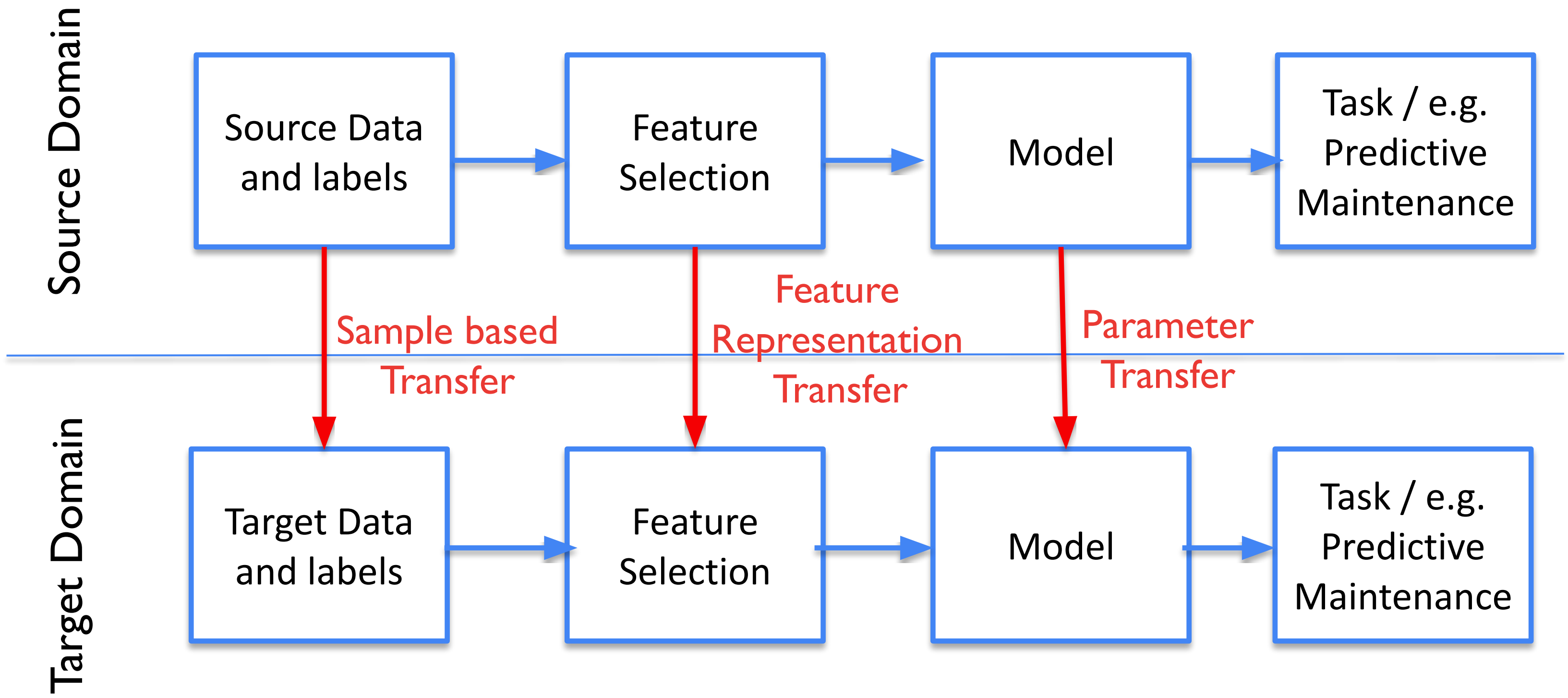
Where did we use transfer learning?

Hybrid buses

Modelling State of Health of Li-Ion Batteries

- Vehicle type
  - single decker, double decker, articulated
- Battery generation
  - Akasol, Samsung, …
- Countries
  - Sweden, UK, …
- Data driven grouping

Practical Approach - Domain adaptation



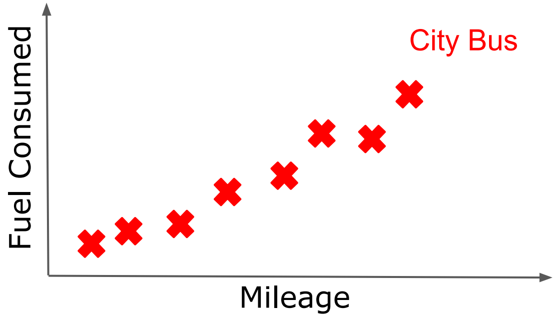
Sample Based Transfer Learning for Predictive Maintenance

- Selecting the right sample population for modelling
    - Physical configurations
    - Deployment areas
    - Operating conditions
    - Usage patterns
    - …
- Finding homogeneous sub-fleets based on similarity
    - Learning the representation to capture the characteristics of the equipment
    - Grouping peers with clustering and a similarity metric



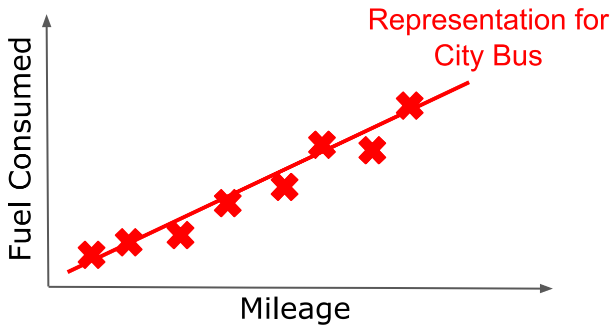
Sample Based TL for Predictive Maintenance

- A data-driven approach for finding the sub-fleet cluster from the whole population
    - Learning the representation of different equipment from sensor data



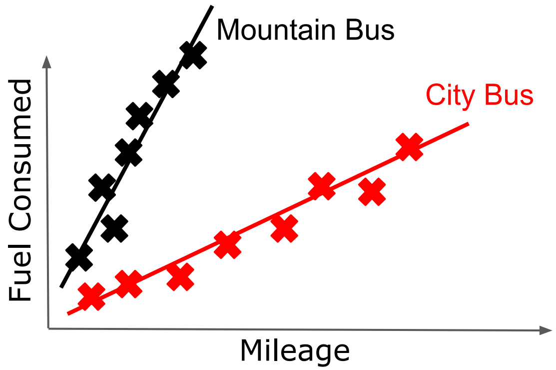
Sample Based TL for Predictive Maintenance

- A data-driven approach for finding the sub-fleet cluster from the whole population
  - Learning the representation of different equipment from sensor data



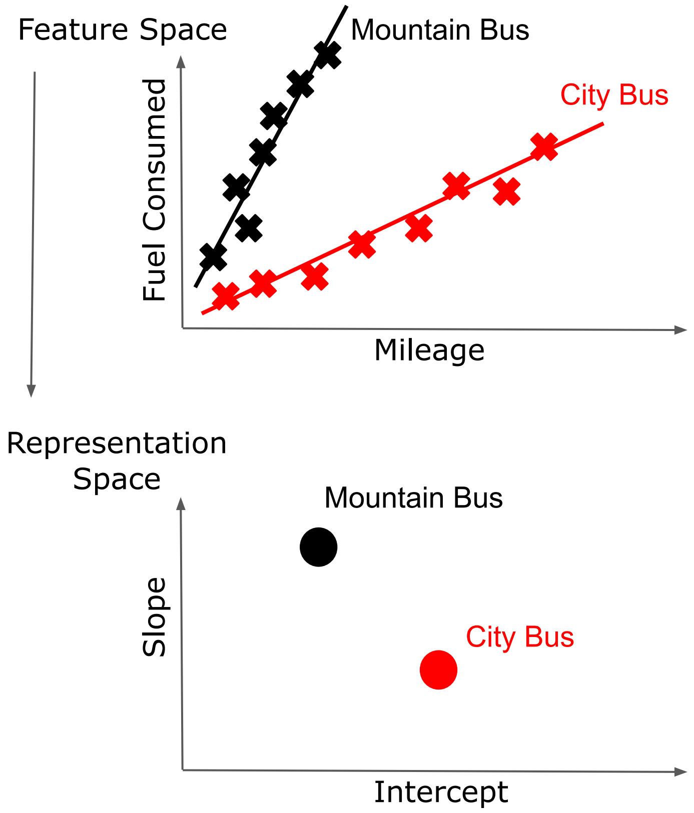
Sample Based TL for Predictive Maintenance

- A data-driven approach for finding the sub-fleet cluster from the whole population
  - Learning the representation of different equipment from sensor data



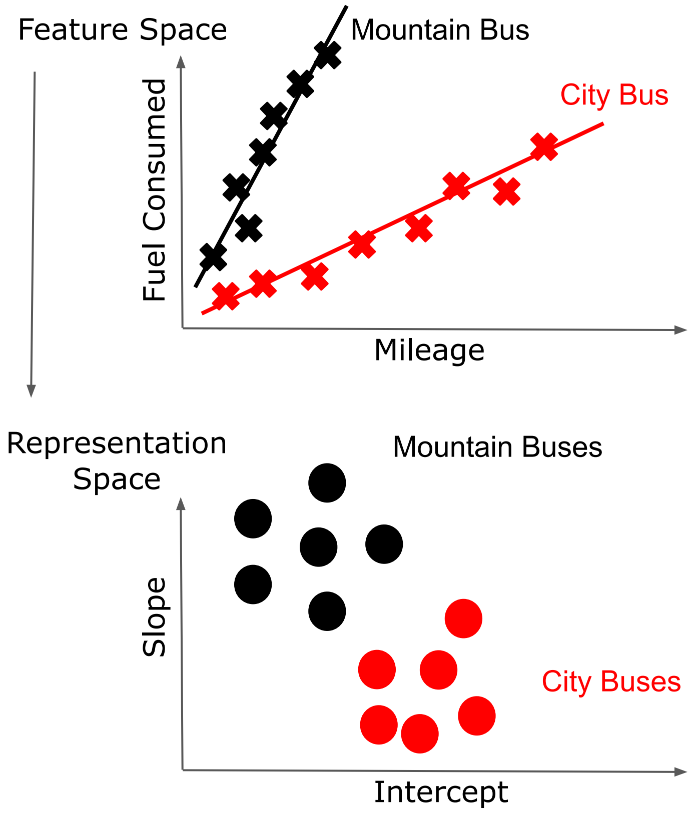
Sample Based TL for Predictive Maintenance

- A data-driven approach for finding the sub-fleet cluster from the whole population
  - Learning the representation of different equipment from sensor data
  - Finding clusters of peers in the representation space



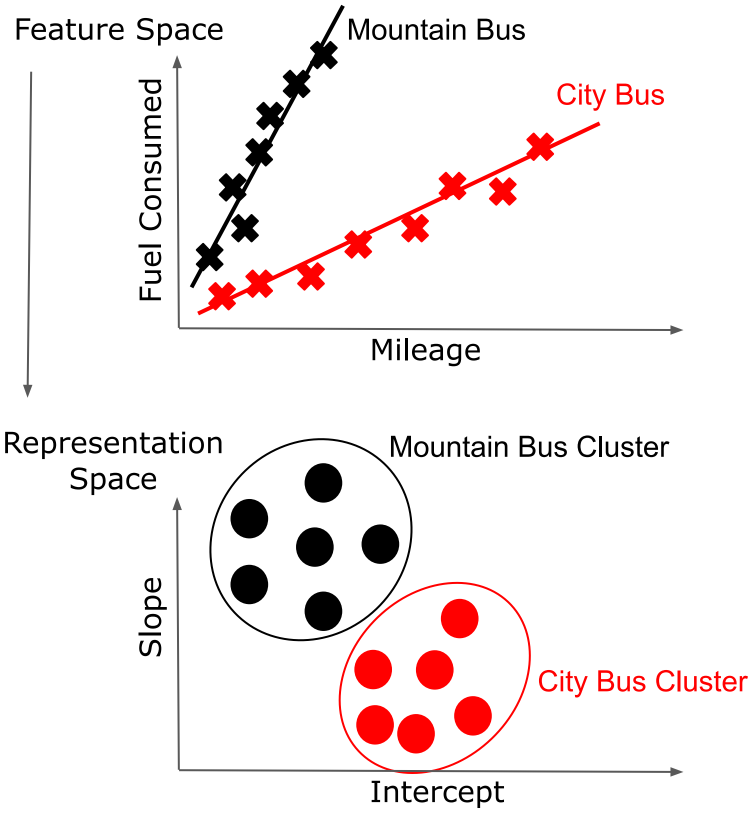
Sample Based TL for Predictive Maintenance

- A data-driven approach for finding the sub-fleet cluster from the whole population
  - Learning the representation of different equipment from sensor data
  - Finding clusters of peers in the representation space



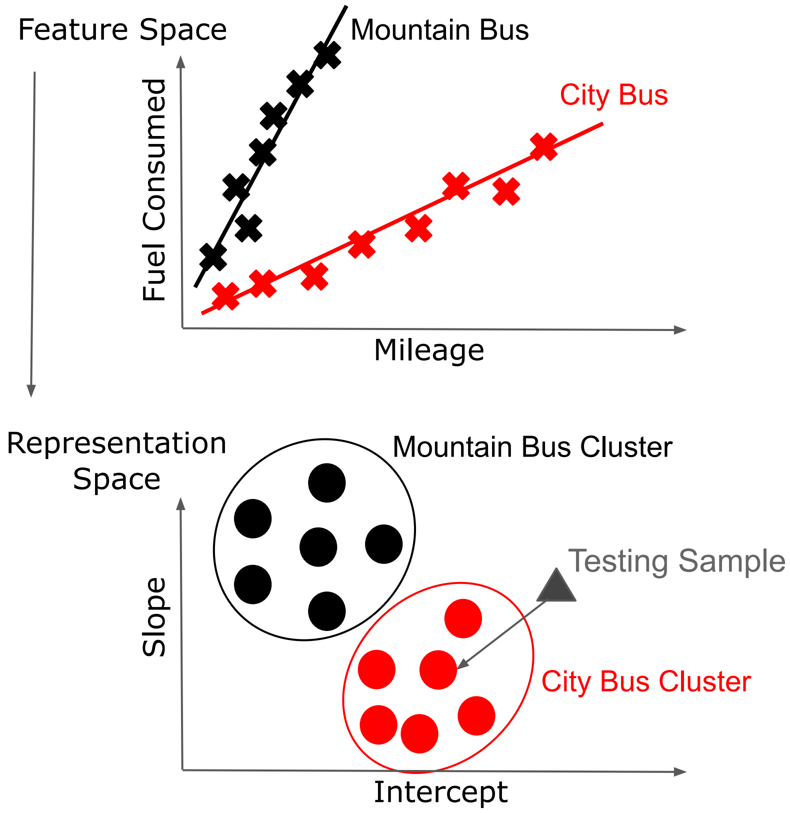
Sample Based TL for Predictive Maintenance

- A data-driven approach for finding the sub-fleet cluster from the whole population
  - Learning the representation of different equipment from sensor data
  - Finding clusters of peers in the representation space
  - Modelling the health indicator for each cluster



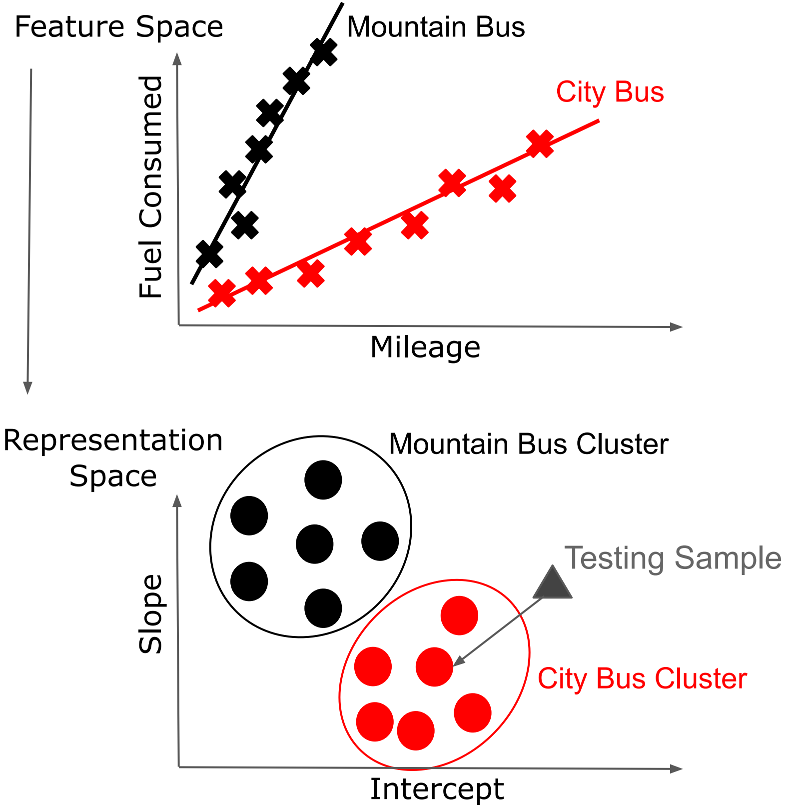
Sample Based TL for Predictive Maintenance

- A data-driven approach for finding the sub-fleet cluster from the whole population
  - Learning the representation of different equipment from sensor data
  - Finding clusters of peers in the representation space
  - Modelling the health indicator for each cluster
  - Matching the testing unit to a cluster and cast predictions on the health indicator



Sample Based TL for Predictive Maintenance

- A data-driven approach for finding the sub-fleet cluster from the whole population
  - Learning the representation of different equipment from sensor data
  - Finding clusters of peers in the representation space
  - Modelling the health indicator for each cluster
  - Matching the testing unit to a cluster and cast predictions on the health indicator
- Configuring the method based on the context information
  - Selecting the representation based on the characteristics of the features (linear or piecewise linear for accumulative variables)
  - The characteristics of battery deterioration varies between different generations



Results of Forecasting SOH

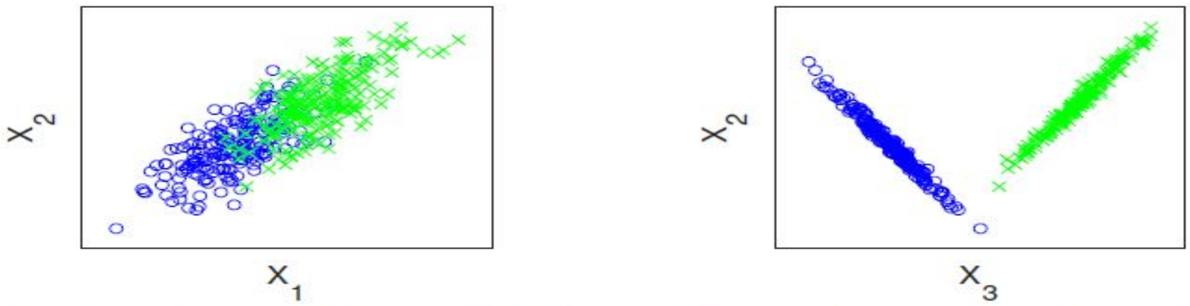
- Given five years of historical data
  - Predict SOH value at the end of the fifth year, given four and a half year of data for training

Methods	Training set vehicles (four years and half data included in the training process)	Leave-out set vehicles (completely excluded from the training process)
Random Forest (conventional approach)	3.2258 ± 0.1168	3.6935 ± 0.3017
FP linear model, k-means clustering, Random Forest	2.2137 ± 0.0826	2.5629 ± 0.322

Feature Based Transfer Learning for Predictive Maintenance

The Problem with Identifying Invariant Features

- ML faces a significant challenge in dynamically evolving environments, where the training conditions (domain) are different from the testing conditions.
- Aim is to identify features that are invariant across different domains.



Magliacane et al, 2018. Domain adaptation by using causal inference to predict invariant conditional distributions. In *Advances in Neural Information Processing Systems* (pp. 10846-10856)

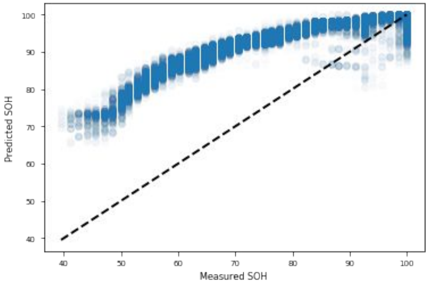
Identification of Invariant Features using GA

- We propose to use a Genetic Algorithm (GA) to select invariant features to transfer across multiple source domains D_S.
- We make a similar assumption to (Magliacane et al, 2018): if a feature subset is invariant across all source domains, then this holds in the target domain.
- The GA is initiated with a population of individuals encoding feature subsets as chromosomes of binary strings.
- The GA evaluates feature subsets according to their performance across all available source domains.

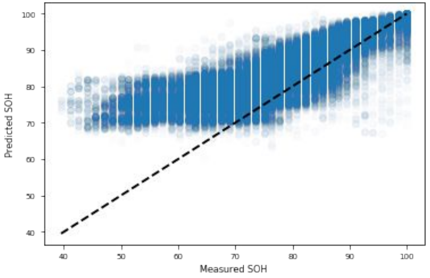
An Application to Li-Ion Batteries

- Our modeling of Li-Ion batteries showed that the hybrid bus battery deterioration processes vary significantly across different bus configuration and operating conditions.
- The GA is used to select invariant features to SoH (State of Health) that can be transferred from source to target domain.
- Our preliminary results identified invariant features under change of:
    - Battery Generation.
    - Chassis Type.
    - Operating Country.

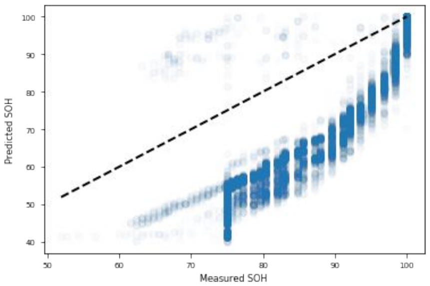
Invariant Feature Selection (Battery Generation)



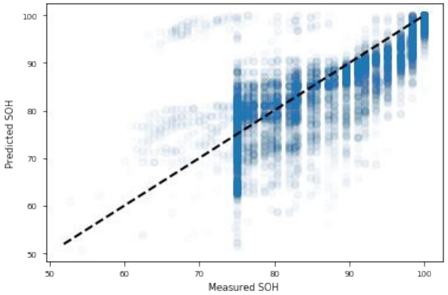
Before Invariant Feature Selection, MAE: **17.04%**



After Invariant Feature Selection, MAE**: 8.78%**



Before Invariant Feature Selection, MAE: **19.63%**



After Invariant Feature Selection, MAE**: 4.84%**

Altarabishi, M.G., et al. Extracting feature selection for predicting State of health of batteries in hybrid buses. DSAA 2021.

Experimental setting

- 1500 hybrid buses
- Different physical configurations
  - Double-Decker, Single-Decker, Articulated
- D
  - C

Scenario #	Source Domain D_S	Target Domain D_T
1	Moderate, Fast	Slow
2	Slow, Fast	Moderate
3	Slow, Moderate	Fast
4	Double-decker, Articulated	Single-decker
5	Single-decker, Articulated	Double-decker
6	Single-decker, Double-decker	Articulated

Performance comparison on all six scenarios

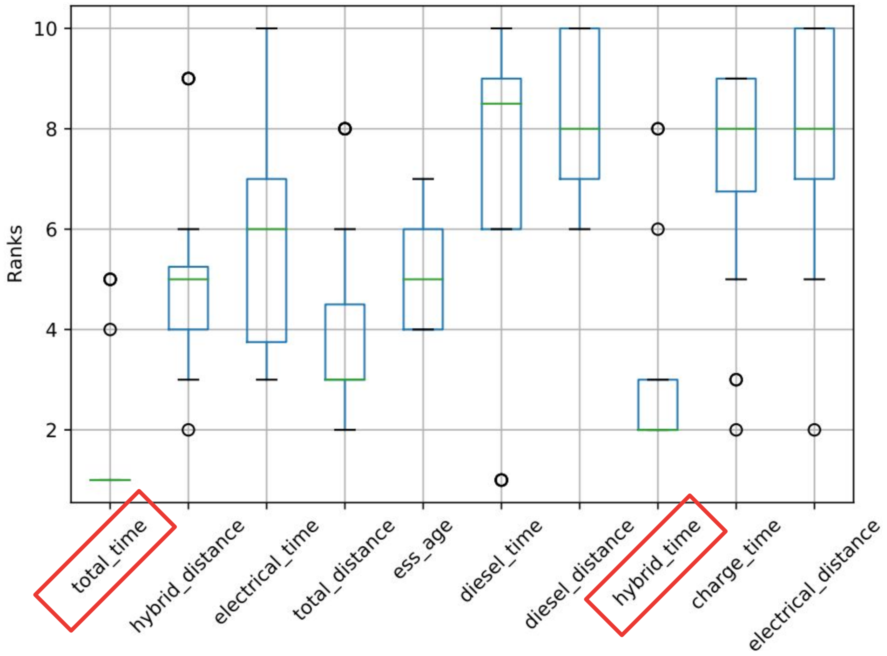
	Slow	Moderate	Fast
GA*	1.50 ± 0.01	1.59 ± 0.01	1.88 ± 0.01
GADIF	1.54 ± 0.02	**1.65 ± 0.01**	**1.96 ± 0.02**
Pearson	1.54 ± 0.06	1.74 ± 0.05	2.02 ± 0.01
RF	1.62 ± 0.07	1.72 ± 0.07	1.99 ± 0.04
LR	1.75 ± 0.06	1.81 ± 0.07	2.12 ± 0.02
SFS	1.54 ± 0.07	1.73 ± 0.07	2.03 ± 0.09
XGB	1.54 ± 0.07	1.74 ± 0.05	1.99 ± 0.11
All Features	**1.53 ± 0.07**	1.72 ± 0.09	2.06 ± 0.14

	Single-decker	Double-decker	Articulated
GA*	2.05 ± 0.00	1.53 ± 0.01	1.48 ± 0.01
GADIF	2.27 ± 0.03	**1.53 ± 0.01**	**1.67 ± 0.03**
Pearson	**2.16 ± 0.12**	1.54 ± 0.03	2.08 ± 0.16
RF	2.19 ± 0.14	1.54 ± 0.03	1.78 ± 0.10
LR	2.29 ± 0.09	1.67 ± 0.01	2.07 ± 0.41
SFS	2.19 ± 0.12	1.60 ± 0.03	2.10 ± 0.15
XGB	2.16 ± 0.12	1.57 ± 0.02	2.08 ± 0.16
All Features	2.19 ± 0.21	1.68 ± 0.05	2.07 ± 0.18

Ranking of features

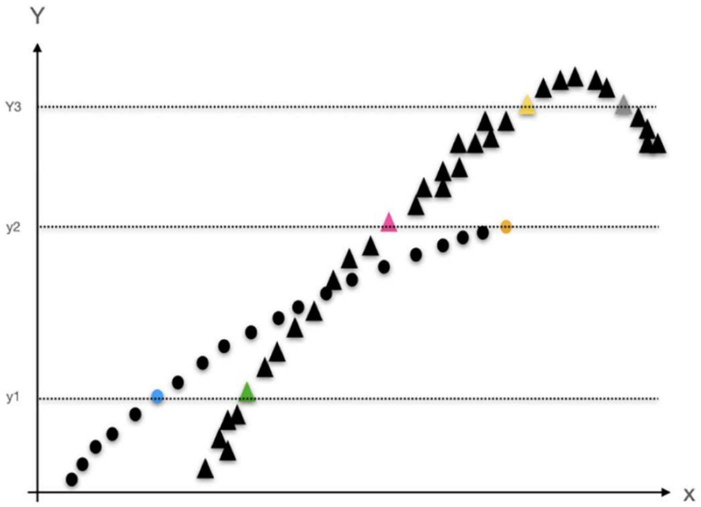
”**total time**” and ”**hybrid time**” are of the highest ranked two features

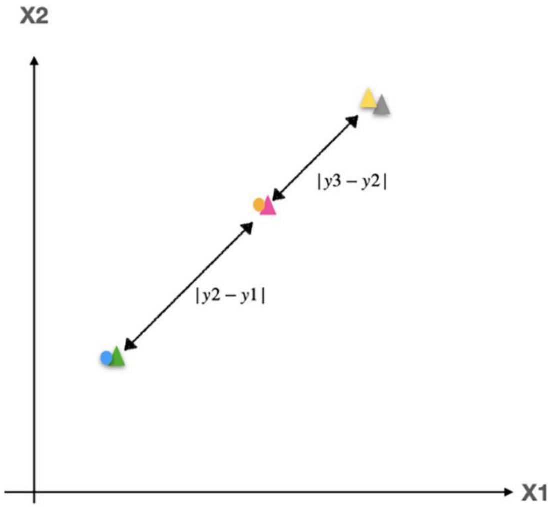
- followed by ”**hybrid distance**”, ”**electrical time**”,”**total distance**” and ”**ESS age**”



Domain Adaptation for Predictive Maintenance

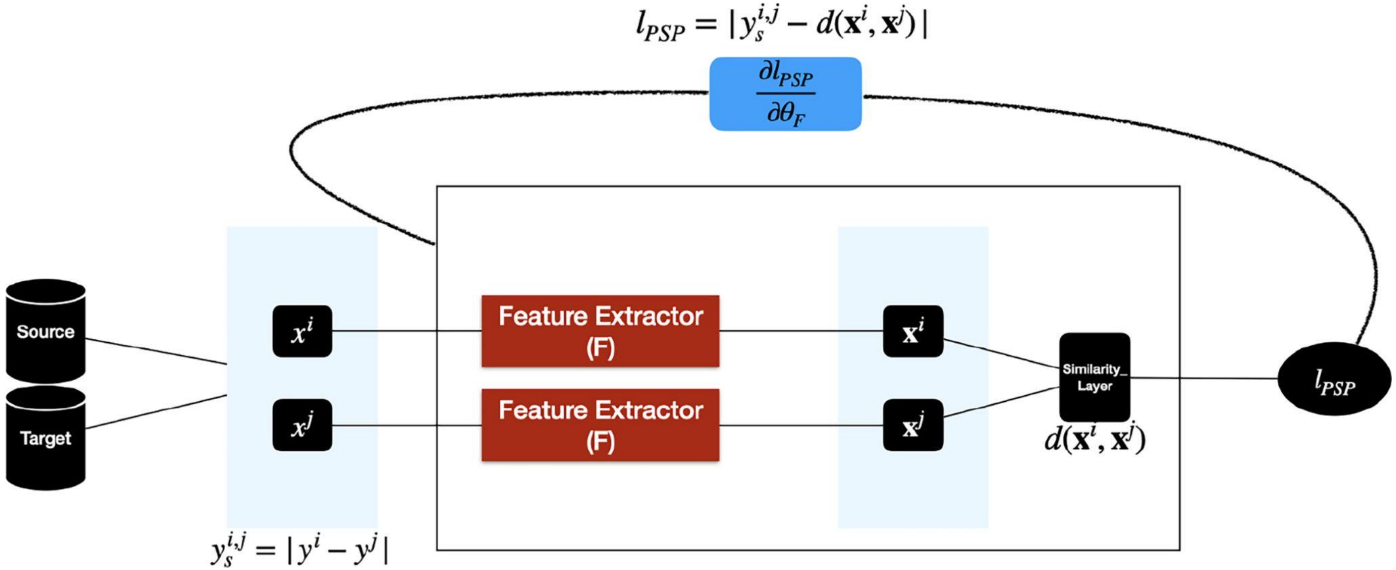
Multi-domain adaptation for regression under conditional distribution shift





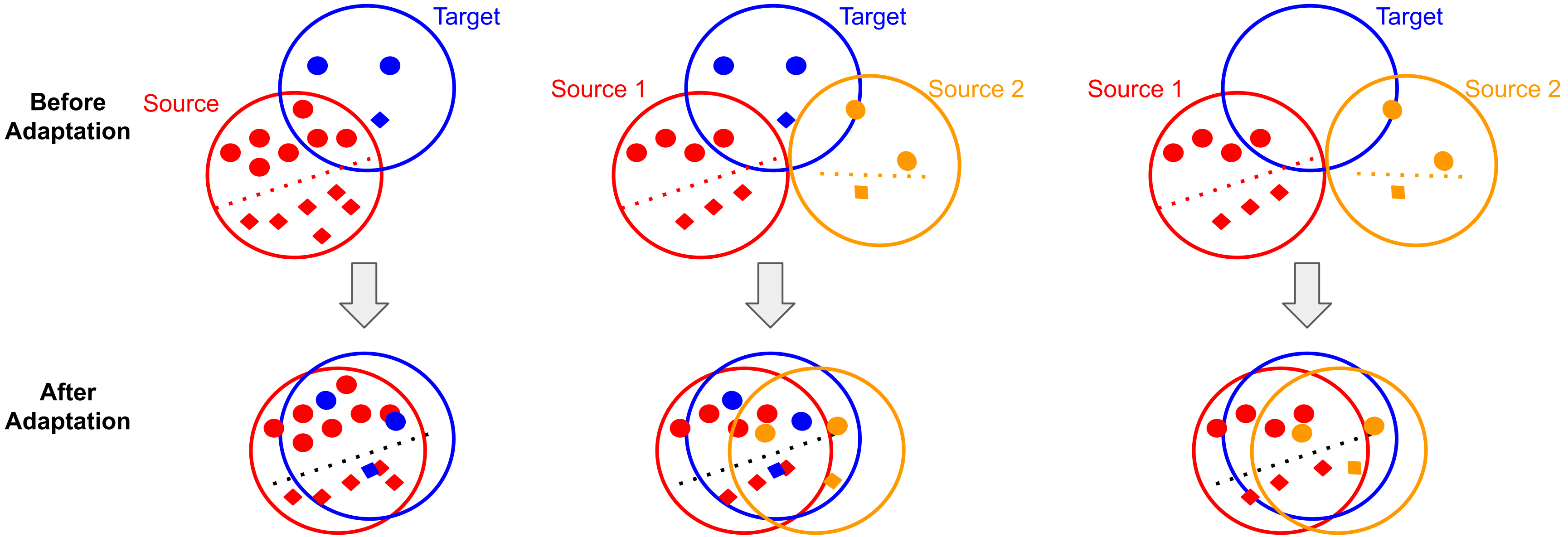
Taghiyarrenani et al. Expert system with application. 2023

Multi-domain adaptation for regression under conditional distribution shift



Taghiyarrenani et al. Expert system with application. 2023

Problem Formulation

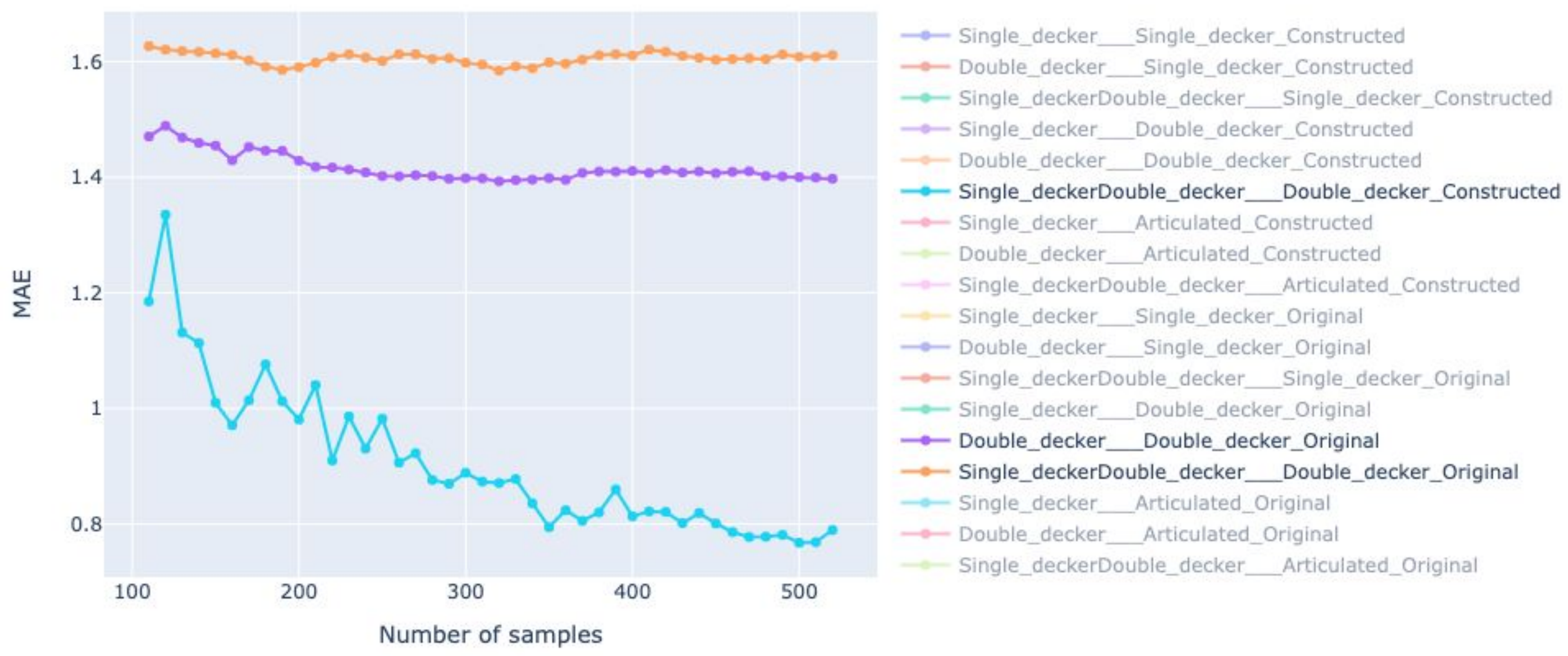


Two domains
- semi-supervised
- enough data in source
- little data in target

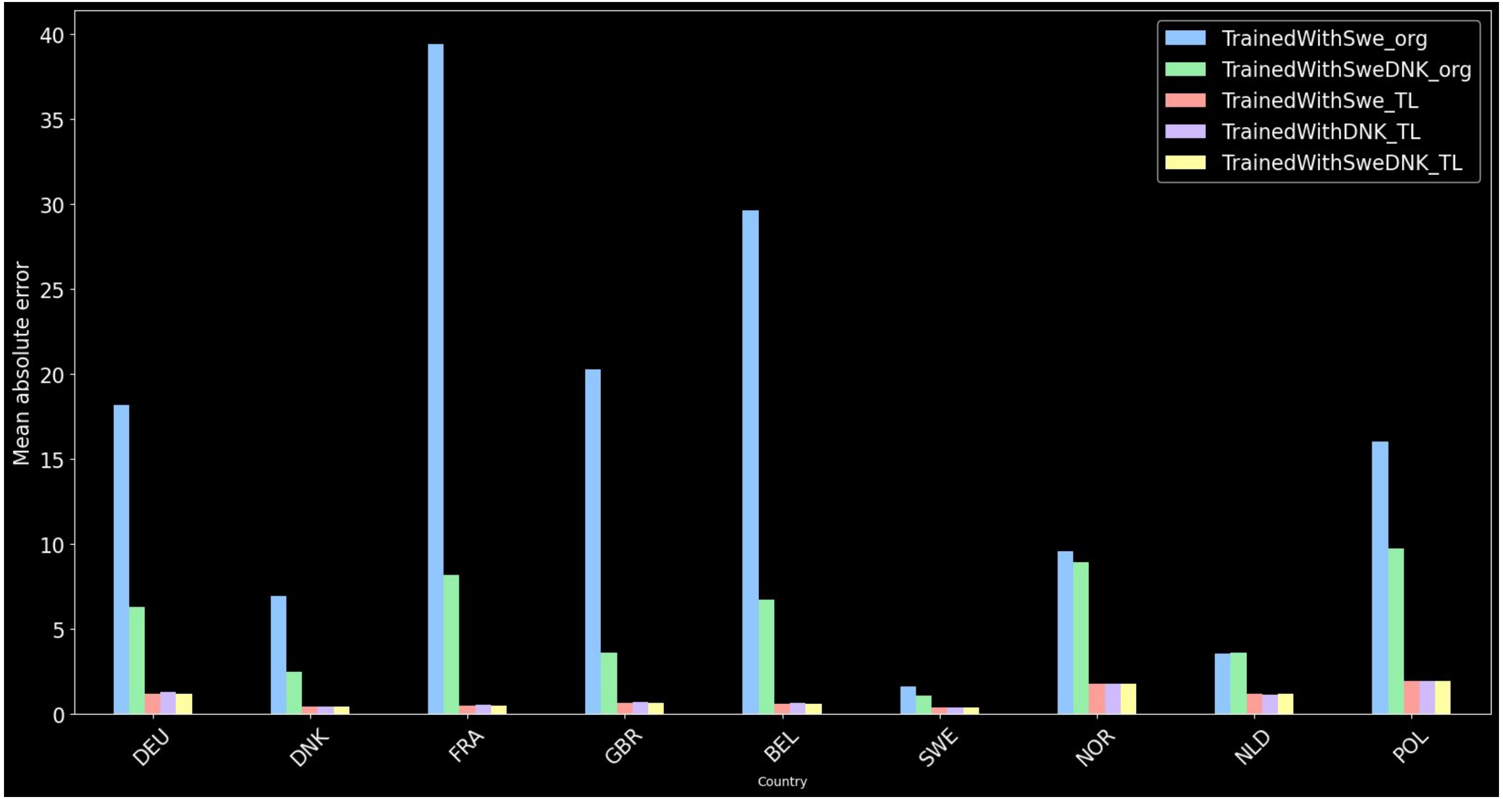
Multiple domains
- Supervised
- Little data in all

Domain Generalisation
- No data in target

Transfer learning from single decker to double decker



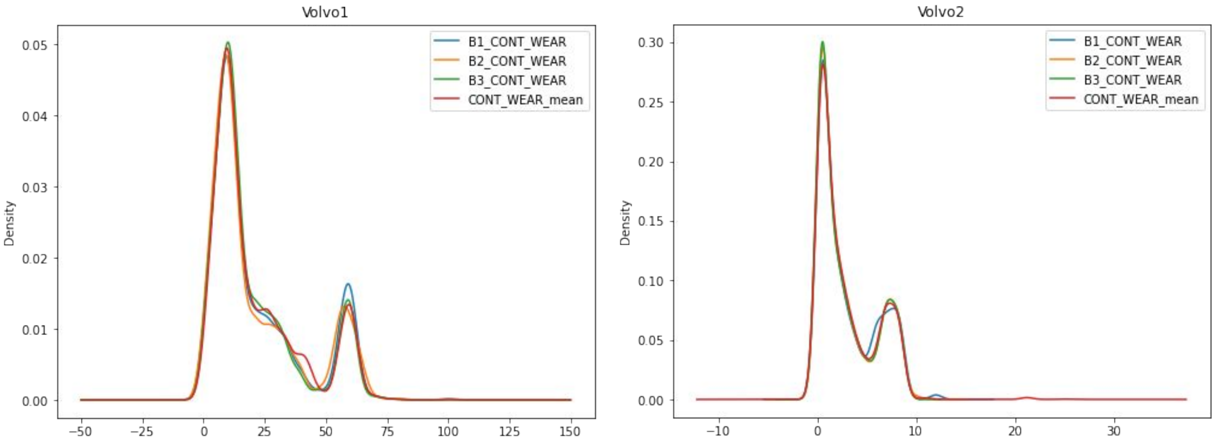
SOH Models trained on Sweden & Norway transfer well to other countries.

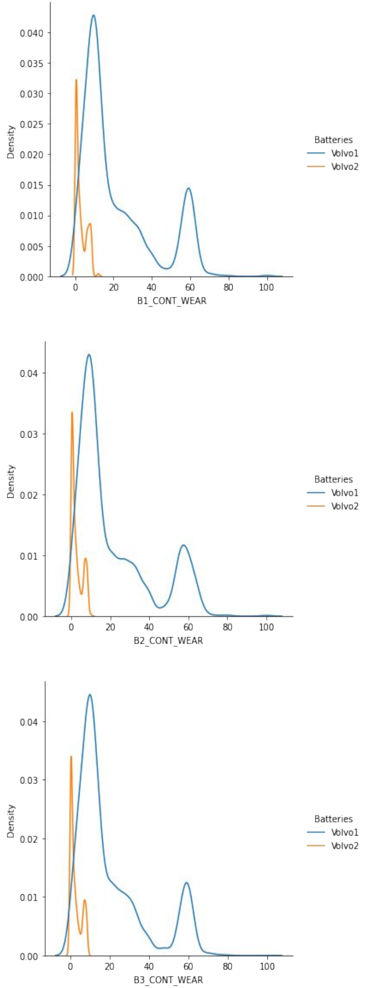


DA doesn't always make sense!

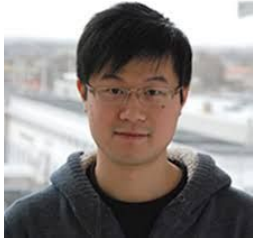
Comparison between Cotactor Types: 1 and 2

Comparison between on-board cotactors





Thank you :)



Yuantao Fan, HH

Klas Thunberg, VBC

Jens Lundström, VGCS

Mohammed Ghaith Altarabichi, HH

Zahra Taghiyar, HH

Daniel Reimhult, VGCS

Sławomir Nowaczyk, HH

And, other colleagues at HH, VBC, VGCS, ...