Classifying time series data via frequency decomposition and manifold techniques

Adam Guo, supervised by Prof. Hrushikesh Mhaskar

Pomona College ('22), Claremont Graduate University

Objectives

Improve time series classification by:

- Extracting instantaneous freqencies through a novel signal separation operator for use as features
- Representing time series data as points on a Grassmann manifold via dynamic models and using kernel methods on manifolds

Introduction

The objective of this research project is to investigate novel solutions to the problem of classifying time series data. Time series classification is a general problem widespread in signal processing and machine learning, from identifying the sources of sound or radar signals, to determining human movements from accelerometer data and detecting algorithmically-generated "deepfake" video and audio.

Many machine learning problems can be broadly split into a feature extraction part, where relevant information is extracted from the data, and a model fitting part, where a numerical model is trained using the extracted features. We investigated two methods for feature extraction: a novel signal separation operator (SSO) that extracts intrinsic mode functions (IMFs) from signals, and a method for representing signals as points on a Grassmann manifold via ARMA parametrisation. These techniques prove to be effective at capturing relevant information from signals, performing well in traditional classification algorithms.

Signal separation

The technique of modeling signals as linear combinations of sinusoids is of fundamental importance in signal processing and used widely across many fields. In the big data era, the analysis of non-stationary sinusoids is becoming increasingly prevalent, formulated by

$$f(t) = \sum_{j=1}^{K} A_j(t) \exp(i\phi_j(t)) + A_0(t) \tag{1}$$

where $A_j(t)$ and $\phi_j(t)$ denote the amplitude and phase functions respectively, and $A_0(t)$ denotes the minimally oscillatory trend of $f(t)$. In [1], Chui and Mhaskar construct a deep network based on a novel signal separation operator that recovers the number K of terms in 1, the instantaneous frequencies (IFs) $\phi_j'(t)$, the amplitude functions $A_j(t)$, and the trend $A_0(t)$.

We use the first (lowest-energy) IFs extracted as features for classification.

Representation on Grassmann manifold

Traditional machine learning techniques largely assume data that lies in Euclidean space. Increasingly, techniques are being developed that account for non-linear geometric constraints on sampled data. One such technique is to interpret popular dynamic models of time series data as constructing points on the Grassmann manifold (the space of d-dimensional subspaces in $\mathbb{R}^n$) [2].

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

$$f(t) = Cz(t) + w(t), \quad w(t) \sim \mathcal{N}(0, R) \tag{2}$$

$$z(t+1) = Az(t) + v(t), \quad v(t) \sim \mathcal{N}(0, Q) \tag{3}$$

where $z \in \mathbb{R}^d$ is the hidden state vector, $d \leq p$ is the hidden state dimension [2]. There are widely-used closed form solutions for estimating the parameters A and C. It can be shown that the expected observation sequence is given by

$$\mathbb{E}[\left(f(0)\ f(1)\ f(2)\ \cdots\right)^T] = \left[C^T\ (CA)^T\ (CA^2)^T \cdots\right]^T z(0) \tag{4}$$

Hence, the expected observations of $f(t)$ lie in the column space of the observability matrix $O_\infty = \left[C^T\ (CA)^T\ (CA^2)^T \cdots\right]^T$, which can be approximated by truncating at the mth block to form the finite observability matrix $O_m \in \mathsf{M}_{mp\times d}$. Thus, the ARMA model yields a representation of a signal as a subspace of Euclidean space, and hence a point on the Grassmann manifold. We store each subspace by orthonormalising O_m. [2] recommends setting d and m to be equal, between 5 and 10.

Kernel methods on Grassmann manifold

Many machine learning algorithms, such as support vector machines (SVMs) and deep neural networks, depend on the Euclidean structure of input data, particularly structures such as norms and inner products [3]. One way to address this incongruency is to use kernel methods that are defined on manifolds. In [3] Jayasumana et al. generalise the Gaussian radial basis function (RBF) kernel $\exp\left(-\gamma\|x-y\|^2\right)$ to manifolds by replacing the Euclidean distance with a positive definite distance function, specifically the projection distance on the Grassmann manifold

$$d_P([Y_1],[Y_2]) = 2^{-1/2}\left\|Y_1Y_1^T - Y_2Y_2^T\right\|_F \tag{5}$$

where $[Y_i]$ is the subspace spanned by the columns of Y_i, and Y_1 and Y_2 are matrices with orthonormal columns. This yields the projection Gaussian kernel on the Grassmann manifold:

$$k_P([Y_1,Y_2]) = \exp\left(-\gamma d_P^2([Y_1],[Y_2])\right) \tag{6}$$

where γ is a hyperparameter. This enables the use of algorithms like SVM to classify time series data in the form of orthonormal matrices representing points on a Grassmann manifold.

Results

Algorithms	Alcohol EEG	Epilepsy EEG	Vehicle audio	Video digits
IF SVM	67%	**95%**		
IF CNN	69%	90%		
IF CNN → SVM	71%	90%		
Grassmann SVM	**99.8%**	66%	63%	**98%**
IF + Grassmann SVM	79%	85%		
Grassmann SVM w/ Hermite kernel	55%	64%	**68%**	40%
IF + Grassmann SVM w/ Hermite kernel	71%	68%		

Table 1: Classification accuracies (best performer for each dataset in bold)

Experiments

We performed two main experiments using several datasets comprising multivariate time series data. First, we performed the signal separation to extract the first instantaneous frequency from each signal in the dataset, which we used as input directly to CNN and SVM classifiers. We also used the trained CNN to extract features from the time series to classify using an SVM.

Secondly, we computed the representation of each signal as a point on the Grassmann manifold according to (4), and trained an SVM using the kernel function described in (6) to perform classification.

We performed these tests on four sets of data, using 60% of each dataset for training and 40% of testing. The optimal hyperparameters used for each are listed.

- SUNY EEG Database Data Set, a set of measurements from 64 electrodes placed on the scalp of alcoholic and non-alcoholic subjects [4]
  $d = m = 9, \gamma = 0.2$
- Epileptologie Bonn EEG data set, a set of single-channel measurements from epileptic and non-epileptic patients [5]
- Audio recordings of different vehicles moving through a parking lot at approximately 15 mph [6]
- Video recordings of people speaking numeric digits 0-9 [7]

Note that due to memory constraints, the SSO algorithm could not be run on the vehicle audio or video digit datasets.

Conclusions

We conclude that the approach of representing time series data as points on a Grassmann manifold via ARMA parameterisation, then applying a manifold kernel SVM for classification is highly effective for datasets with a high number of features. The alcohol EEG dataset comprises 64 channels while the video digits dataset has effectively 3,850 channels, one for each pixel of the frame.

The Grassmann SVM approach performs relatively poorly on datasets with fewer features, such as the epilepsy EEG dataset and the vehicle audio dataset, which comprise 1 and 2 features respectively. In the case of the epilepsy EEG dataset, using instantaneous frequency extraction as an input to other classification algorithms such as SVM and CNN performs relatively well. It is notable that combining the IF and Grassmann SVM approaches splits the difference in performance between the two approaches.

References

[1] Charles K. Chui and H.N. Mhaskar.
Signal decomposition and analysis via extraction of frequencies.
Applied and Computational Harmonic Analysis, 40(1):97–136, January 2016.

[2] P. Turaga, A. Veeraraghavan, A. Srivastava, and R. Chellappa.
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, November 2011.

[3] Sadeep Jayasumana, Richard Hartley, Mathieu Salzmann, Hongdong Li, and Mehrtash Harandi.
Kernel Methods on Riemannian Manifolds with Gaussian RBF Kernels.
IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(12):2464–2477, December 2015.
arXiv: 1412.0265.

[4] Xiao Lei Zhang, Henri Begleiter, Bernice Porjesz, Wenyu Wang, and Ann Litke.
Event related potentials during object recognition tasks.
Brain Research Bulletin, 38(6):531–538, January 1995.

[5] Ralph G. Andrzejak, Klaus Lehnertz, Florian Mormann, Christoph Rieke, Peter David, and Christian E. Elger.
Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state.
Physical Review E, 64(6):061907, November 2001.

[6] Justin Sunu and Allon G. Percus.
Dimensionality reduction for acoustic vehicle classification with spectral embedding.
arXiv:1705.09869 [physics, stat], February 2018.
arXiv: 1705.09869.

[7] Linh Lieu and Naoki Saito.
Signal Ensemble Classification Using Low-Dimensional Embeddings and Earth Mover's Distance.
In Jonathan Cohen and Ahmed I. Zayed, editors, *Wavelets and Multiscale Analysis: Theory and Applications*, Applied and Numerical Harmonic Analysis, pages 227–256. Birkhäuser, Boston, 2011.