# Sticky HDP-HMM for Prediction of Economic Events

**Jacob Mathew and David Allen**
HRL Laboratories, LLC
3011 Malibu Canyon Road
Malibu, CA 90265
dlallen@hrl.com

## 1 Introduction

This research seeks to generate temporal event predictions using the sticky Hierarchical Dirichlet Process - Hidden Markov Model (sticky HDP-HMM) [2], a generalization of the infinite HMM [1]. Hidden Markov Models (HMMs) are one of the most widely used machine learning techniques for analyzing temporal data. One significant limitation of this traditional approach is that the number of states in the HMM, $N$, has to be decided a priori, but for a number of applications it is not possible to hypothesize this accurately. The nonparametric Bayesian solution [3] to this is to remove the dependence on $N$ by effectively making it infinite and specifying a prior over it; such as done in the HDP-HMM model [4]. An extension to the HDP-HMM model, known as the sticky HDP-HMM model [2] additionally contains a bias towards self-transitions.

In [2] the sticky HDP-HMM is introduced with application to speaker diarization, where audio recordings are segmented based on the speaker and simultaneously identify the number of speakers. For that application inference is done to calculate the posterior of the hidden states corresponding to the observations, where the hidden states are interpreted to correspond to the various speakers. The goal is then to identify the number of speakers and which speaker corresponds to each observation.

We believe that another interesting question that can be asked about the HDP-HMM is how effective it is for making predictions for future observations. We consider in this paper an application of HDP-HMM to the prediction of events based on stock market indexes. Economic predictions are a well-researched topic and an area where successful predictions have large impact, so the ability of a model do well in this domain is very significant. We show that HDP-HMM can be successfully applied in such prediction tasks and that the utility of the model extends beyond the inference of hidden states. In addition, we believe that the predictive power of the model is not fully exploited by the use of only economic data to predict economic events. In extensions to this work we are incorporating significantly more dimensions to the input data by using indicators from non-traditional sources such as web search trends, news articles, and social media (such as Tweets and blog posts).

## 2 Background

An HMM is a generative model of a sequence of observations that hypothesizes that every observation (or emission) in the sequence is generated from an underlying discrete hidden state that can transition between time steps. Two main assumptions are made about the dynamics of the process. First, the observation at a given point in the sequence is dependent on only the hidden state at that point. Second, the hidden state at a time $t+1$ is independent of the hidden states up to time $t-1$, give the state at time $t$. The HMM requires that the hidden states belong to a discrete set, but the observations emitted by the hidden states can either be discrete or continuous. For the rest of this report we restrict the discussion to only continuous observations.

A Bayesian approach to learning an HMM requires defining the number of hidden states $N$. The dependence on the number of states $N$ is avoided in the HDP-HMM by making $N$ infinite. Conceptually this means that the transition matrix is infinite, and a prior is placed over these infinite matrices. The basic tool that is used in the HDP-HMM to achieve this is the Dirichlet Process. The Dirichlet Process is a stochastic process that defines a distribution. It has two parameters, $\alpha$ the "concentration" parameter and $H$, a base distribution and is denoted as $DP\left(H,\alpha\right)$. A draw from $DP\left(H,\alpha\right)$ is itself a distribution with the same support as $H$. An important property of the drawn distribution is that it is a discrete distribution, even if $H$ is continuous. The basic idea that is leveraged in HDP-HMM is that if $H$ is a distribution over emission parameters then a draw from $DP\left(H,\alpha\right)$ is a distribution over an infinite discrete subset of the parameters, which can be taken to represent a row of the transition matrix of an infinite HMM. $DP\left(H,\alpha\right)$ therefore can represent a prior over a row of the transition matrix.

As a first approach to modeling an infinite HMM, we might consider modeling each row as a separate draw from DP, thus giving us a prior over the entire transition matrix. However, the Dirichlet process does not guarantee that the support of each draw from $DP\left(H,\alpha\right)$ will be the same. In fact if $H$ is continuous, the support of each draw will have no overlap with probability 1. This effectively means that a transition matrix defined using this method specifies a system where every state transition will be to a state that has never been seen before. This is clearly not a useful model. The hierarchical Dirichlet process solves this problem by replacing the continuous $H$ with a discrete one $H'$ which is taken to be a draw from another Dirichlet process (call it the top-level DP) $H' \sim DP\left(H,\gamma\right)$. This forces the output of the low-level DP to be distributions over the same parameters, which makes it appropriate to use as the rows of the transition matrix.

The HDP-HMM gives us a way to specify priors over the transition matrices that ensure that transition probabilities are similar. This alone is not fully sufficient for many applications. Consider for example the intended application where we are modeling daily stock index values. If the hidden state is intended to indicate market criteria that determine properties such as volatility in the market, it is reasonable to assume that the hidden state does not change every day. If we model the rows of the transition matrix as draws from $DP\left(H,\alpha\right)$, even if $H$ is discrete, the draws will not give us transition probabilities with this property. The sticky HDP-HMM provides a solution to this. Instead of drawing from the same $DP$ process to define the rows of the transition matrix, each draw the (now discrete) base distribution is weighted towards a self-transition.

## 3 Predictions

We believe that the advantages of the nonparametric Bayesian approach taken in the sticky HDP-HMM are applicable to the task of data prediction as well as inference of hidden states. We investigate this idea by using the sticky HDP-HMM in the task of predicting economic events based on stock index values. There are many different types of events of economic interest and impact. Stock market crashes, bubbles, changes to indicators such as unemployment, trading activity are some examples. Of the many ways in which to analyze indicators such as daily changes, percent daily changes, moving averages among others, one approach that has been used in the economic literature has been the standard score, or the z score, which measures how many standard deviations a value is from the mean, $z = \frac{x-\mu}{\sigma}$. We use the z score of stock index values as the basis for defining events that we would like to predict. In particular, let $x_i$ be the change in the index value between day $i$ and day $i-1$, $\mu_i\left(n\right)$ be the average value of the daily change for the $n$ days before $i$, and $\sigma_i\left(n\right)$ be the standard deviation for the daily change for the $n$ days before $i$. Define $n$ days $z$ score at day $i$ to be:

$$z_i^n = \frac{x_i - \mu_i\left(n\right)}{\sigma_i\left(n\right)}$$

We consider an event to have occurred on day $i$ if $\left|z_i^{30}\right| \geq 4$ or $\left|z_i^{90}\right| \geq 3$. Our goal is to predict the occurrence of an event on the next day based on the values of the stock index during the previous days.

The sticky HDP-HMM is an appealing model to use in this case. A generative model where hidden states determine the observations has intuitive justification for the workings of the stock market. The hidden states could correspond to market situations like bubbles, depressions, volatility, among others. The number of such hidden states is not known in advance and it is generally expected that such hidden states do not change quickly, that is, they are sticky. We focus in this paper on the use the percent daily changes of the index values as our observations and assume that the emission distributions are Gaussian.

We learn a sticky HDP-HMM using the percent daily changes of the IGBVL stock index between the days May 3, 2007 and May 2, 2012. In Figure 1 we plot the observations against the resulting inferred hidden states.

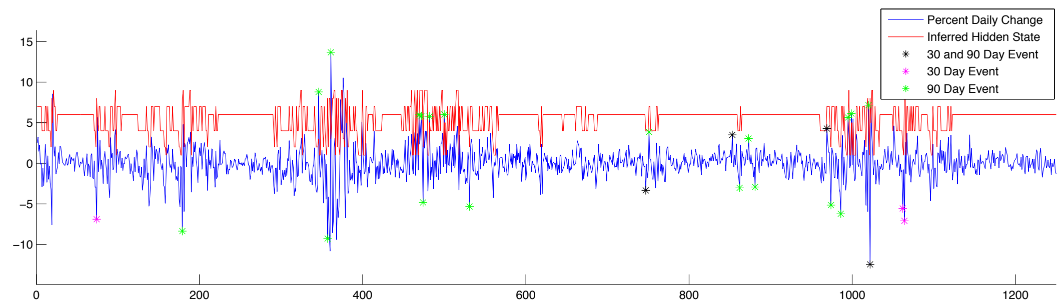
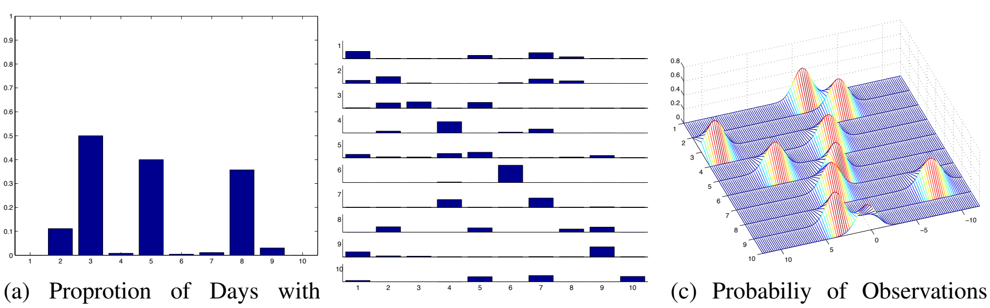



Figure 1: Inferred Hidden States for IGBVL between Sep 30, 2009 and Feb 16, 2012

We can see that the different states represent different dynamics of the market. To see the distinctions between states more clearly we plot in Figure 2.a the proportion of days with events for each inferred hidden state. We see that there is a marked difference between each state. This gives us optimism that the model may provide us with useful predictive power. Figure 2.b shows the actual transition matrix that was learned by the model. One example to note is the non-negligible probability of transitioning from states 10 and 1 to state 5. Both 10 and 1 have low proportion of events while state 5 has a high proportion of events. This is further support for the predictive power of the model. Figure 3.c shows the probability of observations given each state.



(a) Proprotion of Days with Events for each State

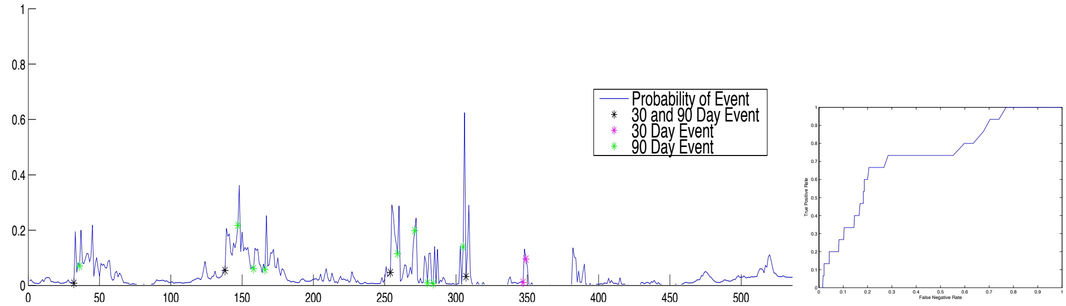
(b) Transition Probabilities

(c) Probabiliy of Observations given each State

Figure 2

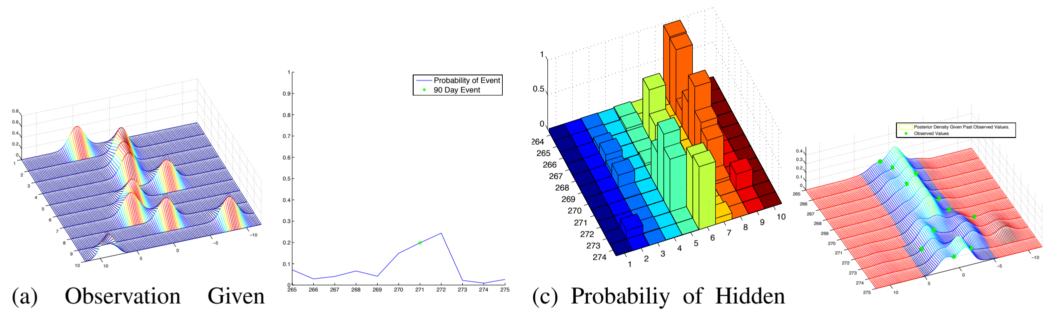
We use the HDP-HMM to define a joint distribution over sequences of observations and then derive the posterior of future observations given the past observations. The observations are percent daily changes and so the posterior of the percent daily changes for the next day can be determined which can then be converted to a posterior over the Z scores from which the probability of an event can be calculated. Figure 3.a shows the result of training the model on the percent daily change of IGBVL from May 03, 2007 to Nov 05, 2009 and using the trained model to calculate the probability of events for the days between Mar 17, 2010 and May 02, 2012. From the figure it is clear that the model provides us with useful predictive power, as most of the events occur on days where the model predicts a non-trivial probability of seeing an event. Figure 3.b shows the ROC curve associated with using different probability thresholds for generating a point estimate of events.

In order to better understand and analyze the predictions provided by the model we provide a detailed view (Figure 4) into the predictions for the days indexed 265 to 275 in Figure 3.a. Figure 4.a shows the probability of observations given each state for the learned model. Figure 4.b shows the probability of an event for the 11 days, Figure 4.c shows the posterior distribution of the hidden states during that period, and Figure 4.d shows the posterior distribution of observations (percent daily changes) along with the actual observed value. In Figure 4.d, the area in red corresponds to values that would lead to a significant z-score event.




(a) Probability of Events for IGBVL between Mar 17, 2010 and May 02, 2012 (b) ROC Curve

Figure 3: Analysis of predictive HDP-HMM.




(a) Observation Given Hidden State (b) Probabiliy of Events (c) Probabiliy of Hidden State (d) Density of Observation

Figure 4: Details of Model Prediction for Days Indexed 265 to 275

# 4 Conclusion

In this work we demonstrated that the sticky HDP-HMM model could be used not only for inferring hidden states, but also for prediction. We demonstrated our initial version of the model against economic data where our results showed were very promising. This provides strong support for extending the work to incorporate higher dimensional input data such as web search trends and social media data.


### Acknowledgments

Supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior National Business Center (DoI / NBC) contract number D12PC00285. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBE, or the U.S. Government.


# References


[1] Matthew J. Beal, Zoubin Ghahramani, and Carl Edward Rasmussen. The infinite hidden markov model. In *Advances in Neural Information Processing Systems 14*, pages 577–584, 2001.

[2] E. B. Fox, E. B. Sudderth, M. I. Jordan, and A. S. Willsky. An HDP-HMM for systems with state persistence. In *Proc. International Conference on Machine Learning*, July 2008.

[3] Samuel J. Gershman and David M. Blei. A tutorial on Bayesian nonparametric models. *Journal of Mathematical Psychology*, 56(1):1–12, February 2012.

[4] Y.W. Teh, M.I. Jordan, M.J. Beal, and D.M. Blei. Hierarchical dirichlet processes. *Journal of the American Statistical Association*, 101(476):1566–1581, 2006.