Explicit-Duration Markov Switching Models

Silvia Chiappa
Statistical Laboratory, University of Cambridge, UK
and
Microsoft Research Cambridge, UK
chiappa.silvia@gmail.com
Editorial Scope

Topics

Foundations and Trends® in Machine Learning publishes survey and tutorial articles on the theory, algorithms and applications of machine learning, including the following topics:

- Adaptive control and signal processing
- Applications and case studies
- Behavioral, cognitive, and neural learning
- Bayesian learning
- Classification and prediction
- Clustering
- Data mining
- Dimensionality reduction
- Evaluation
- Game theoretic learning
- Graphical models
- Independent component analysis
- Inductive logic programming
- Kernel methods
- Markov chain Monte Carlo
- Model choice
- Nonparametric methods
- Online learning
- Optimization
- Reinforcement learning
- Relational learning
- Robustness
- Spectral methods
- Statistical learning theory
- Variational inference
- Visualization

Information for Librarians

Foundations and Trends® in Machine Learning, 2014, Volume 7, 6 issues. ISSN paper version 1935-8237. ISSN online version 1935-8245. Also available as a combined paper and online subscription.
Explicit-Duration Markov Switching Models

Silvia Chiappa\textsuperscript{1}
Statistical Laboratory, University of Cambridge, UK
and
Microsoft Research Cambridge, UK
chiappa.silvia@gmail.com

\textsuperscript{1}The author is currently at Google DeepMind, UK.
# Contents

1 Introduction 2

2 Background 7
   2.1 Hidden Markov Model 8
   2.2 Graphical Models and Belief Networks 10
   2.3 Expectation Maximization 16

3 Explicit-Duration Modelling 18
   3.1 Decreasing Count Variables 21
   3.2 Increasing Count Variables 21
   3.3 Count-Duration Variables 23
   3.4 Explicit-Duration MSMs $p(\sigma_{1:T}, v_{1:T})$ 24
   3.5 Explicit-Duration SLGSSM 38
   3.6 Approximations 60

4 Discussion 64

Acknowledgements 67
Abstract

Markov switching models (MSMs) are probabilistic models that employ multiple sets of parameters to describe different dynamic regimes that a time series may exhibit at different periods of time. The switching mechanism between regimes is controlled by unobserved random variables that form a first-order Markov chain. Explicit-duration MSMs contain additional variables that explicitly model the distribution of time spent in each regime. This allows to define duration distributions of any form, but also to impose complex dependence between the observations and to reset the dynamics to initial conditions. Models that focus on the first two properties are most commonly known as hidden semi-Markov models or segment models, whilst models that focus on the third property are most commonly known as changepoint models or reset models. In this monograph, we provide a description of explicit-duration modelling by categorizing the different approaches into three groups, which differ in encoding in the explicit-duration variables different information about regime change/reset boundaries. The approaches are described using the formalism of graphical models, which allows to graphically represent and assess statistical dependence and therefore to easily describe the structure of complex models and derive inference routines. The presentation is intended to be pedagogical, focusing on providing a characterization of the three groups in terms of model structure constraints and inference properties. The monograph is supplemented with a software package that contains most of the models and examples described. The material presented should be useful to both researchers wishing to learn about these models and researchers wishing to develop them further.

---


DOI: 10.1561/2200000054.

1More information about the package is available at www.nowpublishers.com.
Markov switching models (MSMs) are probabilistic models that employ multiple sets of parameters to describe different dynamic regimes that a time series may exhibit at different periods of time. The switching mechanism between regimes is controlled by unobserved variables that form a first-order Markov chain.

MSMs are commonly used for segmenting time series or to retrieve the hidden dynamics underlying noisy observations.

Consider, for example, the time series displayed in Figure 1.1(a), which corresponds to the measured leg positions of an individual performing repetitions of the actions low/high jumping and hopping on the left/right foot. A segmentation of the time series into the underlying actions could be obtained with a MSM in which each action forms a separate regime, e.g. by computing the regimes with highest posterior probabilities.

As another example, consider the time series displayed with dots in Figure 1.1(b), which corresponds to noisy observations of the positions of a two-wheeled robot moving in the two-dimensional space according to straight movements, left-wheel rotations and right-wheel rotations.\footnote{This example is discussed in detail in §3.5.3}
Figure 1.1: (a): Body-marker recording of an individual performing repetitions of the actions low jumping up and down, high jumping up and down, hopping on the left foot and hopping on the right foot (CMU Graphics Lab Motion Capture Database). (b): Actual positions (continuous line) and measured positions (dots) of a two-wheeled robot moving in the two-dimensional space. The initial actual position is indicated with a star.

The actual positions are displayed with a continuous line. Denoised estimates of the positions could be obtained with a MSM in which the robot movements are described with continuous unobserved variables and in which each type of movement forms a separate regime, e.g. by computing the posterior means of the continuous variables.

In standard MSMs, the regime variables implicitly define a geometric distribution on the time spent in each regime. In explicit-duration MSMs, this constraint is relaxed by using additional unobserved variables that allow to define duration distributions of any form. Explicit-duration variables also allow to impose complex dependence between the observations and to reset the dynamics to initial conditions.

Explicit-duration MSMs were first introduced in the speech community [Ferguson, 1980] and are mostly used to achieve more powerful modelling than standard MSMs through the specification of more accurate duration distributions and dependencies between the observations. In this case, the models are most commonly known with the names of

\[\text{This example is discussed in detail in §3.5 and in Appendix A.4}\]
hidden semi-Markov models or segment models. However, the possibility to reset the dynamics to initial conditions has recently led to the use of explicit-duration variables also for Bayesian approaches to abrupt-change detection, for identifying repetitions of patterns (such as, e.g., the action repetitions underlying the time series in Figure 1.1(a)), and for performing/approximating inference\(^3\)\cite{Fearnhead, Fearnhead and Vasileiou, Chiappa and Peters, Bracegirdle and Barber}. In these cases, the models are most commonly known with the names of changepoint models or reset models.

Explicit-duration MSMs have been used in many application areas including speech analysis\cite{Russell and Moore, Levinson, Rabiner, Gu et al., Gales and Young, Ostendorf et al., Moore and Savic, Liang et al., Chen et al., Gales and Young, Ostendorf et al., Moore and Savic, Liang et al., Yu}, handwriting recognition\cite{Chen et al., Mooney, Gales and Young, Ostendorf et al., Moore and Savic, Liang et al., Chen et al., Gales and Young, Ostendorf et al., Moore and Savic, Liang et al., Yu}, activity recognition\cite{Yu, Kobayashi, Huang et al., Oh et al., Chiappa and Peters, Yu}, musical pattern recognition\cite{Pikrakis et al., Pikrakis et al., Yu}, financial time series analysis\cite{Bulla and Bulla, Bulla and Bulla}, rainfall time series analysis\cite{Sansom and Thomson, Sansom and Thomson}, protein structure segmentation\cite{Schmidtler et al., Schmidtler et al.}, gene finding\cite{Winters-Hilt et al., Winters-Hilt et al.}, DNA analysis\cite{Barbu and Limnios, Fearnhead and Vasileiou, Fearnhead and Vasileiou, Fearnhead and Vasileiou}, plant analysis\cite{Guédon et al., Guédon et al.}, MRI sequence analysis\cite{Faisan et al., Faisan et al.}, ECG segmentation\cite{Hughes et al., Hughes et al.}, and waveform modelling\cite{Kim and Smyth, Kim and Smyth, Kim and Smyth, Kim and Smyth}; see references in Yu\cite{Yu} for more examples.

Explicit-duration MSMs originated from the idea of explicitly modelling the duration distribution by defining a semi-Markov process on the regime variables, namely a process in which the trajectories are piecewise constant functions – with interval durations drawn from an explicitly defined duration distribution – and in which the variables at jump times form a Markov chain. The first and currently standard approach achieves that with variables indicating the interval duration, and derives inference recursions using only jump times\cite{Rabiner, Rabiner, Gales and Young, Ostendorf et al., Ostendorf et al., Yu}. To simplify the derivations of posterior distributions at times that are different

\(^3\)By inference we mean the computation of posterior distributions, namely distributions of unobserved variables conditioned on the observations.
from jump times, Chiappa and Peters [2010] use count variables in addition to duration variables, such that the combined regime and count-duration variables form a first-order Markov chain. Other methods that explicitly model the duration distribution have been proposed with different goals and in different communities. These methods can all be viewed as different ways to define a first-order Markov chain on the combined regime and explicit-duration variables that induces a semi-Markov process on the regime variables.

In this monograph we provide a description of explicit-duration modelling that aims at elucidating the characteristics of the different approaches and at clarifying and unifying the literature. We identify three fundamentally different ways to define the first-order Markov chain on the combined regime and explicit-duration variables, which differ in encoding in the explicit-duration variables the location of (i) the preceding, (ii) the following, or (iii) both the preceding and following regime change or reset. We discuss each encoding in the context of MSMs of simple unobserved structure and of MSMs that contain extra unobserved variables related by first-order Markovian dependence. The models are described using the formalism of graphical models, which allows to graphically represent and assess statistical dependence, and therefore to easily describe the structure of complex models and derive inference routines.

The remainder of the manuscript is organized as follows. Chapter 2 contains some background material. We start with a general description of MSMs and by showing that the regime variables implicitly define a geometric duration distribution. In §2.1 we introduce the hidden Markov model, which represents the simplest MSM, and explain how to obtain a negative binomial duration distribution with regime copies. In §2.2 we introduce the framework of graphical models, and explain how to graphically assess statistical independence in a particular type of graphical models, called belief networks, that will be used for describing the models. In §2.2.1 we illustrate how belief networks can be used to easily derive the standard inference recursions of MSMs. In §2.3 we give a general explanation of the expectation maximization algorithm, which represents the most popular algorithm for parameter
learning in probabilistic models with unobserved variables. In Chapter 3 we describe the different approaches to explicit-duration modelling by categorizing them into three groups. The groups are introduced in §3.1, §3.2 and §3.3. In §3.4 we discuss in detail explicit-duration modelling in MSMs containing only regime variables, explicit-duration variables, and observations. In §3.5 we discuss in detail explicit-duration modelling in a popular MSM containing additional unobserved variables related by first-order Markovian dependence, namely the switching linear Gaussian state-space model, and discuss how the findings generalize to similar models with unobserved variables related by first-order Markovian dependence. The case of more complex unobserved structure is not considered. In §3.6 we describe approximation schemes to reduce the computational cost of inference. In Chapter 4 we summarize the most important points of our exposition and make some historical remarks.
References


References


References


