

Unsupervised domain adaptation for dynamic in weighing motion system of freight rail carriages under varying ballast conditions

Francesco Cannarile

Aramis s.r.l., Milano, Italy. E-mail: francesco.cannarile@aramis3d.com

Michele Compare

Aramis s.r.l., Milano, Italy. E-mail: michele.compare@aramis3d.com

Energy Department, Politecnico di Milano, Italy

Enrico Zio

Aramis s.r.l., Milano, Italy

Energy Department, Politecnico di Milano, Italy. E-mail: enrico.zio@polimi.it

MINES ParisTech, PSL Research University, CRC, Sophia Antipolis, France

Eminent Scholar, Department of Nuclear Engineering, College of Engineering, Kyung Hee University, Republic of Korea

Elisa Duca, Ilario Febi, Marco Renzetti, Massimo Platini

Generali Costruzioni Ferroviarie, Unità Operativa di Rozzano, Via Brodolini 1, Rozzano (MI), Italy

We have recently developed a novel Dynamic Weight In Motion (DWIM) system for freight carriages, which is based on a combination of Discrete Fourier Transform (DFT) and ELastic NET (ELNET) linear regression. To improve this method, we propose an unsupervised domain adaptation method based on subspace alignment. This learns a mapping function to align the features extracted from data of carriages of known load in a given DWIM setting with those extracted from carriages of unknown load in another positions. The application of the proposed method provides promising results.

Keywords: Railway industry; Dynamic Weight In Motion; domain adaptation, subspace alignment.

1. Introduction

Developing effective train weighing systems is of great interest in the railway industry, as accurate load estimates can be used by railroad companies for checking overloads, identifying potentially unsafe carriages and improving railway maintenance planning. Nowadays, train weight measurements are typically taken by static systems, which require the trains to stop on the weight scale. Thus, static weighing is time-demanding, with adverse consequences on the wheeling of the network, the capacity of delivering trains from intermodal terminals for good shipment, etc.

On the contrary, Dynamic Weight in Motion (DWIM) systems avoid carriages stopping, with significant benefits to circulation timing and business (Poulikakos et al. (2008); Mayer et al. (2012); Allotta et al. (2015)).

We have recently developed a novel Dynamic Weight In Motion (DWIM) system for freight carriages. This is based on a combination of Discrete Fourier Transform (DFT) and ELastic NET (ELNET) linear regression (Cannarile et al.

(2020)) and provides very accurate carriage load estimates. However, the applicability of the developed DWIM method is limited to the railway branch where data are collected from: the ballast therein strongly influences the system response to loads and, thus, makes the trained model not applicable to other branches, where training and test data are drawn from different probability distributions.

On the other hand, the solution of training a new DWIM model on the data collected from the specific branch where sensors installed is not practicable, given the obvious difficulties in obtaining the ground truth loads (e.g., by installing a static weighing system).

To extend the applicability of the DWIM system trained on a railway branch to other branches, we investigate the use of an unsupervised domain adaptation method based on subspace alignment (Fernando et al. (2013)). This learns a mapping function aligning the subspace generated by the first \tilde{K} Principal Components (PCs) extracted from data of carriages of known load in a given

DWIM setting (source domain) with those extracted from carriages of unknown load in another location (target domain).

The choice of relying on subspace alignment based on PCA to perform unsupervised domain adaptation is mainly justified by the difficulty in obtaining large training sets of labelled data. In these conditions, recent techniques based on Deep Neural Networks (DNN) cannot be properly exploited, as these require a large number of labelled data for effective training (Kouw and Loog (2019)).

After the alignment, the DWIM system mapping the aligned features of a carriage with the corresponding load, is developed by training an Elastic NET (ELNET) regression model (Zou and Hastie (2005); Cannarile et al. (2019)) on the aligned source data. This can be used to infer the loads of carriages of unknown load in the target domain. The application of the proposed method provides promising results.

The paper is organized as follows: Section 2 states the problem; Section 3 details the proposed solution. The application of the methodology to real data is described in Section 4, whereas Section 5 draws the work conclusions.

2. Problem Statement

We consider a standard supervised learning setting with available training dataset $D_S = \{\mathbf{x}_{S_i}, y_i\}_{i=1, \dots, N_S}$, containing the vectors $\mathbf{x}_{S_i} \in \mathcal{X} \subseteq \mathbb{R}^K$ of K features extracted from strain signals recorded from the i^{th} carriage (hereafter referred to as training instance), and the corresponding labelled load $y_i \in \mathcal{Y} \subseteq \mathbb{R}_0^+$. To build a DWIM system, we have to find a regression function

$$y = f(\mathbf{x}) \quad (1)$$

that associates to any test example \mathbf{x}_{test} from a carriage of unknown load, the corresponding output load $y_{test} = f(\mathbf{x}_{test})$.

We have developed a method to build a DWIM system in the case in which the test data \mathbf{x}_{test} are gathered from the the same sensors installed on the same rail branch of the trainig dataset D_S .

However, this DWIM system may not perform well on test data $D_T = \{\mathbf{x}_{T_i}\}_{i=1, \dots, N_T}$ collected from a different rail branch, as the ballast therein strongly influences the system response to loads. Formally, we can state that training and test data from different rail branches may be drawn from different probability distributions over the same feature-label space pair $\mathcal{X} \times \mathcal{Y}$. These distributions are referred as *source domain* and *target domain*, indicated by $p_S(\mathbf{x}, y)$ and $p_T(\mathbf{x}, y)$, respectively. Our goal is to develop a DWIM system capable of predicting the labels of test instances (i.e., carriages) drawn from a target domain given the

labelled samples drawn from a source domain and unlabeled samples drawn from the target domain itself. This particular learning setting is known as unsupervised domain adaptation.

Notice that the term unsupervised refers to the fact that we assume that labelled instances in the target domain are lacking.

3. Proposed Solution

Our solution relies on the assumption that the different domains contain both domain-specific noise and common feature subspaces. Adaptation, then, consists in finding these common feature subspaces and, thus, mapping the source data to the target data.

In details, we rely on the subspace alignment method proposed in Fernando et al. (2013), which seeks a domain invariant feature space by learning a mapping function that aligns the source subspace with the target one. The proposed solution is pictorially shown in Figure 3.

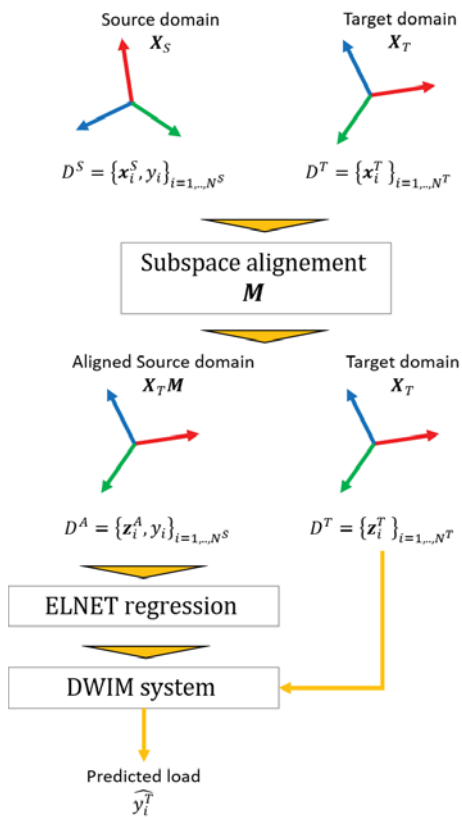


Fig. 1. The proposed unsupervised domain adaptation based DWIM system

We first construct a feature subspace of size

$\tilde{K} < K$, composed of the \tilde{K} most important eigenvectors induced by Principle Component Analysis (PCA). Then, we directly reduce the discrepancy between the two domains by moving the source and target subspaces closer. This is achieved by optimizing a mapping function that transforms the source subspace into the target one. In details, let \mathbf{X}_S and \mathbf{X}_T be the matrices of size $K \times \tilde{K}$ formed by the first \tilde{K} eigenvectors of the covariance matrices estimated from z-normalized source and target data D^s and D^T , respectively. Alignment is achieved by learning a $\tilde{K} \times \tilde{K}$ matrix M such that the aligned source subspace $\mathbf{X}_A = \mathbf{X}_S M$ is as close as possible to the target subspace, from the Bergman divergence perspective (Bregman (1967)). In details, we solve the following minimization problem:

$$M^* = \operatorname{argmin}_M F(M) \quad (2)$$

where we use the Frobenius norm $F(M)$ as Bregman divergence metrics Bregman (1967); Fernando et al. (2013); Gloub and Van Loan (1996):

$$F(M) = \|\mathbf{A}\|_F^2 = \operatorname{Tr}(\mathbf{A}\mathbf{A}^H) \quad (3)$$

where $\mathbf{A} = \mathbf{X}_S M - \mathbf{X}_T$ and \mathbf{A}^H is its conjugate transpose. The Frobenius norm is invariant to orthonormal matrix multiplications. Then, Eq. 3 can be re-written as

$$\begin{aligned} F(M) &= \left\| \mathbf{X}_S^T \mathbf{X}_S M - \mathbf{X}_S^T \mathbf{X}_T \right\|_F^2 \\ &= \left\| M - \mathbf{X}_S^T \mathbf{X}_T \right\|_F^2 \end{aligned} \quad (4)$$

where \mathbf{X}_S^T denotes the transpose of matrix \mathbf{X}_S . From Eq. 4 we have that the minimization problem in Eq. 2 is solved when $M = \mathbf{X}_S^T \mathbf{X}_T$. This entails that the new target aligned source coordinate system is equivalent to $\mathbf{X}_A = \mathbf{X}_S \mathbf{X}_S^T \mathbf{X}_T$. Then, the source data are projected onto the target aligned subspace, \mathbf{X}_A , obtaining the transformed source data $\tilde{D}_A = \{z_{Ai}, y_i\}_{i=1, \dots, N_S}$, whereas the target data are projected onto the target subspace \mathbf{X}_T obtaining the transformed target data $\tilde{D}_T = \{z_{Ti}\}_{i=1, \dots, N_T}$:

$$z_{Ai} = \mathbf{x}_{Si} \mathbf{X}_S \mathbf{X}_S^T \mathbf{X}_T, \quad i = 1, \dots, N_S \quad (5)$$

$$z_{Ti} = \mathbf{x}_i \mathbf{X}_T, \quad i = 1, \dots, N_T \quad (6)$$

3.1. Elastic net linear regression

We learn an elastic-net linear regression model on labelled aligned source data D_A . The developed

regression model is then used to estimate the unknown load \hat{y}_{Ti} of target carriage z_{Ti} . Namely, the aligned source data $D_A = \{z_{Ai}, y_i\}_{i=1, \dots, N_S}$ are used to estimate the parameter of a linear model which is assumed to relate the load y_i with feature vector z_{Ai} ,

$$y_i = \beta_0 + \sum_{k=1}^{\tilde{K}} z_{Ai} [k] \beta_k + \epsilon_i \quad (7)$$

where $\beta = [\beta_1, \dots, \beta_{\tilde{K}}]^T$ are unknown model parameters and ϵ_i is an error term. These parameters are estimated solving the following regularized learning problem

$$\operatorname{argmin}_{\{\beta_0, \beta\}} L(\beta_0, \beta) + \lambda [\alpha P_1(\beta) + (1 - \alpha) P_2(\beta)] \quad (8)$$

where

$$L(\beta_0, \beta) = \frac{1}{2N_S} \sum_{i=1}^{N_S} \left(y_i - \beta_0 - \sum_{k=1}^{\tilde{K}} z_{Ai} [k] \beta_k \right)^2 \quad (9)$$

and where $P_1(\cdot)$ and $P_2(\cdot)$ denote the LASSO and the ridge penalties, respectively (Zou and Hastie (2005); Friedman et al. (2010)):

$$P_1(\beta) = \|\beta\|_1 = \sum_{q=1}^Q |\beta_q| \quad (10)$$

$$P_2(\beta) = \|\beta\|_2^2 = \sum_{q=1}^Q |\beta_q|^2 \quad (11)$$

The effect of the LASSO penalty is that of performing automatically feature selection, by shrinking some coefficients β_k towards zero. This enhances the accuracy and interpretability of the statistical model. The effect of the ridge penalty is that of limiting the magnitude of the model parameters, thus preventing the model to overfit the training data.

In Eq. 8 the hyperparameter $\lambda > 0$ is a user-specified regularization parameter: the smaller its value, the larger the level of regularization imposed (Zou and Hastie (2005); Friedman et al. (2010)).

The hyperparameter $\alpha \in [0, 1]$ is a mixing parameter, which trades off the importance of the LASSO and ridge penalties.

These hyperparameters can be set using a robust cross validation procedure as described in (Canarile et al. (2019); Friedman et al. (2010)).

Finally, the load of a new test example in the target domain z_{Ti} is predicted as

$$\hat{y}_{Ti} = \hat{\beta}_0 + \sum_{k=1}^{\hat{K}} z_{test}[k] \hat{\beta}_k \quad (12)$$

where $\{\hat{\beta}_0, \hat{\beta}\}$ are the parameter estimates resulting from the minimization of the convex program in Eq. 7.

4. Case Study

We have available real data from $N_S = 192$ different carriages and from $N_T = 49$ carriages. These have been collected from pairs of strain sensors installed, respectively, on a freight intermodal terminal near Milano (terminal A), and a train workshop facility close to Alessandria, Italy (terminal B).

In both terminals, the measuring facility is made up of optical sensors installed on a recently tapered straight railtrack section with standard and constant gauge. The section is 10 m long, with proper stiffness and no impairments. The optical sensors are connected by fiber optics and their data are acquired at sampling frequency $f_{rate} = 1KHz$. The acquisition system is a depolarized MOI Hyperion si255, controlled by a CPU; this runs algorithms that identify the train passing on the rail and cut the corresponding signal with 10s time margin before and after.

In this case study, we know the ground truth load of the carriages in both experimental setups. Nonetheless, to validate the proposed methodology we have considered the N_S carriages data collected in terminal A as the source data and the N_T carriages data collected in terminal B as the target data.

The strain signals have been pre-processed and $K = 256$ amplitude features have been extracted using Discrete Fourier Transform (DFT). This way, we obtain the source and target datasets D_S and D_T , respectively. For further details on feature extraction, please refer to our previous work (Cannarile et al. (2020)).

The extracted features have been scaled to have zero mean and unit variance. Yet, notice that the number of features extracted is larger than that of source and target instances. Then, PCA cannot be directly applied. For this, we have resorted to the truncated PCA, keeping the first $\hat{K} = N_T = 49$ principal components. This allows projecting the source and target data in their corresponding subspace using Eqs. 5 and 6, respectively. Finally, the parameters of the linear model have been estimated solving the regularized learning problem in Eq. 7 using only the aligned source data.

All computations have been performed using an Intel Core i7-7770 CPU at 3.60 GHz processor with 8 GB RAM in Python 3.6 environment.

M1	MRE
M2	0.065
	0.280

4.1. Results

Table 1 compares the performance of the proposed method (M1) with that obtained by applying directly the DWIM system based on ELNET linear regression without performing feature alignment (M2). That is, the DWIM system is trained using not the aligned data $D_S = \{x_{Si}, y_i\}_{i=1, \dots, N_S}$. As performance metric, we have reported the Mean Relative Error (MRE) resulting from the application of the developed DWIM systems on target data.

From Table 1, we cannot but conclude that the proposed method drastically improves the performance of a DWIM system, with respect to that applied directly to target data without performing feature alignment.

5. Conclusions

In this work, we have proposed an unsupervised domain adaptation method, which relies on subspace alignment to extend the applicability of a DWIM system to rail branches different from that where data have been collected to develop it. The proposed method has shown to drastically improve the performance of a DWIM system directly applied to target data without performing subspace feature alignment. The promising results pave the path to future investigation.

References

- Allotta, B., P. D'Adamio, L. Marini, E. Meli, L. Pugi, and A. Rindi (2015). A new strategy for dynamic weighing in motion of railway vehicles. *IEEE Transactions on Intelligent Transportation Systems* 16(6), 3520–3533.
- Bregman, L. M. (1967). The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR computational mathematics and mathematical physics* 7(3), 200–217.
- Cannarile, F., M. Compare, P. Baraldi, G. Diodati, V. Quaranta, and E. Zio (2019). Elastic net multinomial logistic regression for fault diagnostics of on-board aeronautical systems. *Aerospace Science and Technology* 94.
- Cannarile, F., M. Compare, E. Zio, E. Duca, I. Febi, M. Platini, and M. Renzetti (2020). A novel dynamic weighing in motion system for freight rail carriages. *Submitted for publication*.
- Fernando, B., A. Habrard, M. Sebban, and T. Tuytelaars (2013). Unsupervised visual domain adaptation using subspace alignment. pp. 2960–2967.

- Friedman, J., T. Hastie, and R. Tibshirani (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software* 33(1), 1.
- Gloub, G. H. and C. F. Van Loan (1996). Matrix computations. *Johns Hopkins University Press, 3rd edition*.
- Kouw, W. M. and M. Loog (2019). A review of single-source unsupervised domain adaptation. *CoRR abs/1901.05335*.
- Mayer, R., L. Poulidakos, A. Lees, K. Heutschi, M. Kalivoda, and P. Soltic (2012). Reducing the environmental impact of road and rail vehicles. *Environmental Impact Assessment Review* 32(1), 25–32.
- Poulidakos, L., M. Arraigada, G. Morgan, K. Heutschi, P. Anderegg, M. Partl, and P. Soltic (2008). In situ measurements of the environmental footprint of freight vehicles in Switzerland. *Transportation Research Part D: Transport and Environment* 13(4), 274–282.
- Zou, H. and T. Hastie (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 67(2), 301–320.