NIPS 2015
Neural Information Processing Systems 2015

Reviews For Paper

Paper ID 360
Title An Infinite Hidden Markov Model With "Local" Transitions

---

Masked Reviewer ID: Assigned_Reviewer_1

Review:

Question	
Quality Score - Does the paper deserves to be published?	5: Marginally below the acceptance threshold
Confidence	4: Reviewer is confident but not absolutely certain
Please summarize your review in 1-2 sentences	An interesting extension to the HDP-HMM for stochastic processes with multiple time scales. Gibbs sampler shows improved accuracy on a suitable "cocktail party" data set.
Comments to author(s). First provide a summary of the paper, and then address the following criteria: Quality, clarity, originality and significance.	In their paper the authors propose an extension to the HDP-HMM which takes the similarity of hidden states into account. Transition rates are rescaled with a similarity function so that local jumps occur more frequently. Experiments using a Gibbs sampler show improvements in accuracy if the data set contains local transitions, but no disadvantage for a data set generated by a standard HDP-HMM. The paper is well written and contains a clear description of the proposed model and the Gibbs sampler used for inference. The model is an interesting extension of the standard HDP-HMM, as "frequent small jumps, rare big jumps" is a realistic assumption for data sets with multiple time scales. However, one aspect is missing from the discussion: do the results indicate in any way if local transitions are present in the data set? I would expect different estimates of lambda, but that parameter seems to be the same for both experiments.

Masked Reviewer ID: Assigned_Reviewer_2

Review:

Question	
Quality Score - Does the paper deserves to be published?	7: Good paper, accept
Confidence	4: Reviewer is confident but not absolutely certain

Please summarize your review in 1-2 sentences	This paper provides an new approach for adopting a state similarity function in rescaling the transition probabilities for the iHMM. The paper is generally well-written but further clarification and discussion is required in the paper on the applicability of the model to real world analysis scenarios that are not covered by the Experiments.
Comments to author(s). First provide a summary of the paper, and then address the following criteria: Quality, clarity, originality and significance.	This paper develops an Infinite HMM which allows for scaling of transition probabilities by a similarity function. This innovation allows preferential jumping between states which are deemed to be "close" in terms of associated parameters. Central to this work is the equivalence between the Normalised Gamma Process representation that can be recast as an equivalent marginalised Markov Jump process that enables efficient posterior inference. Quality - The paper presents an incremental but useful advance on the state-of-the-art. The paper provides a near complete narrative from motivation to the idea and its implementation detail with a practical example. However, the paper seems incomplete as the Experiments section ends abruptly and it contains no Discussion - is this a compilation error? The abstract could be made shorter if space is required but, if there is no compilation error, there is currently a significant amount of white space available and no reason why more text could not have been included. In particular, some further discussion of the behaviour exhibited in Figure 2 would be necessary. The use of local transitions increases the number of states used. The authors attribute this to the fact that with local transitions, there is a greater probability of new states with similar latent features, which I agree with. However, it does beg the question, what interpretable value do the states have? Suppose I apply both methods in a real scenario, where the true states and the process generating them are unknown, how do I evaluate whether the LT or non-LT models are best? In Figure 3, for instance, the LT model achieves a similar log-likelihood to the non_LT model but uses many more states. It would be useful for the authors to comment on practical applied scenarios. Clarity - Other than the seeming incompleteness, the paper is well-written and I understood the main concepts well. Originality - I am not aware of any similar approaches developed elsewhere. This is an interesting and more general alternative to the "Sticky HMM" of Fox et al. Significance - This modelling approach offers a new way of specifying transition probabilities for iHMMs. This new prior maybe usefully applicable in some modelling situations but there others where there maybe a preference for associating nearby similar observations with a single persistent state (the non-LT approach). The work does not fully supercede other related work in the area.

Masked Reviewer ID: Assigned_Reviewer_3

Review:

Question	
Quality Score - Does the paper deserves to be published?	5: Marginally below the acceptance threshold
Confidence	4: Reviewer is confident but not absolutely certain
Please summarize your review in 1-2 sentences	An infinite hidden Markov model that prefers "nearby" states through kernel functions. Necessity of such a restrictive HMM is not clear beyond the speaker diarization problem treated in this paper.
Comments to author(s). First provide a summary of the paper, and then address the following criteria: Quality, clarity, originality and significance.	This paper proposes an infinite hidden Markov model which prefers transitions to "nearby" exemplified by a kernel function between states. For this purpose, authors used a Gamma process representation of HDP-HMM and modifies the scale parameter of the Gamma distribution, which is usually 1, by a kernel function between states. Inferences are generally straightforward, but authors also proposed a representation with "failed" jumps which works as auxiliary variables to facilitate sampling from conjugate posteriors. When I first read this paper, I cannot understand why such a restrictive HMM is actually needed. Reading the experimental section, I could see the motivation of this paper on speaker diarization problems, where the hidden states consists of binary vector corresponding to speaking/silence of each speaker, thus large jump between these binary vectors is improbable. Although the authors compared the proposed model with factorial HMM, the most related model might be the Markov indian buffet process, which has latent binary state vectors and sometimes a switch between binary states occurs. Because this paper does not include experiments with actual data other than couctail party problems, without comparison to more straightforward model for speaker diarization it is not clear that the proposed model is actually necessary and applies to broad problems. While the current model assume that a transition from \theta_i to \theta_j is equally probable to a transition from \theta_j to \theta_i through a kernel function, this is not generally true because some state might be a "ground"

	state that can be reached from many states easily. The scaled Gamma process construction in equation (8) is reminiscent of DILN of Paisley+ (2011), where the scale parameter itself is stochastic and generated from a Gaussian process. Therefore, it might be interesting to extend the current model to employ a Gaussian process, not a fixed kernel function, that promotes local transition but sometimes allows transition to states far away. In short, proposed model is interesting and effective to some extent, but it is not clear where it is useful to apply beyond more straightforward approaches explored so far.

Masked Reviewer ID: Assigned_Reviewer_4

Review:

Question

Quality Score - Does the paper deserves to be published?	5: Marginally below the acceptance threshold
Confidence	3: Reviewer is fairly confident
Please summarize your review in 1-2 sentences	The main idea of the paper is to express local Markov transitions in an HDP-HMM model by augmenting jumps with a 'failure rate' which depends on the parameters of each state. The idea is interesting but write-up and comprarisons could be improved.

Masked Reviewer ID: Assigned_Reviewer_5

Review:

Question

Quality Score - Does the paper deserves to be published?	7: Good paper, accept
Confidence	4: Reviewer is confident but not absolutely certain
Please summarize your review in 1-2 sentences	The paper succeeds in developing the theory and sampling methods for the potentially useful idea of using pairwise similarities to locally modify the transition probabilities of a HDP-HMM. Experiments on synthetic data appear promising, though possibly incomplete, and experiments on real data are missing.

Masked Reviewer ID: Assigned_Reviewer_6

Review:

Question

Quality Score - Does the paper deserves to be published?	4: An OK paper, but not good enough
Confidence	5: Reviewer is absolutely certain
Please summarize your review in 1-2 sentences	An elegant idea for adding additional structure to non-parametric HMMs. The main issue is that the experiments only contained a single application to a problem with a finite state space. An additional application with a continuous latent space is needed both to demonstrate the usefulness of the model and further verify the tractability of inference.