

Simultaneous Feature and HMM Model Learning for Landmine Detection using Ground Penetrating Radar

Xuping Zhang, Seniha Esen Yuksel, Paul Gader, Joseph N. Wilson

Department of Computer and Information Science and Engineering, University of Florida

{xupzhang, seyuksel, pgader, jnw}@cise.ufl.edu

Abstract

Hidden Markov Models (HMMs) have been widely used in landmine detection with Ground Penetrating Radar (GPR) data; however, to the best of our knowledge, there are no other studies that investigated the simultaneous learning of the features and the HMM parameters. In this paper, we present a novel method based on Gibbs sampling that both learns a feature extraction model as well as an HMM model. The new system allows for the training of new features when the sensor systems are different. Experiments show that our algorithm is more robust to initialization and can find better solutions.

1. Introduction

In landmine detection, a moving GPR system collects multi-dimensional signals at regular intervals; and the features extracted from these signals are fed into an HMM model that converts the signals into probabilities [1, 2]. In current mine detection systems, although very noisy, mine signatures look like hyperbolas in 2-D GPR images; and morphological operations [3, 4] are used to evaluate the degree of a diagonal or anti-diagonal shape in small windows.

In these systems, it is very important that the features represent the signal accurately and succinctly. However, often times, the features need to be modified to accommodate a change in sensor systems. Such a modification process is laborious and requires specific domain knowledge. Hence, it is worthwhile to investigate automatic feature learning methods.

In this paper, we present a new learning method based on Gibbs sampling [5] which is one of the Monte Carlo Markov Chain (MCMC) sampling methods [6, 7]. Instead of maximizing an objective function, Gibbs sampling samples the parameters from their conditional probability distributions. Becoming popular in the

Bayesian community, it is more robust to initialization and usually achieves a better solution that is closer to the global optimum.

2. Technical approach

Our goal is to maximize the probability of mine signatures given the parameters of feature extraction and the parameters of the HMM model. In our Bayesian framework, we sample these parameters from their individual full conditional probability distributions when our probability assumptions match the true distributions of real data. In this section, we briefly describe Gibbs sampling, formulate the feature model, and state the assumptions to put the problem into the Bayesian framework.

2.1 Gibbs sampling

Gibbs sampling is used for joint distribution estimation when the full conditional distributions of all the random variables are available. Gibbs sampling provides samples to estimate the joint distribution of the random hidden variables and parameters. Then it uses those samples to estimate the parameters of models or compute the probabilities of new data given the model. Given joint density $p(x_1, x_2, \dots, x_p)$ for a set of random x_1, x_2, \dots, x_p , we sample the variable x_i from the condition distribution $p(x_i | x_j; i \neq j)$. Starting from initial values $x_1^{(0)}, \dots, x_p^{(0)}$, the Gibbs sampler draws samples of random variables as follows:

$$\begin{aligned} x_1^{(\tau+1)} &\sim p\left(x_1 | x_2 = x_2^{(\tau)}, \dots, x_k = x_k^{(\tau)}\right) \\ &\vdots \\ x_i^{(\tau+1)} &\sim p\left(x_i | \left(x_1 = x_1^{(\tau+1)}, \dots, x_{i-1} = x_{i-1}^{(\tau+1)}\right), \left(x_{i+1} = x_{i+1}^{(\tau)}, \dots, x_k = x_k^{(\tau)}\right)\right) \\ &\vdots \\ x_k^{(\tau+1)} &\sim p\left(x_k | x_1 = x_1^{(\tau+1)}, \dots, x_{k-1} = x_{k-1}^{(\tau+1)}\right) \end{aligned}$$

where τ denotes the index of iterations.

It is shown that as $\tau \rightarrow \infty$, the distribution of (x_1, x_2, \dots, x_p) converges to $p(x_1, x_2, \dots, x_p)$. Equivalently, as $\tau \rightarrow \infty$, the distribution of $x_i^{(\tau)}$ converges to $p(x_i)$ for $i = 1, \dots, p$. So the Gibbs sampler treats the samples $x_1^{(\tau)}, \dots, x_p^{(\tau)}$, for $\tau > M$ as a sample from $p(x_1, x_2, \dots, x_p)$ by selecting some large value for M . The initial sampling procedure before convergence is known as the *burn-in procedure*. Now we can calculate the expectation of a function $f(x)$ over the distribution $p(x_i)$ by Monte Carlo integration

$$E_{p(x_i)}[f(x_i)] = \int f(x_i)p(x_i)dx \approx \frac{1}{N} \sum_{\tau=M}^{M+\omega} (f(x_i^{(\tau)}))$$

where τ is the index of iterations in the sampling process, and ω is total number of samples collected.

2.3 Model formulation

A given image A is split into p overlapping $N \times K$ subimages as shown in Figure 1:

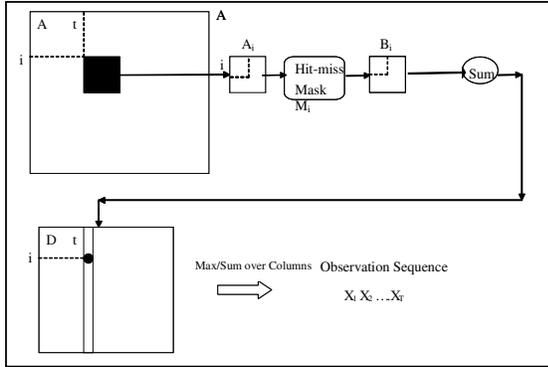


Figure 1: System for feature learning

We transform each subimage to an $NK \times 1$ vector $A_i, i=1, \dots, p$. Letting $l = NK$, for each image A_i , an $l \times 1$ binary hit mask M_i is applied using convolution operation as follows. We define the vector B_i as

$$B_i = A_i \circ M_i + \zeta, i = 1, \dots, p, \zeta \sim N(0, \sigma_\zeta^2 I_{L \times 1})$$

where the symbol \circ denotes point wise (Hadamard) multiplication, and ζ represents a zero mean Gaussian perturbation with covariance matrix $\sum_\zeta = \sigma_\zeta^2 I_{L \times 1}$.

Therefore, the k -th element of B_i can be denoted as:

$$B_{ik} = A_{ik} M_{ik} + \zeta_k, \zeta_k \sim N(0, \sigma_\zeta^2)$$

Now we define D_i as the sum over B plus a zero mean additive Gaussian perturbation with variance σ_η^2 :

$$D_i = \sum_k B_{ik} + \eta, \quad \eta \sim N(0, \sigma_\eta^2)$$

Suppose we have T subimages along the row. For each column of subimages we define one feature $x_t, t = 1, \dots, T$, as the aggregation of D_i with the additive zero mean Gaussian noise ϵ , i.e.,

$$x_t = \sum_i D_i + \epsilon, \quad \epsilon \sim N(0, \sigma_\epsilon^2)$$

Now we assign the label y_t to the feature x_t , given the threshold ξ .

$$y_t = \begin{cases} 1, & x_t > \xi \\ 0, & x_t \leq \xi \end{cases}$$

We assign x_t as the state probability density within HMM model framework. With the transition probability matrix, we can compute the probabilities of image sequences by an HMM algorithm [8].

To put the problem into the Bayesian framework, we assume the probability of binary hit mask M is given by a binomial distribution.

$$P(M_{ik} | p_k) \propto p_k (1 - p_k)$$

We assume the prior of the binary mask is the conjugate prior of binomial distribution. So given the α and β , the prior for the probability p_k is the beta distribution given as

$$p_k \sim \text{beta}(\alpha_k, \beta_k)$$

2.4 Computation by Gibbs Sampling

Since the posterior distribution is not available in explicit form, we use Gibbs sampling to sample all the needed variables M, B, D , and x as follows:

1) Sample the prior p given (M, α, β) from a beta distribution.

$$P(p_k | M_{*k}, \alpha_k, \beta_k) \propto \text{beta}(\alpha_k + n_k, \beta_k + N_k - n_k),$$

where $n_k = \#(M_{*k} = 1)$

2) Sample the Mask M given (p, B, A) .

$$P(M_{ik} = 1 | p_k, B, A) \propto (1 - p_k) \exp\left(-\frac{(B_{ik} - A_{ik} M_{ik})^2}{2\sigma_\zeta^2}\right)$$

After these two values are normalized, we sample M :

$$M_{ik} | p_k, B, A \sim \text{binomial}(P(M_{ik} = 1 | p_k, B, A), P(M_{ik} = 0 | p_k, B, A))$$

3) Sample the variable B given (A, M) .

$$\begin{aligned}
P(B_i | D_i, A_i, M_i) &\propto P(D_i | B_i)P(B_i | A_i, M_i) \\
&\propto \exp\left(-\left(D_i - \sum_{k=1}^L B_{ik}\right)^2 / 2\sigma_\eta^2\right) \\
&\exp\left(-\sum_{k=1}^L (B_{ik} - A_{ik}M_{ik})^2 / 2\sigma_\zeta^2\right)
\end{aligned}$$

Rather than sampling B as a matrix, it is better to sample component-wise from Gaussian distribution.

$$B_{ik} | B_{-ik}, A_{ik}, M_{ik} \sim N\left(\frac{\tau_\zeta A_{ik} M_{ik} + \tau_\eta \left(D_i - \sum_q B_{iq} + B_{ik}\right)}{\tau_\zeta + \tau_\eta}, (\tau_\zeta + \tau_\eta)^{-\frac{1}{2}}\right),$$

where $\tau = 1/\sigma^2$ denotes the precision of the Gaussian.

4) Sample the variable D given (x, B) .

$$\begin{aligned}
P(D | x_i, B_i) &\propto P(D | B_i)P(x_i | D) \\
&\propto \exp\left(-\sum_i \left(D_i - \sum_{k=1}^L B_{ik}\right)^2 / 2\sigma_\eta^2\right) \\
&\exp\left(-\left(x_i - \sum_i D_i\right)^2 / 2\sigma_\varepsilon^2\right).
\end{aligned}$$

Similarly D is sampled component wise from the Gaussian distribution.

$$D_i | D_{-i}, x_i, B_i \sim N\left(\frac{\tau_\eta \sum_{k=1}^L B_{ik} + \tau_\varepsilon \left(x_i - \sum_q D_q + D_i\right)}{\tau_\eta + \tau_\varepsilon}, (\tau_\eta + \tau_\varepsilon)^{-\frac{1}{2}}\right)$$

5) Sample the variable x given (y, D) from the truncated Gaussian distribution with threshold ξ .

$$\begin{aligned}
P(x_i | y_i = 1, D) &\propto N\left(\sum_i D_i, \sigma_\varepsilon^2\right) \text{ truncated at right by } \xi, \\
P(x_i | y_i = 0, D) &\propto N\left(\sum_i D_i, \sigma_\varepsilon^2\right) \text{ truncated at left by } \xi.
\end{aligned}$$

After the burn-in period we can get the Gibbs samples at the s -th iteration as $[p^{[s]}, M^{[s]}, B^{[s]}, D^{[s]}, x^{[s]}]$,

$s = 1, \dots, I$]; and use these samples for prediction and posterior inference.

The Gibbs Sampling Algorithm:

Set hyper parameter $\alpha, \beta, \sigma_\zeta, \sigma_\eta, \sigma_\varepsilon$ and threshold ξ . Start with the initial state sequences $Q^{[0]}$.

Loop

Split all the images sequences as different image segments according to the state that they are associated to.

Loop for each state segments

Start with the initial $[p^{[0]}, M^{[0]}, B^{[0]}, D^{[0]}, x^{[0]}]$.

Loop to sample the state parameters

At the iteration s ,

Sample prior $p^{[s]}$ given $(M^{[s-1]}, \alpha, \beta)$.

Sample $M^{[s]}$ given $(p^{[s]}, B^{[s-1]}, A)$.

Sample $C^{[s]}$ given $(A, M^{[s]})$.

Sample $D^{[s]}$ given $(x^{[s-1]}, B^{[s]})$.

Sample $x^{[s]}$ given $(y, D^{[s]})$.

Stop after the required number of iterations.

End loop of state segments

Compute every state probability density with sampled state parameters for all the images sequences.

Find the best state sequences by Viterbi algorithm

Stop after the fix number of iterations

The parameters were initialized with a random number generator, and the hyper-parameters were initialized based on the experiments.

To predict the label of new sequences, in MCMC approach, $P(y_{new} = 1 | A_{new})$ can be estimated by

$$\begin{aligned}
&P(y_{new} = 1 | A_{new}, M, A) \\
&= \frac{1}{l} \sum_{s=M}^{M+l} P(y_{new} = 1 | A_{new}, B^{(s)}, M^{(s)}, D^{(s)}, x^{(s)})
\end{aligned}$$

However to make the computations easy, the mode of the samples is used as the estimated parameter.

3. Experiments and Results

Experiments were performed on images acquired from an arid test site with the NIITEK GPR [9]. The data set contains two classes (mines and non-mines). In this study, the data was first preprocessed such that each image was normalized and scaled to the [0, 1] interval. Then, the images were binarized and skeletonized to get the clean gray-level image, and a 5x5 window was moved along the x-axis to get the image sequences. For each data sample of nine images, one image sequence was extracted. Twenty-five image sequence samples are shown in Figure 2, where each row is a sequence. Two sequences are separated by a horizontal gray bar, and two adjacent images in one sequence are separated by a vertical gray bar. It can be seen that the sequences consist of ascending-edge and descending-edge images. Two-hundred image

sequences were extracted from landmine data for mine class in the training set.

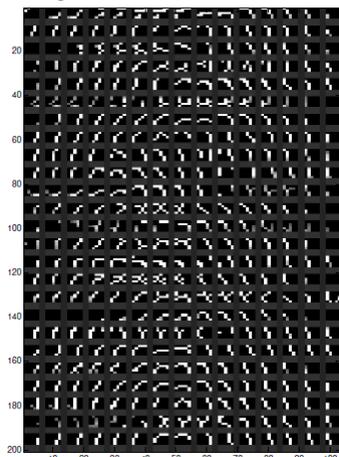


Figure 2: 25 sequences extracted from mine images

Probability of Detection (PD) vs Probability of False Alarm (PFA) is displayed in Figure 3 for (1) baseline HMM (HMM-EM), (2) Our Gibbs sampling HMM algorithm without feature learning, (3) Our Gibbs sampling feature learning with the HMM algorithm. Note that, the baseline HMM algorithm uses the parameters that has been extensively used over the years, and has become the standard in landmine detection. Our algorithm performs comparably to this baseline HMM, exceeding it at 90% detection, and also has the ability to adjust to different sensor configurations.

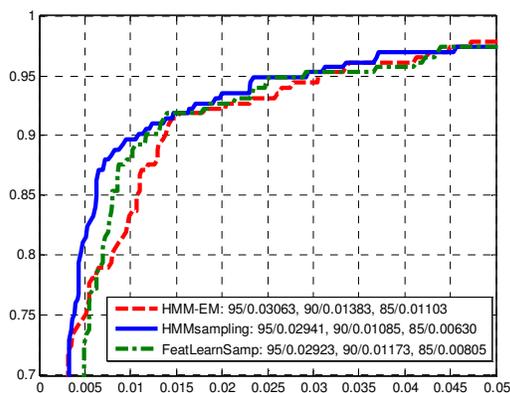


Figure 3: Receiver Operating Characteristic (ROC) curves comparing Gibbs feature learning without HMM learning, Gibbs feature and HMM learning and the standard HMM.

4. Conclusion

A new feature learning method based on Gibbs sampling with HMM model is proposed. It can extract

the features and learn the parameters of the HMM model simultaneously. The Gibbs sampling method is used so that the method is more robust to initialization and achieves a better solution. The experiments show that this method rivals the performance of other methods in landmine detection, and has the ability to adjust to different datasets collected under various environmental conditions. For our future work, we will investigate a composite model of feature and HMM learning, and apply this method to other problems.

5. Acknowledgment

This research was partially supported by NSF Grant No. 0730484.

References

- [1] P. Gader, M. Mystkowski, and Y. Zhao, Landmine Detection with Ground Penetrating Radar Using Hidden Markov Models, *IEEE Trans. Geosciences and Remote Sensing*, 39(6), pp 1231-1244, 2001.
- [2] 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, 2003.
- [3] P.Gader and A.Gillies, Recognition of Handwritten Digits using Template and Model Matching, *Pattern Recognition*, Vol. 24, No. 5, pp. 421-431, 1991.
- [4] P. Gader and M. Khabou, Automatic Feature Generation for Handwritten Digit Recognition, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 18, No 12, pp1246-1261, Dec 1996.
- [5] G. Casella and E. I. George. Explaining the Gibbs sampler. *The American Statistician*, 46(3):167.174, 1992.
- [6] R. M. Neal. Probabilistic inference using Markov chain Monte Carlo methods. *Technical Report CRG-TR-93-1*, University of Toronto, 1993.
- [7] R.E. Kass, B.P. Carlin, A. Gelman and R. M. Neal, Markov Chain Monte Carlo in Practice: A Roundtable Discussion, *The American Statistician*, vol. 52, pp. 93-100, 1998.
- [8] B.-H.Juang and L.R. Rabiner, The segmental K-means algorithm for estimating parameters of hidden Markov models, *IEEE Trans. Acoustics, Speech and Signal Processing*, Vol.38, no.9, pp.1639-1641, 1990.
- [9] W-H. Lee, P. D. Gader, and J. N. Wilson, "Optimizing the Area under a Receiver Operating Characteristic Curve with Application to Landmine Detection", *IEEE Trans. Geoscience and Remote Sensing*, vol. 45, No. 2, pp. 389-398, Feb. 2007