Information Sciences 412–413 (2017) 50–66

Contents lists available at ScienceDirect

Information Sciences

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

A hidden Markov model with dependence jumps for predictive modeling of multidimensional time-series

Anastasios Petropoulos[a], Sotirios P. Chatzis[a,*], Stelios Xanthopoulos[b]

[a] Department of Electrical Eng., Computer Eng., and Informatics, Cyprus University of Technology, 33 Saripolou Str., Limassol 3036, Cyprus
[b] Department of Mathematics, University of the Aegean, Samos, Greece

A R T I C L E I N F O

Article history:
Received 1 December 2016
Revised 16 May 2017
Accepted 22 May 2017
Available online 25 May 2017

Keywords:
Temporal dynamics
Hidden Markov models
Expectation-maximization
Variable order
Dependence jumps

A B S T R A C T

Hidden Markov models (HMMs) are a popular approach for modeling sequential data, typically based on the assumption of a first- or moderate-order Markov chain. However, in many real-world scenarios the modeled data entail temporal dynamics the patterns of which change over time. In this paper, we address this problem by proposing a novel HMM formulation, treating temporal dependencies as latent variables over which inference is performed. Specifically, we introduce a hierarchical graphical model comprising two hidden layers: on the first layer, we postulate a chain of latent observation-emitting states, the temporal dependencies between which may change over time; on the second layer, we postulate a latent first-order Markov chain modeling the evolution of temporal dynamics (dependence jumps) pertaining to the first-layer latent process. As a result of this construction, our method allows for effectively modeling non-homogeneous observed data, where the patterns of the entailed temporal dynamics may change over time. We devise efficient training and inference algorithms for our model, following the expectation-maximization paradigm. We demonstrate the efficacy and usefulness of our approach considering several real-world datasets.

1. Introduction

Modeling sequential data continues to be a fundamental task and a key challenge in the field of machine learning, encountered in a plethora of real-world applications, including bioinformatics, document analysis, financial engineering, speech processing, and computer vision, to name just a few. In this paper, we focus on the problem of *sequence prediction*, dealing with *continuous*, possibly *high-dimensional* observations (time-series). Machine learning literature comprises a rather extensive corpus of proposed prediction algorithms for sequences of continuous observations. Among them, the hidden Markov model (HMM) is one of the most popular methods, used in a great variety of application contexts. This popularity is mainly due to the fact that HMMs are flexible enough to allow for modeling complex temporal patterns and structures in sequential data. Specifically, HMMs are popular for their provision of a convenient way of modeling observations appearing in a sequential manner and tending to cluster or to alternate between different possible components (subpopulations) [10,35].

Most popular HMM formulations are based on the postulation of first-order Markovian dependencies; in other words, only one-step-back temporal dynamics are considered. Such an assumption allows for increased simplicity and low computational complexity of the resulting model training and inference algorithms. However, postulating first-order temporal

* Corresponding author.
E-mail address: sotirios.chatzis@eecei.cut.ac.cy (S.P. Chatzis).

http://dx.doi.org/10.1016/j.ins.2017.05.038
0020-0255/

dynamics does also entail ignoring the possibility of the modeled data comprising longer temporal dynamics. Even though this assumption might be valid in some cases, it is well-known to be unrealistic in several application scenarios, including handwriting recognition, molecular biology, speech recognition, and volatility prediction in financial return series, thus undermining the modeling effectiveness.

To resolve this problem, several researchers have attempted to introduce HMM-type models with higher-order dependencies. Characteristic examples are the methods presented in [29] and [30], with successful applications to the problem of speech recognition, the method presented in [33], applied to handwriting recognition, the method of [21], designed to address challenges related to pattern recognition tasks in molecular biology, and the method presented in [1], which was successfully applied to the field of robotics. However, a major drawback of such higher-order HMM approaches is their considerably increased computational costs, which become rather prohibitive as model order increases. An effort to ameliorate these issues of higher-order HMMs is presented in [22]. In that work, instead of directly training Rth order HMMs on the data, a method of fast incremental training is used that progressively trains HMMs from first to Rth order.

Note, though, that using higher-order HMMs gives rise to a source of significant burden for researchers and practitioners, namely the need to determine the most appropriate order for the postulated models. This procedure entails fitting multiple models to the available data to choose from, and application of some validation procedure, which, apart from computationally cumbersome, is also likely to become prone to overfitting [32]. Finally, another limitation of the existing higher-order HMM formulations concerns their static and homogeneous assumptions, i.e. their consideration that the temporal dynamics order in the modeled data does not change over time. Indeed, sequential data with variable order in the entailed temporal dynamics are quite often encountered in real-world application scenarios [2,11,13,19]. Therefore, allowing for capturing more complex structure of temporal dynamics in the modeled data, where effective dependencies may change over time as a result of dynamic switching between different temporal patterns, is expected to result in much better modeling and predictive performances. Indeed, previous works such as [8] and [15] have proven the efficacy of postulating simple variable-order Markov chains in diverse application settings. However, development of a variable-dynamics HMM-type model has not yet been considered in the literature.

To address these problems of conventional higher-order HMMs, some researchers have proposed appropriate models with variable order Markovian dynamics assumptions. For instance, a variable order Markov model is presented in [2] to address the problem of prediction of discrete sequences over a finite alphabet; the method is successfully applied to three different domains, namely English text, music pieces, and proteins (amino-acid sequences). More recently, [19] presented a simple and effective generalization of variable order Markov models to full online Bayesian estimation. Generalization of variable order Markov models in this context enables perpetual model improvement and enrichment of the learned temporal patterns by accumulation of observed data, without any need for human intervention. Despite these merits, a drawback of both these approaches concerns their inability to model sequential data comprising continuous observations, i.e. sequences each frame of which is a (probably high-dimensional) D-dimensional vector of real values, defined in $\mathbb{R}^D$. Finally, [52] propose a two-stage modeling approach towards variable order HMMs: the first stage consists in discovering repetitive temporal patterns of variable length, while the second stage consists in performing prediction by means of a separate simple HMM fit to the temporal pattern determined to be relevant at each specific time point. Similar to the previous approaches, a major limitation of [52] consists in its incapability to model sequential observations taking *continuous* values in $\mathbb{R}^D$.

In this paper, we address the aforementioned shortcomings, by introducing an HMM variant capable of capturing *jumps* in the temporal *dependence patterns* of modeled sequential data. Specifically, we introduce a hierarchical graphical model comprising two hidden layers: on the*first layer*, we postulate a *chain* of *latent observation-emitting states,* the*dependencies* between which may*change over time*; on the *second layer,* we postulate a *latent first-order Markov chain* modeling the *evolution* of temporal dynamics (*dependence jumps*) pertaining to the first-layer latent process. As a result of this construction, our model allows for effectively modeling non-homogeneous observed data, where the patterns of temporal dependencies may change over time. To allow for tractable training and inference procedures, our model considers *temporal dependencies* taking the form of *variable dependence jumps*, the order of which is *inferred* from the data as part of the model inference procedure.

Our method is designed to allow for modeling *both* discrete and continuous observations; it allows for capturing seasonal effects in the modeled sequences, and enhances modeling in the implied autocorrelation structure of the observed sequences. In addition, contrary to the related methods of [11] and [13], our method does *not* require utilization of any kind of approximation to perform model training and inference. Indeed, both model training and inference can be performed *exactly* and in a computationally efficient way, using elegant algorithms derived under the expectation-maximization paradigm [18]. We demonstrate the efficacy of our approach considering real-world application scenarios.

The remainder of this paper is organized as follows: In Section 2, we introduce our proposed model and derive its training and inference algorithms. In Section 3, we experimentally evaluate our approach, and exhibit its advantages over existing approaches. Finally, in Section 4 we conclude this paper, summarizing and discussing our results.