HMM as Building Block : in Models of Brain, Language, and Mind

Alexander M. Duda
University of Illinois at Urbana-Champaign
Department of Electrical and Computer Engineering
Beckman Institute for Advanced Science and Technology
amduda@illinois.edu

I. Introduction

As shown in the classic Cave-Neuwrith experiments, hidden Markov models (HMMs) have a remarkable ability to discover structure in input data in a self-organizing manner [1]. Additionally, they have displayed strong results in sequence-learning in speech processing [2] and computational biology [3]–[5]. Paired with their general ability to be run backwards and display a kind of "thinking", HMMs are a natural candidate for a building block in abstract models of brain, language, and mind. This paper will focus on a couple of successful models that incorporate HMMs as a building block. Section II will provide a brief background on the HMM. Section III will focus on a HMM cascade model for concept learning, whereby groundbreaking experiments in multi-sensory cognitive robotics are described, which enable development of higher-order multi-modal concepts, as well as generative models for actions in the real world [6]–[8]. Section IV will focus on projects involving hierarchical HMMs that discover structure within a particular data stream at various levels of spatiotemporal granularity. Section V will then discuss the short-term role that HMMs should play in making progress on the brain/mind problem (Section V) and conclude by speculating on the possible role that HMMs could play in the future (Section VI).

II. Background

An HMM is a discrete-time stochastic process with two continuous components $\{X_n, Y_n\}$, defined on probability space $(\Omega, \mathcal{F}, P)$ [6], [9], [10]. We define $\{X_n\}_{n=1}^{\infty}$ to be a discrete-time first-order Markov chain with state space $R = \{1, \ldots, r\}$, where r is a fixed known constant. The model starts in a specific state, $i \in R$, with probability $\pi_i = P(X_1 = i)$. With Π representing the set of r-length stochastic vectors, we define $\boldsymbol{\pi} \in \Pi$ as $\boldsymbol{\pi} = \{\pi_i\}$. Thus, for $i, j \in R$, the Markov chain transition probabilities are given by $a_{ij} = P(X_n = j | X_{n-1} = i)$. With $\boldsymbol{A} = \{a_{ij}\}$, $\boldsymbol{A} \in \mathcal{A}$, where $\mathcal{A}$ represents the set of all $r \times r$ stochastic matrices.

In an HMM, $\{X_n\}$ is not visible; its statistics can only be inferred from the observable random process $\{Y_n\}$, which is a probabilistic function of the former. In other words, given X_n, Y_n will assume values from some space E according to a conditional probability distribution. The resulting conditional density of Y_n is generally assumed to belong to a parametric family of densities $\{b(\cdot;\theta) : \theta \in \Theta\}$, where the density parameter θ is a function of X_n, and Θ is the set of valid parameters for the specified conditional density assumed by the model. The conditional density of Y_n given $X_n = j$ can be written as $b(\cdot;\theta_j)$ or, to be concise, $b_j(\cdot)$, when the θ_j dependence is clear.

With $\Phi = \Pi \times \mathcal{A} \times \Theta$ representing the HMM parameter space, the model $\boldsymbol{\varphi} \in \Phi$ is expressed as $\boldsymbol{\varphi} = \{\pi_1, \ldots, \pi_r, a_{11}, a_{12}, \ldots, a_{rr}, \theta_1, \ldots, \theta_r\}$. The parameters of the model can be accessed through coordinate projections (e.g., $a_{ij}(\boldsymbol{\varphi}) = a_{ij}$). When we are not interested in estimating $\boldsymbol{\pi}$, we let $\Phi = \mathcal{A} \times \Theta$. Occasionally the literature uses other model parameterizations [11], [12].

III. Cascade of HMMs

When referring to a "cascade of HMMs" we do so in the sense described on p.238 of Levinson [10] and illustrated by Fig. 1. With a focus on multi-sensory integration and associative memory, HMMs at the bottom of the cascade are each assigned input from different sensory streams. Thus, there is an HMM for auditory inputs and an HMM for visual inputs. The states of these HMMs serve as the input to HMMs higher up the cascade, like the audio-visual HMM and the audio-tactile HMM. Continuing in this fashion, one can construct levels of HMMs that can discover structure at increasing levels of abstraction and multi-modal integration. In this section we will go into some of the details involved in designing and implementing such a cascade model. In particular, Section III-A will examine some modifications that need to be made to the classic HMM to prepare it for online learning; Section III-B will detail the specifics of the cascade model; Section III-C will explore some of the remarkable results achieved by the model; Section III-D will allude to some more recent extension results.

A. Modifications for Online Learning

Though the classic HMM is a powerful model, one of the drawbacks is that two of the most common methods for parameter estimation (Baum-Welch and methods based on the Viterbi algorithm) both require off-line processing [2], [6]. For real-time learning, an iterative or online method is more appropriate. One can remedy the situation by (1) minimizing the prediction error of the model through recursive methods

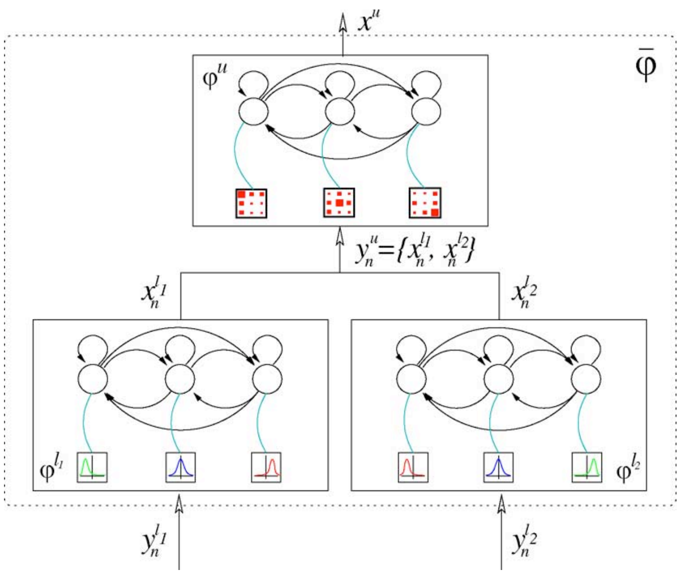

Fig. 1. An example of a "Cascade of HMMs", whereby Models φ^{l_1}, φ^{l_2}, and φ^u are all HMMs [6].)

[11], [13], [14] or (2) recursively maximize the likelihood of the estimated model for an observation sequence [12], [14]–[18]. In particular, the recursive maximum likelihood estimation (RMLE) algorithm of Krishnamurthy and Yin has been successful [18] (as evident by the results described in Section III-C).

B. Design

The "HMM Cascade Model for Concept Learning" has the topology shown in Fig. 1 [6]. Thus, $\bar{\varphi} = \{\boldsymbol{\varphi}^{l_1}, \boldsymbol{\varphi}^{l_2}, \boldsymbol{\varphi}^{u}\}$, where the component models are HMMs defined as in Section II. For $v \in \{l_1, l_2, u\}$, we let $\{X_n^v, Y_n^v\}$ represent the state and observation sequences, respectively, to $\boldsymbol{\varphi}^{l_1}, \boldsymbol{\varphi}^{l_2}, \boldsymbol{\varphi}^{u}$. In the model, observations $Y_k^{l_j}$ of lower models $\boldsymbol{\varphi}^{l_j}$ are generally assumed to be continuous. The observations Y_k^u of upper model $\boldsymbol{\varphi}^u$ are the concatenated state sequences of the lower level models. Thus, $Y_k^u = \left(X_k^{l_1}, X_k^{l_2}\right)$, and $\boldsymbol{\varphi}^u$ models the joint distribution of $X_k^{l_1}$ and $X_k^{l_2}$ for each state $j = 1, \ldots, r^u$, where r^u represents the number of states in $\boldsymbol{\varphi}^u$. Though not strictly required, $X_k^{l_1}$ and $X_k^{l_2}$ are assumed independent to simplify computations. The current state of the upper model, $\boldsymbol{\varphi}^u$, can be viewed as a probabilistic function of the lower models, which implies a compositional model.

There are other *important* aspects to the design of the model that enabled the results to be discussed; e.g., the particular number of hidden states used in each of the HMMs, the initialization of the model parameters, the feature extraction process used for live speech and visual inputs, the pre-programming of various "instincts", the use of a "switching" HMM, the finite-state machine controller (used as a part of the autonomous exploration mode of the robot), the actual details of the robot used as a means of embodiment, as well as other issues that are beyond the scope of this paper. The interested reader is encouraged to study the paper by Squire and Levinson [6]).

C. Results

The primary set of experiments involved a robot and a person looking at the same object (a red ball, green ball, cat, dog); the person spoke the name of the object or some aspect of it, and the robot learned to associate the word/phrase with the visual features of the object. After 30 minutes of training the parameter values were learned to the point where the robot did not make any classification errors.

D. Related Work

More recent work that extends the cascade HMM model has shown remarkable abilities in action-word learning, as well as complex action sequence generation [8]. In the work of Niehaus and Levinson, an advanced humanoid iCub robot was taught basic word-action pairings for arm gestures like "up","down", "left", "right", "raise", "lower". Then, through a series of tutoring sessions, the robot was able to learn a sequence of said word-action pairings– "raise, left, right, left, right, lower" and identify it as a new *complex* gesture: "wave"; similarly, the robot was able to learn another sequence of word-action pairings– "up, right, left, right, left, down" and identify it as a new *complex* gesture: "shake". Furthermore, the robot was able to learn a sequence of these newly learned complex gestures– "wave, shake" and identify it as a new (even more) *complex* gesture:"greet". In total, the robot learned five different sequence gestures.

Moreover, because of the generative nature of the cascade of HMMs, when the robot was asked to repeat its learned gestures (via spoken request) it did not simply "re-play" a recorded motion; it had extracted a model of the gesture and could use that model to generate an instance of the gesture. Thus, upon request, the generated gesture could vary from the originally learned gesture but still be distinctly correct. This exploitation of the model properties to produce ever more complex representations of compositional and hierarchical behaviors is extremely promising and provides another example of the power of cascades of HMMs.

IV. Hierarchical HMMs

By "hierarchical HMM" we mean in the sense of Fine, Singer, and Tishby [19]. Their work was motivated by the ubiquity of complex, multi-scale, recursive, self-similar, and hierarchical structure found in many natural sequences; particularly in language, handwriting, and speech. They successfully extended the standard Baum-Welch algorithm and derived an efficient procedure for estimating model parameters from unlabeled data. They primarily desired to "enable better modeling of different stochastic levels and length scales", as well as be able to "infer correlated observations over long periods in the observation sequence through higher levels of the hierarchy". They demonstrated the success of their model by learning a multi-resolution structure of English text, whereby different levels in the hierarchy of the model were able to learn different resolutions of the text. In this section we will go into some of the details involved in designing and implementing such a hierarchical HMM model. In particular, Section IV-A will detail the specifics of the hierarchical model; Section IV-B will explore some of the results achieved by the model; Section IV-C will allude to some more recent extension results.

A. Design

HHMMs generalize standard HMMs by making each of the hidden states a HHMM. Thus, the states of an HHMM emit sequences, as opposed to a single symbol. This notion can be formalized. Let Σ be a finite alphabet and Σ^* be the set of all possible generated strings. An observation sequence will then be a finite string, $\bar{O} = o_1 o_2 \cdots o_T \in \Sigma^*$. A state of an HHMM is denoted by q_i^d where $d \in \{1, \ldots, D\}$, with state index i and hierarchy index d (where $d = 1$ for the root and $d = D$ for production states). The number of sub-states of internal state q_i^d is denoted by $|q_i^d|$, where it is recognized that internal states are not required to have the same number of sub-states. Whenever ambiguity does not result, for brevity, the state index is omitted so that q^d denotes a state at level d. An HHMM is characterized by its model topology and the state transition probability between the internal states and the output distribution vector of production states. Therefore, for each internal state q_i^d (where $d \in \{1, \ldots, D\}$), there exists a state transition probability matrix $A^{q^d} = (a_{ij}{}^{q^d})$, where $a_{ij}{}^{q^d} = P\left(q_j^{d+1} | q_i^{d+1}\right)$ is the probability of making a horizontal transition from state i to state j, both of which are sub-states of q^d. Likewise, the initial distribution vector over the sub-states of q^d, which is the probability that state q^d will initially activate state q_i^{d+1}, is denoted by $\Pi^{q^d} = \{\pi^{q^d}\left(q_i^{d+1}\right)\} = \{P\left(q_i^{d+1} | q^d\right)\}$. If q_i^{d+1} is an internal state then $\pi^d\left(q_i^{d+1}\right)$ can be viewed as the vertical transition probability, whereby parent state q^d transitions to sub-state q_i^{d+1}. Each production state q^D is parametrized by its output probability vector $B^{q^D} = \{b^{q^D}(k)\}$, where $b^{q^D}(k) = P\left(\sigma_k | q^D\right)$ is the probability that the production state q^D will output symbol $\sigma_k \in \Sigma$. Therefore, if we define $\{1, \ldots, D\} = [D]$, the HHMM can be denoted by $\lambda = \{\lambda^{q^D}\}_{d \in [D]} = \{\{A^{q^d}\}_{d \in [D-1]}, \{\Pi^{q^d}\}_{d \in [D-1]}, \{B^{q^d}\}\}$. Fig. 2 illustrates a simple four level HHMM.

The whole process can be described in a succinct fashion. Beginning at the root state, a string is generated and one of the root's sub-states is selected randomly according to Π^{q^1}. Likewise, each internal state q that is entered has one of q's sub-states selected randomly according to Π^q. The process continues with the selected sub-state which will recursively activate one of its sub-states. This process continues until a production state, q^D, is reached; at this time a single symbol will be generated according to B^{q^D} and control will return to activated state q^D. Once a recursive string is generated, the recursion-initiating state will select the next state (in the same level) according to the level's state transition matrix; the newly selected state will then start a new recursive string generation process. Each level (excluding the top) has a terminating state, q_{end}^d, which terminates the stochastic state activation process. When a terminating state is reached, the parent state of the entire hierarchy regains control. The top level (root state) is the only level which does not have a terminating state. Therefore, when control of all recursive activations return to the root state, the observation sequence generation is complete, at which time the root state can trigger the generation of a new stochastic string. The model is strongly connected. [1]

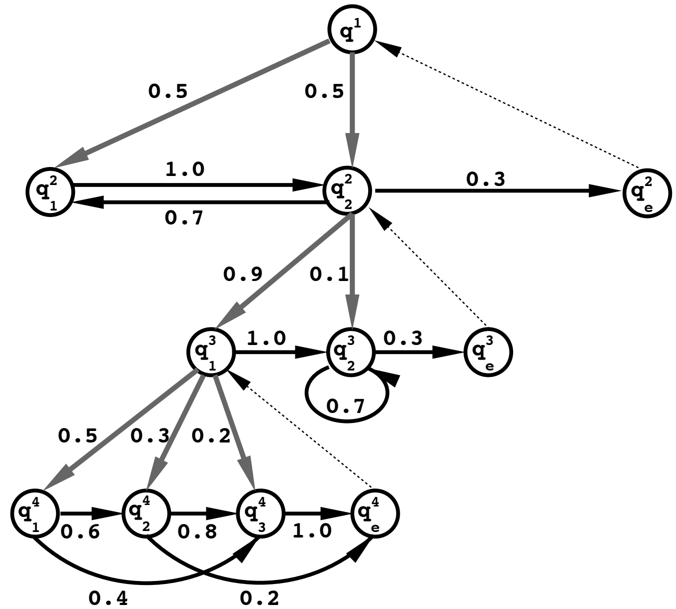

Fig. 2. An example of a "hierarchical HMM" with four levels [19]. Vertical transitions are denoted by gray arrows; horizontal transitions are denoted by black arrows; thin dashed arrows denote the forced returns from the terminating state of each level back to the parent state's level. Productions states are not included, to make the figure simpler.)

As with the classic HMM, there are three primary computations that should take place for the HHMM: (1) calculation of the likelihood of a sequence, (2) discovery of the most probable state sequence, and (3) estimation of the model parameters. The details of how these computations are carried out extend beyond the scope of this brief paper; the interested reader is advised to study p.45-49 of the paper by Fine, Singer, and Tishby [19].

B. Results

An unbalanced three-level HHMM was designed to extract a multi-level structure for English text. The HHMM had a variable number of sub-states at each internal state and production states at all levels (which made it "unbalanced"). Two primary results were observed:

1) The induced distribution had high variance across the different sub-states. The set of most probable strings to be produced by each state had very little in common.
2) The most probable strings produced by states low in the hierarchy corresponded to phonetic units like *ing*, *th*, *wh*, and *ou*. In higher levels of the hierarchy (second and third level) the most probable strings were words and

[1]Strongly connected means all states can be reached in a finite number of steps, beginning at the root state.

phrases like *is not*, *will*, and *where*. The most probable strings at the top of the hierarchy (at the root), displayed sentence-level strings whereby such strings would be likely to end with a punctuation mark.

These results are fascinating and intuitive.

C. Related Work

With a desire to further understand HHMMs, the author naively read Kurzweil's latest book, where he claims to have discovered "how to create a mind" and it involves using HHMMs [20]. Considering Google recently hired Kurzweil to be Director of Engineering and provide him with "basically unlimited" resources to create this "mind", it seemed that reading his book would be a good idea, to the extent that it might provide insight into the project. This section will provide a brief review of the technical content, without delving into issues of writing style or citation decisions. The review is by no means meant to be exhaustive.

1) Neural Module: Kurzweil believes that there is a particular cortical structure (a roughly 100-neuron population) that can be viewed as the important "neural module" in the brain and it has particular topological and synaptic adaptive properties[2]. These ideas are based on recent fascinating studies. In a 2011 paper by Markram et al., he asserts that he was "search[ing] for evidence of Hebbian assemblies at the most elementary level of the cortex." However, what he discovered were "elusive assemblies [whose] connectivity and synaptic weights are highly predictable and constrained." He concludes that "these findings imply that experience cannot easily mold the synaptic connections of these assemblies" and conjectures that "they serve as innate, Lego-like building blocks of knowledge for perception and that the acquisition of memories involves the combination of these building blocks into complex constructs" [21], [22]. Thus, according to Markram's studies, it seems that the brain is made up of neural modules whose synapses are basically static; their strengths and connections are genetically determined, while synapses that span modules can be adaptive. Kurzweil goes on to discuss a recent study of the brain's connectivity that seemed to show that it holds to a very regular grid pattern [23]. However, unacknowledged by Kurzweil, some have criticized the study and indicated that the results are "an artifact attributable to the limitations of their method" [24]. Combining these results among others, Kurzweil, then proceeds to discuss his "pattern recognizers".

2) Pattern Recognizer: A "pattern recognizer" to Kurzweil is either a HMM or a HHMM. His view of it seems fairly conventional and he does not go into much technical detail. From his perspective such an abstract model is functionally equivalent to a neural module, as described in the previous section. Using a large collection of such models, one could design a brain/mind.

3) Opinion: Having read the book, a lot of what Kurzweil is saying is reasonable[3]. However, it is not necessarily new, though his book might give a different impression. Moreover, he is just providing another "functionally equivalent" model that claims to carry out the same role as a small group of 100 neurons; this ends up being the author's main criticism of the work. We do not know enough about neurons, how they interact with one another, how they store information, etc. to create an abstract model that can claim to be functionally equivalent. More work needs to be done in simulating high resolution models of neurons, synapses, and cortical regions within an embodied framework that includes access to real-time, complex, noisy signals. Many of Kurzweil's ideas are in line with the work done in Levinson's lab. But, unlike Levinson's work, which has started to focus more on understanding how computation and information-processing occurs in actual neurons (with a desire to have a healthy interplay between abstract models and biological models), Kurzweil has *a priori* abstracted out the details of a 100-neuron module and decided that a HMM can take its place. There is no reason to think this might be the case. Section VI discusses this topic a bit more and speculates on the role of HMMs in future models of brain, language, and mind.

[2]"Neural Module" is not Kurzweil's phrasing, but the author's.

[3]Even if, unfortunately, the author did not learn anything new about HMMs or HHMMs from the book.

V. Short-Term Role of HMMs in Brain/Mind Problem

Though it seems that the HMM (either by itself or as a building block within a more complex model) is not necessarily the *complete* answer to building a model of brain, language, and mind, it is a fantastic tool that, in the author's opinion, could serve a new role in making progress toward the goal. A new model of associative memory based on the dynamics of simulated spiking neural networks is in the process of being designed and implemented [25], [26]. The dynamics operate in a very high dimensional space; even when it is projected down to a lower dimensional space the meta-stable attractor dynamics will likely be complex, chaotic, and fractal.

In order to say meaningful, quantitative statements about said dynamics, principled methods must be applied (and perhaps developed). In particular, as it turns out, the problem of quantifying the extent to which the model has meta-stable states, is information-preserving, and exhibits associativity is isomorphic to characterizing and comparing phase portraits, or more specifically, establishing distance metrics that can be used to consistently quantify the similarity of a pair of phase portraits. Due to noise in the input signals fed into the model, as well as randomness in the neuron-neuron communication, treating the phase portraits as stochastic is reasonable, which means the HMM (and possibly the HHMM could be relevant).

To begin, it will be illustrative to consider an example system [27]:

$$\begin{aligned} \dot{x} &= -bx + y - y^3 \\ \dot{y} &= -by + z - z^3 \\ \dot{x} &= -bz + x - x^3 \end{aligned}$$

The system shows a range of attractors for different values of the parameter b. For example, when b is set to $b = 0.30$, six limit cycles result, as shown in Figure 3. Thus, if a small perturbation was provided (that did not change the system parameters), the system could fall into any of the six different orbits. However, with a larger external perturbation, that results in a small adjustment of the parameter, to $b = 0.28$, the system now exhibits two different chaotic attractors, as shown in Figure 4. Furthermore, when the parameter is adjusted to $b = 0.235$, a single chaotic attractor results, as shown in Figure 5.

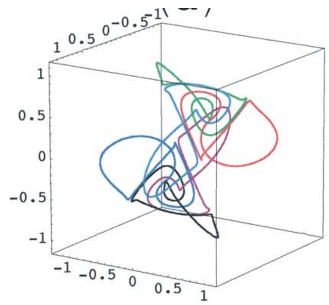

Fig. 3. b=0.30

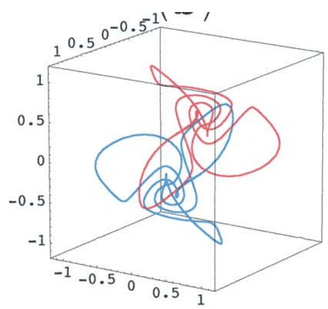

Fig. 4. b=0.28

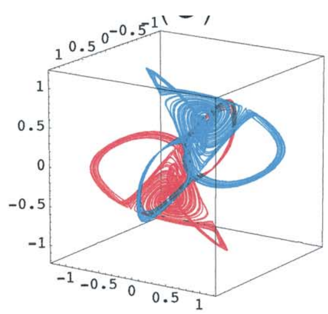

Fig. 5. b=0.235

Each resultant phase portrait is quite complex. However, one promising way to deal with the complexity is the following. Create a frame-of-reference centered at the current time, discretize the space in the surrounding area (to within a certain quantization), label these regions in space: $\{1, \ldots, n\}$, for n possible discrete regions, and allow these labels to serve as the symbol set for an HMM. An example illustration with $n = 8$ is shown in Fig 6. Thus, the HMM could learn, for a given phase portrait, a model to describe the structure/grammar of the dynamics. With such a representation, the distance between phase portraits could simply be the Kullback-Leibler divergence (or possibly Fisher information metric) [28]. Having an information-based metric would be extremely useful. More research needs to be conducted on the topic but it is clear that using an HMM or HHMM to build a model of the phase portrait's structure could be very promising and possibly bring order and clarity to a complex problem.

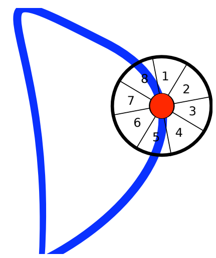

Fig. 6. Example illustration of discretizing space about frame-of-reference centered at current point (as represented by red dot). Here, eight different regions have been identified, relative to the current location. Such a set of regions could be considered a symbol set that an HMM could use to learn a model for trajectories.

We are not the first to suggest that HMMs could be useful in modeling (or capturing structure) in a nonlinear (possibly chaotic) system. Myers, Singer, et al., focused on modeling and prediction of a chaotic system with one-dimensional observations of a quantized Henon map, that appears to have worked well showing that an increase in the number of states increases the log-likelihood [29]. As they assert, "by their very nature, dynamical systems are Markov processes and the presence of an attractor in a chaotic system imposes a natural probabilistic measure on the state space– the invariant density", the estimation of which was developed by Marteau and Abarbanel [30]. Myers, et al. were inspired by ideas from Fraser, who showed that a HMM can capture some of the aspects of a chaotic system [31]. In a more general setting, Mees, et al. explore the relationship between modeling nonlinear dynamical systems and statistics, though on p.382 there is a brief discussion of using HMMs [32]. Stamp and Wu also investigated using HMMs to model the logistic and henon maps, showing promising results [33]. It appears that less work has been done on arbitrary nonlinear dynamical systems, especially of higher dimension, or as estimated from the phase portrait, especially with a view towards *comparison* of different systems. Such experiments should be carried out.

VI. Conclusion

The HMM is a useful and profound model that has enjoyed success on its own and as a building block in a number of models closely linked to brain, language, and mind. Of particular interest have been cascades of HMMs that included promising multi-sensory integration and action-word generation experiments; hierarchical HMMs have also shown promise in being able to represent structure in a given data stream at different spatiotemporal scales. Some researchers have claimed that the HMM is the ideal model and can be used to build a brain/mind. Such claims seem premature. More research needs to be done to understand the way that populations of neurons represent and process information, with a special focus on the dynamics of simulated spiking neural networks (SSNNs). It appears that even in such research, HMMs could play an important role in disentangling the complex hierarchical structure that appears in the form of dynamical phase portraits and provide a bridge between dynamics and information measures. Perhaps such experiments will eventually reveal that SSNNs do, in fact, represent information in ways that can be captured by abstract models that include HMMs. The author is not an "armchair debater" [4]; therefore, it is imperative that real-world experiments are carried out to confirm or deny such claims. Fortunately, a series of beginning experiments is already underway.

REFERENCES

[1] R. L. Cave and L. Neuwirth, "Hidden markov models for english," *Proceedings of the Symposium on the Application of Hidden Markov Models to Text and Speech*, pp. 16 –56, 1980.
[2] L. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition*. Prentice Hall PTR, 1993.
[3] P. Boufounos, S. El-Difrawy, and D. Ehrlich, "Hidden markov models for dna sequencing," 2002.
[4] A. Krogh, M. Brown, I. S. Mian, K. Sjolander, and D. Haussler, "Hidden markov models in computational biology: Applications to protein modeling," *Journal of molecular biology*, vol. 235, no. 5, pp. 1501–1531, 1994.
[5] D. Haussler, D. Kulp, M. G. Eeckman, and F. H. Reese, "A generalized hidden markov model for the recognition of human genes in dna," in *Proc. Int. Conf. on Intelligent Systems for Molecular Biology, St. Louis*, 1996, pp. 134–142.
[6] K. Squire and S. Levinson, "Hmm-based concept learning for a mobile robot," *Evolutionary Computation, IEEE Transactions on*, vol. 11, no. 2, pp. 199–212, 2007.
[7] L. Niehaus, "Autonomous learning of action-word semantics in a humanoid robot," *University of Illinois Press*, 2011.
[8] L. Niehaus and S. Levinson, "Online learning and integration of complex action and word lexicons for language grounding," in *Development and Learning and Epigenetic Robotics (ICDL), 2012 IEEE International Conference on*, 2012, pp. 1–6.
[9] A. B. Poritz, "Hidden markov models: A guided tour," *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASP)*, vol. 1, pp. 7–13, 1988.
[10] S. E. Levinson, *Mathematical Models for Speech Technology*. West Sussex, England: Wiley, 2005.
[11] I. B. Collings, V. Krishnamurthy, and J. B. Moore, "On-line identification of hidden markov models via recursive prediction error techniques," *Signal Processing, IEEE Transactions on*, vol. 42, no. 12, pp. 3535–3539, 1994.
[12] U. Holst and G. Lindgren, "Recursive estimation in mixture models with markov regime," *Information Theory, IEEE Transactions on*, vol. 37, no. 6, pp. 1683–1690, 1991.
[13] A. Arapostathis and S. I. Marcus, "Analysis of an identification algorithm arising in the adaptive estimation of markov chains," *Mathematics of Control, Signals and Systems*, vol. 3, no. 1, pp. 1–29, 1990.
[14] F. LeGland and L. Mevel, "Recursive estimation in hidden markov models," in *Decision and Control, 1997., Proceedings of the 36th IEEE Conference on*, vol. 4. IEEE, 1997, pp. 3468–3473.
[15] V. Krishnamurthy and J. B. Moore, "On-line estimation of hidden markov model parameters based on the kullback-leibler information measure," *Signal Processing, IEEE Transactions on*, vol. 41, no. 8, pp. 2557–2573, 1993.
[16] T. Rydén, "On recursive estimation for hidden markov models," *Stochastic Processes and their Applications*, vol. 66, no. 1, pp. 79–96, 1997.
[17] F. LeGland and L. Mével, "Recursive identification of hmms with observations in a finite set," in *Decision and Control, 1995., Proceedings of the 34th IEEE Conference on*, vol. 1. IEEE, 1995, pp. 216–221.
[18] V. Krishnamurthy and G. G. Yin, "Recursive algorithms for estimation of hidden markov models and autoregressive models with markov regime," *Information Theory, IEEE Transactions on*, vol. 48, no. 2, pp. 458–476, 2002.
[19] S. Fine, Y. Singer, and N. Tishby, "The hierarchical hidden markov model: Analysis and applications," *Machine learning*, vol. 32, no. 1, pp. 41–62, 1998.
[20] R. Kurzweil, *How to Create A Mind: The Secret of Human Thought Revealed*. Viking, 2012.
[21] H. Markram and R. Perin, "Innate neural assemblies for lego memory," *Frontiers in Neural Circuits*, vol. 5, no. 6, 2011.
[22] R. Perin, T. K. Berger, and H. Markram, "A synaptic organizing principle for cortical neuronal groups," *Proceedings of the National Academy of Sciences*, vol. 108, no. 13, pp. 5419–5424, 2011.
[23] V. J. Wedeen, D. L. Rosene, R. Wang, G. Dai, F. Mortazavi, P. Hagmann, J. H. Kaas, and W.-Y. I. Tseng, "The geometric structure of the brain fiber pathways," *Science*, vol. 335, no. 6076, pp. 1628–1634, 2012.
[24] M. Catani, I. Bodi, and F. DellAcqua, "Comment on the geometric structure of the brain fiber pathways," *Science*, vol. 337, no. 6102, p. 1605, 2012. [Online]. Available:
[25] A. M. Duda and S. E. Levinson, "Complex networks of spiking neurons: Collective behavior characterization," *Proceedings of the Eighth International Conference on Complex Systems*, pp. 1627–1629, 2011.
[26] ——, "Nonlinear dynamical multi-scale model of associative memory," in *Proceedings of the 2010 Ninth International Conference on Machine Learning and Applications*, ser. ICMLA '10. Washington, DC, USA: IEEE Computer Society, 2010, pp. 867–872.
[27] M. Kaufman and R. Thomas, "Emergence of complex behaviour from simple circuit structures," *C. R. Bio.*, vol. 326, pp. 205–214, 2003.
[28] T. M. Cover and J. A. Thomas, *Elements of information theory*. Wiley-interscience, 2012.
[29] C. Myers, A. Singer, F. Shin, and E. Church, "Modeling chaotic systems with hidden markov models," in *Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on*, vol. 4, 1992, pp. 565–568.
[30] P. Marteau and H. D. Abarbanel, "Noise reduction in chaotic time series using scaled probabilistic methods," *Journal of Nonlinear Science*, vol. 1, no. 3, pp. 313–343, 1991.
[31] A. M. Fraser, "Chaotic data and model building," *NATO ASI SERIES B PHYSICS*, vol. 256, pp. 125–130, 1991.
[32] A. I. Mees, *Nonlinear dynamics and statistics*. Birkhauser, 2001.
[33] D. I. Stamp and Q. Wu, "Prediction of chaotic time series using hidden markov models," in *Control '98. UKACC International Conference on (Conf. Publ. No. 455)*, 1998, pp. 1420–1425 vol.2.

[4]See p.240 of Levinson [10]