

Inertial Hidden Markov Models: Modeling Change in Multivariate Time Series

George D. Montañez

Machine Learning Department
Carnegie Mellon University
Pittsburgh, PA USA
gmontane@cs.cmu.edu

Saeed Amizadeh

Yahoo Labs
Sunnyvale, CA USA
amizadeh@yahoo-inc.com

Nikolay Laptev

Yahoo Labs
Sunnyvale, CA USA
nlaptev@yahoo-inc.com

Abstract

Faced with the problem of characterizing systematic changes in multivariate time series in an unsupervised manner, we derive and test two methods of regularizing hidden Markov models for this task. Regularization on state transitions provides smooth transitioning among states, such that the sequences are split into broad, contiguous segments. Our methods are compared with a recent hierarchical Dirichlet process hidden Markov model (HDP-HMM) and a baseline standard hidden Markov model, of which the former suffers from poor performance on moderate-dimensional data and sensitivity to parameter settings, while the latter suffers from rapid state transitioning, over-segmentation and poor performance on a segmentation task involving human activity accelerometer data from the UCI Repository. The regularized methods developed here are able to perfectly characterize change of behavior in the human activity data for roughly half of the real-data test cases, with accuracy of 94% and low variation of information. In contrast to the HDP-HMM, our methods provide simple, drop-in replacements for standard hidden Markov model update rules, allowing standard expectation maximization (EM) algorithms to be used for learning.

Introduction

“Some seek complex solutions to simple problems; it is better to find simple solutions to complex problems.” - Soramichi Akiyama

Time series data arise in different areas of science and technology, describing the behavior of both natural and man-made systems. These behaviors are often quite complex with uncertainty, which in turn require us to incorporate sophisticated dynamics and stochastic representations to model them. Furthermore, these complex behaviors can *change* over time due to some external event and/or some internal systematic change of dynamics/distribution. For example, consider the case of monitoring one’s physical activity via an array of accelerometer body sensors over time. A certain pattern emerges on the time series of the sensors’ readings while the person is walking; however, this pattern quickly changes to a new one as she begins running. From the data analysis perspective, it is important to first

detect these *change points* as they are quite often indicative of an “interesting” event or an anomaly in the system. We are also interested in characterizing the new *state* of the system (e.g. running vs. walking) which reflects its mode of operation. Change point detection methods (Kawahara, Yairi, and Machida 2007; Xie, Huang, and Willett 2013; Liu et al. 2013; Ray and Tsay 2002) have been proposed to address the first challenge while Hidden Markov Models (HMMs) can address both.

One crucial observation in many real-world systems, natural and man-made, is that the behavioral changes are typically infrequent; that is, the system takes some (unknown) time before it changes its behavior to a new *modus operandi*. For instance, in our earlier example, it is unlikely for a person to rapidly fluctuate between walking and running, making the durations of different activities over time relatively long and highly variable. We refer to this as the *inertial property*, alluding to the physical property of matter such that it continues along a fixed course unless acted upon by an external force. Unfortunately, classical HMMs are not equipped with sufficient mechanisms to capture this property and often produce high rates of state transitioning with subsequent false positives when detecting change points.

Few solutions exist in the literature to address this problem. In the context of Markov models, Fox *et al.* (Fox et al. 2011; Willsky et al. 2009) have recently proposed the *sticky hierarchical Dirichlet process hidden Markov model (HDP-HMM)* which uses a Bayesian non-parametric approach with appropriate priors to promote self-transitioning (or *stickiness*) for HMMs. Despite its elegant theoretical foundation, the sticky HDP-HMM is not a practical solution in many real-world situations. In particular, the performance of the HDP-HMM tends to degrade as the dimensionality of the problem increases beyond ten dimensions. Moreover, due to iterative Gibbs sampling for its learning, the sticky HDP-HMM can become computationally prohibitive. In practice, the most significant drawback of the sticky HDP-HMM originates with its non-parametric Bayesian nature: due to the existence of many hyperparameters, the search space for initial tuning is exponentially large and significantly affects the learning quality for a given task.

In this paper, we propose a regularization-based framework for HMMs, called *Inertial hidden Markov models* (Inertial HMMs), which are biased towards the inertial state-

cient training data, again without an emphasis on inertial transitioning between states. Similarly, Johnson (Johnson 2007) proposed using Dirichlet priors on multinomial hidden Markov models as a means of enforcing sparse emission distributions.

Fox *et al.* (Fox et al. 2011) developed a Bayesian sticky HMM to provide inertial state persistence. They presented a method capable of learning a hidden Markov model without specifying the number of states or regularization-strength beforehand, using a hierarchical Dirichlet process and truncated Gibbs sampling. As discussed, their method requires a more complex approach to learning the model and specification of several hyperparameters for the Bayesian priors along with a truncation limit. In contrast, our models only require the specification of two parameters, K and ζ , whereas the sticky HDP-HMM requires analogous truncation level L and κ parameters to be chosen, in addition to the hyperparameters on the model priors.

Conclusions

For modeling changes in multivariate time series data, we introduce two modified forms of hidden Markov models that effectively enforce state persistence. Although the derived methods are simple, they perform well and are computationally tractable. We have shown that inertial models are easily implemented, add almost no additional computation cost, run efficiently, and work well on data with moderate dimensions. Their simplicity is thus a feature and not a bug.

Furthermore, a simple method was developed for automated selection of each regularization parameter. Our experiments on synthetic and real-world data show the effectiveness of inertial HMMs, giving large improvements in performance over standard HMMs and the sticky HDP-HMM.

The simplicity of our models pave the way for natural extensions, such as incremental parameter learning and changing the form of the class conditional emission distributions to incorporate internal dynamics. Such extensions are the focus of future work.

Acknowledgments

The authors would like to thank Emily Fox and Erik Sudderth for their discussions, feedback and assistance with use of the HDP-HMM toolbox. GDM is supported by Yahoo Labs and the National Science Foundation Graduate Research Fellowship under Grant No. 1252522. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors alone and do not necessarily reflect the views of the National Science Foundation or any other organization.

References

Altun, K.; Barshan, B.; and Tunçel, O. 2010. Comparative study on classifying human activities with miniature inertial and magnetic sensors. *Pattern Recogn.* 43(10):3605–3620.

Bishop, C. M. 2007. *Pattern Recognition and Machine Learning*. Springer. 616–625.

Fox, E. B., and Sudderth, E. B. 2009. HDP-HMM Toolbox. <https://www.stat.washington.edu/~ebfox/software.html>. [Online; accessed 20-July-2014].

Fox, E. B.; Sudderth, E. B.; Jordan, M. I.; Willsky, A. S.; et al. 2011. A sticky HDP-HMM with application to speaker diarization. *The Annals of Applied Statistics* 5(2A):1020–1056.

Gauvain, J.-l., and Lee, C.-h. 1994. Maximum A Posteriori Estimation for Multivariate Gaussian Mixture Observations of Markov Chains. *IEEE Transactions on Speech and Audio Processing* 2:291–298.

Gini, C. 1936. On the measure of concentration with special reference to income and statistics. In *Colorado College Publication*, number 208 in General Series, 73–79.

Johnson, M. 2007. Why doesn't EM find good HMM POS-taggers. In *EMNLP*, 296–305.

Kawahara, Y.; Yairi, T.; and Machida, K. 2007. Change-point detection in time-series data based on subspace identification. In *Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on*, 559–564. IEEE.

Liu, S.; Yamada, M.; Collier, N.; and Sugiyama, M. 2013. Change-point detection in time-series data by relative density-ratio estimation. *Neural Networks* 43:72–83.

Meilä, M. 2003. Comparing clusterings by the variation of information. In Schölkopf, B., and Warmuth, M. K., eds., *Learning Theory and Kernel Machines*, volume 2777 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg. 173–187.

Neukirchen, C., and Rigoll, G. 1999. Controlling the complexity of HMM systems by regularization. *Advances in Neural Information Processing Systems* 737–743.

Rabiner, L. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* 77(2):257–286.

Ray, B. K., and Tsay, R. S. 2002. Bayesian methods for change-point detection in long-range dependent processes. *Journal of Time Series Analysis* 23(6):687–705.

Wikipedia. 2014. Gini coefficient — Wikipedia, the free encyclopedia. [Online; accessed 8-June-2014].

Willsky, A. S.; Sudderth, E. B.; Jordan, M. I.; and Fox, E. B. 2009. Nonparametric Bayesian learning of switching linear dynamical systems. In *Advances in Neural Information Processing Systems*, 457–464.

Xie, Y.; Huang, J.; and Willett, R. 2013. Change-point detection for high-dimensional time series with missing data. *Selected Topics in Signal Processing, IEEE Journal of* 7(1):12–27.