Signal Processing 114 (2015) 61–74

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Jeffrey's divergence for state-space model comparison

C. Magnant [a,b,*], E. Grivel [b], A. Giremus [b], B. Joseph [a], L. Ratton [c]

[a] Thales Systemes Aeroportes S.A. 25, Avenue Gustave Eiffel, 33600 Pessac, France
[b] Université de Bordeaux—Bordeaux INP ENSEIRB-MATMECA—IMS—UMR CNRS 5218, 351, Cours de la Libération, 33405 Talence, Cedex, France
[c] Thales Systemes Aeroportes S.A. 2, Avenue Gay-Lussac, 78990 Elancourt, France

ARTICLE INFO

Article history:
Received 15 July 2014
Received in revised form
26 December 2014
Accepted 8 February 2015
Available online 18 February 2015

Keywords:
Kalman filtering
Model selection
Jeffrey's divergence
Classification
Tracking

ABSTRACT

Optimal filters such as Kalman filters are used in a wide range of applications from speech enhancement to biomedical applications. They are based on an *a priori* state-space model describing a dynamic system. If this model is not well-suited, the accuracy of the state vector may be poor. Therefore, several estimators based on different models can be combined. However, state-space models that are dissimilar enough must be chosen. To our knowledge, there are no guidelines to select them, we hence address this issue in this paper. Given an initial model set, our aim is to determine subsets of similar models by using Jeffrey's divergence between the distributions of the state-vector time paths based on the different models. Our approach operates with the following steps: the so-called dissimilarity matrix composed of Jeffrey's divergences between model pairs is created. Then, this matrix is transformed to get the same properties as a correlation matrix and an eigenvalue decomposition is performed. Subsequently, we propose an interpretation of the predominant eigenvalues which is then used to deduce the number of model subsets and their cardinals. A classification algorithm can then be considered to determine which models belong to which subsets.

1. Introduction

In the field of signal processing, a common problem is to recursively estimate a hidden process $\{x_n\}_{n \in \mathbb{N}}$ from an observed process $\{y_n\}_{n \in \mathbb{N}}$. In the framework of radar processing, once a target has been detected, tracking requires estimating its position in a specific coordinate system by using noisy radar information. Another example is GPS navigation where the aim is to locate the GPS receiver given the positions of the satellites in view and the noisy pseudo-range measurements to the satellites. Once again, the GPS positions have to be estimated. In video sequences for traffic control, moving objects have also to be tracked. To address the above issues, Bayesian approaches such as Kalman filters [1] and particle filters [2] are commonly used. Their performance in terms of location accuracy is mainly related to the choice of the motion model that is *a priori* used [3]. Although a constant velocity (CV) motion [4,5] often describes the motion of a tanker for instance, some model parameters such as the variance of the measurement noise are not necessarily known. Therefore, some model parameters have to be adjusted [3]. When tracking maneuvering targets, the motion is inherently uncertain. The motion class can change over time; for instance, the object can first follow a CV model [4,5], then a uniformly accelerated motion (UAM) model [4,5], then a Singer model [6]. In addition, kinematic parameters such as the target-jerk variance, the acceleration or the velocity may vary over time. In these cases, using a single estimator based on a unique state[1]-space representation (SSR) of the system [7] leads to poor performance.

* Corresponding author.
E-mail addresses: clement.magnant@fr.thalesgroup.com (C. Magnant), eric.grivel@ims-bordeaux.fr (E. Grivel), audrey.giremus@ims-bordeaux.fr (A. Giremus), bernard.joseph@fr.thalesgroup.com (B. Joseph), laurent.ratton@fr.thalesgroup.com (L. Ratton).

[1] A state of a system stores the current values of the internal elements of the system. The state variables that are in the state vector depend on the application. When dealing with a motion model, the state vector can be composed of the target position, the velocity and the acceleration.

http://dx.doi.org/10.1016/j.sigpro.2015.02.006

More generally in signal processing, defining the SSR of the system can be a difficult task. To avoid the above constraints, multiple-model (MM) based methods can be considered. Three generations exist the first of which uses a finite number of estimators associated to different SSRs, but there is no interaction between them. In the second generation which gave rise to the interactive multiple model (IMM) [8], a cooperation strategy is included at the input of the MM structure [9]. Although the IMM has become popular over the last decades, defining the number of models to be used and selecting their SSRs are still open topics. Usually, the user aims to choose dissimilar models to ensure diversity [5]. Nevertheless, it was shown in [10] that using too many models does not necessarily improve the performance of a single-Kalman-filter based estimator. In the third generation, the main feature is that the model number varies over time. This leads to the variable-structure IMM (VS-IMM) and its variants [11]. In that case, unlikely models are continually removed while more relevant are activated. Therefore, given an initial set of representative models of the system, the user has to define model subsets and then deduce adjacencies[2] between models. Evaluating similarities between models thus plays a crucial role in this lastest generation. A preliminary study to classify models is hence of interest.

In this paper, unlike the approaches based on the Akaike informationcriterion (AIC) [12,13], the Bayesian information criterion (BIC) [14] and the divergence information criterion (DIC) [15], our purpose is not to develop a data-driven selection of the system's SSR. The originality of our work is that the dissimilarity is addressed in terms of distributions on the whole paths and not in terms of the paths themselves. Indeed, we compare models and not some of their realizations. For this purpose, our approach is based on Jeffrey's divergence (JD) which is the symmetric version of the Kullback–Leibler (KL) divergence [16,17]. Note that the KL divergence has already been proposed to compare Gaussian processes as described in [18], but only asymptotic expressions are obtained. Recent works also derive an expression of this divergence for the specific cases of first-order moving-average (MA) models [19] and Markov sources [20]. Our work addresses the more general case of linear Gaussian state spaces. In [21], we derived a recursive expression of the JD between the joint distributions of the state vector at successive instants based on two distinct models in order to measure the degree of mismatch between both state models. In this paper, our contributions are the following:

1. We aim to evaluate the dissimilarities between two or more state models. Thus, considering a given set of models, JDs between models that are chosen pairwise can be computed and organized in a matrix form to create the so-called dissimilarity matrix [22]. However, the number of models to be compared can be high. In that case, there are too many JDs to be studied simultaneously. Therefore, referring to various results analyzing the correlation matrices,[3] the dissimilarity matrix is transformed to obtain the same properties as a correlation matrix. Then, at each instant, the eigenvalues of the resulting matrix are computed in order to study how they evolve over time. Given an interpretation of these eigenvalues, we deduce the number of model subsets and their cardinals from the initial model set. Subsequently, a classification algorithm is used to determine which models belong to which model subsets.
2. Examples are given in the field of target tracking but others could be considered. These examples can be useful when selecting a reduced number (e.g. 2–3) of competing dissimilar models, before using optimal filters in an MM structure.

The outline of this paper is as follows: in Section 2, background information is recalled. More specifically, the linear SSR of the system is introduced. In Section 3, we derive the recursive expression of the JD between two state models. In Section 4, we extend our approach to a set of two or more models. In Section 5, we analyze its relevance in various cases. A toy example consists in comparing two random-walk processes. Then, we apply our method to object tracking. For this latter application, the SSRs for the UAM and the Singer motion are recalled and both model families are compared. Finally, starting from a given initial model set composed of both previous model families, model subsets are determined and the relevance of this classification is then studied regarding an MM-based algorithm. Conclusions and perspectives are finally drawn in Section 6.

2. Background

2.1. About the state-space representation

In the fields of control and signal processing, Bayesian recursive approaches can be used. They are based on the SSR of the system, which is a mathematical model of the dynamic system [24,25]. It usually consists of two equations[4]: one related to the evolution model, whereas the other describes the relationship between the state vector and the observation vector. In the following, let us focus on the linear case. If M models are considered, the ith model is defined as follows:

$$\boldsymbol{x}_{k+1} = \Phi^i \boldsymbol{x}_k + \boldsymbol{u}_k^i \tag{1}$$

$$\boldsymbol{y}_k = H^i \boldsymbol{x}_k + \boldsymbol{b}_k^i \tag{2}$$

where $\boldsymbol{x}_k$ is the state vector at time k of size l and Φ^i is the transition matrix. The model noise $\boldsymbol{u}_k^i$ is zero-mean white and Gaussian with covariance matrix Q^i. Here, Q^i is assumed to be invertible. $\boldsymbol{y}_k$ denotes the observation and H^i the observation matrix. In addition, $\boldsymbol{b}_k^i$ is a zero-mean Gaussian white noise with covariance matrix R^i and is uncorrelated with $\boldsymbol{u}_k^i$.

Note that an alternative probabilistic formulation of Eqs. (1) and (2) is[5]

$$p^i(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k) = \mathcal{N}(\Phi^i \boldsymbol{x}_k, Q^i) \tag{3}$$

[2] A model is said adjacent to another if the transition probability from this model to the other is different from zero.

[3] Namely the Karhunen–Loeve transform, the subspace decomposition methods based on the eigenvalue decomposition of the correlation matrix [23].

[4] When dealing with H_∞ filtering [26], a third equation is considered, namely $\boldsymbol{z}_k = L^i \boldsymbol{x}_k$, where L^i denotes a matrix combining state variables.

[5] $\mathcal{N}(m, C)$ denotes the Gaussian distribution with mean m and covariance matrix C.