2008 American Control Conference
Westin Seattle Hotel, Seattle, Washington, USA
June 11-13, 2008

Model Reduction For Reduced Order Estimation in Traffic Models

Joseph S. Niedbalski, Kun Deng, Prashant G. Mehta, and Sean Meyn

Abstract— **This paper is concerned with model reduction for a complex Markov chain using state aggregation. The work is motivated in part by the need for reduced order estimation of occupancy in a building during evacuation. We propose and compare two distinct model reduction techniques, each of which is based on the potential matrix for the Markov semigroup. The first method is based on spectral graph partitioning where the weights are defined by the entries of the potential matrix. The second approach is based on aggregating states with similar long term uncertainty, where uncertainty is captured using conditional entropy. It is shown that entropy can be conveniently expressed in terms of the potential matrix.**

In application to the building model, the entries of the potential matrix correspond to the mean time an individual occupies a given cell. Numerical results are described, including a simulation study of the reduced order estimator.

I. Introduction

Grid based methods, including cellular automata [1], are computationally attractive for simulating traffic in buildings, planes, and outdoor walkways [2], [3]. The power of these methods lies in their ability to simulate agent-based behavior by having each agent follow a set of heuristic rules that determine how agents move and interact with one another. The discrete-time/discrete-state models favored in these papers are useful for efficient computer simulations for the purposes of performance assessment and off-line design, but their application for real-time control and estimation has been limited on account of complexity.

The focus of this paper is on Markovian models, and the particular application of interest is the modeling of occupancy evolution in a large building. Of particular interest is *egress*. That is, we consider a transient regime in which the occupants will leave the building eventually. Egress may be due to an emergency, or the end of a work day. In this case the Markov model contains a single absorbing state that represents an empty building. The floors in the building are divided into cells to create a large grid, as illustrated in Fig. 1.

A sensor network comprising of several sensors scattered throughout the building monitor specific regions of the grid. Subject to conditions on the occupancy evolution and the observation process, the system can be described as a hidden Markov model (HMM) [4], [5]. The most natural estimator in this context is the Bayesian estimator, which is optimal with respect to an ℓ_2 criterion [6].

Model reduction is essential in this application since common models explode in complexity with building size and topology. For example, a natural Markov model for the simple building illustrated in Fig. 1 has 400 nodes. With a single person, this leads to 400 states for the Markov chain. For a crowd of people with a limit of one person per node, this grows to 2^{400} possible states.

The authors are with the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, 1308 West Main Street, Urbana, IL 61801. Email: niedbals@uiuc.edu, kundeng2@uiuc.edu, mehtapg@uiuc.edu, and meyn@control.csl.uiuc.edu

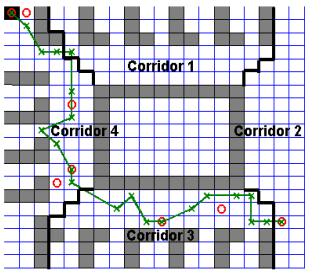

Fig. 1. Layout of Office Model where × corresponds to the sample path of the agent, ○ are the sensors, the heavy black lines show the divisions between the corridors, gray grid spaces are walls, and the black grid space is the exit. The agent starts in the bottom-right corner and proceeds toward the exit according to simple probabilistic rules.

Perhaps the most natural approach to model reduction is to aggregate states into *super-states* to obtain a Markov model on the aggregated state space. This approach can be justified using the notion of nearly completely decomposable Markov chains introduced by Ando and Fisher [7], which is closely related to quasi-stationarity [8]. This is the background for the decomposition results using singular perturbation in [9]. For a certain restricted class of Markov chains, aggregation has also been considered using the concept of *lumpability* [10]. This approach can offer substantial computational savings in the estimation [11], as well as the control.

An important new tool for understanding multi-scale phenomenon is based on the spectral theory of Markov models. For a finite state-space ergodic Markov chain the second eigenvalue is precisely the rate of convergence to stationarity. A more recent contribution to the theory of Markov chains is the use of the second eigenvector (or eigenfunction) to obtain intuition regarding dynamics, as well as methods for aggregation in complex models. This technique was introduced as a heuristic in [12], [13] to obtain a state-space decomposition based on an analysis of the *Perron cluster of eigenvalues* for the generator of a Markov process. The technique has been applied in diverse settings: [13] considers analysis of the nonlinear chaotic dynamics of Chua's circuit model, [12] concerns molecular models, [14] considers model validation for combustion dynamics, and [15] treats transport phenomena in building systems. In each of these papers it is shown through numerical examples

978-1-4244-2079-7/08/$25.00 ©2008 AACC.

that the associated eigenvectors carry significant information regarding dynamics. In particular, its sign structure can be used to obtain a decomposition for defining super-states. Theory to support this aggregation technique is contained in [16], based on a change of measure similar to what is used to establish large deviations asymptotically. In this way, these results may be regarded as an extension of the classical Wentzell–Freidlin theory [17].

In this paper we are motivated by the problem of constructing multi-scale models for the purposes of estimation in transient systems, such as the application to egress. We find that a valuable tool for analysis and the construction of a reduced order model is found in the *potential matrix* of the Markov model. The potential matrix is the solution to a Lyapunov equation for an associated linear system.

Based on the potential matrix, we consider two distinct methods for aggregating states.

The first is based on spectral graph partitioning, in which the Markov model defines a graph, and the weights on the edges of the graph are defined by the entries of the potential matrix. In application to egress, large weights correspond to a pair of cells that have a high level of interaction on average.

The second method is based on aggregating states with similar long term uncertainty, where uncertainty is captured using the entropy $\mathscr{H}(x_0) \doteq H(x_1^n|x_0)$ of the Markov sequence $x_1^n = \{x_1, \ldots, x_n\}$ with initial condition x_0. It is shown that entropy can be conveniently expressed in terms of the potential matrix. State aggregation is achieved by grouping together states with similar entropy.

Either of the proposed methods yields a coarse partition consisting of aggregated states, called 'super-states'. Based on this partition, a reduced order Markov model is defined whose state space consists of super-states. The resulting model can then be used for estimation, or other purposes.

The remainder of the paper is organized as follows. In Section II we describe the two aggregation methods, and in Section III the theory is illustrated using the example illustrated in Fig. 1. Section IV presents results on reduced order estimation for this example . Conclusions and directions for future research are summarized in Section V.

II. Aggregation Methods for Markov Model

A. Markov Model

We consider a finite state-space Markov chain with a single absorbing state (see [18] for terminology). The following notation is adopted throughout the paper: The state value at time t is denoted as $X(t)$, the initial condition $X(0)$ is denoted as x_0, and the sequence $\{X(1), X(2), \ldots, X(n)\}$ is denoted as X_1^n. The transition matrix is defined for each i, j via,

$$P_{ij} = \text{Prob}(X(t+1) = d_j | X(t) = d_i), \tag{1}$$

A transition matrix defines a directed graph,

$$G = \{V, E\}, \tag{2}$$

whose vertices $V = \{d_j\}$ are states, and edge set E contains those pairs (i, j) for which $P_{ij} > 0$.

Although the theory is far more general, for the purposes of exposition we consider the following special case in which X is a non-homogeneous, transient random walk on the nodes of a graph. We denote by $\mathsf{D} = \{d_i\}_{i=1}^M$ the nodes of the graph, and $d_\infty \in \mathsf{D}^c$ the state at which the chain is eventually absorbed (i.e., $P(d_\infty, d_\infty) = 1$). Hence the state space of the Markov chain is the union $\mathsf{D} \cup \{d_\infty\}$. In application to egress, this corresponds to a model consisting of a *single agent*; The state at time t is the location of the agent in the building if $X(t) \in \mathsf{D}$, and the agent has left the building if $X(t) = d_\infty$.

Although, we account for both probability of movement and noise in the Markov model, we do not account for congestion effects in model reduction and estimation considerations relevant to this paper; cf. [19] for modeling and estimation for cellular automata models with congestion.

Sensors are assumed memoryless: Letting S denote a sequence of observations we have,

$$\text{Prob}\{S(t) = s \mid X_0^t\} = \text{Prob}\{S(t) = s \mid X(t)\}, \tag{3}$$

where $s \in \{s_0, s_1, \ldots, s_N\}$ are the possible observation values corresponding to N sensors, and s_0 corresponds to no observation. A Bayesian estimator can be constructed based on the observation matrix defined by

$$O_{ij} = \text{Prob}\{S(t) = s_i \mid X(t) = d_j\}. \tag{4}$$

In this paper we focus on a model reduction framework without explicitly considering the sensors. The two initial conditions x_0 and x_0' are considered similar if the resulting probabilistic behavior of X from the respective initial conditions is similar. Algorithms for constructing reduced order models for both P and O as well as reduced order estimator are described in Section II-E.

B. Potential Matrix

The *potential matrix* is defined as the sum,

$$R = \sum_{t=0}^{\infty} P^t \tag{5}$$

where P^t is the t-fold matrix product of P. For each initial condition we define the row vector with values $\mu_{x_0}(j) = R_{ij}$ for $i = x_0$, and $j = 1, \ldots, M$. We have $P_{ij}^t = \text{Prob}(X(t) = d_j | X(0) = d_i)$, and hence the potential matrix has the following interpretation,

$$R_{ij} = \mathsf{E}[\text{Number of times } X(t) = d_j \text{ prior to exit} \mid X(0) = d_i].$$

In building examples we find that initial conditions in separate corridors will give rise to approximately singular measures $\mu_{x_0}, \mu_{x_0'}$, while initial conditions in a common room result in nearly equal values.

The potential matrix is the solution to a linear equation analogous to the Lyapunov equation for LTI systems,

$$R - PR = I. \tag{6}$$

It follows that the vector μ_{x_0} is the solution to the linear equation,

$$\mu_{x_0} - \mu_{x_0} P = \delta_{x_0}, \tag{7}$$

where δ_{x_0} is the Dirac delta measure supported at the initial condition x_0.

In the following two sections we desribe two methods that employ the potential matrix to construct super-states.

C. Spectral Method

The first approach is to use the potential matrix to identify and group cells with a high degree of interaction. The potential matrix provides information on degree of average interaction amongst cells: the entry R_{ij} is the expected number of visits to the j^{th} cell starting from the i^{th} cell. Hence, if R_{ij} or R_{ji} are relatively large, then the two states exhibit significant interaction. Aggregation based on the level of interaction can be posed as a spectral graph partitioning problem.

We assign weights to the edges of the graph (2) defined as

$$W_{ij} = \frac{R_{ij} + R_{ji}}{2}. \tag{8}$$

The matrix W is symmetric, and each entry W_{ij} is a measure of the average time passed between nodes d_i and d_j.

For aggregation purposes we use the $NCut_k$ algorithm [20], [21]. For a given collection of k super-states comprising of k disjoint and nonempty subsets $\{D_1, ..., D_k\} \doteq \Delta$ taken from the vertex set V, the $NCut_k$ metric is defined as

$$Ncut_k(\Delta) = \frac{cut(D_1, V - D_1)}{assoc(D_1, V)} + ... + \frac{cut(D_k, V - D_k)}{assoc(D_k, V)}, \tag{9}$$

where $cut(E,F) = \sum_{i\in E, j\in F} W_{ij}$ and $assoc(E,V) = \sum_{i\in E, j\in V} W_{ij}$. The $NCut_k$ algorithm seeks to choose Δ that minimizes the $NCut_k$ metric. In effect, the $NCut_k$ algorithm not only balances the size of the cuts, but also the sizes of the clusters.

The ideas are best explained for the special case of $k = 2$, where the solution is also most easily apparent in terms of spectral bisection. We begin by defining a diagonal matrix D as

$$D_{ii} = \sum_{j\in V} W_{ij}. \tag{10}$$

Using D and W, the Laplacian matrix L is defined as

$$L = D - W. \tag{11}$$

With these matrices, the $NCut_2$ reduces to a solution of the following bisection problem:

$$\begin{aligned} &\min_{y} \ \frac{y^T L y}{y^T D y} \\ &\text{s.t.} \ \ y(i) \in \{1, -b\}, \ y^T D \mathbf{1} = \mathbf{0}, \end{aligned} \tag{12}$$

where $y(i)$ can take values from a discrete set; cf., [20]. These values also indicate the optimal aggregation that serves to minimize the $NCut_2$ metric. This problem is often relaxed by allowing y to take on any real value. In this case, the solution simplifies to solving a generalized eigenvalue problem

$$Lf = \lambda_2 D f, \tag{13}$$

where λ_2 is the second lowest eigenvalue of L and f is the corresponding eigenvector. The super-states are defined according to

$$\begin{aligned} D_1 &= \{d_i : f_i > 0\} \\ D_2 &= \{d_i : f_i \le 0\}. \end{aligned} \tag{14}$$

This gives the standard spectral bisection method. The method can also be applied recursively whereby a superstate at any stage is split by application of spectral bisection method.

Apart from recursive spectral bisection, there are several alternate approximation methods for the general $NCut_k$ problem. In this paper we utilize the Meila-Shi algorithm to make the multiple cuts [21]. The method utilizes the first k eigenvectors of the generalized eigenvalue problem together with an application of the k-means algorithm. We refer the reader to [21] for details on the algorithm.

D. Entropy Method

In the second approach we propose to use *entropy* as an analogue of *energy* to generalize the Hankel norm approach to model reduction. In particular, much like energy is used for model reduction in linear systems (initial states with low energy are discarded), entropy can be used for model reduction via aggregation, where initial states with a similar level of long-term uncertainty are aggregated.

For a given initial condition we denote the entropy of the distribution of the Markov sequence X_1^t by

$$H(X_1^t \mid x_0) = -\mathsf{E}[\log p(X_1^t \mid x_0)], \tag{15}$$

where $p(X_1^t | x_0)$ denotes the joint probability distribution [22]. The limit as $t \to \infty$ is the infinite-horizon entropy considered here,

$$\mathscr{H}(x_0) = \lim_{t\to\infty} H(X_1^t \mid x_0). \tag{16}$$

The limit is finite since the Markov chain is assumed to be absorbed at d_∞ with probability one from each initial condition.

The total entropy (16) can be expressed in terms of the potential matrix. We first observe that Equation (16) can be expressed as

$$\begin{aligned} \mathscr{H}(x_0) &= \lim_{t\to\infty} \sum_{i=1}^{t} H(X(i) \mid X_1^{i-1}, x_0) \\ &= \lim_{t\to\infty} \sum_{i=1}^{t} H(X(i) \mid X(i-1)), \end{aligned} \tag{17}$$

where the first step is due to the chain rule for entropy and the second step is a result of the Markov assumption [22]. Writing the recursion,

$$\begin{aligned} \pi_{t+1} &= \pi_t P, \\ \pi_0 &= \delta_{x_0}, \end{aligned} \tag{18}$$

we obtain a compact representation for entropy:

$$\begin{aligned} \mathscr{H}(x_0) &= -\sum_t \sum_{i,j} \pi_t(i) P_{ij} \log(P_{ij}) \\ &= -\sum_{i,j} \mu_{x_0}(i) P_{ij} \log(P_{ij}). \end{aligned} \tag{19}$$

State aggregation is performed by sorting the initial conditions of X based on the corresponding values of entropy $\mathscr{H}(x_0)$. Depending on the desired number of super-states, the ordered states are partitioned into states of roughly equal size.

E. Reduced Order Markov System

Aggregation leads to a coarse graph with nodes $\{D_i\}$ that correspond to aggregated states. The next step is to define a transition matrix on this graph. We present two approaches.

1) Uniform Smearing: Here a probability is uniformly smeared over a superstate. The reduced order Markov matrix is denoted as $P^{\text{red,unif}}$, and obtained using

$$P_{ij}^{\text{red,unif}} = \frac{\sum_{u\in D_i}\sum_{v\in D_j} P_{uv}}{K_i}, \tag{20}$$

where $u \in D_i$ denotes the cells (nodes) of the original graph that have been aggregated to form the super-state D_i and K_i is a normalization constant.

To define a reduced order observation matrix, we follow the heuristic of uniformly smearing the probability. As a result, there is now a probability of detection

$$o_{D_j} = \frac{\#\text{ of sensors in } D_j}{\#\text{ of cells in } D_j} \tag{21}$$

for an agent in (one of the cells contained in) D_j. This probability is used to construct a reduced order observation matrix $O^{\text{red,unif}}$ with respect to the sensors present. We note that the number of observation states for a single agent can at most be the number of aggregated states, plus the additional unobserved state.

2) Weighted Smearing: For the weighted smearing case a weighting vector, w, must first be defined with respect to the nodes of the original Markov chain. For an ergodic Markov matrix, w would be the stationary distribution. With absorbing Markov chains, such a weighting is clearly not applicable. We define the weighting vector according to the proportion of time that is spent in a specific cell contained in a given super-state:

$$w_k = \frac{\sum_{i\in V} R_{ik}}{\sum_{k\in D}\sum_{i\in V} R_{ik}}, \tag{22}$$

where V is the vertex set, R is the Potential matrix, k is a particular cell in the grid, and D is the super-state in which state k lies. With the weight vector, the reduced order Markov matrix is obtained using

$$P_{ij}^{\text{red,weight}} = \frac{\sum_{u\in D_i} w_u \sum_{v\in D_j} P_{uv}}{K_i}, \tag{23}$$

where K_i is once again a normalization constant.

Just as in the uniform case, a reduced order observation matrix must be defined. As before, there is now a probability of detection

$$o_{D_j} = \sum_{k\in S(D_j)} w_k, \tag{24}$$

where $S(D_j)$ is the set of sensors contained in D_j. This probability is used to construct a reduced order observation matrix $O^{\text{red,weight}}$ with respect to the sensors present.

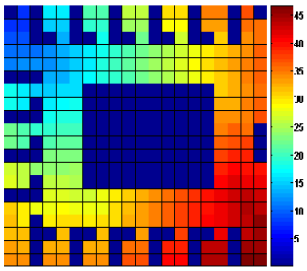

Fig. 2. The entropy defined by Equation (19) for each grid location. Notice the increased uncertainty with farther distances from the exit.

III. Example of Aggregation Method

In this section we consider the large scale office building example illustrated in Fig. 1. This is a model of a single floor in a building. The exits are denoted by black, and walls by gray. Unless otherwise stated, it is assumed that only the top-left exit is in use.

The motion of the agent is described by a Markov chain on a graph consisting of the 400 cells shown, plus the state corresponding to an empty building. The probabilistic model is constructed so that an agent has a higher probability of moving toward an exit.

Based on the geometry seen in Fig. 1 one would guess that a single (spectral) split of the states would occur by dividing along the diagonal from the top-left to bottom-right corner. Fig. 3 (top left) illustrates the results of the Meila-Shi algorithm for a single cut. The single cut result is consistent with the aggregation obtained via spectral bisection.

For a model as large as this example, multiple cuts are necessary. Recursive spectral bisection, while simple to implement, does not always give good cuts because this method does not take all of the information of the system into account. It is in these cases that the Meila-Shi algorithm is much more effective. Fig. 3 shows the results of aggregating the original 400 states into 2, 4, and 9 super-states. For the most part, the super-states are grouped by states that are all adjacent to one another. Intuitively, the aggregates make sense because corridors are grouped with the connected offices. Aggregating states in this way will greatly alleviate the computational burden of estimation.

We now turn to the entropy based aggregation. Figure 2 depicts the entropy plot ($\mathscr{H}(x_0)$ for initial conditions on the grid). Entropy increases as the grid locations are farther away from the exit. This is expected because entropy is a measure of uncertainty; with increased distance from the exit, more uncertainty results. Locations with zero entropy correspond to the walls.

Aggregation into two super-states based on entropy is carried out as described in Sec. II-D. For a particular initial condition x_0, the entropy of the cells in the support of μ_{x_0} are displayed in ascending order in Fig. 5. By assigning a single threshold at the mean entropy value, we aggregate the cells into two super-states. The graphical interpretation of this aggregation technique is displayed in Fig. 4. To construct additional super-states, we use multiple threshold values to

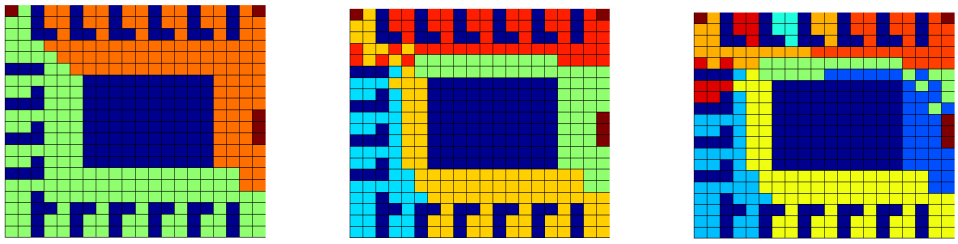

Fig. 3. Results of the Meila-Shi algorithm for a 1, 3, and 8 cut, respectively. Notice how cells in super-states are not always adjacent.

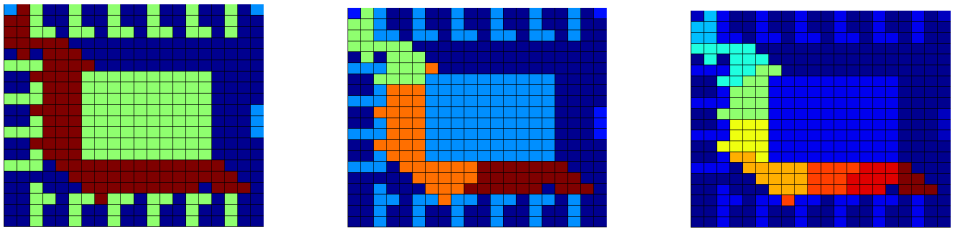

Fig. 4. Depictions of 2, 4, and 9 aggregated states grouped using entropy on the support of μ_{x_0}. The bold lines indicate the different super-states.

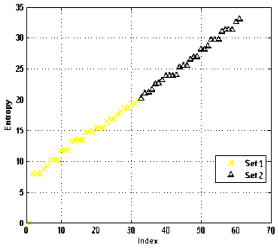

Fig. 5. The entropy that is supported by μ_{x_0} sorted in ascending order.

split the entropy of the cells on the support of μ_{x_0}. Fig. 4 uses 3 (8) threshold values for 4 (9) aggregated states. One interesting observation regarding the partition with 9 super-states is that cells near the exits of the individual offices are grouped together. It is expected that the super-states would approximate the system relatively well because of this aggregation of cells.

IV. Full and Reduced Order Estimation

We now apply these model reduction techniques to state estimation.

A. Problem Statement and Estimation

A Markov model for the building example in Section III describes the evolution of a single agent moving toward the exit. We now include sensors to estimate the location of the agent, which is modeled by a memoryless observation process satisfying (4). We consider Bayesian estimation using the full and reduced order Markov models, and compare their respective performance.

The *non-normalized* recursive estimation equation is given by

$$\hat{v}_{t+1} = \hat{v}_t P \mathrm{diag}(O_{Y_{t+1}}), \tag{25}$$

where $\hat{v}_t$ is the current estimate, P is the Markov matrix, and $\mathrm{diag}(O_{Y_{t+1}})$ is a diagonal matrix with diagonal constructed from the row of the observation matrix corresponding to the observation at $time = t+1$. The optimal Bayesian estimate is obtained by normalizing $\hat{v}_{t+1}$ to be a probability vector. It represents the conditional probability that the agent occupies a node given all past and present observations.

For the purposes of reduced order estimation, super-states are created by aggregating states, as described in Section III. Aggregation leads to a coarse graph with nodes $\{D_i\}$ for which reduced order Markov and observation matrices are defined. Once P^{red} and O^{red} are defined (as in Section II-E), the reduced-order estimation is carried out following Equation (25). The results are described next.

B. Estimation Results

We applied the reduced order model to estimate the correct corridor for the agent as she exits the building starting from the location depicted in Fig. 1. The full order estimation used the original 400 state model while reduced order estimation was carried out with models consisting of a varying number $(4, 9, 16)$ of aggregated states.

The estimation results were post-processed to obtain the probability of agent occupying a corridor. For the spectral method, these results are shown in Fig. 6. Overall, the probability from a reduced order estimate approximately follows the trend of the full order estimate (which is much more accurate). The probability estimated for corridors 1 and 2 is zero and hence is not shown. This is consistent with the dynamics of the Markov chain. At times where there is a detection by a sensor (times 8 and 13), the estimate jumps (and improves) but the quality of the estimate declines with time if there are no additional observations. The estimates typically improves with finer aggregation (full order estimator's predictions are better than predictions with

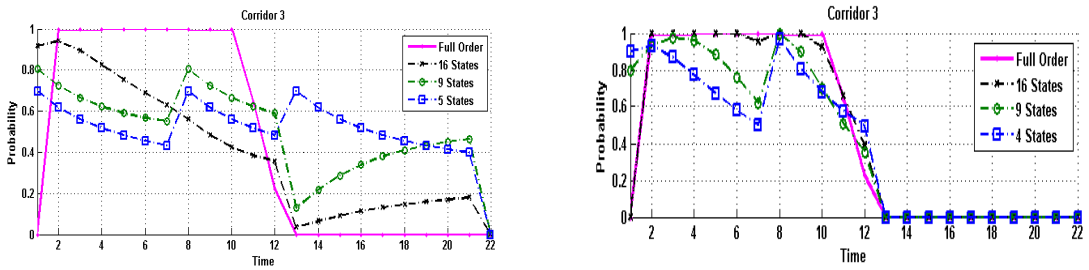

Fig. 6. Estimation results with spectral (left) and entropy (right) methods: the plots show the estimate probabilities of being located in Corridors 3. Full order estimation as well as reduced order estimation using a varying number of super-states are displayed.

16 states, which in turn are better than 9). Exceptions arise when cells within a superstate overlap between the corridors.

Next, we discuss the estimation results based on model reduction using the entropy approach. Simulation results are shown on the right in Fig. 6. Although many of the previous comments apply to these results too, overall the entropy method gives much better estimates than the spectral method. As before, the estimate improves with increasing number of states. A crucial observation is that estimates based on 16 super-states are virtually indistinguishable from the full order estimates. This suggests a remarkable reduction in complexity. Also as before, the estimates improve at the time of detection and decline with no detection.

V. Conclusion

In this paper we demonstrated two methods for model reduction of a large Markov model. Both methods utilize the potential matrix for state aggregation. The first method is essentially spectral graph partitioning (see [12], [13], [15]) with connectivity between aggregated states defined in terms of the potential matrix. The second method is based upon aggregating states with similar long-term uncertainty, where entropy is proposed as a metric.

An example was used to demonstrate the proposed model reduction methods. The models were used for the purposes of reduced order estimation of an agent's location as he (or she) exits a large building. The reduced order estimators gave acceptable estimates in most cases. The reduced order models obtained using the entropy based method gave better results than obtained using the spectral method. With the former, a 16 state model provided estimates comparable with the full order (400 state) model.

These simulation results point to the potential of aggregation for reduced order estimation in large Markov models. There are many open issues. In particular, there is the issue of complexity faced in *construction* of the reduced order model. One also needs error bounds for the reduced-order estimator. This is the subject of continuing research.

References

[1] S. Wolfram. *Theory and applications of cellular automata*. World Scientific, Singapore, 1986.

[2] A. Schadschneider. Cellular automation approach to pedestrian dynamics-theory. In Michael Schreckenberg and Som Deo Sharma, editors, *Pedestrian and Evacuation Dynamics*, pages 75–86. Springer, 2002.

[3] S. Gwynne, E.R. Galea, M. Owen, P.J. Lawrence, and L. Filippidis. A systematic comparison of building EXODUS predictions with experimental data from the Stapelfeldt trials and the Milburn House evacuation. *Applied Mathematical Modelling*, 29:818–851, 2005.

[4] L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, February 1989.

[5] Y. Ephraim. Hidden Markov processes. *IEEE Trans on Information Theory*, 48(6), 2002.

[6] R. J. Elliott, L. Aggoun, and J. B. Moore. *Hidden Markov Models: Estimation and Control*. Springer-Verlag New York, Inc., New York, NY, first edition, 1995.

[7] H. A. Simon and A. Ando. Aggregation of variables in dynamic systems. *Econometrica*, 29:111–138, 1961.

[8] J. N. Darroch and E. Seneta. On Quasi-Stationary Distributions in Absorbing Discrete-Time Finite Markov Chains. *J. Appl. Probab.*, 2(1):88–100, 1965.

[9] R. G. Phillips and P. V. Kokotovic. A singular pertubation approach to modeling and control of Markov chains. *IEEE Transactions on Automatic Control*, 26(5), 1981.

[10] L. B. White, R. Mahony, and G. D. Brushe. Lumpable hidden Markov models-model reduction and reduced complexity filtering. *IEEE Transactions on Automatic Control*, 45(12), 2000.

[11] M. Huang and S. Dey. Distributed state estimation for hidden Markov models by sensor networks with dynamic quantization. *Proceedings of the 2004 Intelligent Sensors, Sensor Network and Information Processing Conference, ISSNIP '04*, pages 355–360, 2004.

[12] Ch. Schütte, A. Fischer, W. Huisinga, and P. Deuflhard. A direct approach to conformational dynamics based on hybrid Monte Carlo. *J. Comput. Phys., Special Issue on Computational Biophysics*, 151:146–168, 1999.

[13] O. Junge and M. Dellnitz. Almost invariant sets in Chua's circuit. *Int. J. Bif. and Chaos*, 7:2475–85, 1997.

[14] I. Mezic and A. Banaszuk. Comparison of systems with complex behavior. *Physica D*, 197:101–133, 2004.

[15] P. G. Mehta, M. Dorobantu, and A. Banaszuk. Graph-based multi-scale analysis of building system transport models. In *Procs. of American Controls Conference*, pages 1110–1115, Minneapolis, 2006.

[16] W. Huisinga, S. Meyn, and C. Schütte. Phase transitions and metastability in Markovian and molecular systems. *Ann. Appl. Probab.*, 14(1):419–458, 2004.

[17] A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein. Metastability in stochastic dynamics of disordered mean-field models. *Probab. Theory Related Fields*, 119(1):99–161, 2001.

[18] S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. Springer-Verlag, London, second edition, 1993. online: `http://black.csl.uiuc.edu/~meyn/pages/book.html`.

[19] J. Niedbalski and P. G. Mehta. Simulation and estimation of evacuation dynamics in a building. In *Procs. of American Control Conference*, 2007.

[20] J. Shi and J. Malik. Normalized cuts and image segementation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

[21] M. Meila and J. Shi. A random walks view of spectral segmentation. *AISTATS*, 2001.

[22] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., first edition, 1991.