REGRESSING KERNEL DICTIONARY LEARNING

Kriti Kumar*, Angshul Majumdar** M Girish Chandra* and A Anil Kumar*

* TCS Research and Innovation, Bangalore, India

** IIIT Delhi, New Delhi, India

ABSTRACT

In this work, we present a kernelized dictionary learning framework for carrying out regression to model signals having a complex nonlinear nature. A joint optimization is carried out where the regression weights are learnt together with the dictionary and coefficients.

Relevant formulation and dictionary building steps are provided. To demonstrate the effectiveness of the proposed technique, elaborate experimental results using different real-life datasets are presented. The results show that non-linear dictionary is more accurate for data modeling and provides significant improvement in estimation accuracy over the other popular traditional techniques especially when the data is highly non-linear.

RATIONALE AND USEFULNESS

- We are living in the era of data deluge. Even though visual data has dominated/dominating this deluge, the equation is changing with huge data emanating from the Internet of Things (IoT)/ Machines.
- In order to understand the data and make effective use of them, it is necessary to have appropriate data-driven methods to capture the nature of data. With this understanding, one can carry out different inference tasks like, classification, clustering and regression.
- For any data analysis, it is necessary to identify dependent variables also known as responses or predicands, and independent variables or predictors. The relationship between the predictors and responses is described by a regression function.
- This function approximation approach is useful to model the data, to characterize different states of the data generating source.
- The changes can be further leveraged towards arriving at appropriate predictive and prescriptive analytics results.

PROPOSED METHOD

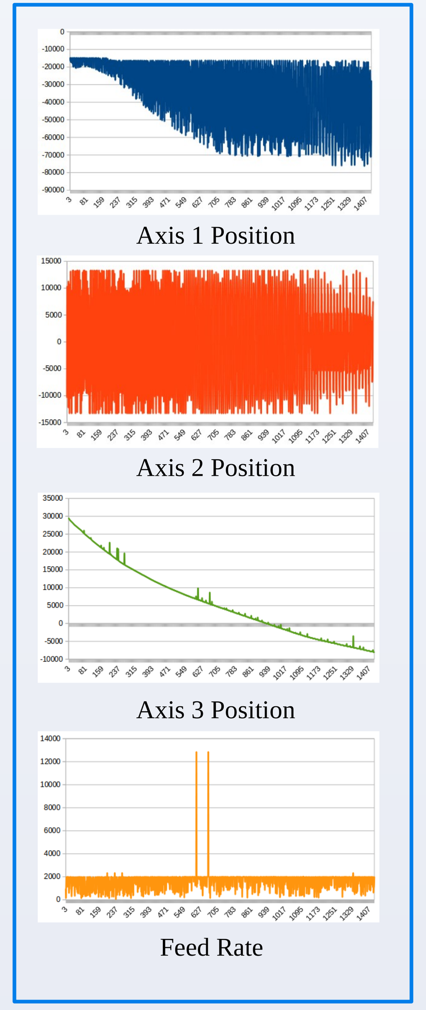
Given a multivariate data of N samples, let $X \in R^{L\times N}$ represents the independent variables of feature vector length L and $y \in R^{N}$ represent dependent variable. We propose to incorporate a ridge regression penalty into the kernel dictionary learning framework for carrying out a joint optimization where the dictionary atom, coefficients and the regression weights are learnt together.

Kernelization takes care of the non-linearities in the system and hence a simple linear regression formulation is sufficient after the transformation.

Mathematically, the proposed formulation is given as:

$$\min_{A,Z,w} \|\varphi(X) - \varphi(X)AZ\|_F^2 + \lambda\|y - wZ\|_2^2 + \mu\|w\|_2^2$$

TRAINING PHASE



Input Space Multi-modal Sensor Data, X

CNC Machine

Non-linear Mapping

$\kappa(x_i, x_j) = \varphi(x_i)^T\varphi(x_j)$ for all $i, j = 1, \dots, N$

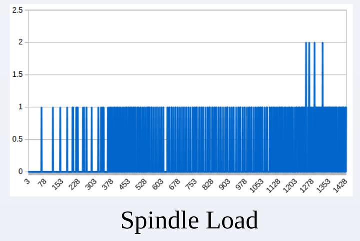
Kernel Space Based on Kernel function κ $K(X, X) \in R^{N\times N}$

Learning Data Model Using Joint Optimization

Dictionary $A \in R^{N\times K}$

Weights $w \in R^{K}$

Output y



Update Dictionary A → Update Common Coefficients $Z \in R^{K\times N}$ → Update Regression Weights w → Check for Convergence

No → Update Dictionary A

Yes → Stop

Sub Problems to Solve:

$$A \leftarrow \min_{A} \|\varphi(X) - \varphi(X)AZ\|_F^2$$

$$Z \leftarrow \min_{Z} \|\varphi(X) - \varphi(X)AZ\|_F^2 + \lambda\|y - wZ\|_2^2$$

$$w \leftarrow \min_{w} \lambda\|y - wZ\|_2^2 + \mu\|w\|_2^2$$

Closed Form Solution for Update:

$$A = Z^T(ZZ^T)^{-1}$$

$$(A^T K(X,X)A + \lambda w^T w)Z = A^T K(X,X) + \lambda w^T y$$

$$w(\lambda ZZ^T + \mu I) = \lambda y Z^T$$

TESTING PHASE

Input Space Multi-modal Sensor Test Data, x_{test}

Non-linear Mapping

Kernel Space Based on Kernel function κ $K(x_{test}, X)$

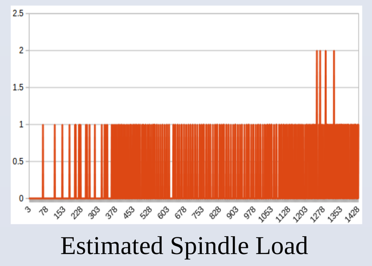
$K(x_{test}, X) = [\kappa(x_{test}, x_1), \dots, \kappa(x_{test}, x_N)]$

Estimating Common Coefficients z_{test}

Dictionary A

Weights w

Output Prediction $\hat{y}_{test} = w z_{test}$



Sub Problems to Solve:

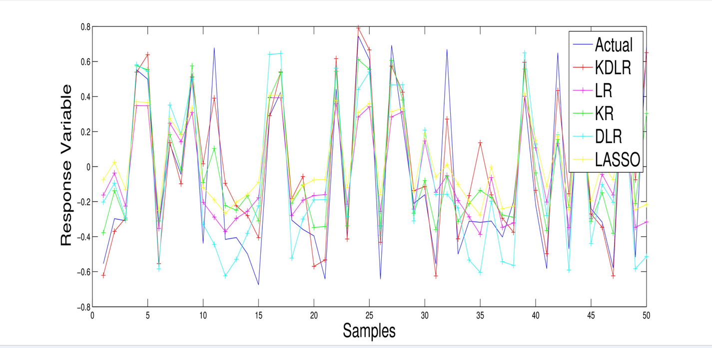
$$z_{test} \leftarrow \min_{z_{test}} \|\varphi(x_{test}) - \varphi(X)Az_{test}\|_F^2$$

Closed Form Solution:

$$A^T K(X,X)Az_{test} = A^T K(x_{test}, X)^T$$

EXPERIMENTAL RESULTS

Synthetic Dataset

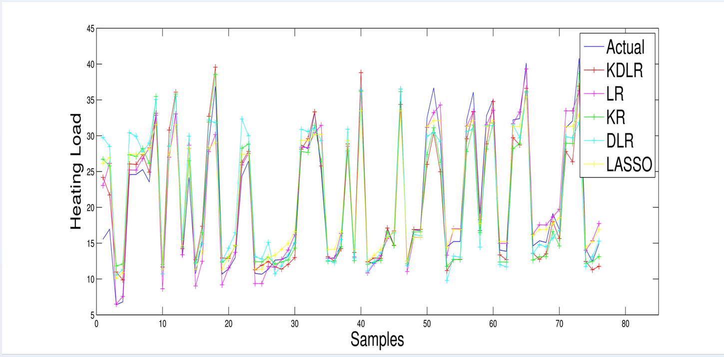


Estimation Accuracy

Algorithm	MSE	PCC
KDLR ($K = 20, \sigma = 0.5$)	$\mathbf{0.036 \pm 0.010}$	**0.918**
KR	0.054 ± 0.013	0.873
LR	0.135 ± 0.019	0.607
DLR ($K = 3$)	0.134 ± 0.019	0.608
LASSO	0.137 ± 0.012	0.607

The non-linear data simulated for evaluation by taking 3 predictors and 1 response variable

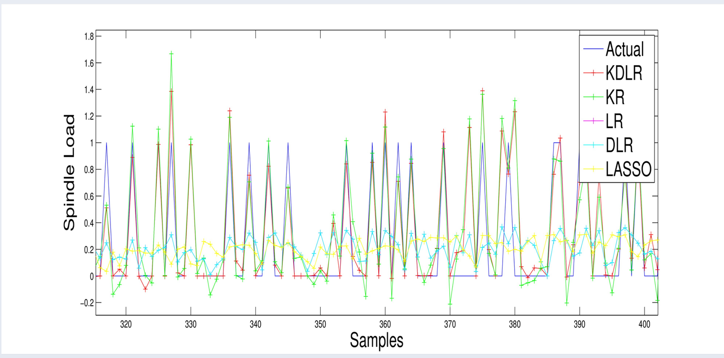
Public Dataset



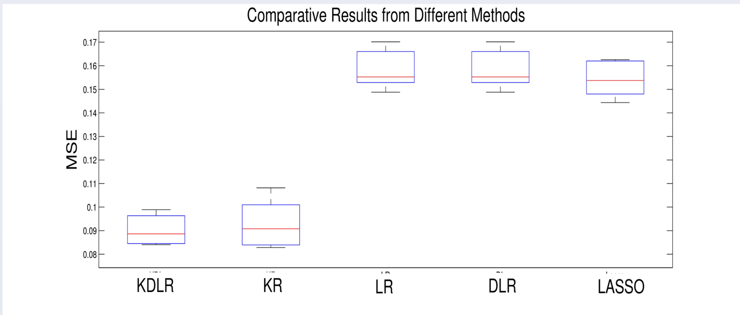
Algorithm	MSE	PCC
KDLR ($K = 20, \sigma = 6$)	$\mathbf{7.985 \pm 1.123}$	**0.962**
KR	9.164 ± 1.652	0.955
LR	9.617 ± 2.407	0.953
DLR ($K = 5$)	19.133 ± 4.310	0.905
LASSO	13.382 ± 3.807	0.945

Energy Efficiency: To assess the heating load requirements of buildings as a function of building parameters (Relative compactness, Surface area, Wall area, Roof area, Overall height, Orientation, Glazing area, Glazing area distribution)

Factory Dataset



Box Plot for Spindle Load Estimation Accuracy



CONCLUSIONS

- Useful experimental results are obtained with different real-life datasets demonstrate the potential of the proposed algorithm in effective modeling of the data.
- This technique offers significant improvement in estimation accuracy over the other popular traditional techniques.
- The work can be extended to handle multiple response variables. Also, one can consider deep dictionaries for more accurate modeling to represent the data.
- Additionally, one can also think of working out kernelized regressors using graph signal based dictionaries to effectively capture the complex inter-relationships among the multi-variate data samples.

CONTACT

kriti.kumar@tcs.com, m.gchandra@tcs.com, achannaanil.kumar@tcs.com
angshul@iiitd.ac.in