A Preliminary Investigation of Hierarchical Hidden Markov Models for Tutorial Planning

Kristy Elizabeth Boyer[*], Robert Phillips[1], Eun Young Ha, Michael D. Wallis[1],
Mladen A. Vouk, and James C. Lester
Department of Computer Science, North Carolina State University
[1]Dual Affiliation with Applied Research Associates, Inc., Raleigh, North Carolina
[*]Corresponding author: keboyer@ncsu.edu

For tutorial dialogue systems, selecting an appropriate dialogue move to support learners can significantly influence cognitive and affective outcomes. The strategies implemented in tutorial dialogue systems have historically been based on handcrafted rules derived from observing human tutors, but a data-driven model of strategy selection may increase the effectiveness of tutorial dialogue systems. Tutorial dialogue projects including CIRCSIM-TUTOR [1], ITSPOKE [2], and KSC-PAL [3] have utilized corpora to inform the behavior of a system. Our work builds on this line of research by directly learning a hierarchical hidden Markov model (HHMM) for predicting tutor dialogue acts within a corpus. The corpus was collected during a human-human tutoring study in the domain of introductory computer science [4]. We annotated the dialogue moves with dialogue acts (Table 1). The subtask structure and student problem-solving action correctness were also annotated manually.

Table 1. Dialogue act annotation scheme

Dialogue Act Tag	Description
ASSESSING QUESTION (AQ)	Request for feedback on task or conceptual utterance.
EXTRA-DOMAIN (EX)	Asides not relevant to the tutoring task.
GROUNDING (G)	Acknowledgement/thanks.
LUKEWARM CONTENT FEEDBACK (LCF)	Negative assessment with explanation.
LUKEWARM FEEDBACK (LF)	Lukewarm assessment of task action or conceptual utterance.
NEGATIVE CONTENT FEEDBACK (NCF)	Negative assessment with explanation.
NEGATIVE FEEDBACK (NF)	Negative assessment of task action or conceptual utterance.
POSITIVE CONTENT FEEDBACK (PCF)	Positive assessment with explanation.
POSITIVE FEEDBACK (PF)	Positive assessment of task action or conceptual utterance.
QUESTION (Q)	Task or conceptual question.
STATEMENT (S)	Task or conceptual assertion.

We trained first-order Markov (bigram) models, HMMs, and HHMMs on the annotated sequences. In ten-fold (five-fold for the HHMMs due to data sparsity) cross-validation, the HHMM (partially depicted in Figure 1) predicted tutor dialogue acts with an average 57% accuracy, significantly higher than the 27% accuracy of bigram models ($p<0.0001$) and better than the 48% accuracy of HMMs without hierarchical structure ($p<0.05$).[3]

[3] p-values reported are from one-tailed two sample t-tests for equality of means with pooled variance

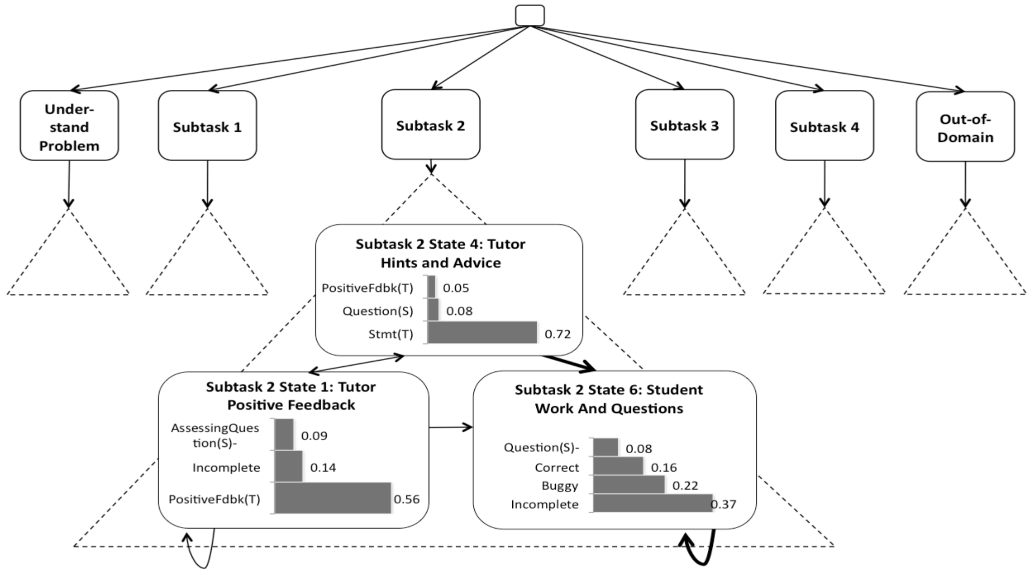

Figure 1. Subset of learned HHMM

Because of HHMMs' capacity for explicitly representing hidden dialogue structure and hierarchical task structure, they perform better than bigrams and HMMs for predicting tutor moves in our corpus. The models' performance points to promising future work that includes utilizing additional lexical and syntactic features along with fixed student characteristics within a hierarchical hidden Markov modeling framework. More broadly, the results highlight the importance of considering task structure when modeling a complex domain such as those that often accompany task-oriented tutoring. Finally, a key direction for data-driven dialogue management is to learn unsupervised dialogue act and task classification models.

Acknowledgments. This work is supported in part by the NC State Department of Computer Science and NSF grants CNS-0540523, REC-0632450, IIS-0812291, and a Graduate Research Fellowship. Any opinions, findings, conclusions, or recommendations expressed in this report are those of the participants, and do not necessarily represent official views, opinions, or policy of the National Science Foundation.

References

[1] M. Evens and J. Michael. *One-on-One Tutoring by Humans and Computers.* Mahwah, New Jersey: Lawrence Erlbaum Associates, 2006.

[2] K. Forbes-Riley and D. Litman. Adapting to Student Uncertainty Improves Tutoring Dialogues, *Proceedings of the 14th International Conference on Artificial Intelligence and Education*, pp. 33-40, 2009.

[3] C. Kersey, B. Di Eugenio, P. Jordan and S. Katz. KSC-PaL: A Peer Learning Agent that Encourages Students to take the Initiative, *Proceedings of the NAACL HLT Workshop on Innovative Use of NLP for Building Educational Applications*, pp. 55-63, 2009.

[4] K.E. Boyer, R. Phillips, A. Ingram, E. Y. Ha, M. D. Wallis, M. A. Vouk and J. C. Lester (in press). Characterizing the Effectiveness of Tutorial Dialogue with Hidden Markov Models, to appear in *Proceedings of the 10th International Conference on Intelligent Tutoring Systems*, 2010.