In Proceedings of ICCM - 2007- Eighth International Conference on Cognitive Modeling. 145 - 150. Oxford, UK: Taylor & Francis/Psychology Press.

Learning Cognitive Load Models for Developing Team Shared Mental Models

Xiaocong Fan (XFAN@Psu.Edu)
School of Engineering, The Behrend College
The Pennsylvania State University, Erie, PA 16563 USA

Po-Chun Chen (PZC123@Psu.Edu)
Department of Computer Science and Engineering
The Pennsylvania State University, University Park, PA 16802 USA

John Yen (JUY1@Psu.Edu)
College of Information Sciences and Technology
The Pennsylvania State University, University Park, PA 16802 USA

Abstract

Cognitive studies indicate that members of a high performing team often develop shared mental models to predict others' needs and coordinate their behaviors. The concept of shared mental models is especially useful in the study of human-centered collaborative systems that require humans to team with autonomous agents in complex activities. We take the position that in a mixed human/agent team, agents empowered with cognitive load models of human team members can help humans develop better shared mental models. In this paper, we focus on the development of realistic cognitive load models. Cognitive experiments were conducted in team contexts to collect data about the observable secondary task performance of human participants. The data were used to train hidden Markov models (HMM) with varied number of hypothetical hidden states. The results indicate that the model spaces have a three-layer structure. Statistical analysis reveals some characteristics of top layer models, which can be used in guiding the selection of HMM-based cognitive load models.

Introduction

Human-centered multi-agent teamwork has attracted increasing attentions in multi-agent systems field (Bradshaw et al., 2002; Norling, 2004). Human-centered teamwork, involving both humans and software agents, is about collaboratively establishing situation awareness, developing shared mental models as situation evolves, and appropriately adapting to mixed-initiative activities. Humans and autonomous agents are generally thought to be complementary: while humans are limited by their cognitive capacity in information processing, they are superior in spatial, heuristic, and analogical reasoning; autonomous agents can continuously learn expertise and tacit problem-solving knowledge from humans to improve system performance.

However, the foundation of human-agent collaboration keeps being challenged because of nonrealistic modeling of mutual awareness of the state of affairs. In particular, few researchers look beyond to assess the principles of modeling shared mental constructs between a human and his/her assisting agent. Moreover, human-agent relationships can go beyond partners to teams (Fan, Yen, & Volz, 2005). Many informational processing limitations of individuals can be alleviated by having a group perform tasks. Although groups also can create additional costs centered on communication, resolution of conflict, and social acceptance, it is suggested that such limitations can be overcome if people have *shared cognitive structures* for interpreting task and social requirements (Lord & Maher, 1990). Therefore, there is a clear demand for investigations to broaden and deepen our understanding on the principles of shared mental modeling among members of a mixed human-agent team.

There are lines of research on multi-agent teamwork, both theoretically and empirically. For instance, Joint Intention (Cohen & Levesque, 1991) and SharedPlans (Grosz & Kraus, 1996) are two theoretical frameworks for specifying agent collaborations. One of the drawbacks is that, although both have a deep philosophical and cognitive root, they do not explicitly accommodate the modeling of human team members. Cognitive studies suggested that teams which have shared mental models are expected to have common expectations of the task and team, which allow them to predict the behavior and resource needs of team members more accurately (Rouse, Cannon-Bowers, & Salas, 1992; Klimoski & Mohammed, 1994). Cannon-Bowers et al. (Rouse et al., 1992) explicitly argue that team members should hold compatible models that lead to common "expectations". We agree on this and believe that the establishment of *shared mental models* among human and agent team members is a critical step to advance human-centered teamwork research.

The long-term goal of our research is to understand how shared cognitive structures can enhance human-agent team performance. In particular, we argue that to favor human-agent collaboration, an agent system should be designed to allow the estimation and prediction of human teammates' (relative) cognitive loads, and use that to offer improvised, unintrusive help. Ideally, being able to predict the cognitive/processing capacity curves of teammates could allow team members to help *the right party at the right time* (Fan & Yen, 2007), avoiding unbalanced work/cognitive loads among the team.

The specific objective of the work reported here is to develop a computational cognitive capacity model to facilitate the establishment of shared mental models. The rest of the paper is organized as follows. In Section 2 we review studies on cognitive load and its measurements. Section 3 gives our motivation of using HMM-based approach to modeling human cognitive loads. Section 4 describes the cognitive task design and the experiment conducted to collect observable measures of secondary task performance in a team context. Section 5 reports the methodology of learning Hidden Markov Models (HMM) and the principles of selecting appropriate HMM models for an agent to estimate its human partner's dynamic cognitive load.

Background on Cognitive Load Studies

People are information processors. Cognitive load studies (Miller, 1956; Lord & Maher, 1990; Baddeley, 1992) are, by and large, concerned about working memory capacity and how to circumvent its limitations in human problem-solving activities such as learning and decision making.

According to the cognitive load theory (Paas & Merrienboer, 1993), *cognitive load* is defined as a multidimensional construct representing the load that a particular task imposes on the performer. It has a causal dimension including causal factors that can be characteristics of the subject (e.g. expertise level), the task (e.g. task complexity, time pressure), the environment (e.g. noise), and their mutual relations. It also has an assessment dimension reflecting the measurable concepts of mental load (imposed exclusively by the task and environmental demands), mental effort (the cognitive capacity actually allocated to the task), and performance.

Lang's information-processing model (Lang, 2000) consists of three major processes: encoding, storage, and retrieval. The encoding process selectively maps messages in sensory stores that are relevant to a person's goals into working memory; the storage process consolidates the newly encoded information into chunks, forming associations and schema to facilitate subsequent recalls; the retrieval process searches the associated memory network for a specific element/schema and reactivates it into working memory. The model suggests that processing resources (cognitive capacity) are independently allocated to the three processes. In addition, working memory is used both for holding and for processing information (Baddeley, 1992). Due to limited capacity, when greater effort is required to process information, less capacity remains for the storage of information. Hence, the allocation of the limited cognitive resources has to be balanced in order to enhance human performance. This comes to the issue of measuring cognitive load, which has proven difficult for cognitive scientists.

Cognitive load can be assessed by measuring mental load, mental effort, and performance using rating scales, psychophysiological, and secondary task techniques (Paas, Tuovinen, Tabbers, & Gerven, 2003). Self-ratings may appear questionable and restricted, especially when instantaneous load needs to be measured over time. Although physiological measures are sometimes highly sensitive for tracking fluctuating levels of cognitive load, costs and work place conditions often favor task- and performance-based techniques, which involve the measure of a secondary task as well as the primary task under consideration. Secondary task techniques are based on the assumption that performance on a secondary task reflects the level of cognitive load imposed by a primary task (Sweller, 1988). From the resource allocation perspective, assuming a fixed cognitive capacity, any increase in cognitive resources required by the primary task must inevitably decrease resources available for the secondary task (Lang, 2000). Consequently, performance in a secondary task deteriorates as the difficulty or priority of the primary task increases. The level of cognitive load can thus be manifested by the secondary task performance: the subject is getting overloaded if the secondary task performance drops.

A secondary task can be as simple as detecting a visual or auditory signal but requires sustained attention. Its performance can be measured in terms of reaction time, accuracy, and error rate. However, one important drawback of secondary task performance, as noted by Paas (Paas et al., 2003), is that it can interfere considerably with the primary task (competing for limited capacity), especially when the primary task is complex. To better understand and measure cognitive load, Xie and Salvendy (2000) introduced a conceptual framework, which distinguishes instantaneous load, peak load, accumulated load, average load, and overall load. It seems that the notation of instantaneous load, which represents the dynamics of cognitive load over time, is especially useful for monitoring the fluctuation trend so that free capacity can be exploited at the most appropriate time to enhance the overall performance in human-agent collaborations.

Modeling Cognitive Loads Using HMM

A hidden Markov model (HMM) (Rabiner, 1989) is a statistical approach to modeling systems that can be viewed as a Markov process with unknown hidden parameters. The hidden state variables are not directly visible, but influenced by certain observable variables. Each hidden state has a probability distribution over the possible observable symbols. Therefore the sequence of observable states can be used to make inference on the sequence of hidden states of a HMM. A HMM is denoted by $\lambda = \langle N, V, A, B, \pi \rangle$, where N is a set of hidden states, V is a set of observation symbols, A is a set of state transition probability distributions, B is a set of observation symbol probability distributions (one for each hidden state), and π is the initial state distribution. Hidden Markov models have been widely applied in bioinformatics and temporal pattern recognition (such as speech, handwriting, and gesture recognition).

An intelligent agent being a cognitive aid, it is desirable that the model of its human partner implemented within the agent is *cognitively-acceptable*, if not descriptively accurate. However, building a cognitive load model that is *cognitively-acceptable* is not trivial. A HMM-based approach is used in this study for several reasons. First, cognitive load has a dynamic nature. As we mentioned above, being able to monitor the dynamics of a human partner's cognitive load over time is very useful for an agent to proactively identify collaboration opportunities in human-centered teamwork. The inference of the instantaneous cognitive load can be cast as a temporal pattern recognition problem, which is especially suitable to adopt a HMM.

Second, the HMM approach demands that the system being modeled (here, human's cognitive capacity) has both observable and hidden state variables, and the hidden variables should be correlated to the observable variables. As discussed above, there is ample evidence supporting secondary task performance as a highly sensitive and reliable technique for measuring human's cognitive load (Paas et al., 2003). We thus can use the secondary task performance as observable signals to estimate the hidden cognitive load state. For example, the secondary task performance can be measured in terms of the number of items correctly recalled. According to Miller's 7 ± 2 rule, the observable states will take integer values from 0 to 9 (assume it is 9 when the number of items correctly recalled is no less than 9). Hence, the strong tie, as uncovered in cognitive studies, between human's cognitive load and his/her

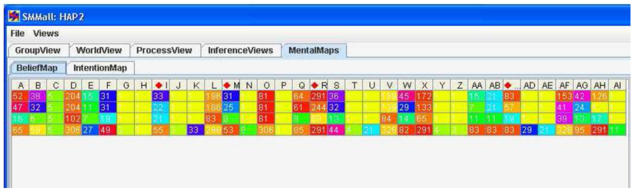


Figure 1: Shared Belief Map.

secondary task performance also justifies the use of a HMM approach.

In the study reported below, we employed an experimental approach to collect realistic data of secondary task performance in a collaborative setting. We then used the data to learn HMM models of various number of hidden states, trying to understand the properties of HMM models that are acceptable for modeling human cognitive load.

Cognitive Task Design and Data Collection

To study the dynamics of cognitive loads when humans are working in collaborative settings, we developed a system simulating a dynamic battlefield infosphere. A team can have several team members; each of them has limited observability (say, covering only a portion of the battlefield). The goal of a team is to selectively share information among members in a timely manner to develop global situation awareness (e.g., for making critical decisions).

Team members share information through a GUI with a shared belief map as shown in Figure 1. A shared belief map is a table with color-coded *info-cells*—cells associated with information. Each row captures the belief model of one team member, and each column corresponds to a specific information type (all columns together define the boundary of the information space being considered). Thus, info-cell C_{ij} of a map encodes all the beliefs (instances) of information type j held by agent i. Color coding applies to each info-cell to indicate the number of information instances held by the corresponding agent.

The concept of shared belief map facilitates the development of global situation awareness. It helps maintain and present a human partner with a synergy view of the shared mental models evolving within a team. Information types that are semantically related (e.g., by inference rules) can be closely organized in the map. Hence, nearby info-cells can form meaningful plateaus (or contour lines) of similar colors. Colored plateaus indicate those sections of a shared mental model that bear high overlapping degrees. In addition, the perceptible color (hue) difference manifested from a shared belief map indicates the information difference among team members, and hence visually represents the potential information needs of each team member.

We designed a primary task and a secondary task for the human subjects. The *primary task* of a human subject is to share the right information with the right party at the right time. Every time step (about 15 seconds), simulated spot reports (situational information) will be generated and randomly dispatched to team members. An info-cell on a person's belief map will be flashed (for 2 seconds) whenever that person gets new information of the type represented by that cell. The flashed cells are exactly those with newly available information that should be shared among teammates at that time step. An info-cell is frozen (the associated information is no longer sharable) when the next time step comes. Hence, a human subject has to share the newly available information with other team members under time stress. To share the information associated with an info-cell, a human subject needs to click the right mouse button on the cell to pop up a context menu, and select the receiving teammate(s) from the pop-up menu. Because information is randomly dispatched to team members, to each participant, the flashed info-cells vary from time to time, and there can be up to 12 info-cells flashed at each time step.

To choose an appropriate secondary task for the domain problem at hand is not trivial, although the general rationale is that the secondary task performance should vary as the difficulty of the primary task increases. Typically, a secondary task requires the human subjects to respond to *unpredictable stimuli* in either overt (e.g., press a button) or covert (e.g., mental counting) ways. Just for the purpose of estimating a human subject's cognitive load, any artificial task can be used as a secondary task to force the subject to go through. However, in a realistic application, we have to make sure that the selected secondary task interacts with the primary task in meaningful ways, which is not easy and often depends on the domain problem at hand. Specific to this study, the *secondary task* of a human subject is to *remember and mark* the cells being flashed (not necessarily in the exact order). Secondary task performance at step t is thus measured as the number of cells marked correctly at t. The more number of cells marked correctly, the lower the subject's cognitive load.

While the experiment is designed in a collaborative setting with a meaningful primary task, we here especially focus on the secondary task performance. We would like to collect realistic data of secondary task performance and use the data to learn and understand the properties of HMM models of human cognitive loads. We randomly recruited 30 human subjects from undergraduate students and randomly formed 10 teams of size three. We ran the simulation system 9 times for each team and collected the secondary task performance of each team member: the number of info-cells marked correctly at each time step. Each run of the experiment has 45 time steps. We thus collected $10 \times 3 \times 9 = 270$ observation sequences of length 45.

Learning Cognitive Load Models

Learning Procedure

With the collected observation sequences, we took a two-phase approach. We first applied a window method to learn HMMs from sub-sequences and then evaluated each learned model with the original 270 observation sequences. Specifically, we assume human cognitive load can be modeled by HMMs with n hypothetical hidden states where $3 \leq n \leq 10$. To train a n-state HMM, we applied a window of width n on the original observation sequences to extract sub-sequences as training data. For example, to learn a 5-state HMM, a window of width 5 was used, which produced $270 \times 41 = 11,070$ training samples.

The training samples then were fed to the Baum-Welch algorithm (Rabiner, 1989) to learn HMMs. The training

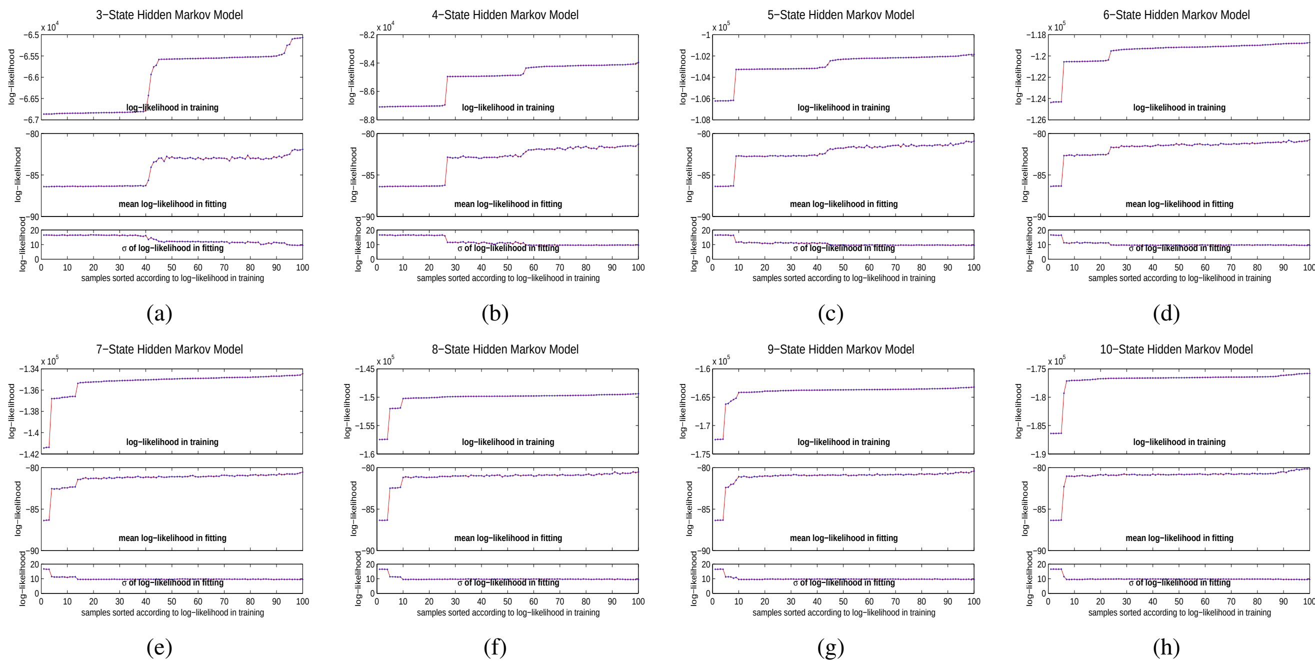


Figure 2: Each subfigure has top, middle, and bottom components, which plot the log-likelihoods of models after training, the log-likelihoods in testing, and the standard deviation of the log-likelihoods in testing. It clearly indicates (1) Maxima of each model space (from 3 to 10) form a 3-layer structure; (2) Better trained models lead to better testing log-likelihoods; and (3) Better trained models incur lower deviations. Model space also varied: as the number of hidden states increased from 3 to 10, the fraction of models at the middle and bottom levels reduced with the fraction of models at the top level increased.

process was terminated upon the convergence of its log-likelihood. Given the possibility of convergence at local maxima, we randomly generated initial guesses of parameters (A, B, π) and repeated 100 times for each hypothetical n-state model. Consequently, we obtained 8 model spaces, each has 100 HMMs with n hidden states $(3 \leq n \leq 10)$. The top component of each subfigure in Figure 2 plots ascendingly the final log-likelihoods of the learned models of the corresponding model space.

In the second phase, for each learned HMM, we used the Forward procedure (Rabiner, 1989) to evaluate its performance by computing the occurrence probabilities (log-likelihoods) of the original 270 observation sequences, which produced 270 log-likelihoods. The middle component of each subfigure in Figure 2 plots, for each corresponding model plotted in the top component, the mean of the 270 log-likelihoods resulted from fitting (testing) the model with the original observation sequences. The bottom components plot the standard deviations of each model in fitting.

The Model Space of Cognitive Load

Each subfigure in Figure 2(a-h) has top, middle, and bottom components, which plot the log-likelihoods of models after training, the log-likelihoods in testing, and the standard deviation of the log-likelihoods in testing. It clearly indicates that the model spaces with the number of hidden states ranging from 3 to 10 share some common properties. First, each model space (from 3 to 10) has a 3-layer structure, which means the log-likelihood maxima are clustered in three levels (models at the middle and bottom levels converged to local maxima). Second, better trained models performed better in testing: the trend of the log-likelihoods in fitting is consistent with the trend of the log-likelihoods in training (as ordered ascendingly in the top component). Third, better models produced lower deviation in fitting. Also, as the number of hidden states increased from 3 to 10, the fraction of models at the middle and bottom levels reduced with the fraction of models at the top level (converged at global maxima) increased. Extremely, most of the models in the space of 3-state HMMs are ‘bad’ models, while most of the models in the space of 10-state HMMs are ‘good’ models.

Table 1: Means of the longest hidden-state jumps (LHSJ) and mean fractions of transition pairs with stronger backward jumps (FSB) for HMMs with states from 3 to 10.

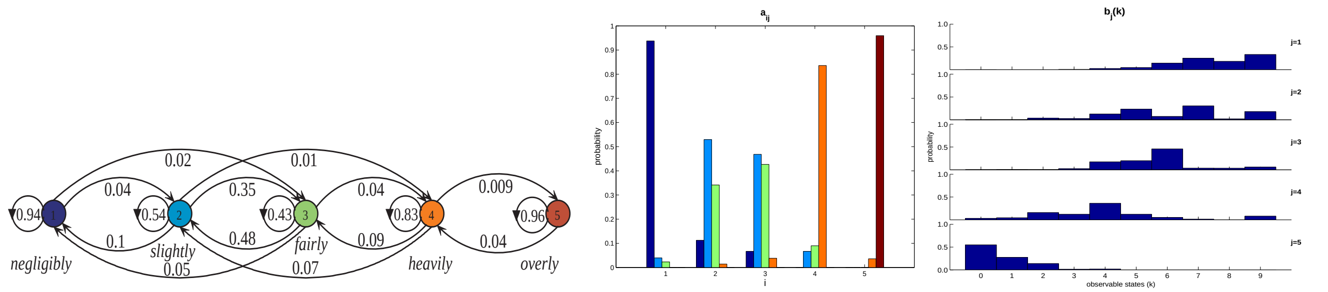
states	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
LHSJ	1.33	1.76	2.10	2.56	2.83	3.30	3.82	4.19
FSB	1.00	0.76	0.66	0.62	0.60	0.58	0.56	0.54

LHSJ = 0.0906 + 0.407 states

Properties of ‘Good’ Cognitive Load Models

We may wonder whether there are any properties shared by the ‘good’ models as appeared at the top layer of each model space. We first examine B, the observation probability distributions. There is a strong evidence that the B’s of models at the top layer demonstrated more distinguishable peaks, compared with those at lower layers which typically had indistinguishable peaks or mixed distributions. Figure 3(c) gives the B of one 5-state model at the top layer.

There are several statistics to check on the model parameter A, state transition probability distributions. With only



(a) (b) (c)

Figure 3: (a) An example 5-state HMM; (b) Transition probability distributions A; (c) Observation probability distributions B.

Table 2: For each state, the mean number of state transitions and the mean of initial state probability π_i.

states	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
3	1.00/0.64	2.00/0.31	*2.00/***0.045**							
4	1.89/0.44	2.20/0.26	*2.75/0.27*	**1.86/0.037**						
5	2.32/0.41	2.70/0.20	*3.04/0.18*	2.98/0.19	**1.70/0.028**					
6	2.87/0.33	3.60/0.18	*3.66/0.14*	3.65/0.15	3.14/0.17	**1.51/0.025**				
7	3.24/0.29	3.92/0.18	4.40/0.12	*4.44/0.11*	4.07/0.12	**2.68**/0.16	**1.23/0.019**			
8	3.52/0.26	4.27/0.15	4.64/0.12	4.92/0.10	*5.01/0.09*	4.34/0.11	**2.56**/0.15	**1.16/0.017**		
9	4.09/0.24	4.89/0.12	5.41/0.11	5.60/0.10	*5.70/0.09*	5.58/0.08	5.12/0.09	**2.66**/0.16	**0.93/0.016**	
10	4.45/0.21	5.21/0.12	5.62/0.10	5.85/0.09	6.27/0.08	*6.37/0.07*	6.14/0.08	4.77/0.09	**2.61**/0.14	**0.79/0.015**

cell content: number of state jumps/π

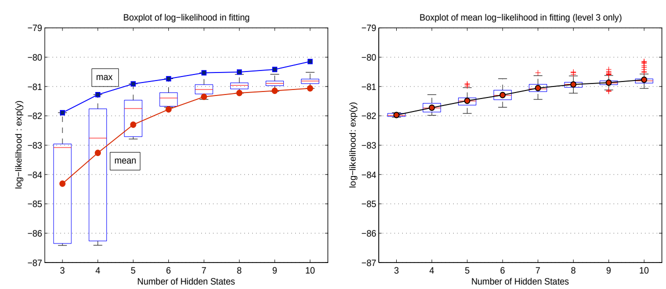
top-layer models considered, Table 1 gives the means of the longest hidden-state jumps (LHSJ) and the mean fractions of transition pairs with stronger backward jumps (FSB). For example, for HMMs with 5 states, on average state transitions have jumps no more than 2.1 states, and of all the possible state transition pairs (A_{ij}, A_{ji}) where $1 \leq i < j \leq 5$ and state j represents a higher cognitive load than state i, 66% have stronger backward transitions ($A_{ij} < A_{ji}$). Of all the models with states from 3 to 10, the means of LHSJ, ranging from 1.33 to 4.19, linearly related to the number of states with slope $0.407 \approx 2/5$. Interestingly, all models have relatively more transition pairs with stronger backward jumps. This seems to suggest that humans can more easily recover from than switch to a higher cognitive load state.

For each state and each model, Table 2 gives the mean number of transitions and the mean of initial state probability. For each category except models with 3 states, it seems that the highest one (4–6) or two (7–10) states have much few number of transitions than the other states. It may suggest that humans, once become "overly" loaded, can not easily return to cognitively favorable states. The highest state of each model category also assume extremely lower initial state probability, and the lower states have relatively higher initial state probability. This is intuitively true because humans seldomly get overloaded in the beginning. For each model category, Table 2 also indicates such a trend: as hidden state changed from low to high, the mean number of state transitions increased to its maximum (in italic), while π_i decreased to minimum (with the highest state ignored). This may indicate that the more active a state is, the less likely a human can be initially in that state.

The Number of Hidden States

A crucial question is, how many hidden states are appropriate for modeling cognitive load using HMMs?

Figure 4(a) gives the Boxplot of likelihoods in fitting for all the models in each space, and Figure 4(b) gives the Boxplot of likelihoods in fitting for top-layer models only. Fig. 4(a) says that the model variance is small enough when the number of hidden states is no less than 4. Fig. 4(b) shows that there is a linear improvement on the top-layer models as state number increases. However, because the observable measure of secondary task performance ranges from 0 to 9, models with too many hidden states may overfit the given data. In addition, the trained models with more than 8 states demonstrated 'strongly-connected' sub-structures. Figure 5 gives a top-layer model with 10 states, where links on the upper part represent forward transitions and those on the lower part are backward ones (state transition probabilities are visualized in color densities). It is clear that this 10-state model can be reduced to a 5-state model if we view (1,2,3) and (4,5,6,7) as two compound states.



(a) (b)

Figure 4: Boxplot of model log-likelihoods.

Practically, it is better to pick from top-layer models with 7 states, which have a mean likelihood e^{-81}–only slightly lower than 10-state models. It is also acceptable to choose from the top-layer models with 5 states; indeed, the performance could be as good as 7-state models if the best of 5-state models were picked. In addition, there is ample evi-

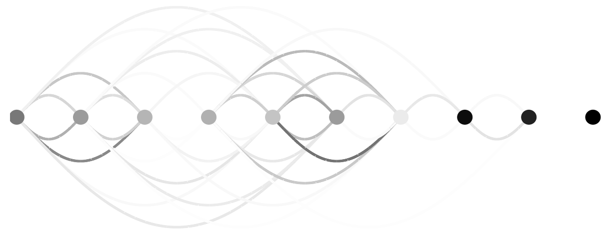


Figure 5: A 10-state model of cognitive load.

dence suggesting that human cognitive load is a continuous function over time and does not manifest sudden shifts unless there are fundamental changes in tasking demands. If we place a constraint on the state transition coefficients such that no jumps of more than 3 states are allowed, then models with 5, 6, or 7 states are best choices (ref. Table 2). Moreover, models with fewer states are more interpretable. Like the 5-state model in Fig. 3, it is easier to assign meaningful names to the hidden states.

In sum, with all the above factors considered, it seems we can follow 7 ± 2 rule to choose the number of hidden states of a HMM-based cognitive load model. To examine the efficacy of such a principle, we used 5-state HMM models and conducted a study (Fan & Yen, 2007) involving two team types: ten of the teams performed with cognitive load estimation available from agents and ten teams with no such estimation. The result indicated that teams with load estimation performed significantly better than teams with no load estimation. The reason is that being able to estimate other team members' cognitive load allows them to share the needed information with the right party at the right time.

Conclusion

An agent empowered with a cognitive load model of its human peer can be beneficial in offering trustable autonomy and unintrusive help. We used Hidden Markov Models to capture cognitive load in a way that can be used in team contexts to make predictions about other team members' workload.

To develop realistic cognitive load models, we conducted cognitive experiments to capture human's observable secondary task performance and used that to train hidden Markov models (HMM) with varied number of hypothetical hidden states. The results indicate that each model space has a three-layer structure, and it is suggested to choose models with 7 ± 2 hidden states. With all the constraints considered, it is recommended that HMMs with 5, 6, or 7 states are best choices for modeling human cognitive load. Statistical analysis revealed that good models also share some common properties: (1) observation probability distributions have distinguishable peaks for different states; (2) highest hidden states have extremely lower initial state probabilities; and (3) the longest state jumps are linearly related to the number of states with a slope $2/5$, and there are more transition pairs with stronger backward jumps. These can be used in guiding the selection of HMM-based cognitive load models.

References

Baddeley, A. D. (1992). Working memory. *Science*, *255*, 556-559.

Bradshaw, J., Sierhuis, M., Acquisti, A., Gawdiak, Y., Prescott, D., Jeffers, R., et al. (2002). What we can learn about human-agent teamwork from practice. In *Workshop on teamwork and coalition formation at AAMAS 02*. Bologna, Italy.

Cohen, P. R., & Levesque, H. J. (1991). Teamwork. *Nous*, *25*(4), 487-512.

Fan, X., & Yen, J. (2007). Realistic cognitive load modeling for enhancing shared mental models in human-agent collaboration. In *Proceedings of the sixth international joint conference on autonomous agents and multiagent systems* (p. 383-390). ACM Press.

Fan, X., Yen, J., & Volz, R. A. (2005). A theoretical framework on proactive information exchange in agent teamwork. *Artificial Intelligence*, *169*, 23–97.

Grosz, B., & Kraus, S. (1996). Collaborative plans for complex group actions. *Artificial Intelligence*, *86*, 269-358.

Klimoski, R., & Mohammed, S. (1994). Team mental model: Construct or metaphor? *Journal of Management*, *20*(2), 403-437.

Lang, A. (2000). The limited capacity model of mediated message processing. *Journal of Communication*, *Winter*, 46-70.

Lord, R. G., & Maher, K. J. (1990). Alternative information processing models and their implications for theory, research, and practice. *The Academy of Management Review*, *15*(1), 9-28.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *63*, 81-97.

Norling, E. (2004). Folk psychology for human modelling: Extending the BDI paradigm. In *AAMAS '04: International conference on autonomous agents and multi agent systems* (pp. 202–209).

Paas, F., & Merrienboer, J. V. (1993). The efficiency of instructional conditions: an approach to combine mental-effort and performance measures. *Human Factors*, *35*, 737-743.

Paas, F., Tuovinen, J. E., Tabbers, H., & Gerven, P. W. M. V. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, *38*(1), 63-71.

Rabiner, L. R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*, 257-286.

Rouse, W., Cannon-Bowers, J., & Salas, E. (1992). The role of mental models in team performance in complex systems. *IEEE Trans. on Sys., man, and Cyber*, *22*(6), 1296-1308.

Sweller, J. (1988). Cognitive load during problem solving: effects on learning. *Cognitive Science*, *12*, 257-285.

Xie, B., & Salvendy, G. (2000). Prediction of mental workload in single and multiple task environments. *International Journal of Cognitive Ergonomics*, *4*, 213-242.