LANDMINE DETECTION WITH MULTIPLE INSTANCE HIDDEN MARKOV MODELS

Seniha Esen Yuksel[a,b], *Jeremy Bolton*[c], *Paul D. Gader*[c]

[a] Middle East Technical University Northern Cyprus Campus, Mersin 10, Turkey
[b] Department of Materials Science and Engineering, University of Florida
[c] Department of Computer and Information Science and Engineering, University of Florida

ABSTRACT

A novel Multiple Instance Hidden Markov Model (MI-HMM) is introduced for classification of ambiguous time-series data, and its training is accomplished via Metropolis-Hastings sampling. Without introducing any additional parameters, the MI-HMM provides an elegant and simple way to learn the parameters of an HMM in a Multiple Instance Learning (MIL) framework. The efficacy of the model is shown on a real landmine dataset. Experiments on the landmine dataset show that MI-HMM learning is very effective, and outperforms the state-of-the-art models that are currently being used in the field for landmine detection.

***Index Terms*—** Multiple instance learning, hidden Markov models, Metropolis-Hastings sampling, landmine detection, ground penetrating radar, time series data,

1. INTRODUCTION

In standard learning techniques, an algorithm is typically presented with exemplar samples from some number of classes, and its goal is to construct a characterization for each class. Many of these methods attempt to identify areas in a feature space that contain a dense number of target exemplars and a low density of non-target exemplars. However, in some learning situations, class labels are not readily available for each sample in the training data. For example, in content-based image classification an image may contain multiple objects, but it might not be easy to identify which of these objects are the relevant ones [1–3]. This type of data is also known as ambiguous data, and learning from ambiguous data remains a hard problem [4, 5]. One of the areas where ambiguous data is encountered is landmine detection using ground penetrating radar (GPR). In radar images produced by GPR sensors, there are areas (subimages or feature sets) in an image that contain a target and areas that do not. However, ground truth is provided only per image and not for the subimages. Therefore this learning scenario provides one class label for multiple instances (multiple feature sets).

To combat this issue, researchers may segment the images manually or semi-automatically to extract target and non-target exemplars for training [6–9]. An example is shown in Fig. 1 on a GPR image. However, this is not only an arduous task, but also is prone to errors resulting from GPR echoes and ground truthing errors, and furthermore, ambiguity still remains. Therefore, rather than struggling against the ambiguous nature of this learning problem, it may be best to use a model that explicitly account for ambiguous data.

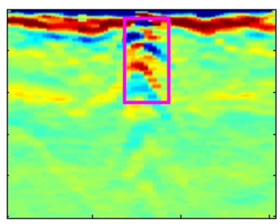


Fig. 1. Ground penetrating radar data with a landmine signature that looks like a hyperbola. Traditionally, a bounding box is placed on the landmine signature and training sets are formed from these signatures. Other signals in this image are the ground, the GPR echo, and the signals from the soil itself.

One solution to learning from ambiguous data is Multiple Instance Learning (MIL). In the MIL scenario, class labels of all of the training data are not available, thus it is not possible to present an algorithm with exemplar samples from each class [10–12]. Instead, an algorithm is presented with a collection of bags, or sets of samples, that are labeled positive or negative. Bags are labeled positive if there exists at least one sample that induces a target concept and are labeled negative if every sample is from the non-target class. This view is illustrated in Fig. 2 where the traditional classifiers are compared to MIL classifiers.

The MIL model has also recently been used in landmine detection to eliminate the problems associated with the bounding box approach, and has shown considerable success [13, 14]. However, neither these studies nor the other

This research was partially supported by the Army Research Office grant W911NF0510067 and the NSF Grant No. 0730484.

978-1-4673-1026-0/12/$31.00

MIL models in the literature could utilize time series data. Since hidden Markov model (HMM) based algorithms that utilize time series data are known to be very useful in landmine detection [15–18], an MIL based model that can model the temporal properties of the data is proposed.

In this study, a novel multiple instance hidden Markov model (MI-HMM) that uses MIL for time series data is developed. In MI-HMM, labels are attached to the bags like in an MIL, but a bag is a set of sequences and these sequences can be of different lengths as shown in Fig. 2(c). The MI-HMM provides an elegant and simple way to learn the parameters of an HMM with a Metropolis-Hastings sampler that rejects or accepts the parameters using the MIL algorithm. Experiments on the landmine dataset have shown that MI-HMM learning is very effective, and outperforms the state-of-the-art models that are being used in the field for landmine detection.

In the remainder of this paper, first standard MIL learning is described. Then, notation for HMMs is introduced. Next MI-HMM model is proposed. Finally, the MI-HMM results on landmine GPR data are presented and discussed.

2. STANDARD MULTIPLE INSTANCE LEARNING

Let x denote a feature vector and Y_x denote the label of this vector. In the MI scenario, a learner is presented with *N* sets (referred to as *bags* in the literature) of m_N vectors, or samples. For the purposes of learning, a set, $X \subset \mathscr{R}^d$, is labeled target ($Y_X = 1$) if there exists at least one target sample within the set. A set *X* is labeled negative ($Y_X = 0$) if all constituent samples are non-target. That is, $\exists x \in X : Y_x = 1 \Rightarrow Y_X = 1$ and $\forall x \in X : Y_x = 0 \Rightarrow Y_X = 0$. With this learning paradigm, *the idea of uncertainty is incorporated using the set (or bag) structure*; and learning the target concept from these bags of samples is called the MIL problem.

Maron et al. developed the Diverse Density (DD) [10] approach which provides a statistical solution to the MIL problem based on Pearl's Noisy OR-Gate model [19]. Most MIL solutions adopt this Noisy OR-Gate Model, which assumes that only one target sample within a bag is necessary and sufficient for a bag to induce a target concept. In standard DD approaches, a target concept, which is characterized by a feature vector f, is learned given a collection of positively labeled bags, $\mathbf{B}^+$ and a collection of negatively labeled bags $\mathbf{B}^-$. Assuming observed sets *X* are independent, the target concept, f, is chosen to maximize the expression in Eq. 1 [10]

$$\hat{f} = \text{argmax}_f \prod_{X \in \mathbf{B}^+} P(f|X) \prod_{X \in \mathbf{B}^-} P(\neg f|X) \quad (1)$$

where $\hat{f}$ is the desired target concept, and $\neg f$ are the samples that are not the targets. Assuming a Noisy OR-Gate model [19], the posterior probability factors in Eq. 1 can be calculated in terms of the constituent samples in each bag ($x \in X$) as follows :

$$P(f|X) = 1 - \prod_{x \in X} (1 - P(f|x)) \quad (2)$$

$$P(\neg f|X) = \prod_{x \in X} (1 - P(f|x)) \quad (3)$$

In Eq. 1, the idea is to increase the probability of the target concept in the positive bag and to also increase the probability of the non-target concepts in the negative bags. With the noisy-OR assumption, in Eqs. 2 and 3, the right hand side of the equations have been described solely in terms of the target concept f.

3. HIDDEN MARKOV MODELS

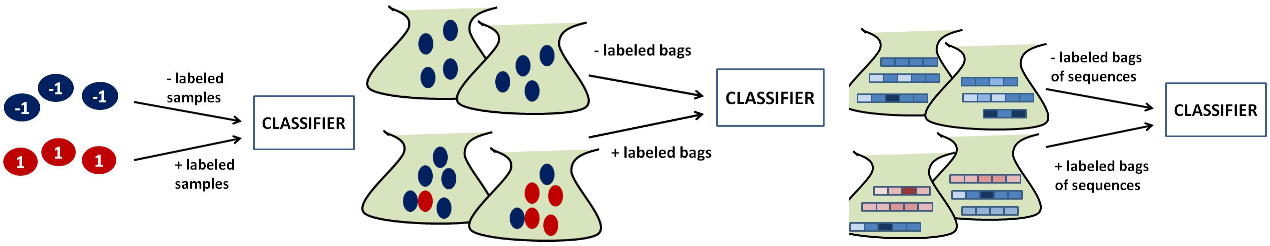
Hidden Markov Model (HMM) is a very popular tool to represent time-series data as it considers the statistical dependence among samples. In this section, we will only provide the very basics and notations for HMM as they are used in Sec. 4. The notation for HMMs is as follows:

- W = number of states.
- M = number of symbols in the codebook.
- T = length of observation sequence.
- $V = \{v_1, ..., v_M\}$ the discrete set of observation symbols.
- $\check{x} =< x_1 x_2 ... x_T >$ denotes an observation sequence, where $x_t \in V$ is the observation at time t.
- $Q = q_1 q_2 ... q_T$ is a fixed state sequence, where q_t is the state at time t .
- $S = \{S_1, S_2, ..., S_W\}$ are the individual states.
- $\Theta = \{\pi, A, B\}$ is the notation for an HMM model
- The initial state distribution $\pi = \{\pi_r\}_{r=1}^W$, where $\pi_r = P(q_1 = S_r)$
- The state transition probability $A = \{\{a_{rj}\}_{r=1}^W\}_{j=1}^W$, where $a_{rj} = P(q_{t+1} = S_j | q_t = S_r)$
- The emission probability $B = \{\{b_j(m)\}_{j=1}^W\}_{m=1}^M$, where $b_j(m) = P(v_m \texttt{ at } t | q_t = j)$

Given an HMM model, the probability of a sequence is computed as:

$$P_{HMM}(\check{x}|\Theta) \quad = \quad \sum_Q \pi_{q_1} \prod_{t=1}^{T-1} a_{q_t q_{t+1}} b_{q_t}(\check{x}_t). \quad (4)$$

Parameters of the HMM model are typically learned generatively by the maximum likelihood (ML) criterion [20, 21], or discriminatively by the the minimum classification error (MCE) criterion [16, 22, 23]. The ML criterion, when implemented by the expectation-maximization (EM) algorithm leads to a local optimum in the parameter space. Similarly, MCE models are generally learned using gradient-based approaches which also converge to a local minimum. One possibility to get around this problem is to use Markov chain Monte Carlo (MCMC) sampling methods to estimate the parameters of an HMM [24, 25].



(a) Traditional classifier (b) MIL classifier with bags of vectors (c) MI-HMM classifier with bags of sequences

Fig. 2. Traditional, MIL and MI-HMM classifiers. (a) Traditional classifier with labeled samples. In traditional supervised learning algorithms, a label is attached to each training sample, and the classifier is trained with these labeled samples. (b) In MIL, training class labels are attached to bags. A bag is a set of samples, and the samples within each bag are called instances. A bag is labeled as positive if and only if at least one of its instances is positive; otherwise, it is labeled as negative [10]. The bags on the right contain at least one red (positive) sample, which makes them positively labeled. (c) In MI-HMM, a bag is a set of sequences. These sequences can be of different lengths. A bag is labeled positive if and only if at least one of its sequences is positive. The bags on the right contain at least one red (positive) sequence, which makes the bags positively labeled.

4. MI-HMM: FROM BAGS OF SEQUENCES

MI-HMM is a tool that permits the learning of sequence models under the MI learning scenario. In Sec. 4.1, we formulate the MI-HMM which is a discriminative model. Learning the parameters in a discriminative HMM model is generally difficult, and using gradient based approaches are commonly subject to learning locally optimal parameter sets [16,22,23]. Furthermore, due to the Noisy-OR formulation of MIL models, standard optimization methods will not yield a closed form solution. Therefore in Sec. 4.2 we describe Metropolis-Hastings sampling to update the parameters of MI-HMM.

4.1. Formulation

Assume $\breve{x} =< x_1, ..., x_i, ..., x_T >= \{x_i\}_{i=1}^T$ is a sequence, and each observation in the sequence, $x_i \in V \subset \mathscr{R}^d$, is a d-dimensional vector. A bag, i.e. a set of sequences, X, is labeled target ($Y_X = 1$) if there exists at least one target sequence with in the set. A set X is labeled negative ($Y_X = 0$) if all constituent sequences are non-target. That is, $\exists \breve{x} \in X : Y_{\breve{x}} = 1 \Rightarrow Y_X = 1$ and $\forall \breve{x} \in X, Y_{\breve{x}} = 0 \Rightarrow Y_X = 0$. Given a collection of positively labeled bags, $\mathbf{B}^+$ and a collection of negatively labeled bags $\mathbf{B}^-$, a standard approach would be to learn the HMM parameters given the following objective:

$$\Theta = \text{argmax}_\Theta P(Y_{X_1}, ..., Y_{X_N}, X_1, ..., X_N|\Theta) \qquad (5)$$

The objective in Eq. 5 is to maximize the joint probability of the bags of sequences and the corresponding class labels for the bags. Assuming independence between the bags and assuming the Noisy-OR relationship between the sequences within each bag, Eq. 5 can be expanded as follows:

$$\Theta = \text{argmax}_\Theta \prod_{X \in \mathbf{B}^+} P(Y = 1|X, \Theta) \prod_{X \in \mathbf{B}^-} P(Y = 0|X, \Theta) \qquad (6)$$

where

$$\begin{aligned} P(Y = 1|X, \Theta) &= 1 - \prod_{x \in X} (1 - P_{HMM}(\breve{x}|\Theta)) \\ P(Y = 0|X, \Theta) &= \prod_{x \in X} (1 - P_{HMM}(\breve{x}|\Theta)) \end{aligned} \qquad (7)$$

It is worthwhile to emphasize that in Eqs. 6 and 7, there are no parameters due to MI learning. The only parameters to learn are the HMM parameters from Eq. 4. Therefore, no additional parameters have been introduced by providing the MI learning of HMMs, but rather a new model has been introduced that does not require individual labels for target sequences. We also alter Eq. 6 to be a likelihood ratio as in Eq. 8, which helps to further separate the values of the probability of positive bags and negative bags.

$$\hat{\Theta} = \text{argmax}_\Theta P^*(\Theta) = \text{argmax}_\Theta \frac{\prod_{X \in \mathbf{B}^+} P(Y = 1|X, \Theta)}{\prod_{X \in \mathbf{B}^-} P(Y = 1|X, \Theta)}. \qquad (8)$$

4.2. Parameter Learning

The noisy-OR framework is not easy to solve with gradient based algorithms, and the optimization task presented in Eq. 8 is quite difficult. Therefore a sampling based learning scheme is proposed. The general idea behind the proposed sampling scheme is to draw parameters Θ from a proposal density and then perform a Metropolis rejection step based on our objective P^* in Eq. 8. However, in this case, direct sampling of $\Theta = \{\pi, A, B\}$ is also not easy, and it may consist of a large number of parameters depending on the number of states and number of distinct observations. The estimates of π are obtained from A using a ratio of transition counts. For the other parameter estimates, a Metropolis-Hastings / Gibbs sampling scheme is proposed. In this scheme, samples are generated from a simpler distribution, the so-called *proposal* *density* [26], and are used to search the parameter space of the objective in Eq. 8.

Note that the rows in the state transition matrix $A_{\bullet j}$, and rows in the emission matrix $B_{i\bullet}$ are all multinomial distributions. Therefore, an intuitive choice for the proposal density is the Dirichlet distribution $\mathscr{D}(\alpha)$ [27]. Our proposed method assumes a mixture of Dirichlet distributions as given in Eq. 9

$$\begin{aligned} \forall j,\ A_{\bullet j} &\sim c_1\mathscr{D}(k_1\alpha) + c_2\mathscr{D}(k_2\alpha) \\ \forall i,\ B_{i\bullet} &\sim c_1\mathscr{D}(k_1\alpha) + c_2\mathscr{D}(k_2\alpha) \end{aligned} \tag{9}$$

where parameters c_1 and c_2 are the mixture components, and k_1, and k_2 determine the "focused" or "random" nature of the component in sampling. Using Eq. 9, new samples of the parameters are drawn from the Dirichlet mixture. Then a Metropolis step (accept or reject decision) is performed after each Gibbs draw (or iteration), thus the Dirichlet mixture is our Metropolis proposal distribution. For the Metropolis step, variable θ' forthwith denotes new draws (known in the literature as the *tentative new states*) of either $A_{\bullet j}$ or $B_{i\bullet}$ and θ^c denotes the sample accepted at iteration c (known in the literature as the *current state*). The new state, θ', is accepted or rejected based on the ratio r at iteration $c+1$ [26]. This ratio is computed as:

$$r_{c+1}(\theta') = \min\left\{1, \frac{P^*(\theta')}{P^*(\theta^c)}\frac{\mathscr{D}(\theta^c;\theta')}{\mathscr{D}(\theta';\theta^c)}\right\} \tag{10}$$

Therefore, θ' is accepted if $r_{c+1}(\theta')$ is equal to or larger than one. Otherwise, θ' is accepted with probability $r_{c+1}(\theta')$. Due to this accept/reject property, the Metropolis-Hastings sampling training of MI-HMM is able to evolve with new parameters and can avoid getting stuck in local minimum.

Although the notations are similar, it is important to notice that in Eq. 9, random samples are generated from the Dirichlet distribution which requires only one set of parameters, whereas in Eq. 10 the Dirichlet distribution is being evaluated which requires two sets of parameters.

5. LANDMINE DETECTION

In the following, a real-world landmine dataset is tested. In GPR images, scanning from left to right, a landmine signature would appear as a rising edge followed by a falling edge as shown in Fig. 3. Therefore, edge features are computed from GPR images and edge feature sequences are constructed for each horizontal image scan. The goal is to learn the horizontal patterns indicative of a landmine signature using an HMM model. In the following, first the dataset is described, then MI-HMM is compared to a state-of-the-art HMM [17]; a benchmark approach that is currently used in the field, which is referred to hereafter as the "standard HMM".

A NIITEK Inc. landmine detection system with a GPR sensor was used to collect data from various test sites consisting of gravel and dirt roads containing buried landmines and clutter objects. Typical landmine and clutter signatures are shown in Fig. 3. Subsurface objects appear as hyperbolic signatures within the GPR data. To lessen the computational burden of more complex algorithms, a standard prescreening algorithm is run to identify areas of interest, also called alarms. The resulting data collection consists of approximately 1,000 target alarms and 2,500 non-target alarms. A GPR alarm is a 3-dimensional data cube: 416 samples in depth, 61 samples downtrack (down the road) and 24 samples crosstrack (for each channel in the GPR antenna). Several feature extraction steps are performed in order to construct the observation sequences, as explained in detail in [8, 28]. With these steps, horizontal image scans are converted into feature vector sequences that indicate the presence of various edge types. As a result, at each fixed depth there is a horizontal sequence of edge feature vectors of length 15. Each image has 416 potential training sequences but has only one class label associated with the image. Therefore, to reduce the number of non-target sequences from a target image, the selection of training samples is aided using a Markov Random Field (MRF) "bounding box" [29]. The goal of the MRF is to bound the subimage with the highest energy – the target. This is a standard and automated procedure to reduce the arduousness of the task and the enormity of data typically used. Though this initial step eliminates many of the non-target sequences within a target image, it is not perfect and much ambiguity during training remains.

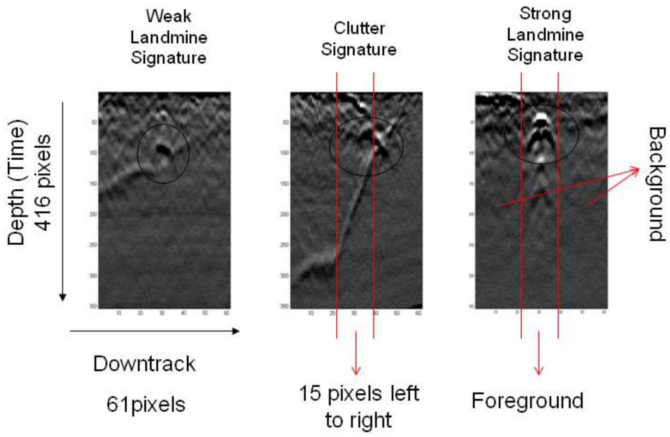


Fig. 3. Three GPR downtrack images showing landmine and clutter signatures (which are circled). The foreground and background areas used for preprocessing are also shown.

5.1. Experimental Results

The proposed MI-HMM is compared to the standard HMM [17] using the aforementioned landmine data. The MI-HMM uses discretized sequences, so the feature vectors were discretized (uniformly) to one of 25 different symbols. The standard HMM uses non-discretized sequences as an input. Both algorithms use a Gibbs sampling optimization schedule: the proposed method uses an MIL objective; the standard HMM uses a joint probability objective and two HMMs (target model and background model). Training for the MI-HMM is as follows: for each target image, five evenly spaced sequences were selected from within the MRF bounding box and placed into positive bags, and five randomly selected sequences were chosen from non-target images and placed into negative bags. The standard HMM algorithm makes use of two HMMs. It trains a target HMM using a training set of sequences from target images and trains a non-target HMM using a training set of sequences from non-target images. For a fair comparison, the same sequences used in MI-HMM were used to train the two HMMs of the standard HMM. Testing using the MI-HMM is performed by summing the log of the probabilities of each of the 416 sequences in each of the 24 images for each alarm. This accumulated value is considered the target confidence for each alarm. Testing for the standard HMM is the log of the ratio of the probability of the target model over the probability of the non-target model.

In addition to the above comparison, we designed a second experiment where we construct oracles for both the MI-HMM and the standard HMM. Simply speaking, the main target concept that the HMMs should be learning is a sequence with a rising edge followed by a falling edge. Therefore, the oracle simply sifts through all of the test images and disregards all sequences that do not have a strong rising edge followed by a strong falling edge. These oracles show the upper bounds of what an MI-HMM or standard HMM can achieve.

Classification results comparing the MI-HMM and the Standard HMM are presented via a Receiver Operating Characteristic (ROC) curve in Fig. 4 for ten fold cross validation. A ROC curve is a plot of the probability of detection (PD) vs. false alarm rate (FAR). Each ROC curve is shown with error bars which show the 95% confidence interval assuming a binomial distribution on the PD. The results show a significant improvement in classification results using the proposed MI-HMM vs. the standard HMM. FAR results are decreased by 75% at PDs of 70 and 80 and show a greater than 95% FAR reduction at a PD of 90%. In fact, the ROC for the MI-HMM dominates the ROC for the standard HMM at all operating thresholds above a PD of 70. Furthermore, the MI-HMM ROC dominates well outside the 95% confidence interval which indicates improved classification that is statistically significant. The comparison of the HMMs to their respective oracles is also quite notable. When the oracle algorithm is used, both models have statistically similar performances. This indicates that the standard HMM algorithm had the same potential performance upperbound as the MI-HMM, but failed to achieve similar performance results. On the other hand, the MI-HMM could perform near to its oracle even without the oracle algorithm.

6. CONCLUDING REMARKS

An HMM with a MI learning scheme has been presented. MI-HMM has a very clean and elegant mathematical model since there is no addition of parameters, but rather an assumption of the learning scenario. Within the landmine data experiments, the MI-HMM significantly outperformed a standard HMM algorithm which made use of two HMMs (twice the parameters). It also performed near to an "oracle" version of itself. Given the results of these experiments, it is clear that the use of an MI learning scheme when an MI scenario is present can increase classification results.

7. REFERENCES

[1] Q. Zhang, S. A. Goldman, W. Yu, and J. E. Fritts, "Content-based image retrieval using multiple-instance learning," in *Proc. of the 19th Int. Conf. on Machine Learning*, July 2002.

[2] Jun Yang, "Review of multi-instance learning and its applications," Online, 2008.

[3] Zhi-Hua Zhou, Min-Ling Zhang, Sheng-Jun Huang, and Yu-Feng Li, "Multi-instance multi-label learning," *Artificial Intelligence*, vol. 176, no. 1, pp. 2291–2320, 2012.

[4] Z.H. Zhou, "Multi-instance learning: A survey," Tech. Rep., AI Lab, Department of Computer Science and Technology, Nanjing University, Mar. 2004.

[5] Thomas G. Dietterich, Richard H. Lathrop, Tomas Lozano-Perez, and Arris Pharmaceutical, "Solving the multiple-instance problem with axis-parallel rectangles," *Artificial Intelligence*, vol. 89, pp. 31–71, 1997.

[6] H. Frigui, P. Gader, and D. Ho, "Real time landmine detection with ground penetrating radar using discriminative and adaptive hidden Markov models," *EURASIP Journal on Applied Signal Processing*, vol. 1, pp. 1867–1885, January 2005.

[7] J.N. Wilson, P. Gader, Wen-Hsiung Lee, H. Frigui, and K.C. Ho, "A large-scale systematic evaluation of algorithms using ground-penetrating radar for landmine detection and discrimination," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 8, pp. 2560 –2572, Aug. 2007.

[8] H. Frigui and P. Gader, "Detection and discrimination of land mines in ground-penetrating radar based on edge histogram descriptors and a possibilistic K-nearest neighbor classifier," *IEEE Transactions on Fuzzy Systems*, vol. 17, no. 1, pp. 185–199, Feb. 2009.

[9] H. Frigui, Lijun Zhang, and P.D. Gader, "Context-dependent multisensor fusion and its application to land mine detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 6, pp. 2528–2543, June 2010.

[10] Oded Maron and Toms Lozano-Prez, "A framework for multiple-instance learning," in *Advances in Neural Information Processing Systems (NIPS)*. 1998, pp. 570–576, MIT Press.

[11] Vikas C. Raykar, Balaji Krishnapuram, Jinbo Bi, Murat Dundar, and R. Bharat Rao, "Bayesian multiple instance learning: Automatic feature selection and inductive transfer," in *Proceedings of the 25th International Conference on Machine Learning*, pp. 808–815. 2008.

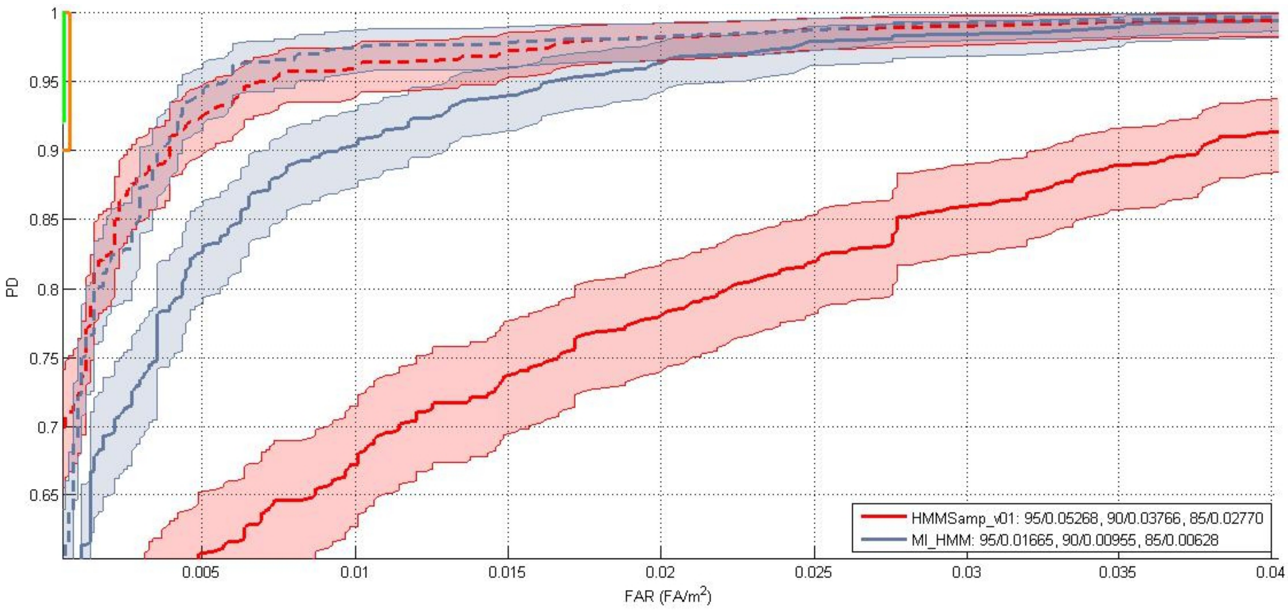


Fig. 4. ROC curves for MI-HMM and standard HMM shown with solid lines, and MI-HMM oracle and standard HMM oracle shown with dashed lines.

[12] C. Bergeron, G. Moore, J. Zaretzki, C. Breneman, and K. Bennett, "Fast bundle algorithm for multiple instance learning," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, , no. 99, pp. 1, 2011.

[13] J. Bolton and P. Gader, "Cross entropy optimization of the random set framework for multiple instance learning," in *20th International Conference on Pattern Recognition (ICPR)*, Aug. 2010, pp. 3907–3910.

[14] Jeremy Bolton, Paul Gader, Hichem Frigui, and Pete Torrione, "Random set framework for multiple instance learning," *Inf. Sci.*, vol. 181, no. 11, pp. 2061–2070, June 2011.

[15] P. Gader, M. Mystkowski, and Y. Zhao, "Landmine detection with ground penetrating radar using hidden Markov models," *IEEE Trans. Geosciences and Remote Sensing*, vol. 39, no. 6, pp. 1231–1244, 2001.

[16] Y. Zhao, P. Gader, P. Chen, and Y. Zhang, "Training dhmms of mine and clutter to minimize landmine detection errors," *IEEE Trans. Geosciences and Remote Sensing*, vol. 41, no. 5, pp. 1016–1024, May 2003.

[17] Xuping Zhang, Seniha Esen Yuksel, Paul Gader, and Joseph Wilson, "Simultaneous feature and HMM model learning for landmine detection using ground penetrating radar," in *6th IAPR Workshop on Pattern Recognition in Remote Sensing (PRRS)*, Aug. 2010, pp. 1–4.

[18] O. Missaoui, H. Frigui, and P. Gader, "Land-mine detection with ground-penetrating radar using multistream discrete hidden Markov models," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 6, pp. 2080–2099, June 2011.

[19] Judea Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1988.

[20] Lawrence R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," in *Proceedings of the IEEE*, 1989, pp. 257–286.

[21] M.A. Mohamed and P. Gader, "Generalized hidden Markov models. I. Theoretical frameworks," *Fuzzy Systems, IEEE Transactions on*, vol. 8, no. 1, pp. 67–81, Feb 2000.

[22] Biing-Hwang Juang, Wu Hou, and Chin-Hui Lee, "Minimum classification error rate methods for speech recognition," *IEEE Transactions on Speech and Audio Processing*, vol. 5, no. 3, pp. 257–265, May 1997.

[23] A. Biem, "Minimum classification error training for online handwriting recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 7, pp. 1041–1051, july 2006.

[24] P.M. Djuric and Joon-Hwa Chun, "An MCMC sampling approach to estimation of nonstationary hidden Markov models," *Signal Processing, IEEE Transactions on*, vol. 50, no. 5, pp. 1113–1123, may 2002.

[25] J. Paisley and L. Carin, "Hidden Markov models with stick-breaking priors," *Signal Processing, IEEE Transactions on*, vol. 57, no. 10, pp. 3905–3917, oct. 2009.

[26] D.J.C. MacKay, *Information theory, inference, and learning algorithms*, Cambridge Univ Press, 2003.

[27] Christopher M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[28] P. Gader, Wen-Hsiung Lee, and J.N. Wilson, "Detecting landmines with ground-penetrating radar using feature-based rules, order statistics, and adaptive whitening," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 11, pp. 2522–2534, Nov. 2004.

[29] Leslie Torrione, Peter A.; Collins, "Application of markov random fields to landmine detection in ground penetrating radar data," in *Proceedings of the SPIE, Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XIII*, 2008, vol. 6953, pp. 69531B–69531B–12.