MULTI-ASPECT TARGET CLASSIFICATION AND DETECTION VIA THE INFINITE HIDDEN MARKOV MODEL

Kai Ni, Yuting Qi and Lawrence Carin

Department of Electrical and Computer Engineering
Duke University, Durham, NC 27708-0291

ABSTRACT

A new multi-aspect target detection method is presented based on the infinite hidden Markov model (iHMM). The scattering of waves from multiple targets is modeled as an iHMM with the number of underlying states treated as infinite, from which a full posterior distribution on the number of states associated with the targets is inferred and the target-dependent states are learned collectively. A set of Dirichlet processes (DPs) are used to define the rows of the HMM transition matrix and these DPs are linked and shared via a hierarchical Dirichlet process (HDP). Learning and inference for the iHMM are based on an effective Gibbs sampler. The framework is demonstrated using measured acoustic scattering data.

Index Terms— Hierarchical Dirichlet processes, hidden Markov models, acoustic signal detection

1. INTRODUCTION

In many sensing scenarios the target is observed from multiple target-sensor orientations (or aspects), and the underlying acoustic or electromagnetic scattered waveforms are highly dependent on the aspect. This angle-dependence property is largely due to the physical composition of the target. It is often difficult to achieve reliable classification based on a single view of the target; this is because the waveforms emitted from two different targets may be similar at certain angles and easily confused for classification. This motivates using a sequence of multiple aspect-dependent looks at a single target, with this *sequential* information offering the potential to substantially improve identification performance.

Hidden Markov models (HMMs) have been successfully applied to model such sequential data for multi-aspect target detection and classification problems [1, 2]. Each HMM "state" represents a set of generally contiguous target-sensor orientations over which the signal statistics are relatively invariant. A length T observation sequence implicitly samples a sequence of T target states. The probability of transitioning from one state to another on consecutive measurements is modeled as a Markov process. Furthermore, because the target is usually concealed or distant, the target-sensor orientation is unknown and the actual sampled state sequence is hidden, motivating an HMM.

In the context of target classification using HMMs, a key issue is to develop a methodology for defining an appropriate set of states. Ideally, an HMM state should be defined by a set of angles for which a particular class of underlying physics dominates. However, in many problems it is not possible to easily find a relationship between the physics and the underlying HMM state. In previous work the state decomposition has been performed in an *ad hoc* manner [1, 2], requiring trial and error to manually select the model structure, (e.g., number of states). In the work reported here we investigate the idea of an *infinite* hidden Markov model (iHMM), which by construction has an infinite number of hidden states and the proper number of states associated with a target is inferred automatically. The iHMM is constituted by a *hierarchical Dirichlet process* (HDP) [3], a nonparametric Bayesian prior for sharing clusters among related groups. However, the HDP assumes a fixed partition of groups of data, while the group partition is random in an iHMM. In addition, the iHMM must utilize the assumed underlying temporal information of the sequential data, which is not considered in the original HDP.

Another limitation of target classification algorithms is that they usually require substantial training data, assumed to be similar to the data on which the algorithm is tested. Unfortunately, in many sensing applications one often has limited training data from a target. In acoustic target classification problems, for instance, one may have multiple sets of limited scattering data, with each data set collected from different but related targets. Rather than building models for each target individually, as adopted in [1, 2], it is desirable to appropriately share the information among these related targets, thus offering the potential to improve overall classification performance. The iHMM effectively solves this problem using the HDP framework.

Teh *et al.* [3] give a brief introduction to applying HDP for the iHMM. Here we perform a more expansive study of information sharing between multiple HMMs using the iHMM. We give examples of a new application, in particular multi-aspect target detection using the iHMM. We demonstrate how the iHMM may be used to build a single model for a class of targets, and how the learned model yields information about the physical characteristics and the relationships between the targets within the class. These ideas are demonstrated using measured acoustic scattering data, with comparison to more traditional techniques.

2. BACKGROUND

2.1. Dirichlet Process Mixture Models

Let H be a measure on a space Θ, and let γ be a positive real number. A *Dirichlet process* (DP) [4] is a distribution for a *random density function* $G(\theta)$, denoted by $G \sim DP(\gamma, H)$. The "base" distribution H provides the prior information of G with $E[G] = G_0$, and the concentration parameter γ controls how similar G is to H. Samples drawn from $DP(\gamma, H)$ are discrete with probability one [4], a property made explicit by the *stick-breaking construction* [5]

$$G(\theta) = \sum_{k=1}^{\infty} \beta_k \delta_{\theta_k^*} \quad \beta_k = \beta_k^{'} \prod_{l=1}^{k-1}(1-\beta_l^{'}) \quad \beta_k^{'} \sim \mathrm{Beta}(1, \gamma), \tag{1}$$

where $\delta_{\theta_k^*}$ is a discrete measure concentrated at θ_k^*. The countably infinite random parameters $\{\theta_k^*\}_{k=1}^{\infty}$ are independently sampled from H, and the weight variables $\{\beta_k\}_{k=1}^{\infty}$ are defined by a Beta distribution that partitions a unit-length "stick".

A Dirichlet process is commonly used as a nonparametric prior distribution for a mixture model with unbounded number of components [6]. Assume the observation x_i is generated from a distribution F with parameter θ_i. The density function on the θ's is G, which is assumed to be drawn from $DP(\gamma, H)$, and thus we have a *Dirichlet process mixture model*. A graphical representation of a DP mixture model is given in Fig. 1(a). Indicator variable z_i denotes the mixture component generating the data point $x_i \sim F(\theta_{z_i}^*)$, i.e., $\theta_i = \theta_{z_i}^*$. With DP as a prior on G the number of mixture components is treated as infinite and the actual (finite) components used by the mixture model are inferred automatically from the data (via the data-dependent likelihood function).

2.2. Hierarchical Dirichlet Processes

The HDP considers problems of multiple related groups of data, where each group is described by an infinite mixture model and the mixture components are shared across groups. Assume we have J groups of data. To construct an HDP, a global probability measure $G_0 \sim DP(\gamma, H)$ is first drawn to define the mixture components, and then $G_j \sim DP(\alpha, G_0)$ is sampled independently for each group. The discreteness of G_0 (as shown in (1)) guarantees that the G_j's will reuse the same set of shared mixture components defined in G_0 but with different proportions [3]:

$$G_0 = \sum_{k=1}^{\infty} \beta_k \delta_{\theta_k^*} \quad G_j = \sum_{k=1}^{\infty} w_{jk} \delta_{\theta_k^*} \quad \boldsymbol{w_j} \sim \mathrm{DP}(\alpha, \boldsymbol{\beta}), \tag{2}$$

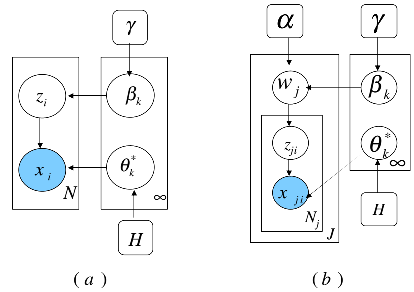

Fig. 1. (a) Graphical model representation of a DP mixture model using (1). (b) Graphical model representation of a hierarchical Dirichlet process (HDP) mixture model using (2).

where $\boldsymbol{w_j}$ is infinite-dimensional probability mass function.

The HDP can be used to model J groups of coupled infinite mixture models. The graphical model of an HDP mixture is shown in Fig. 1(b), where datum x_{ji} in group j is generated by first drawing $\theta_{ji} \sim G_j$, then sampling $x_{ji} \sim F(\theta_{ji})$. Parameters $\{\theta_k^*\}_{k=1}^{\infty}$ are the set of shared mixture components defined in G_0, and z_{ji} is the indicator variable for which $\theta_{ji} = \theta_{z_{ji}}^*$.

3. INFINITE HIDDEN MARKOV MODEL

3.1. HMM for Multi-Aspect Target Sensing

Using HMMs for multi-aspect target detection and classification has been described in [1, 2]. In many applications, the HMM structure is given from previous practical experience or by "experts", and model learning focuses on estimating the parameters from the data. Typically maximum-likelihood parameter learning is performed via the EM algorithm [7], with the number of states assumed known. However the HMM is only a model for the real target scattering, and therefore there may not be a single "correct" HMM structure, i.e., a fixed number of states. Rather than performing model selection to select a fixed number of states [8], we employ a fully Bayesian approach in which the number of states is not fixed *a priori*. The nonparametric HDP is employed to implicitly construct an HMM with an infinite number of states (iHMM), even though in reality the posterior on the number of states is usually peaked about a finite number of states characteristic of the sequential data used for model training.

3.2. Learning iHMM via HDP

An N-state HMM can be considered as a set of N coupled finite mixture models (each with N mixture components). That is, the value of previous state s_{t-1} indexes a specific row of the transition matrix serving as the mixture weights for choosing current state s_t. Given the value of s_t, the observation o_t is sampled from the observation model (mixture component)

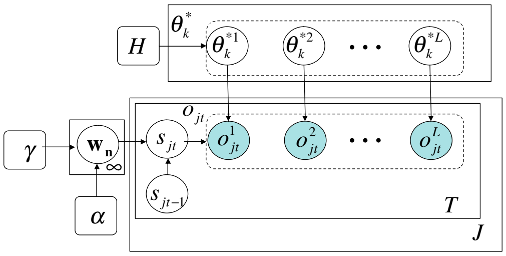

Fig. 2. The infinite hidden Markov model interpreted as an HDP. For observation o_{jt}, s_{jt-1} defines the mixture model to be used and s_{jt} selects the mixture component according to infinite dimension weight vector $w_{s_{jt-1}}$.

indexed by s_t. To consider an infinite number of states, it is natural to use a set of state-specific DPs, one for each value of the state. Furthermore, these DPs must be shared because they use the same set of mixture components defined in the observation matrix. This is similar to the HDP mixture model but with a key difference being that the data (observations) to group partition is fixed in the HDP while the group partition is random (indexed by the hidden previous state) in the iHMM.

The HDP construction of the iHMM for multi-aspect target sensing is shown in Fig. 2, with parameters defined as

$$\begin{aligned} & o_{jt} \mid s_{jt}, \{\theta_k^*\}_{k=1}^{\infty} \sim F(\theta_{s_{jt}}^*) \qquad \{\theta_k^*\}_{k=1}^{\infty} \mid H \sim H \\ & s_{jt} \mid s_{jt-1}, \{\boldsymbol{w_n}\}_{n=1}^{\infty} \sim \mathrm{Mult}(\boldsymbol{w_{s_{jt-1}}}) \\ & \{\boldsymbol{w_n}\}_{n=1}^{\infty} \mid \alpha, \boldsymbol{\beta} \sim \mathrm{DP}(\alpha, \boldsymbol{\beta}) \qquad \boldsymbol{\beta} \mid \gamma \sim \mathrm{Stick}(\gamma), \end{aligned} \tag{3}$$

where $\boldsymbol{w_n}$ corresponds to the row of transition matrix A, and $F(\theta_k^*)$ corresponds to the observation model $b_k(\cdot)$. Each sampled scattered waveform (observation) is represented with an L-dimensional feature vector $o_{jt} = [o_{jt}^1, \ldots, o_{jt}^L]$ and the feature vector is assumed to be generated from a signal model F. For a continuous HMM, F is an L-dimensional Gaussian and θ represents the mean vector and covariance matrix, while for a discrete HMM θ represents the parameters of a multinomial distribution. When the data sequences correspond to scattering from a *class* of targets, rather than build HMMs for each target separately, we may use the iHMM to learn the state structure underlying the data from all targets within the class. A class-based iHMM generalizes well to detection of a target not seen when training, as long as the new target residues within the class of training targets.

The iHMM learning is based on an extension of the HDP Gibbs sampler [3] with temporal information considered. Let $\boldsymbol{S^{-t}}$ denote the hidden state sequence excluding s_t, and $\boldsymbol{O}$ denotes the observation sequence. Using the first order Markov property, we have

$$p(s_t = k \mid \mathbf{S}^{-t}, \boldsymbol{O}) = p(s_t = k | s_{t-1} = j, s_{t+1} = l, o_t) \propto$$
$$\begin{cases} (n_{jk}^{-js_t'} + \alpha\beta_k) \frac{\alpha\beta_l + n_{kl}}{\alpha + \sum_{k'=1}^{K} n_{kk'}} f_k^{-o_t}(o_t), & \text{if } k \in \{1, \ldots, \mathrm{K}\}; \\ (0 + \alpha\beta_u)\beta_{s_{t+1}} f_{k^{new}}^{-o_t}(o_t), & \text{if } k = k^{\mathrm{new}}, \end{cases} \tag{4}$$

with s_t' the previous sampled value of s_t, n_{jk} the count of transitions from state value j to state value k and β_k the mixing weight for state (mixture component) k. The first two terms are contributed from the HDP prior and the third term is from the data likelihood. The number of states in this algorithm is automatically inferred from the data.

4. EXPERIMENTAL RESULTS ON TARGET MODELING AND DETECTION

We consider a target detection problem based on multi-aspect measured acoustic scattering data. Five shell targets are considered as "targets of interest", and six clutter items are considered as false targets. The details of the shell targets and the clutter are described in [1, 2, 9]. The goal is to test if a new target is in the family of shell targets or not. To setup the experiment, we train on four shell targets and the item under test is the remaining one shell target (label 1) or one of the six false targets (label 0). All six of the false targets and the held out shell target were not seen when training. When testing we consider each possible test item, as viewed from all possible initial angles.

We compare development of a single HMM (SHMM) for each of the four training targets, and a single HMM designed for all four training targets (class-based iHMM), and in both cases the iHMM is used for model design. We note that for the SHMM, there are counterpart models that may be designed using the conventional EM algorithm, one model for each shell. The number of HMM states may be inferred for the target-dependent HMMs using physical considerations (note that for the EM solution the number of states is set *a priori*, and fixed). By contrast, for the class-based iHMM (single iHMM for all four training targets), it is difficult to have an analog computed using the EM algorithm, because the proper number of states is difficult to determine (some states associated with the single targets may be shared, and others not shared, and this decomposition is difficult to determine *a priori*). A significant advantage of the iHMM is that it adaptively determines a posterior on the proper number of states, automatically learning the appropriate state sharing structure across the four training targets. We also note that this class-based iHMM, by sharing data across the training targets, will not model any one of the training targets, but will rather infer relationships between the multiple training targets within this class. This generalization is beneficial when testing on targets not seen while training, as long as the testing targets of interest are within the class of training data.

This generalization is reflected in the results presented in Fig. 3, for which the class-based iHMM consistently outperforms the iHMMs designed for each of the training targets separately. The results are measured by Receiver Operation Characteristic (ROC) curves. We show in Fig. 4 the posterior distribution on the number of states associated with the class-based iHMM (for case 1 in Fig. 3), running 100000

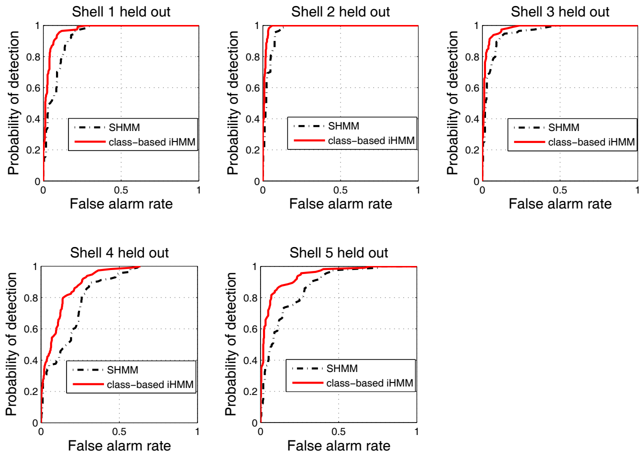

Fig. 3. Detection results for different sets of targets using the SHMM and the class-based iHMM. The results are presented with the respective shell target held out of the training set.

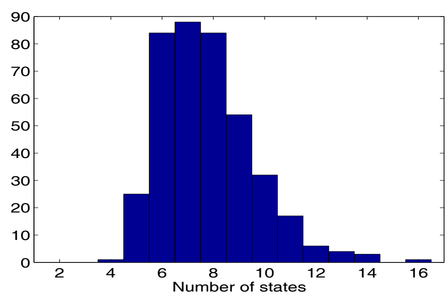

Fig. 4. Posterior distribution on model structure (the number of states) for case 1 in Fig. 3

Gibbs iterations after a 10000 burnin step, collecting 400 posterior samples with a spacing of 250 samples. To see the relationship of target-dependent states across the shell targets, we train the class-based iHMM on the collection of *all* five shell targets and plot the hidden state distribution based on the last Gibbs iteration, with the result depicted in Fig. 5. We observe that the underlying states are shared over all targets, showing the utility of using the class-based iHMM over the single HMMs.

5. CONCLUSION

The infinite hidden Markov model (iHMM) is proposed to solve the fundamental problem of model selection in hidden Markov models as well as the problem of state information sharing between HMMs. Promising iHMM results have been obtained using scattering data from real elastic targets. It is shown that the iHMM not only provides a full posterior distribution on the number of hidden states, but also shares the underlying state information among different targets, with

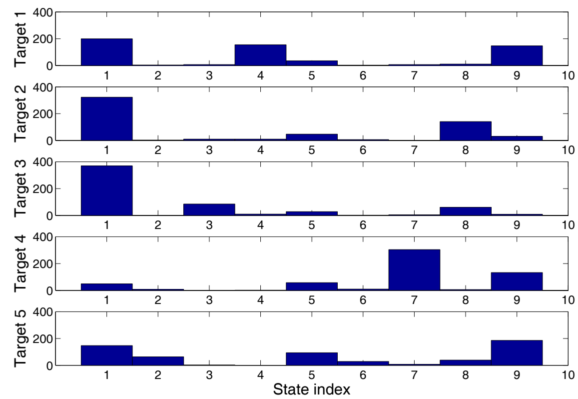

Fig. 5. Hidden state distribution across all five shell targets. Note that many targets within the same class share states

this providing better generalization performance compared to other competing methods.

6. REFERENCES

[1] P. Runkle, P. K. Bharadwaj, L. Couchman, and L. Carin, "Hidden Markov models for multi-aspect target classification," *IEEE Transactions on Signal Processing*, vol. 47, pp. 2035–2040, 1999.

[2] P. Runkle, L. Carin, L. Couchman, T. J. Yoder, and J. A. Bucaro, "Multi-aspect identification of submerged elastic targets via wave-based matching pursuits and hidden Markov models," *Journal of the Acoustical Society of American*, vol. 104, pp. 937–946, 1999.

[3] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei, "Hierarchical Dirichlet processes," *Journal of the American Statistical Association*, 2006, to be appear.

[4] T. S. Ferguson, "A Bayesian analysis of some nonparametric problems," *Annals of Statistics*, vol. 1, no. 2.

[5] J. Sethuraman, "A constructive definition of Dirichlet priors," *Statistica Sinica*, vol. 4, pp. 639–650, 1994.

[6] M. D. Escobar and M. West, "Bayesian density estimation and inference using mixtures," *Journal of the American Statistical Association*, vol. 90, pp. 577–588, 1995.

[7] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, pp. 275–285, 1989.

[8] A. Stolcke and S. Omohundro, "Hidden Markov model induction by Bayesian model merging" *Advances in Neural Information Processing Systems*, vol. 5, pp. 11-18, 1993.

[9] N. Dasgupta, P. Runkle, L. Carin, L. Couchman, T. Yoder, J. Bucaro, and G. Dobeck, "Class-based target identification with multi-aspect scattering data," *IEEE Journal of Oceanic Engineering*, vol. 28, pp. 271–282, 2003.