DETECTABILITY PREDICTION OF HIDDEN MARKOV MODELS WITH CLUTTERED OBSERVATION SEQUENCES

Karl Granström, Peter Willett, Yaakov Bar-Shalom

Department of Signals and Systems, Chalmers University of Technology, SE-41296 Gothenburg, Sweden
Email: karl.granstrom@chalmers.se
Department of Electrical and Computer Engineering, University of Connecticut, Storrs, 06269 CT, USA
Email: willett@engr.uconn.edu, ybs@engr.uconn.edu

ABSTRACT

There is good reason to model an asymmetric threat (a structured action such as a terrorist attack) as an HMM whose observations are cluttered. Recently a Bernoulli filter was presented that can process cluttered observations ("transactions") and is capable of detecting if there is an HMM present, and if so, estimate the state of the HMM. An important question in this context is: when is the HMM-in-clutter problem feasible? In other words, what system properties allow for a solvable problem? In this paper we show that, given a Gaussian approximation of the pdf of the log-likelihood, approximate detection error bounds can be derived. These error bounds allow a prediction of the detection performance, i.e. a prediction of the probability of detection given an "operating point" of transaction-level false alarm rate and miss probability. Simulations show that our analysis accurately predicts detectability of such threats. Our purpose here is to make statements about what sort of threats can be detected, and what quality of observations are necessary that this be accomplished.

Index Terms— Asymmetric threat, Hidden Markov Models, Bernoulli filter, detectability.

1. INTRODUCTION

The term *asymmetric threat* refers to tactics employed by, e.g., terrorist groups to carry out attacks on a superior opponent, while trying to avoid direct confrontation. Analysis of prior terrorist attacks suggests that a high magnitude terrorist attack requires certain enabling events to take place. In this paper terrorist activities are modeled as Hidden Markov (HMM). Excellent tutorials on HMMs can be found in [1, 2]. The applicability of HMMs for terrorist activity modeling and other national security problem situations has been illustrated in previous work, see e.g. [3, 4, 5, 6, 7, 8].

A number of different terrorist plan HMMs are proposed in [5, 6, 7, 8], including models for a truck bombing, Figure 1, and production of weapons grade material, Figure 2.

Submitted to ICASSP 2016.

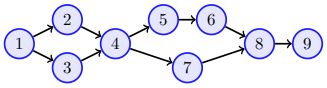

Fig. 1. Markov chain network modeling the planning of a truck bombing. Refer to [5] for additional details.

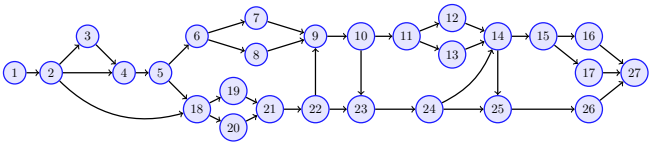

Fig. 2. Markov chain network modeling the production of weapons grade material. Refer to [8] for additional details.

These HMMs include multiple paths from plan conception to plan completion, following the intuition that there are multiple ways to, e.g., hijack a plane. An empirical HMM can be constructed using available prior information, or with the help from experienced intelligence analysts (SMEs) [5]. For example, the HMM for *development of a nuclear weapons program* (DNWP) in [7] is gleaned using the open sources [9, 10, 11, 12, 13].

The basic motivation for modeling terrorist activities via HMMs is twofold. First, carrying out a terrorist activity requires planning and preparations, following steps that form a pattern, admittedly with some "options" modeled as parallel paths within the HMM. This pattern of actions can be modeled using a Markov chain. Second, the terrorists leave detectable clues about these enabling events in the observation space. The clues are not direct observations of the planning and preparations, but are rather related to them, meaning that the states in the Markov model (hence the name) are hidden. For example, an observation of a purchase of chemicals could be indicative of intentions to produce a chemical weapon. However, a purchase of chemicals could very well be a benign event, which motivates inclusion of a model of observations that are unrelated to the HMM. Following the target tracking literature, see e.g. [14], such observations are

here designated as clutter observations.

Ultimately the task is to find out if there is an activity being planned, and if so, find what stage the planning is in. A Bernoulli filter that can solve this problem was recently presented in [15]. Specifically, the Bernoulli filter jointly estimates the probability of HMM existence, denoted $q_{k|k}$, and the pmf, denoted $P_{k|k}(\cdot)$.

Given the Bernoulli filter [15], an important question to consider is: for what system properties is the HMM-in-clutter problem feasible? In this paper we approach this question by focusing on the detection part of the problem, corresponding to the estimated probability of HMM existence in the Bernoulli filter. The ultimate aim is to be able to make statements about maximum levels of clutter allowable; maximum intervals between relevant observations; and a minimum level of complexity.

Specifically, the error probabilities are evaluated, i.e., the probabilities of missed detection and of false detection. Using error probability approximations it is possible to, for a set of parameters that govern the properties of the HMM-in-clutter problem, predict what the detection rate will be for a given false alarm rate. A comparison to empirical results show that the prediction is accurate. A preliminary detectability analysis was presented in [16, 17], this paper extends the work by including a thorough comparison to empirical data.

2. ASYMMETRIC THREAT MODELING

Let $\zeta_k \in \mathcal{S}$ denote the HMM state at time t_k, where $\mathcal{S}$ is a discrete state space with N_s states, $\mathcal{S} = \{S_1, S_2, \ldots, S_{N_s}\}$. In the variant of HMMs used here the observations become available only upon state transitions, and the HMM state transitions follow a first order Markov chain. The observations $\mathbf{z}_k \in \mathcal{Z}$ are discrete random variables, where $\mathcal{Z}$ is a discrete state space with N_z states, $\mathcal{Z} = \{Z_1, Z_2, \ldots, Z_{N_z}\}$. If an HMM state transition has happened, then with probability of detection $p_{\mathrm{D}} \in (0, 1)$ the HMM generates an observation $\mathbf{z}_k$. The HMM observation process is defined by the probability mass function (pmf) $g_s(\mathbf{z}_k|\zeta_k)$. There are also clutter observations (false alarms) super-imposed on the true HMM observations. In each time-step, with probability $p_{\mathrm{FA}} \in (0, 1)$ a clutter observation is generated as a random sample from a process with pmf $g_{\mathrm{FA}}(\mathbf{z}_k)$.

3. DETECTABILITY PREDICTION

Given a sequence of time steps, for some of the time steps there will be an observation, denoted $\mathbf{z}_k$, and for some there will be no observation. The time is assumed to be discretized in small enough increments such that there is never more than a single observation per time step. Further, the scope of the paper is limited by the assumption that the parameters of the HMM and the clutter process are known. An important topic for future work is to consider modeling errors, i.e. unknown parameters.

In this section we present a detectability prediction method. We consider the following two hypotheses:

H_0: The observations were generated by a clutter process. This means that there is no structure in the sequence of observations, it is random.

H_1: The observations were generated by an HMM-in-clutter process. This means that among the random clutter observations, there are observations caused by the HMM that has some degree of structure.

To decide between the hypotheses H_0 and H_1 we employ a decision rule

$$\delta_\ell = \begin{cases} 1 & > \\ \gamma & \text{if } \ell = \tau \\ 0 & < \end{cases} \tag{1}$$

where ℓ is the log likelihood ratio, i.e., if $\delta_\ell = 1$ we choose H_1 and if $\delta_\ell = 0$ we choose H_0. To analyze the detectability we focus on the conditional error probabilities P_F and P_M defined by

$$P_F(\delta) = P_0(\delta \text{ chooses } H_1) \tag{2a}$$

$$P_M(\delta) = P_1(\delta \text{ chooses } H_0) \tag{2b}$$

Given probability density functions (pdfs) $p(\ell|H_0)$ and $p(\ell|H_1)$ it is easy to compute the errors for given model parameters and a given threshold τ. However, expressing the pdfs $p(\ell|H_i)$ analytically is prohibitively difficult and complex in the general case.

To alleviate this complexity we will consider a simplified type of HMM. Then using a Gaussian approximation of $p(\ell|H_1)$ an exact expression for the probability of miss P_M can be computed, and for the probability of false alarm P_F an upper bound can be computed. Lastly, using the upper bound for the false alarm probability, it is possible to derive the log likelihood threshold that gives a certain false alarm probability, and subsequently it is possible to compute a prediction for the probability of detection.

3.1. Daisy chain HMMs

We consider a very simple kind of HMM– the daisy chain – see Figure 3. It is defined by four model parameters:

1: N_S: the number of states in the chain. This parameter is related to the level of complexity of the HMM; generally higher N_S implies higher complexity.

2: P_T: the probability of transitioning to the next state. The transition probabilities are assumed uniform, i.e. in each time step the probability of remaining in the same state for one more time step is equal for all states; note that this assumption only applies to this HMM, it is not in general necessary for our analysis. Because the HMM observations are modelled as only becoming available upon state transitions,

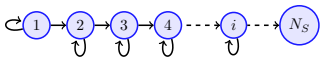

Fig. 3. Daisy chain Markov network

P_T is important as $1/(1-P_T)$ describes the expected length of the intervals between the observations that are relevant to the HMM.

3: p_{D}: the probability of an HMM generated observation, given that there was a state transition.

4: p_{FA}: the probability of a clutter generated observation, i.e. an observation that is not related to the HMM. This parameter governs the level of clutter.

In the examples in Figures 1 and 2 each way from the first to the last state forms a daisy chain, and therefore the comparatively simple daisy chain can be used as an approximation of more complex HMMs. It is assumed that the size of the observation state space is equal to the HMM state space, i.e. $N_z = N_s$, and that the observation pmf is

$$g_s(\mathbf{z} = Z_j | \mathbf{s}_k = S_i) = \begin{cases} P_{obs} & \text{if } j = i \\ \frac{1-P_{obs}}{N_z - 1} & \text{otherwise} \end{cases} \tag{3}$$

where $0 \ll P_{obs} \lesssim 1$ (i.e. P_{obs} is close to one). Additionally it is assumed that the clutter is uniformly distributed

$$g_{\mathrm{FA}}(\mathbf{z}) = {N_z}^{-1} = {N_s}^{-1} \tag{4}$$

i.e. it is equiprobable for all the N_z possibilities. This observation model means that each state has a unique observation, and it is unlikely that, given that a state transition is detected, it is reported incorrectly. For example, if there are two states, 1) rent apartment and 2) buy fertilizer, then if apartment rental is detected it is unlikely to be reported as fertilizer purchase, and vice versa. The likelihood ratio for a Daisy Chain HMM is given in [16, Eq. 15 - 16].

3.2. Approximate error probabilities

For the daisy chain type HMM described above previous work [16] has shown that, under hypothesis H_1, we can approximate the true pdf over ℓ with a Gaussian pdf

$$p(\ell | H_1) \approx \mathcal{N}(\ell\,;\, \hat{\mu}, \hat{\sigma}) \tag{5}$$

where the mean $\hat{\mu}$ and standard deviation $\hat{\sigma}$ are functions of the number of states N_S, the transition probability P_T, the probability of HMM observation p_{D}, and the probability of a clutter observation p_{FA}. Please refer to [16] for details that are too lengthy to repeat here.

The probability of miss $P_M(\delta)$ for a given threshold τ is then given by

$$P_M(\delta) = \mathrm{P}(\ell < \tau) = \int_{-\infty}^{\tau} \mathcal{N}(\ell\,;\, \hat{\mu}, \hat{\sigma})\, \mathrm{d}\ell = F_\ell(\tau) \tag{6}$$

where $F_\ell(\cdot)$ is the Gaussian cumulative distribution function (cdf).

An approximation of the true pdf over ℓ, under hypothesis H_0 has been attempted. However, empirical results showed that the Gaussian approximation of $p(\ell|H_0)$ is not sufficiently accurate. When the pdf for the log-likelihood under H_0 is unknown the probability of false alarm cannot be directly computed. However, using the pdf under H_1 we can derive the Chernoff bound for the probability of false alarm. The upper bound is, see e.g. [18],

$$P_F(\delta_\ell) \leq \exp\left(\mu_{\ell,0}(s) - s\tau\right) \tag{7}$$

for all $s > 0$, where $\mu_{\ell,i}$ is the cumulant generating function of ℓ under H_i. The upper bound (7) can be minimized over $s > 0$ to find the tightest bound. Using the relationship $\mu_{\ell,0}(s) = \mu_{\ell,1}(s-1)$, see e.g. [18], and a variable substitution $t = s - 1$, we can rewrite the bound (7) as

$$P_F(\delta_\ell) \leq \exp\left(\mu_{\ell,1}(t) - (t+1)\tau\right) \tag{8}$$

for all $t > -1$. Inserting the Gaussian cumulant generating function and maximizing w.r.t. t we get the minimum error bound

$$P_F(\delta_T) \leq \exp\left(t_F\hat{\mu} + \frac{1}{2}t_F^2\hat{\sigma}^2 - (t_F+1)\tau\right) \tag{9}$$

$$t_F = \max\left\{-1\,,\ \frac{\tau - \hat{\mu}}{\hat{\sigma}^2}\right\} \tag{10}$$

Note that a property of the Chernoff bound is that it may be trivial, i.e. for some $\hat{\mu}$, $\hat{\sigma}$ and τ the error bound is larger than one, see e.g. [18].

3.3. Predicting the detectability performance

Using the upper bound (9) for the probability of false detection it is possible to find a likelihood threshold τ_α that gives a probability of false detection less than or equal to α. Letting $t_F = (\tau - \hat{\mu})/\hat{\sigma}^2$ and solving (9) for τ we get

$$\tau_\alpha \leq \left(\hat{\mu} - \hat{\sigma}^2 + \sqrt{\hat{\sigma}^4 - 2\hat{\mu}\hat{\sigma}^2 - 2\hat{\sigma}^2\log(\alpha)}\right) \tag{11}$$

For some values of α, $\hat{\mu}$ and $\hat{\sigma}$ the solution will be an imaginary number. In this case τ is trivially given by setting $t_F = -1$, which gives $\tau = \hat{\mu} - \hat{\sigma}^2$. This gives a predicted detection $F_\ell(\tau_\alpha)$ for a given combination of the parameters N_S, P_T, p_{D} and p_{FA}.

Note that in general HMMs designed for asymetric threats are not daisy chains, nor do they have equal transition probability for all state. However, the detectability prediction can still be used, as will be shown in the next section.

4. SIMULATION RESULTS

Intelligence observation data of the kind considered here is inherently secret, and for this reason results for real observa-

tion data records are unavailable, and could not be published if they were. Instead we present results for simulated data.

Five different HMMs were simulated:

1) Planning of a truck bombing, see [5] for details.

2) Production of weapons grade material, see [8] for details.

3) Planning and strategy, see [6] for details.

4) Collection of resources, see [6] for details.

5) Preparations for a hijacking, see [6] for details.

In the remainder of this section we will refer to the models as HMM 1, HMM 2, and so on. Neither of the models is a simple daisy chain; on the contrary all five have more complex structure, as shown for HMM 1 in Figure 1, and for HMM 2 in Figure 2. Because of page length constraints, illustrations of remaining three HMMs are omitted.

The first two models have uniform transition probabilities. The last three models do not have this property, instead the transition probabilities are specified in the models, see [6]. For the HMM state transitions in models 1 and 2, three different probabilities of transition were simulated $P_T \in \{0.10, 0.20, 0.30\}$.

The HMM observation pmf (3) and clutter pmf (4) were used for all five models, with $P_{obs} = 0.99$. Different probabilities of HMM observation and probabilities of clutter observation were simulated, $p_{\mathrm{D}} \in \{0.10, 0.20 \ldots 0.90\}$ and $p_{\mathrm{FA}} \in \{0.10, 0.20 \ldots 0.90\}$, with $P_{obs} = 0.99$. Empirically we have found that it is not necessary to also simulate multiple values for P_{obs} because it is the product $P_{obs}p_{\mathrm{D}}$ that is important, i.e. it is sufficient to simulate different values of p_{D}. For each HMM and each parameter combination 100 Monte Carlo simulations were run. In Figure 4 the empirical detection rate at 10% empirical false alarm rate is shown[1], and the line along which the empirical detection rate is 50% is highlighted with a white line.

For each HMM and each parameter combination a detectability prediction was computed using the method outlined in Section 3. Because none of the five simulated HMMs is a daisy chain, there is no direct correspondence between the number of states N_S of the HMM, and the value N_S used in the prediction. Empirically we have found that in the prediction N_S should be set to the expected value of the number of states that the HMM passes through from first to last state. The reason for this is that, in order to pass from the first to the last state, it is not necessary to pass through each state. For example, in HMM 1 there are four different ways to go from state 1 to state 9, see Figure 1. The expected value of the number of states that are passed is 6.5, which is rounded down to $N_S = 6$.

Additionally, in HMMs 3, 4 and 5, the transition probabilities are not uniform, which is an assumption in the detectability prediction. In the detectability predictions for HMMs 3, 4 and 5 we have set the transition probability to the mean of the HMM transition probabilities.

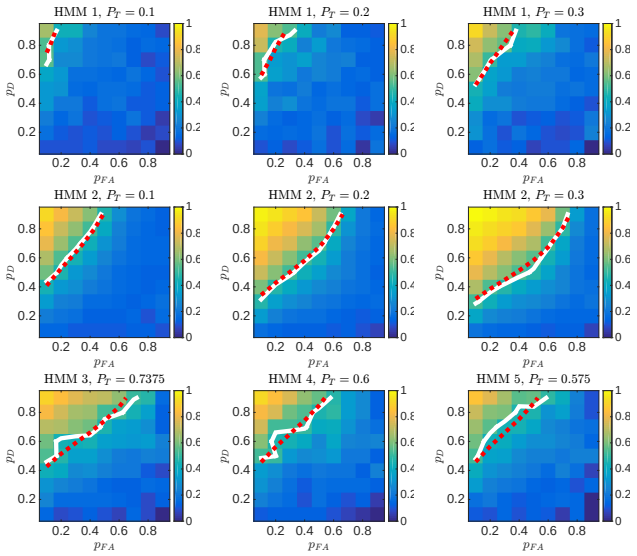

Fig. 4. The heat maps show empirical detection rate at 10% empirical false detection alarm rate for different combinations of the parameters p_{D} and p_{FA}; the title of each plot indicates the HMM and the probability of transition. The white lines indicate 50% empirical detection rate, and the red dotted lines indicate the corresponding predicted performance.

The predicted 50% detection at 10% false alarm line, as computed by the detectability prediction, is shown with a dotted red line in Figure 4. The results clearly show that the detectability prediction is quite accurate, especially for HMM 2, which is the most complex of the five simulated models.

5. CONCLUSIONS AND FUTURE WORK

Recently a Bernoulli filter was presented that can process a sequence of cluttered observations and determine if there is an underlying structure to the observations caused by a terrorist plan. This paper presented a detectability analysis of the problem that, given a set of model parameters, allows the probability of false alarm and probability of detection to be predicted. A comparison to empirical false alarm rate and detection rate show that the prediction is quite accurate. This is important because it will allow us to answer questions such as, for a given HMM, what is the maximum level of clutter that can be handled? Important topics for future work includes consideration of modelling errors, i.e., considering what the performance is when the parameters of the process causing the observations are unknown.

[1]Empirical detection rate is, for all time steps that an HMM existed, the % time steps that the estimated probability of existence was larger than the threshold $\tau_{0.1}$ (i.e. $\alpha = 0.1$). Empirical false alarm rate is, for all time steps that an HMM did not exist, the % time steps that the estimated probability of existence was larger than the threshold $\tau_{0.1}$.

6. REFERENCES

[1] L. Rabiner and B.-H. Juang, "An introduction to hidden Markov models," *IEEE ASSP Magazine*, vol. 3, no. 1, pp. 4–16, Jan. 1986.

[2] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[3] P. Schrodt, "Pattern recognition of international crises using hidden Markov models," in *Political Complexity: Nonlinear Models of Politics*, D. Richards, Ed., pp. 296–328. University of Michigan Press, Ann Arbor, MI, USA, 2000.

[4] T. Coffman and S. Marcus, "Dynamic classification of groups through social networks and HMMs," in *Proceedings of IEEE Aerospace Conference*, Big Sky, MT, USA, Mar. 2004, pp. 3197–3205.

[5] S. Singh, H. Tu, J. Allanach, J. Areta, P. Willett, and K. Pattipati, "Modeling threats," *IEEE Potentials*, pp. 18–21, Aug./Sept. 2004.

[6] H. Tu, J. Allanach, S. Singh, K. Pattipati, and P. Willett, "Information integration via hierarchical and hybrid Bayesian networks," *IEEE Transactions on Systems, Man, and Cybernetics–Part A: Systems and Humans*, vol. 30, no. 1, Jan. 2006.

[7] S. Singh, W. Donat, H. Tu, J. Lu, K. Pattipati, and P. Willett, "An advanced system for modeling asymmetric threats," in *Proceedings of 2006 IEEE International Conference on Systems, Man, and Cybernetics*, Taipei, Taiwan, Oct. 2006.

[8] S. Singh, H. Tu, W. Donat, K. Pattipati, and P. Willett, "Anomaly detection via feature-aided tracking and hidden Markov models," *IEEE Transactions on Systems, Man, and Cybernetics–Part A: Systems and Humans*, vol. 39, no. 1, pp. 144–159, Jan. 2009.

[9] F. Barnaby, *How to Build a Nuclear Bomb and Other Weapons of Mass Destruction*, Nation Books, New York, NY, USA, 2004.

[10] R. Paternoster, "Nuclear weapon proliferation indicators and observables," Tech. Rep. LA-12430-MS, Los Alamos National Laboratory, Dec. 1992.

[11] F. Settle, "Nuclear chemistry, nuclear proliferation," .

[12] L. Spector and J. Smith, *Nuclear Ambitions: The Spread of Nuclear Weapons 1989–1990*, Westview Press, Boulder, CO, USA, 1990.

[13] U.S. Congress, Office of Technology Assessment, "Technologies underlying weapons of mass destruction," Tech. Rep. OTA-BP-ISC-115, U.S. Printing Office, Washington, DC, USA, Dec. 1993.

[14] Y. Bar-Shalom, P. K. Willett, and X. Tian, *Tracking and Data Fusion: A Handbook of Algorithms*, YBS Publishing, 2011.

[15] K. Granström, P. Willett, and Y. Bar-Shalom, "A Bernoulli filter approach to detection and estimation of hidden Markov models using cluttered observation sequences," in *Proceedings of the IEEE Conference on Acoustics, Speech and Signal Processing*, Brisbane, Australia, Apr. 2015.

[16] K. Granström, P. Willett, and Y. Bar-Shalom, "Detectability analysis of detection and estimation of structured action from cluttered data," in *Proceedings of the International Conference on Information Fusion*, Washington, DC, USA, July 2015, pp. 173–179.

[17] K. Granström, P. Willett, and Y. Bar-Shalom, "Asymmetric threat modeling using HMMs: Bernoulli filtering and detectability analysis," *IEEE Transactions on Signal Processing*.

[18] H. V. Poor, *An Introduction to Signal Detection and Estimation, Second Edition*, Springer, 1994.