Time Series Data Classification on Grassmann Manifold

Waheeb Al-Samhi1*, Mohammed Al-Soswa2 and Yunes Al-Dhabi3
1School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, Jiangsu, 210094, China
2School of Information Engineering, Chang’an University, Xi’an, 710064, China
3Software College, Northeastern University, Shenyang, Liaoning, 110819, China
*Corresponding author’s e-mail: waheeb.samhi@njust.edu.cn

Abstract. Time Series Classification has become one of the most challenging problems in many signal processing and machine learning applications, e.g., audio/video signal processing and EEG signal processing. We propose a novel method of extracting features from data for classification and representing data on a Grassmann manifold by parameterizing it using the autoregressive moving model (ARMA). Then we perform classification on these extracted features by training support vector machines (SVM), with appropriate kernels on the Grassmann manifold. We performed tests on several publicly available datasets. We found that an SVM with a proper kernel on the Grassmann manifold consistently performs better than an SVM using a typical Gaussian kernel that acts on the data in Euclidean space. Furthermore, we found that the Grassmann SVM technique overperforms the literature in some high-dimensional datasets, without the need for any other preprocessing techniques. This work demonstrates the power of data-based manifold techniques in improving the performance of existing algorithms.

1. Introduction
Times series data occurs in many fields and applications of machine learning, including signal processing, image classification, medicine and finance. A standard approach to time series issues typically requires feeding features of time series in a machine learning algorithm and achieve goal results [7]. Engineering of Features usually required domain information of discipline from which data originated from. For example, when we were dealing with EEG signals classification, the possible features would involve different frequency bands and various other advanced statistical characteristics. It would involve power spectra.

Time Series Classification (TSC) is a challenging problem, and the features of the time series are more unstable than image features, so extracting key features from them is difficult. And the final classification effect is affected by feature extraction of the time series, so the accuracy of the results is still not very high. With the expansion of time arrangement information accessibility, many TSC calculations have been proposed. A couple has thought about Deep Neural Networks (DNNs) to play out this undertaking among these techniques. This is astounding as profound learning has seen exceptionally effective applications in the most recent years. DNNs have for sure upset the field of computer vision, particularly with the approach of novel more profound designs, for example, Remaining and Convolutional Neural Organizations. Aside from pictures, consecutive information, for example, text and sound, can likewise be handled with DNNs to arrive at cutting edge execution for record characterization and discourse acknowledgment. At present, researchers also split the classification of time series into three categories: time point classification [8], form classification, and
transition classification. Distance could be utilized for the resemblance of time series, including such Euclidean distance. With DTW, the spatial resemblance can be used. The similarity of the process of generation of data can be used.

Traditional SVM relapse (SVR) approaches have traditionally used non-linear system recognition based on the support vector machine (SVM), which in certain recurrent pieces of Hilbert space can be seen as a certain non-linear autoregressive and moving normal (ARMA) model (RKHS). In this paper, we first use the autoregressive-moving average (ARMA) model to parametrize time-series data, and then it represents the ARMA parameters as points on a Grassmann manifold (collection of Euclidean subspaces). Then we utilize a support vector machine (SVM) equipped with kernel methods to perform classification defined on Grassmann manifolds.

2. Related Work

Time series classification problems are separated from customary characterization issues in light of the fact that the traits are requested. If the requesting is by time is indeed unimportant. The significant trademark is that there might be prejudicial highlights subject to the requesting. Preceding 2003, there were at that point in any event 100 papers proposing TSC calculations. However, as [17] brought up that year, in a persuasive and exceptionally referred to paper, the nature of exact assessment would in general, be too helpless compared with the remainder of the AI people group. Most of TSC calculations were tried on a solitary dataset, and most of these datasets were manufactured and made by proposing creators for the motivation behind exhibiting their calculations. The presentation of the College of California, Riverside (UCR) time arrangement grouping and bunching vault [18] was intended to moderate these issues. The prepared accessibility of openly accessible datasets is definitely in any event, somewhat liable for the quick development in the number of distributions proposing time arrangement order calculations. Before the late spring of 2015, more than 3000 specialists have downloaded the UCR chronicle and it has been referred to a few hundred times. The storehouse has added to expand the nature of assessment of new TSC calculations. Most analyses include an assessment on more than forty informational indexes, frequently with refined criticalness testing and most creators discharge source code. This level of assessment and reproducibility is by and large in a way that is better than most territories of AI and information mining research.

Nonetheless, there are still some essential issues with distributed TSC research that we intend to address. Initially, essentially all assessments are performed on a solitary train/test split. The first inspiration for this was maybe honorable. The first makers of the document had noticed that a few creators would at times accidentally cripple standard classifiers. For instance, [25] didn’t z-standardize the time arrangement after-effects for straightforward Euclidean distance coordinating, causing the Euclidean distance to perform basically arbitrarily on an undertaking where it very well may be relied upon to perform impeccably. This model was outwardly clear by a cautious examination of the figures in [25]. However, in a straightforward table of results, such mistakes would likely never be recognized. Along these lines, the single train/test split was intended to secure the correlations with a known exhibition benchmark.

Grassmann manifold properties have been widely used in some of these computer vision problems involving subspace limits, such as [20], which tackles optimization for informative predictions about the Grassmann manifold. In order to perform affine invariant clustering of forms, the Grassmann manifold structure of the affine shape space is also used in [21]. By using Mercer kernels on the Grassmann manifold, Hamm and Lee [22] perform discriminative classification over subspaces for object recognition tasks. In [23], a face image and its disturbances are approximated as a linear subspace due to registration errors and are thus embedded as points on a Grassmann manifold. The multiplicity of Grassmann is or is tuned to unique domains that lack generality. The geometry of the Grassmann manifold for subspace tracking in array signal processing applications was exploited by Srivastava and Klassen [24]. Similarly, the geometry of the Stiefel manifold was found, in addition to the subspace structure, to be useful in applications where the special choice of base vectors is also important [25].
The collection of instruments that rely on the Grassmann manifold's Riemannian geometry are also present appropriate algorithms to execute these computations alongside the mathematical formulations.

The Riemannian manifold and the Grassmann manifold of SPD matrices have a number of applications that are defined by SPD matrices in the computer vision [10]. For example, in object detection [12], texture classification [11], object monitoring, behavior recognition and human recognition [18], the covariance area descriptors are used. One of the leading areas for the development of non-linear algorithms on SPD matrices was diffusion tensor imaging (DTI) [10]. Structure tensors are also used to encode significant image characteristics, such as texture and motion [14], in optical flow estimation and motion segmentation. In single image segmentation, structure tensors were also used [19].

3. Dynamic model parameterization

During the time spent recognizable proof (in the feeling of particular and assessment) of powerful direct frameworks, rather muddled issues associated with the parametrization of the model happen. A specific piece of these intricacies emerges just in the multivariable case, and this is one motivation behind why recognizable proof of multivariable frameworks isn't yet a norm task in applications.

We are worried about the parametrization of (discrete-time), direct, time-invariant and limited dimensional frameworks here, both in (vector-) contrast condition and in state-space structure, emphasizing the principal case. For the outcomes introduced, it isn't basic whether the data sources are imperceptibly repetitive sound, regardless of whether there are what's more noticed information sources (or, as a third case, whether the framework is deterministic). In this way for notational accommodation, we will primarily talk about the situation where the data sources are surreptitiously repetitive sound.

An ARMA framework is an arrangement of straight contrast conditions of the structure:

\[
\sum_{i=0}^{p} A(i) y(t - i) = \sum_{i=0}^{q} B(i) s(t - i),
\]

where the unobserved inputs are white noise, i.e.

\[
E \varepsilon(i) = 0, E \varepsilon(s) \varepsilon(t) = \delta_{st} \Sigma,
\]

Where \( y(t) \) are the (observed) ouputs and where \( A(i), B(i) \forall \in \mathbb{R}^{5x5} \) are parameter matrices. Let \( a(z) = \sum_{i=0}^{p} A(i) z^i, b(z) = \sum_{i=0}^{q} B(i) z^i \)

The autoregressive-moving-average (ARMA) is a well-known dynamic model for time series data that parameterizes a signal \( f(t) \) by the equations

\[
f(t) = Cz(t) + w(t), w(t) \sim N(0, R)
\]

\[
z(t + 1) = Az(t) + v(t), v(t) \sim N(0, Q)
\]

Where \( z \in \mathbb{R}^d \) is hidden state vector, \( f: \mathbb{R} \rightarrow \mathbb{R}^p \), and \( d \leq p \) is the hidden state dimension [1]. There are widely-used closed form solutions for estimating. The parameters \( A \) and \( C \). It can be shown [1] that expected observation sequence is termed as.

\[
O_{\infty} = \mathbb{E} \left[ \begin{bmatrix} f(0) \\ f(1) \\ f(2) \\ \vdots \\ f(\infty) \end{bmatrix} \right] = C A^T \mathbb{E} z(0)
\]

4. Representation on Grassmann manifold

A point on Grassmann manifold \((n, r)\), where \( n > r \) is the Euclidean \( n - dimensional \) space's \( r \)-dimensional linear subspace. Such a point usually specifies the \( n \times r \) matrix, the columns of which bear an orthonormal subspace basis. With this representation, the subspace stretched by columns \( Y \) or \( span(Y) \) that we specify by \( [Y] \) is the point on the Grassmann manifold. We may use \( G \) to denote \((n, r)\) Grassmann manifold. A multiple known as \((n, r)\) Stiefel manifold space is created by \( n \times r \) \((n > r)\)
matrices with orthonormal columns; a canonical Riemannian metric inherits the \((n, r)\) Stiefel manifold. Points on \(\text{Gr}_n\) are orthonormal column equivalence groups of \(n \times r\) matrices, where if their columns occupy the same \(r\)-dimensional subspace, all of the matrices are equal. Thus, by multiplying isometries on the right (change of orthogonal base), the orthogonal group \(O_r\) acts on the Stiefel manifold, and Grassmann can be defined as a set of orbits of this operation. As this action is both proper and free, Grassmann forms manifold, and by equipping it with the usual standard Riemannian metric derived from metric on Stiefel manifold, it is given a Riemannian structure. In this article, it shows that the expected \(f(t)\) outcomes lie in the column space of the observability matrix \(O_\infty\) and we approximate \(O_\infty\) to form \(O_m \in \mathbb{M}_{\text{mpxd}}\) by truncating on the \(m\)th block. The ARMA model, therefore, produces a signal representation as a Euclidean subspace (given by column space of \(O_m\)), and thus a point on the Grassmann manifold \(\text{Gr}(d, mp)\).

5. Kernel methods on Grassmann manifold (SVM)

Kernels are simply a way to represent given data samples flexibility to compare the samples in a particular complex space. Kernels are essential in comparing images of variant sizes, object 3D structures, and even text documents of different lengths and formats [9]. A kernel is a random function that enables one to map objects in a tricky space to a high dimension space, making a comparison of the complex features an easy task. The Kernels are applied in functions like the support vector machines. The classifier finds the linear separator between the data points projected into feature space, maximizing the two classes' data points [10]. Kernel-based methods are used widely in manifold learning. Jayasumana et al. [2] define the analog of the Gaussian RBF kernel for the Grassmann manifold:

\[
\begin{align*}
    d_P([Y_1], [Y_2]) &= 2^{-1/2} \|Y_1^T Y_1^T - Y_2^T Y_2^T\|_F \\
    k_P([Y_1], [Y_2]) &= \exp(-\gamma d_P^2([Y_1], [Y_2]))
\end{align*}
\]

Where \([Y_i]\) is subspace spanned by columns of \(Y_i\), \(Y_1\) and \(Y_2\) are matrices with orthonormal Columns, and \(\gamma\) is a hyperparameter. We use the representation of a time series to point on Grassmann manifold to do classification based on Support Vector Machines (SVMs) utilizing the RBF kernel.

6. Algorithm

**Algorithm 1** The supervised learning algorithm on Grassmann Manifold

**Input:** A training set \(\{X_i\}_{i=1}^n\) list of train labels \(\{y_i\}_{i=1}^n\) list of test signals

1. For \(i=1, 2, \ldots, n\).
   - Compute parameters \(C\) and \(A\) for \(X_i\).
   - Compute \(O_m\) for \(X_i\).
   - Orthonormalize \(O_m\) and store as \(U_i\).
2. Using kernel \(k_P\), fit SVM on \(\{U_i\}_{i=1}^n, \{y_i\}_{i=1}^n\)
3. Predict SVM on \(\{y_i\}_{i=1}^m\).

**Output:** List of predicted test labels

7. Experiments and Results

Experiments were performed on three data sets, which include SUNY EEG database [3], Lip videos [5], and vehicle audio recordings [4]. In the SUNY EEG database, we predefined the Train/test split at 48% test data, and one trial is performed with the used parameter are set to \(d=m=10, \gamma=0.2\). In the vehicle audio recordings [4], the audio recordings of different vehicles mobbing in a parking lot at around 15 mph. We defined the Train/test split at 50% for testing, and we performed 20 trials. We used the last 6 seconds of every recording where the car is near the microphone, and we set the parameters to \(d=2, m=10, \gamma=10\). In the lip videos [5], the video recordings are of a person speaking the digits 1-5, and we
set the Train/Test split at 40% for testing, and we performed 20 trials with parameters set to $d=m=10$, $\gamma=0.2$. Since the videos are not equally long, for Euclidean SVM, we perform principal component analysis across the frames to extract 30 principal vectors. We implemented Algorithm 1 for every dataset separately.

**Table 1. Classification accuracies**

| Dataset         | Ours   | Euclidean | Literature |
|-----------------|--------|-----------|------------|
| Alcohol EEG     | 98.8%  | 79.9%     | 97.1% [6]  |
| Vehicle audio   | 61.5%  | 50.8%     | 88.2% [4]  |
| Video digits    | 96.0%  | 75.6%     | 94.7% [5]  |

As we can see from the above table 1, we can find that the Grassmann SVM approach performed better than the Euclidean SVM on time series with high dimensional representation, which proves the effectiveness of using the proposed method of parameterizing signal in the Grassmann manifold. We also found that the Grassmann SVM approach on raw data for both the EEG and video digit datasets performed better than the literature shown in table 1, which uses extensive preprocessing. Overall, classification is better using the manifold approach.

### 8. Conclusion

In this paper, we have presented a novel method for extracting data using the autoregressive-moving average (ARMA) and representing it as points on the Grassmann manifold for classification. We then used several manifold techniques, including the Grassman support vector machine (SVM) with the proper kernels for classification defined on the Grassmann manifold. Performing the proposed method on three sets of data, including the most commonly used datasets on Time Series Classification, our proposed method results have shown that the Grassmann SVM performed better on high-dimensional raw data. And it performs significantly better (10-20% more) than straightforward Euclidean SVM, which proves the effectiveness of parametrizing signals in the Grassmann manifold. For video digit datasets, the Grassmann SVM approach on raw data sometimes overperformed the benchmark results in the literature for these high-dimensional datasets, without the need for any other preprocessing techniques as used in the literature. With this project, we have gained familiarity with developing techniques in the fields of manifold learning, signal processing, and function approximation. The project enhanced our understanding of novel algorithms for machine learning, encouraging us to do further research in these directions.

### References

[1] Turaga, P., Veeraraghavan, A., Srivastava, A., & Chellappa, R. (2011). Statistical computations on Grassmann and Stiefel manifolds for image and video-based recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(11), 2273-2286.

[2] Jayasumana, S., Hartley, R., Salzmann, M., Li, H., & Harandi, M. (2015). Kernel methods on Riemannian manifolds with Gaussian RBF kernels. *IEEE transactions on pattern analysis and machine intelligence*, 37(12), 2464-2477.

[3] Zhang, X. L., Begleiter, H., Porjesz, B., Wang, W., & Litke, A. (1995). Event-related potentials during object recognition tasks. *Brain research bulletin*, 38(6), 531-538.

[4] Sunu, J., & Percus, A. G. (2018, March). Dimensionality reduction for acoustic vehicle classification with spectral embedding. In 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC) (pp. 1-5). IEEE.

[5] Lieu, L., & Saito, N. (2011). Signal Ensemble Classification Using Low-Dimensional Embeddings and Earth Mover’s Distance. In Wavelets and Multiscale Analysis (pp. 227-256). Birkhäuser Boston.

[6] Sharma, M., Deb, D., & Acharya, U. R. (2018). A novel three-band orthogonal wavelet filter bank method for automated identification of alcoholic EEG signals. *Applied Intelligence*, 48(5),
6

[7] Marteau, P. F. (2008). Timewarp edit distance with stiffness adjustment for time series matching. IEEE transactions on pattern analysis and machine intelligence, 31(2), 306-318.

[8] Hyndman, R. J., & Athanasopoulos, G. (2015). 8.9 Seasonal ARIMA models. Forecasting: principles and practice., OTexts. Retrieved, 19.

[9] Jayasumana, S., Hartley, R., Salzmann, M., Li, H., & Harandi, M. (2015). Kernel methods on Riemannian manifolds with Gaussian RBF kernels. IEEE transactions on pattern analysis and machine intelligence, 37(12), 2464-2477.

[10] Shvarts, E. (2013). Learning time series classification by clustering on the manifold. Claremont Graduate University.

[11] Tuzel, O., Porikli, F., & Meer, P. (2006, May). Region covariance: A fast descriptor for detection and classification. In European conference on computer vision (pp. 589-600). Springer, Berlin, Heidelberg.

[12] Tuzel, O., Porikli, F., & Meer, P. (2008). Pedestrian detection via classification on Riemannian manifolds. IEEE transactions on pattern analysis and machine intelligence, 30(10), 1713-1727.

[13] Goh, A., & Vidal, R. (2008, June). Clustering and dimensionality reduction on Riemannian manifolds. In 2008 IEEE Conference on Computer Vision and Pattern Recognition (pp. 1-7). IEEE.

[14] Caseiro, R., Martins, P., Henriques, J. F., & Batista, J. (2012). A nonparametric Riemannian framework on the tensor field with application to foreground segmentation. Pattern Recognition, 45(11), 3997-4017.

[15] Carreira, J., Caseiro, R., Batista, J., & Sminchisescu, C. (2012, October). Semantic segmentation with second-order pooling. In European Conference on Computer Vision (pp. 430-443). Springer, Berlin, Heidelberg.

[16] Li, P., Wang, Q., Zuo, W., & Zhang, L. (2013). Log-Euclidean kernels for sparse representation and dictionary learning. In Proceedings of the IEEE international conference on computer vision (pp. 1601-1608).

[17] Chen, S., McKinney, G. J., Nichols, K. M., Colbourne, J. K., & Sepulveda, M. S. (2015). Novel cadmium responsive microRNAs in Daphnia pulex. Environmental science & technology, 49(24), 14605-14613.

[18] Harandi, M. T., Sanderson, C., Hartley, R., & Lovell, B. C. (2012, October). Sparse coding and dictionary learning for symmetric positive definite matrices: A kernel approach. In European Conference on Computer Vision (pp. 216-229). Springer, Berlin, Heidelberg.

[19] Lin, D., Yan, S., & Tang, X. (2006, June). Pursuing informative projection on Grassmann manifold. In 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06) (Vol. 2, pp. 1727-1734). IEEE.

[20] Begelfor, E., & Werman, M. (2006, June). Affine invariance revisited. In 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06) (Vol. 2, pp. 2087-2094). IEEE.

[21] Hamm, J., & Lee, D. D. (2008, July). Grassmann discriminant analysis: A unifying view on subspace-based learning. In Proceedings of the 25th international conference on Machine learning (pp. 376-383).

[22] Lui, Y. M., & Beveridge, J. R. (2008, October). Grassmann registration manifolds for face recognition. In European conference on computer vision (pp. 44-57). Springer, Berlin, Heidelberg.

[23] Srivastava, A., & Klassen, E. (2004). Bayesian and geometric subspace tracking. Advances in Applied Probability, 43-56.

[24] Lui, Y. M., Beveridge, J. R., & Kirby, M. (2009, September). Canonical Stiefel quotient and its application to generic face recognition in illumination spaces. In 2009 IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems (pp. 1-8). IEEE.

[25] Ge, X., & Smyth, P. (2000, August). Deformable Markov model templates for time-series pattern
matching. In Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 81-90).