Static models of derivative-coordinates phase spaces for multivariate time series classification: an application to signature verification

Jonas Richiardi[1], Krzysztof Kryszczuk[2], and Andrzej Drygajlo[1]

[1] Speech Processing and Biometrics Group, Laboratory of IDIAP
Swiss Federal Institute of Technology Lausanne (EPFL)
[2] IBM Zurich Research Laboratory

Abstract. Multivariate time series are sequences, whose order is provided by a time index; thus, most classifiers used on such data treat time as a special quantity, and encode it structurally in a model. A typical example of such models is the hidden Markov model, where time is explicitly used to drive state transitions. The time information is discretised into a finite set of states, the cardinality of which is largely chosen by empirical criteria. Taking as an example task signature verification, we propose an alternative approach using static probabilistic models of phase spaces, where the time information is preserved by embedding of the multivariate time series into a higher-dimensional subspace, and modelled probabilistically by using the theoretical framework of static Bayesian networks. We show empirically that performance is equivalent to state-of-the-art signature verification systems.

1 Introduction

Multivariate time series appear in a vast array of engineering applications, such as computational finance, biosignal processing (e.g. EEG, EMG), bioinformatics, or brain-machine interfacing (which can have tens of data channels). Automatic classification of these time series into pattern classes is often of interest, for example for earthquake detection (which typically is based on two or more data channels) or behavioural biometrics (e.g. speech, signature, gait).

Probabilistic state-space methods such as hidden Markov models (HMMs) offer several advantages over classical time-series models (e.g. ARMAX), for instance handling of hidden variables, continuous and discrete spaces for random variables, and specification of priors [10]. HMMs are top-performing classifiers for signature verification [1, 3]. However, while they are very commonly used, HMMs suffer some important drawbacks, for example feature vectors are deemed conditionally independent given the sequence of states. To relax this unrealistic assumption, techniques have been proposed to model the "trajectories" of the time series across states, for instance in speech recognition [8]. It is also possible to use more general dynamic probabilistic models such as dynamic Bayesian networks, and to model the inter-state relationships between the variables of the time series.

Conversely, many discriminative classification methods that perform very well with static[1] multivariate data, such as random forests [2] or multilayer perceptrons, do not yield good results when applied directly to the modelling of time series, since variability in onset or duration is not handled by default.

In this paper, we propose a middle way between static discriminative classifiers and dynamic generative models, by using a static generative model together with specific feature extraction steps. We use a specific topology of static Bayesian network that is equivalent to a Gaussian mixture model (GMM) to perform classification of multivariate time series, as applied to a handwritten signature verification task[2]. We show that, by using appropriate features extracted from the time series, the framework of static Bayesian networks offers an elegant approach to classification of multivariate time series. The underlying principle is to consider that the observable time series data available are in fact a projection from the phase space of higher dimension of an original dynamic system.

The remainder of this paper is organised as follows: In Section 2, we briefly review the theory of phase space reconstruction, drawing attention to the work of Povinelli et al. [14]. In Section 3 we propose a general framework for classifying features that can be extracted from a multivariate time series, linking in more specifically with signature verification. In Section 4, we propose the use of delta features for phase space reconstruction, and show that they posess significant advantages over the more classically used method of delays. Section 5 presents experimental results on two signature databases, which are discussed in Section 6.

2 Phase space reconstruction

Phase space reconstruction can be approached through the Takens theorem [18]: Given a (regularly sampled, univariate) time series, it states that, under certain conditions, it is possible to reconstruct a multidimensional phase space which corresponds to the same dynamical system as the one which originally generated the time series, but in a different system of coordinates: there is a diffeomorphism between the two spaces. In other words, the values of the (univariate) time series are seen as a one-dimensional projection of the trajectory in multivariate phase space of the dynamical system under consideration.

One sufficient condition entailed by the Takens theorem is that, for a map to exist between the original and the reconstructed space, we must have

$$D \geq 2\mathcal{D} + 1, \tag{1}$$

where D is the embedding dimension of the reconstructed space, and $\mathcal{D}$ is the box-counting dimension of the trajectory in the original space. The principle

---

[1] In this paper, *static* data means that there is no time dependence between feature vectors. For models, it means that time is not part of the model structure.

[2] This paper offers a theoretical explanation for the performance of our original proposal of using Gaussian mixture models for signature verification [15]

behind this theorem is that two points should lie close together in D dimenions not because D is too small, but because their proximity is a property of the set of points. The box-counting dimension is difficult to compute in some cases, when not much data is available and/or dimensionality is high [5], and can be approximated by methods such as the false nearest-neighbour algorithm [9] or the correlation integral approach [5].

The coordinates used for embedding must also be chosen. Typically, the coordinates used are the time series $o(t)$ itself, and integer multiples of the lagged time series $(o(t + \tau), o(t + 2\tau), \ldots)$. While in principle, with an infinite number of noiseless datapoints, the delay τ could be chosen almost arbitrarily [18], in practice heuristics are applied to select this parameter. Typically, either the first zero of the autocorrelation function (corresponding to linear independence of the coordinates) or the first minimum of the automutual information function (a more general notion of independence) are used to select the optimal τ.

Recently, phase space reconstruction has been applied to statistical classification of time series. Imhoff et al. [6] proposed using statistical models of reconstructed phase spaces in medicine, and Povinelli et al. [14] used the method of delays for ECG classification and speech recognition. As we will show in Section 4, delay coordinates have several shortcomings which can be addressed by the use of other coordinates.

3 Features for multivariate time series classification

Approaching multivariate time series classification from a statistical pattern recognition viewpoint, an important family of methods is to transform the original signal into another feature space before attempting classification.

For many applications, the goal is to build the most discriminative classifier possible: the semantics associated to states (e.g. for evolution of a disease over time) are of no interest. Indeed, it is even possible that the time series be transformed from a sequence into an unordered set, while maintaining or improving discriminative abilities in comparison to dynamic classifiers.

To recognise patterns in time series, the signal representation and/or the classifier must be made resilient to changes in the aspect of the pattern, for example in amplitude, onset, or duration. To this end, several feature extraction schemes are possible:

Local features are extracted at the same rate as the sampling of the incoming signal: that is, each input sample data vector corresponds to a local feature vector. *Segmental features* are extracted once the time series has been cut into segments or windows. The segmentation paradigms vary, but a segment typically consists of a sequence of points for which some definition of coherence holds. *Global features* summarise some property of the complete observed time series; for instance the total duration of the signal.

In all three cases, features can be computed from single or multiple variables in the multivariate time series. In signature verification, all three types of features have been used [17], and they can be combined in multiple-classifier systems [16].

4 Delta features for phase space reconstruction

The first and second time derivative of local features, referred to in speech processing as delta (Δ) and delta-delta ($\Delta\Delta$) features, can be used to improve the discriminative ability of the base feature set.

Since the signal is sampled at discrete time intervals, we use a numerical approximation of a first order derivative. By definition:

$$df(\cdot) = lim_{\epsilon \to 0} \frac{f(\cdot + \epsilon) - f(\cdot)}{\epsilon}, \tag{2}$$

which we replace by second order regression using the central difference approximation. Boundary effects are avoided by switching to a forward difference approximation when necessary. Thus, for the tth term in the vector sequence of length T we obtain:

$$df(\cdot_t) \approx \frac{f(\cdot_{t+1}) - f(\cdot_{t-1})}{2} + \frac{f(\cdot_{t+2}) - f(\cdot_{t-2})}{4}. \tag{3}$$

We posit that the key to the effectiveness of *static* probabilistic models for classifying multivariate time series lies in the use of delta features and features that are functions of derivatives of the raw sampled data.

Support for this hypothesis is to be found in a particular method of phase space reconstruction: [12, 18] have shown that, in addition to the method of delays, phase space reconstruction can be performed by using first, second, and successive derivatives of a (univariate) time series $\frac{do}{dt}, \frac{d^2o}{dt^2}, \dots$. Under condition that the Takens theorem is satisfied, this set of coordinates yields an embedding.

Thus, by using delta features, we are in fact reconstructing a phase space $\mathcal{S}$. This means that the dynamics of the time series (e.g. signature) can asymptotically be fully modelled without the need to resort to dynamic models. Because we also use other features (e.g. pressure), we obtain a feature space $\mathcal{S}'$ that is of a higher dimension than the corresponding reconstructed phase space $\mathcal{S}$, and embeds it. We can then view the training of a generative model on the vectors in this space as the computation of a statistical model of distribution of features in $\mathcal{S}'$. This approach was proposed in [14], using delay coordinates.

Reconstructing the phase space using delta features overcomes some original restrictions of the method of delays to univariate time series. Indeed, in order to compute an embedding into phase space for multivariate time series using the method of delays, it is necessary to estimate the optimal embedding lag (optimal delay coordinate) $\hat{\tau}_d$ for each of the dimension d in the multivariate time series independently. If these optimal embedding lags are not equal, the result is that the phase spaces reconstructed for each of the variables of the original time series will differ in length. Then, the problem of combining the different phase spaces is not trivial to solve [4]. It is also possible to consider (potentially suboptimal) equal time lags for each of the variables in the time series [14]. Using delta features, all the phase spaces reconstructed have the same length.

Also, if the signal available is small (e.g. 60 samples), and the optimal lag computed is large with respect to the signal duration, higher-order lags (e.g. 3τ

or 4τ) may actually not be computable. Using delta coordinates alleviates this problem: phase space coordinates can be computed on small amounts of data.

One potential disadvantage of delta coordinates is noise amplification with successive derivatives,but in practice this can be overcome by low-pass filtering and/or robust derivative extraction (e.g. by fitting Legendre polynomials and taking an analytic derivative).

5 Experiments and results

5.1 Signature databases and verification systems

We use two different signature databases for our experiments.

The MCYT database [11] contains signature and fingerprint data for 330 users. We use a 100-users subset of this database, called MCYT-100.

The BMEC 2007 development database contains 50 users and is part of the larger BioSecure DS3 dataset. Signatures are acquired on a low-power mobile platform (Ipaq PDA). For both databases, skilled forgeries are used (between levels 6 and 8 [7]).

The classifier used is a static Bayesian network functionally equivalent to the Gaussian mixture model proposed in [15], and is denoted BN/GMM. The number of diagonal covariance Gaussian mixture components is varied, but the initialisation is always performed using the k-means method, followed by the expectation-maximisation algorithm until convergence. The feature set depends on the experiment and database. The first 5 authentic signatures of the first session are used as training samples for each user.

5.2 Classification with delta features

In this first set of experiments the goal is to empirically demonstrate the benefits of modelling reconstructed phase spaces with delta coordinates instead of only the base feature vector, and to show the superior performance when compared to delay coordinates.

For the first experiment, the base feature spaces considered is a univariate feature space using only the coordinate x_t (univariate time series). The second feature space consists of the phase-space reconstruction $(x, \dot{x}, \ddot{x})$, which is an embedding dimension of 3. For both feature spaces, the parameters of a BN/GMM classifier, using 40 Gaussian mixture components, are trained using expectation-maximisation. As can be seen in Figure 1, training a statistical classifier on the reconstructed phase space yields better significantly better results (McNemar test on classification decisions, thresholded *a posteriori*, p < 0.01) than classification on the initial time series itself, for both phase space reconstruction coordinates used. While not reported here, the results obtained with other univariate coordinates (e.g. y) and on other databases are comparable.

For the second experiment, the classification performance over a derivative-coordinates phase space $(x, \dot{x}, \ddot{x})$ reconstructed from a univariate time series is

compared to the classification performance over a delay-coordinates phase space $(x_t, x_{t+\tau}, x_{t+2\tau})$. The lag is chosen as the mode of the distribution of lags for each user's training, respectively testing, signatures, as per [14]. Figure 1 shows that on this task, a substantial reduction in error rate can be obtained by using phase space reconstruction, and preferring delta coordinates to delay coordinates (statistically significant at p < 0.01).

The third experiment reconstructs a phase space from a multivariate time series (x_t, y_t), resulting in a phase space $(x, \dot{x}, \ddot{x}, y, \dot{y}, \ddot{y})$ for derivative coordinates, and $(x_t, x_{t+\tau}, x_{t+2\tau}, y_t, y_{t+\tau}, y_{t+2\tau})$ for delay coordinates. Again, it can be observed from Figure 1 that both coordinate systems yield improvements over the base feature space, and that more improvement results from the use of derivative coordinates (statistically significant at p < 0.01).

Our experiments on time series datasets not related to biometrics or signature verification (e.g. ECG, gestures, machine control) show that these trends are generally confirmed: that is, training a BN/GMM classifier in the reconstructed phase space tends to yield significantly better classification accuracy than what is obtained in the base space, and delta coordinates yield better results than lag coordinates.
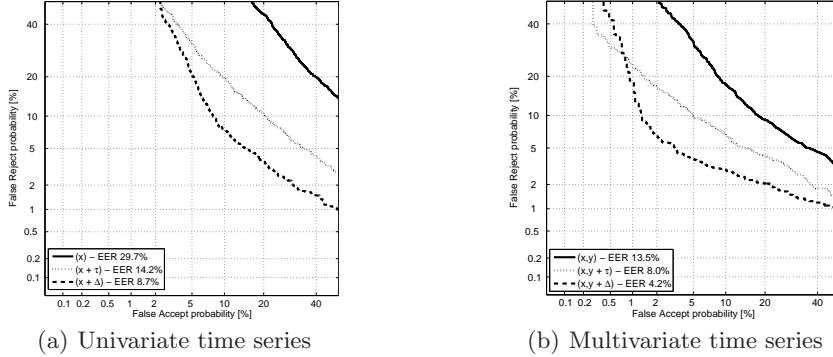

(a) Univariate time series   (b) Multivariate time series

Fig. 1. DET curves for classification of univariate (x, left panel) and multivariate ((x, y), right panel) time series in the phase space with delay coordinates (τ) and delta coordinates (Δ).

5.3 Comparison with dynamic models

In this series of experiments we compare our static BN/GMM classifier with hidden Markov models of equivalent model complexity: the number of free parameters is kept in the same range. It was chosen to compare 5- and 2- states HMMs to the BN/GMM baseline system, using diagonal covariance matrices. The HMMs have a strict left-to-right topology. Both the BN/GMM and the HMM models are initialised using k-means clustering.

In order to assess statistical significance of the difference between classifiers, we compute the EER threshold a posteriori on the classifier outputs, apply it to the output, yielding decision vectors on which we then perform the McNemar significance test with $p = 0.05$.

On MCYT-100, the feature vector used is

$$\mathbf{o}_t = [x_t, y_t, p_t, \theta_t, v_t + \Delta + \Delta\Delta]' = [x_t, y_t, p_t, \theta_t, v_t, \dot{x}_t, \dot{y}_t, \dot{p}_t, \dot{\theta}_t, \dot{v}_t, \ddot{x}_t, \ddot{y}_t, \ddot{p}_t, \ddot{\theta}_t, \ddot{v}_t]', \tag{4}$$

where x_t and y_t are instantaneous position on the horizontal plane, p_t is instantaneous pressure, θ_t is instantaneous pen trajectory tangent, v_t is instantaneous velocity.

The classifiers compared are a BN/GMM model with 30 Gaussian mixture components (929 free parameters), a two-states, 15-Gaussian components HMM (930 free parameters), and a five-states, 6-Gaussian components HMM (933 free parameters). Figure 2 shows that results are very slightly worse for the BN/GMM than the HMM classifiers at EER, but not statistically significantly so ($p = 0.05$).

On BMEC 2007, the feature vector used is $\mathbf{o}_t = [x_t, y_t, \dot{x}_t, \dot{y}_t, \ddot{x}_t, \ddot{y}_t]'$. The classifiers compared are the BN/GMM model with 20 Gaussian mixture components(259 free parameters), a two-states, 10-Gaussian components HMM (260 free parameters), and a five-states, 4-Gaussian components HMM (263 free parameters). Figure 2 shows that the difference in error rates between these models is not statistically significant ($p = 0.05$). As a comparison point, the BioSecure reference system (based on HMMs) on the same data achieves 15% EER.
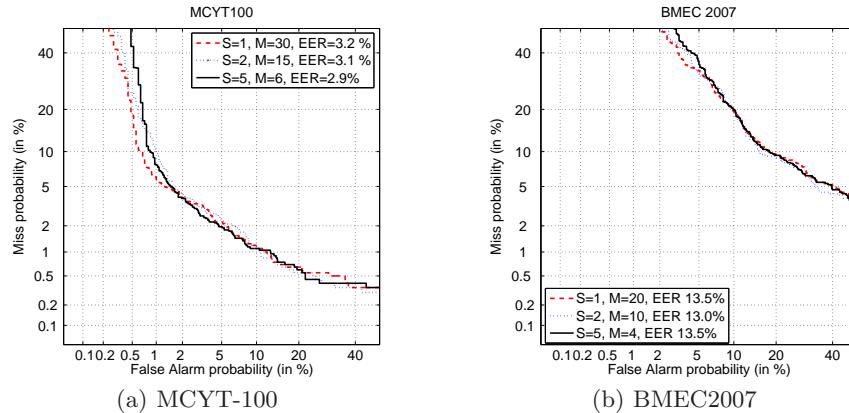

(a) MCYT-100   (b) BMEC2007

Fig. 2. Comparison between the BN/GMM model and HMM models with equivalent number of parameters (S, number of states, and M, number of Gaussian components in the mixture) on MCYT-100 and BMEC 2007.

The same trends are observed on SVC 2004 database [19], where the comparison of a 50-mixture components BN/GMM to a 2-states, 25-components per

state and a 5-states, 10 components per state HMM yield insignificantly different results ($p = 0.05$);

6 Discussion

The use of delta features in speech and signature verification using HMMs was originally motivated by the need to relax the Markov assumption between feature vectors of close states. Delta features have been a common occurence in static models of speech such as used in speaker verification. However, we think that considering the use of delta features in the light of phase space reconstruction theory yields interesting insights into their impressive effectiveness in improving discrimination between classes.

In signature verification, this can be attributed to the fact that the dynamics of the signature are unfolded by embedding into a higher-dimensional subspace, where the specific topology of each user's signature is more apparent in the phase space than when projected down to a single dimension. Since a handwritten signature is the result of a high-order graphomotor process controlled by several parameters [13], it is reasonable to attempt this reconstruction.

The better results obtained for delta coordinates over delay coordinates do not seem to stem from the choice of method used for selecting the optimal lag. We have used both the "first minimum of automutual information" and the "first zero crossing ot the autocorrelation function" heuristics to estimate the optimal lag, and arrived at generally better results with the automutual information. However, both tend to yield small lags, and the resulting reconstructed orbit in phase space is generally stretched along the identity line: the representation in phase space may be needlessly complicated. Secondly, the heuristics mentioned are not necessarily optimal for classification, as observed in [14].

Finally, comparing results with HMMs shows that, for classification tasks on signature data, static models using phase-space signal representations may perform as well as dynamical models. This can be attributed to the fact that time information is embedded in the orbit of the signal in reconstructed phase space. Indeed, for signature time series, static BN/GMM models do not perform on-par with HMMs when no time-dependent information (such as delta features or other features based on derivatives) is included in the feature space.

7 Acknowledgements

Bibliography

[1] Ly Van Bao, Sonia Garcia-Salicetti, and Bernadette Dorizzi. On using the Viterbi path along with HMM likelihood information for online signature verification. *IEEE Trans. on Systems, Man, and Cybernetics, Part B*, 37(5):1237–1247, 2007.

[2] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, October 2001.

[3] Julian Fierrez, Javier Ortega-Garcia, Daniel Ramos, and Joaquin Gonzalez-Rodriguez. HMM-based on-line signature verification: Feature extraction and signature modeling. *Pattern Recognition Letters*, 28(16):2325–2334, December 2007.

[4] Sara P. Garcia. *Phase space reconstruction of multivariate time series from molecular and field Biology*. PhD thesis, Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Oeiras, Portugal., 2005.

[5] Peter Grassberger and Itamar Procaccia. Characterization of strange attractors. *Physical Review Letters*, 50(5):346–349, Jan 1983.

[6] M. Imhoff, M. Bauer, U. Gather, and D. Löhlein. Statistical pattern detection in univariate time series of intensive care on-line monitoring data. *Intensive Care Medicine*, 24(12):1305–1314, December 1998.

[7] ISO/IEC JTC 1/SC 37 Biometrics. TR 19795-3, biometric performance testing and reporting, part 3: Modality specific testing. Technical report, International Standards Organization, 2007.

[8] R. Iyer, H. Gish, M.-H. Siu, G. Zavaliagkos, and S. Matsoukas. Hidden Markov models for trajectory modeling. In *Proc. Int. Conf. on Spoken Language Processing (ICSLP)*, pages 891–894, 1998.

[9] Matthew B. Kennel, Reggie Brown, and Henry D. I. Abarbanel. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys. Rev. A*, 45(6):3403–3411, March 1992.

[10] K. Murphy. *Dynamic Bayesian networks: representation, inference and learning*. PhD thesis, University of California at Berkeley, July 2002.

[11] J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, J. Gonzalez, M. Faundez-Zanuy, V. Espinosa, A. Satue, I. Hernaez, (J.-J.) Igarza, C. Vivaracho, D. Escudero, and (Q.-I.) Moro. MCYT baseline corpus: A bimodal biometric database. In *IEE Proceedings - Vision, Image and Signal Processing*, volume 150, pages 395–401, 2003.

[12] N.H. Packard, J.P. Crutchfield, J.D. Farmer, and R.S. Shaw. Geometry from a time series. *Physical Review Letters*, 45(9):712–716, 1980.

[13] R. Plamondon and F. J. Maarse. An evaluation of motor models of handwriting. *IEEE Trans. on Systems, Man and Cybernetics*, 19(5):1060–1072, Sept.–Oct. 1989.

[14] Richard J. Povinelli, Michael T. Johnson, Andrew C. Lindgren, Felice M. Roberts, and Jinjin Ye. Statistical models of reconstructed phase spaces for

signal classification. *IEEE Trans. on Signal Processing*, 54(6):2178–2186, June 2006.

[15] J. Richiardi and A. Drygajlo. Gaussian mixture models for on-line signature verification. In *Proc. ACM SIGMM Multimedia, Workshop on Biometrics methods and applications (WBMA)*, pages 115–122, Berkeley, USA, Nov. 2003.

[16] Jonas Richiardi and Andrzej Drygajlo. Diversity in gaussian mixture ensembles: an application to signature verification. In *Proc. First European Workshop on Biometrics and Identity Management (BIOID)*, number 5372 in LNCS. Springer, 2008.

[17] Jonas Richiardi, Hamed Ketabdar, and Andrzej Drygajlo. Local and global feature selection for on-line signature verification. In *Proc. IAPR 8th International Conference on Document Analysis and Recognition (ICDAR 2005)*, volume 2, pages 625–629, Seoul, Korea, August-September 2005.

[18] F. Takens. Detecting strange attractors in turbulence. In *Proc. Dynamical Systems and Turbulence*, Lecture Notes in Mathematics 898, pages 366–381. Springer, 1980.

[19] Dit-Yan Yeung, Hong Chang, Yimin Xiong, Susan George, Ramanujan Kashi, Takashi Matsumoto, and Gerhard Rigoll. SVC2004: First international signature verification competition. In *Proceedings 2004 Biometric Authentication: First International Conference, (ICBA 2004)*, pages 16–22, Hong Kong, China, July 2004.