

Sparse Representation for Classification of Posture in Bed

1st Michaela Mesárošová

Department of Control and Instrumentation
Brno University of Technology
Brno, Czech Republic
xmesar01@vut.cz

2nd Ondrej Mihálik

Department of Control and Instrumentation
Brno University of Technology
Brno, Czech Republic
ondrej.mihalik1@vut.cz

Abstract—Redundant dictionaries, also known as frames, offer a non-orthogonal representation of signals, which leads to sparsity in their representative coefficients. As this approach provides many advantageous properties it has been used in various applications such as denoising, robust transmissions, segmentation, quantum theory and others. This paper investigates the possibility of using sparse representation in classification, comparing the achieved results to other commonly used classifiers. The different methods were evaluated in a real-world classification task in which the position of a lying patient has to be deduced based on the data provided by a pressure mattress of 30×11 sensors. The investigated method outperformed most of the commonly used classifiers with accuracy exceeding 92%, while being less demanding on design and implementation complexity.

Index Terms—sparse representation, linear regression, LASSO, redundant basis, SRC, classification

I. INTRODUCTION

Signals are often represented in different systems, typically bases, which provide a non-redundant signal representation. Any corruption or loss of transform coefficients can lead to serious information loss about the original signal. In environments in which the integrity of the signal representation cannot be reliably guaranteed, redundant systems provide useful properties to combat the threat of information loss. Furthermore, the popularity and applications of redundant systems have been increasing over the last years [1]–[3]. This paper focuses on the application of redundant dictionaries for signal processing. We verify and assess the possibility of using this representation in classification tasks, in which the position of a person lying on the pressure-sensitive mattress needs to be determined.

The outline of this paper is as follows. Section II introduces the necessary mathematical apparatus. Section III provides an analysis of the dataset with its representation using redundant coefficients. Section IV discusses the classification criteria and their results. Lastly, in Section V, all achieved results are compared against commonly used classifiers.

II. LINEAR REGRESSION MODELS

As stated in [4], linear regression has been widely used for decades and still constitutes an important pillar in statistics. It is simple and often provides easily interpretable description of the relation between the inputs and output of an analysed

system. In general, regression uses a pool of mathematical and statistical tools to determine the estimation of the relationship between variables. Specifically, we can consider the general relationship among the independent variables \mathbf{x} , the dependent variables y , and the vector of unknown parameters β :

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

where f represents regression function, typically prescribed, and β are parameters found through optimization of the goodness-of-fit of this function. [5] Successful optimization is immensely dependent on the expression we want to optimize. As an input we assume p discrete linear inputs

$$\mathbf{x} = \{x_1, x_2, \dots, x_p\}^T \quad (2)$$

and we aim at the prediction of the output y . Thus, the regression model has the following form:

$$\hat{y} = \hat{\beta}_0 + \sum_{j=1}^p x_j \hat{\beta}_j \quad (3)$$

where \hat{y} is marking predicted value and $\hat{\beta}_0$ represents the bias [4].

A. LASSO

Generally, regression model is more interpretable, if it has fewer descriptive coefficients that bear on the outcome. Least absolute shrinkage and selection operator (LASSO) adds an ℓ_1 penalty term to regularize the least-squares regression. As a result we are able to effectively penalize outliers and alleviate overfitting of the model to given data. Thus we get the following formula:

$$\hat{\beta} = \arg \min_{\beta} \|\mathbf{X}\beta - \mathbf{y}\|_2^2 + \lambda \|\beta\|_1 \quad (4)$$

where λ parameter directly affects sparsity of a resulting model. By increasing λ we restrict the outcome to fewer coefficients. However, decreasing this parameter allows more terms in the estimated model until the obtained result corresponds to the least-squares model when $\lambda = 0$ [5], [6]. In order to determine the optimal value of λ , we conducted a numerical experiment as follows. Squared sum of the residuals of the outcome and the time needed for computation of the model was compared for different λ values. Results are plotted in Fig.

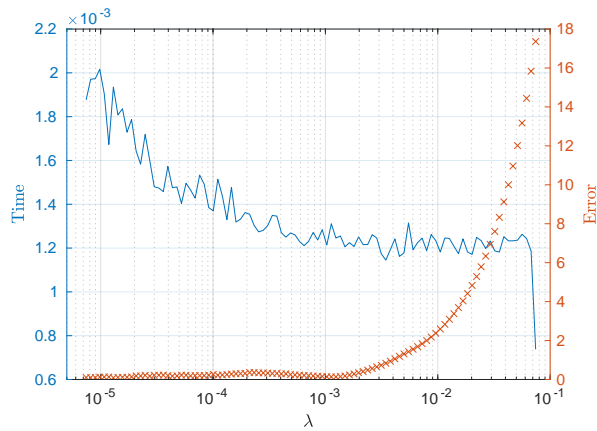


Fig. 1. Dependency of computational complexity and model error on λ

1. Based on the results, we chose $\lambda = 5 \times 10^{-3}$, for the model error lies in the stable interval, and there is no appreciable change in computational time when λ is increased further. In other words, the advantageous properties of a sparse model are retained, yet we do not have to sacrifice the estimation precision of the model. The same regularisation constant value is being used for the remainder of the paper.

III. SPARSE REPRESENTATION OF MEASUREMENTS

Signals analyzed in this paper are measurements obtained from a pressure-sensitive mattress stored as 30×11 matrices. Information associated with each observation is subject number and position number. Subject labels measured data by participating person whilst position number distinguishes between four possible lying positions, namely:

- on the back – class 1,
- on the left side – class 2,
- on the right side – class 3,
- on the stomach – class 4.

The dataset is similar to the set presented in [7], yet larger. It comprises 18 different test subjects in varying positions, containing 290 pressure map images in total. For the sake of clarity, we present a few data examples in Fig. 2.

The strategy chosen to train classification models and evaluate their accuracy is cross-validation [4]. Data are split into folds; each fold comprising pressure images of one individual. This yields 18 folds: 17 are used for training and one for validation. Therefore, there are 18 possible ways of data partition, which yield 18 classifiers of the same type, each validated using a different fold.

Training images are reshaped into column vectors, normalized to unit ℓ_2 norm, and stacked into a new matrix so as to form the redundant dictionary \mathbf{X} . The dictionary is appended by an identity matrix, which affords an effective means of dealing with occlusions [3], [7].

The first part of the algorithm consist in decomposing a validation signal by means of (4), thereby obtaining sparse coefficients $\hat{\beta}$, which form the signal representation as seen

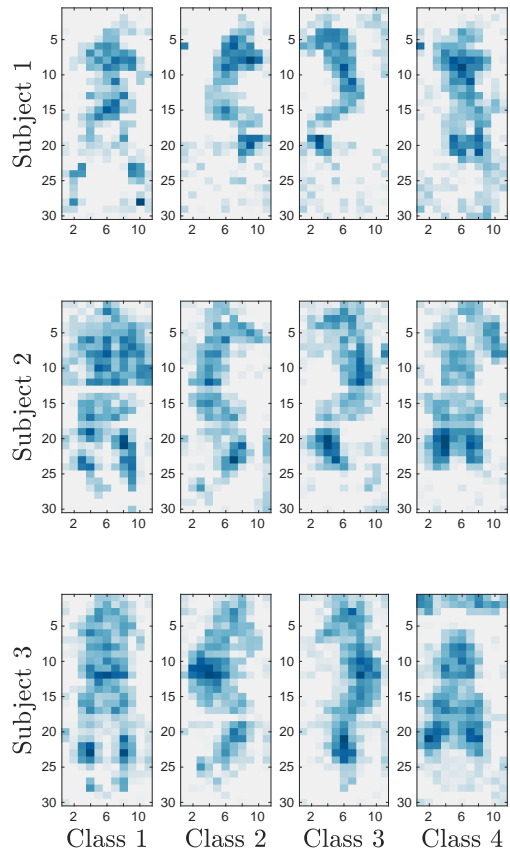


Fig. 2. Data visualisation for three exemplar subjects in all four possible positions

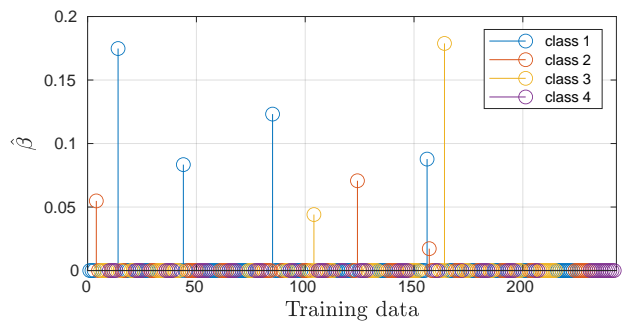


Fig. 3. Sparse coefficients for measurement 20 belonging to subject No. 1 from class No. 1

in Fig. 3. Although the signal decomposition may consist of nearly 250 observations, only a few were recognized by the LASSO algorithm as relevant to description and recreation of the validation signal. Furthermore, the dominant coefficients (blue) suggest that the image belongs to the first class. However, the the biggest coefficient (yellow) may signify that the observation belongs to the third class.

IV. SPARSE REPRESENTATION CLASSIFICATION

To be able to tell which measurement corresponds to which class, marked as k , it is crucial to find suitable decision-making

scheme. To address this problem we adopted three methods presented in paper [8]:

- Maximum coefficient (MC) – The predicted class label corresponds to the class label of the training sample with the largest coefficient of the sparse solution

$$\hat{k} = \arg \max_k (\max_{i \in \mathbf{x}_k} \beta_i). \quad (5)$$

- Maximum sum of class coefficients (MSCC) – The class whose sum of β coefficients is maximized is considered as the predicted class label. In other words, for each class k we take the sum of the coefficients β , and we classify the observation into the class with the largest sum.

$$\hat{k} = \arg \max_k \left(\sum_{i \in \mathbf{x}_k} \beta_i \right). \quad (6)$$

- Minimum class residual (MCR) – This method predicts the class label \hat{k} based on the value of class residual. Whichever class minimizes the residual is chosen as prediction label.

$$\hat{k} = \arg \min_k \|\mathbf{y} - \mathbf{X}_k \beta_k\|_2. \quad (7)$$

where \mathbf{X}_k and β_k contain only observations and coefficients, respectively, corresponding to class k .

Calculation of all model coefficients with the cross-validation setup introduced in III allows creating reconstructions of original images as follows:

$$\mathbf{y} = \mathbf{x}_1 \beta_1 + \mathbf{x}_2 \beta_2 + \dots + \mathbf{x}_p \beta_p \quad (8)$$

For the sake of clarity, the behaviour of (5), (6), and (7) will be illustrated on measurement No. 20 whose sparse coefficients are depicted in Fig. 3.

A. Maximum Coefficients

MC is the simplest out of the three methods introduced above. Fig. 4 shows value of maximal coefficient for each class separately, where it is very easy to unveil the disadvantages of the MC approach. Fig.3, despite the prevailing occurrence of first class coefficients, the decision is made only on the basis of one element. Other coefficients are not involved in the classification process.

B. Maximum Sum of Class Coefficients

It is obvious from Fig. 4, that MSCC has successfully eliminated the unfavourable effect of a single strong coefficient, as in MC. The classification is correct and the difference between the MSCC values is larger, so the result is decided with more certainty.

C. Minimum Class Residual

MCR may be seen as a conceptual extension of the MSCC. If a class makes the highest contribution to forming the signal approximation, then a reconstruction image created by this class coefficients is likely to be the most similar to the original image. Therefore it can be expected for MCR and MSCC to deliver the same or very similar results of classification.

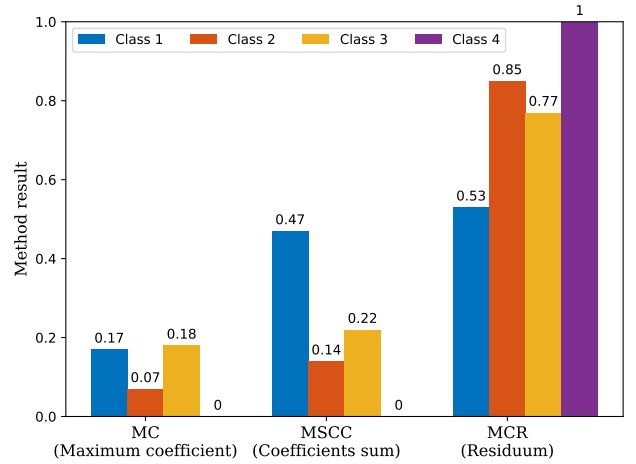


Fig. 4. Results obtained by MC, MSCC and MCR algorithms applied to each class separately using measurement No. 20 from subject No. 1 belonging to the 1st class.

The residual value reflects the error of the reconstructed observation; therefore, classes with zero-valued coefficients will reach 100 % reconstruction error as shown in Fig. 4.

D. Occlusions

In various applications there is a possibility of malfunctioning equipment or corrupted data. Therefore it is important to investigate behaviour of SRC in such cases. We will consider the fault being missing rows as it is a consequence of a non-functional sensor. To make classification possible despite errors in the data, new elements are appended to the dictionary \mathbf{X} representing potential errors [7]. However the fault definition is interchangeable if a user has more information about typical fault character in specific case. Classification with modified dictionary was applied on original as well as on occluded data and results are listed in Table I in the subsequent section.

V. RESULT COMPARISON WITH OTHER CLASSIFIERS

To evaluate classification results for the whole dataset, algorithm is run in the manner described in Section III. Overall success rate is judged based on the TPR value.

$$\text{TPR} = \frac{\text{Correctly classified into the class}}{\text{Total number of class members}} \quad (9)$$

Global TPR value is conveniently calculated from the confusion matrices, see Fig. 5, where partial classification rates for each approach and class is displayed. The global TPRs obtained by MC, MSCC, and MCR are summarized in Table I. To evaluate the viability of SRC more objectively than in [8], the results obtained were compared to several commonly used machine learning approaches and classification methods: Convolutional neural network (CNN), K-nearest neighbours (KNN), linear discriminant analysis (LDA) and bagged classification trees, to name a few.

Most of the classifiers allow selection of a training parameter, such as the regularisation γ for (LDA). A broad range of

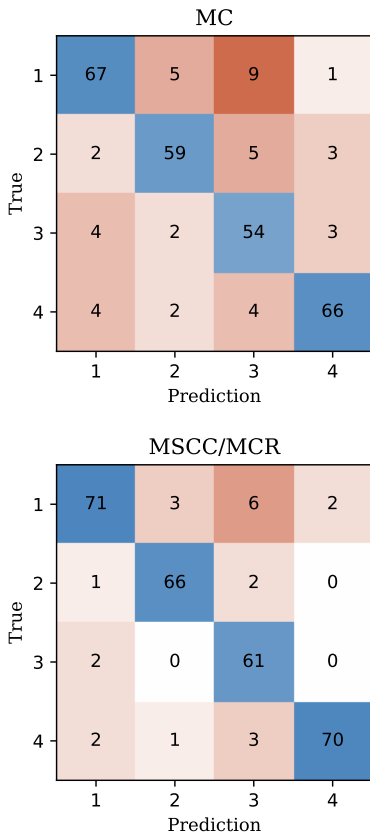


Fig. 5. Confusion matrix overview of the partial classification success. Since MSCC and MCR yield the same results we included them in one common graph.

parameters was tested and the classifiers yielding the highest validation TPRs were used in the table.

Quadratic discriminant analysis (QDA) could not be trained using the dataset without occlusions because the training algorithm failed due to singular covariance matrices. The CNN requires a proper selection of its structure. The following layers turned out to yield the best validation TPRs:

TABLE I
CLASSIFICATION SUCCESS RATES

Classification Approach	TPR			
	trained using clean dataset		trained to deal with occlusions	
	original validation data	occluded validation data	original validation data	occluded validation data
SRC, MC	0.8483	0.8303	0.8552	0.8502
SRC, MSCC	0.9241	0.9136	0.9172	0.9177
SRC, MCR	0.9241	0.9136	0.9172	0.9177
CNN (2 convolutional layers)	0.9310	0.9183	0.9344	0.9232
LDA, $\gamma = 0.76$	0.9034	0.8945	0.8483	0.8389
QDA, $\gamma = 0$	—	—	0.7897	0.7882
regularised logistic regression	0.8931	0.8817	0.8414	0.8300
SVM, box constraint = 0.9	0.8897	0.8776	0.8207	0.8122
KNN, $K = 1$	0.8586	0.8455	0.8483	0.8486
KNN, $K = 10$	0.8379	0.8216	0.8621	0.8554
50 bagged classification trees	0.8207	0.8144	0.8759	0.8720
Sole classification tree	0.5207	0.5077	0.5621	0.5606

- 1) 3 convolutional filters of size 3×3 ,
- 2) leaky relu layer,
- 3) max pooling layer of size 2×2 with stride 2,
- 4) 6 convolutional filters of size 3×3 ,
- 5) leaky relu layer,
- 6) max pooling layer of size 2×2 with stride 2,
- 7) fully connected layer.

The proper structure, however, was established after an extensive search through different structures. Hyper-parameters such as the numbers and sizes of layers, filters and pooling layers were tried before the aforementioned structure was found.

VI. CONCLUSION

This paper discussed possibilities of signal representation via redundant basis and its potential in solving classification tasks – classification of the patients’ lying posture. SRC performance was compared to standard classification methods. As SRC requires but one parameter selection and no feature extraction is necessary, the implementation complexity of SRC is reduced. The accuracy achieved is not sensitive to regularisation parameter in a broad range around optimal value whereas the other compared methods required careful parameter selection using cross-validation, had we tested their accuracy on test data, their TPRs would very likely be smaller. If occlusions are taken into account, performance of SRC is changed only marginally, delivering higher TPR values than most of the common classifiers tested in this study. SRC TPRs were outperformed only by the CNN, but the extra effort expended in the design of CNN is scarcely worth the marginal increase in TPR.

ACKNOWLEDGMENT

The completion of this paper was made possible by the grant No. FEKT-S-23-8451 – “Research on advanced methods and technologies in cybernetics, robotics, artificial intelligence, automation and measurement” financially supported by the Internal science fund of Brno University of Technology.

REFERENCES

- [1] S. Mallat, *A wavelet tour of signal processing: the sparse way*, 3rd ed. Boston: Amsterdam: Academic Press; Elsevier, 2009.
- [2] J. Kovacevic, A. Chebira, *An introduction to frames*, Foundations and Trends® in Signal Processing, Vol. 2, No. 1, pp. 1–94, 2008.
- [3] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, “Robust Face Recognition via Sparse Representation,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 2, pp. 210–227, 2009.
- [4] T. Hastie, R. Tibshirani, J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, 2nd ed. New York: Springer, 2009.
- [5] S. Brunton, J. N. Kutz, *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*, Cambridge: Cambridge University Press, 2019.
- [6] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 58, No. 1, pp. 267–288, 1996.
- [7] O. Mihálik, T. Sýkora, M. Husák, and P. Fiedler, “In-bed posture classification based on sparse representation in redundant dictionaries,” *IFAC-PapersOnLine*, Vol. 55, No. 4, pp. 374–379, 2022.
- [8] J. J. Liu, “Sleep posture analysis using a dense pressure sensitive bedsheet,” *Pervasive and Mobile Computing*, Vol. 10, pp. 34–50, 2014.