Dynamic Bayesian Network Models for Peer Tutor Interactions

Yoav Bergner

Center for Advanced Psychometrics
Educational Testing Service

ybergner@ets.org

Innovative Assessment of Collaboration
Washington D.C., Nov 3-4, 2014

Confidential and Proprietary. Copyright © 2013 by Educational Testing Service.
All rights reserved.

Plan

HMMs and DBNs

Automated Peer Tutoring Assistant

Model 1: estimating the assistance value of tutor inputs

Model 2: a discriminative HMM to classify learning gains

Summary and future work

Confidential and Proprietary. Copyright © 2013 by Educational Testing Service.
All rights reserved.

Some attributes of hidden Markov models

- been around for a long time
  - (e.g., Gilbert, 1959; Baum & Petrie 1966)
- model discrete sequence data using a hidden state which undergoes a Markov process
  - widely applied in computer science, signal processing
- dynamic generalization of finite mixture model with Markov property
- special cases of dynamic Bayesian networks (DBNs)
  - can be extended within this framework
  - the linear Gaussian state-space model (aka Kalman filter) is a continuous-state analogue of the HMM

Confidential and Proprietary. Copyright © 2013 by Educational Testing Service.
All rights reserved.

Some applications of HMMs

Speech recognition, NLP, and computer vision:
observe sequence of wave-forms or pixel maps,
tag underlying (hidden) phoneme, word, or object identity

In education:

- Bayesian knowledge tracing: observe correctness on practice items, identify hidden mastery-learning state, adapt (Atkinson, 1971; Corbett & Anderson, 1995; Reye, 2004)
- Effective knowledge sharing groups (Soller & Stevens, 2007)
- Markov decision processes (Almond, 2007)
  observe test scores, model student growth, and decide interventions
- Expert tutoring styles (Boyer et al., 2011)

Confidential and Proprietary. Copyright © 2013 by Educational Testing Service.
All rights reserved.

Adaptive Peer Tutoring Assistant

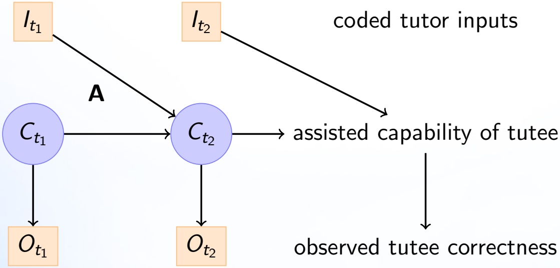
Walker, Rummel & Koedinger (2009-2011)

- Tutee attempts to solve algebra exercises in a cognitive tutor
- a peer tutor (remotely) views each step, and
- the two communicate via a chat window (the peer tutor is provided with hints about the correct step)
- APTA is designed to promote and scaffold positive tutoring behaviors i.e. intelligent support for collaborative learning
- peer tutor chats are automatically coded according to production rules and can trigger feedback for the peer tutor

(clip art credit: DiscoverySchool.com)

Confidential and Proprietary. Copyright © 2013 by Educational Testing Service. All rights reserved.

Model 1: estimating the assistance value of tutor inputs with an IOHMM



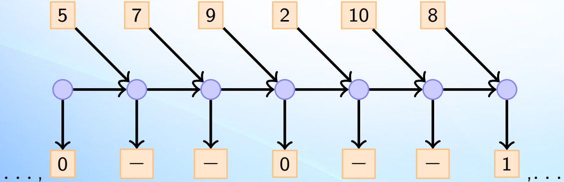
This application is inspired by Bayesian knowledge tracing, although the interpretation is quite different (cf. Beck et al., 2008)

Assistance is a parameter derived from the transition matrix **A**

Confidential and Proprietary. Copyright © 2013 by Educational Testing Service.
All rights reserved.

Model 1: estimating the assistance value of tutor inputs

output code	tutee's action	tutor's action	input code
INCORRECT	[divide rt]		
		u did that last step wrong, u need to divide both sides by r+v	helpAfterIncorrect, helpWithHelpStarters [5][7]
	now cuz i didnt slove for t yet		
	[undo]		
		NO! listen to what i am saying.	startersWithNoHelp [9]
INCORRECT	[divide r+t]		
	i did now what do i have to do		
		its divide by r+t	lowLevelHelp, noStartersWithHelp
		i mean r+v	noHelpWithNoStarters [2][8][10]
	[undo]		
CORRECT	[divide r+v]		



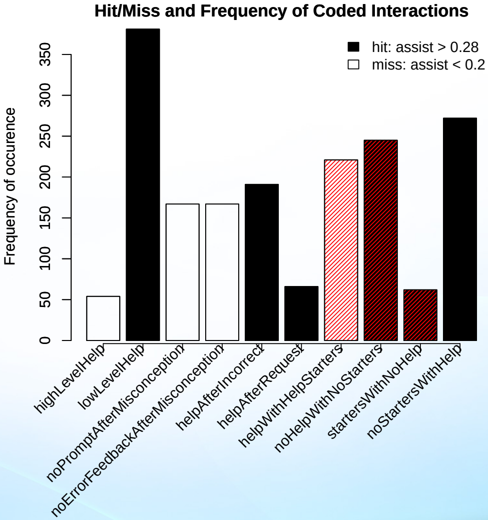
Confidential and Proprietary. Copyright © 2013 by Educational Testing Service.
All rights reserved.

Results from Model 1 study - hit/miss frequency

Now compare the hits and misses for external validity with the APTA codings

Starters are sentence classifiers that tutors were encouraged to use, e.g. "explain why wrong," "ask why"

Appear to have two false positives and one false negative



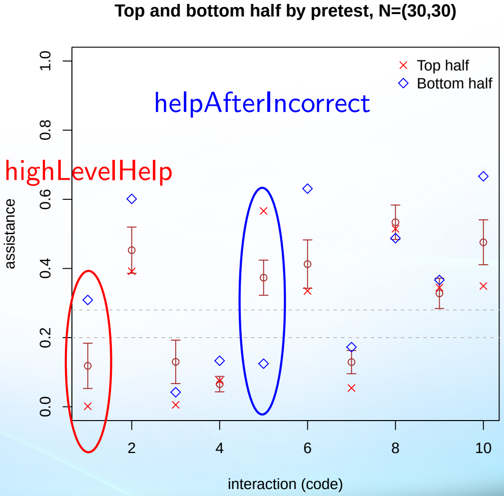
Confidential and Proprietary. Copyright © 2013 by Educational Testing Service. All rights reserved.

Results from Model 1 study - differential effects

pre-test scores were available for all tutees

assistance parameters were trained separately on the upper and lower half of the tutee population by pretest

two interactions flipped classes (hit, miss) in the case of the low-scoring tutees



Confidential and Proprietary. Copyright © 2013 by Educational Testing Service. All rights reserved.

Main findings from Model 1 study

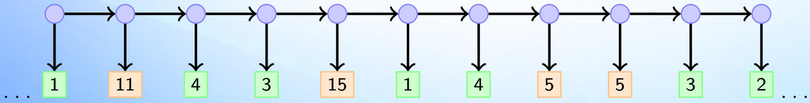
- Assistance values were estimated consistently, appeared to fall into two classes, i.e. hits and misses
- Found differential assistance of some tutor inputs (high level help, help after incorrect) for high and low pretest groups
- Model results revealed inaccuracies in automatic codings
- Drawbacks of the implementation
  - Missing data not missing at random
  - Not clear if immediate gains correlate with learning
  - No accounting for non-cognitive factors

Confidential and Proprietary. Copyright © 2013 by Educational Testing Service.
All rights reserved.

Model 2: classifying sequences for high/low learning gains

combine all tutor and tutee outputs into one output sequence

tutee codes	tutee's action	tutor's action	tutor code (plus affect)
INCORRECT	[divide rt]		
		u did that last step wrong,	wrongstep.feedback,
		u need to divide both sides by r+v	unelaborated.explanation,
CHAT	now cuz i didnt slove for t yet		
UNDO	[undo]		
		NO! listen to what i am saying.	Impoliteness, Rudeness
INCORRECT	[divide r+t]		
CHAT	i did now what do i have to do		
		its divide by r+t	unelaborated.explanation
		i mean r+v	unelaborated.explanation
UNDO	[undo]		
CORRECT	[divide r+v]		



Confidential and Proprietary. Copyright © 2013 by Educational Testing Service.
All rights reserved.

Model 2: The HMM classifier

$$\begin{tikzcd}
& \boxed{\lambda_h = \{\pi_h, A_h, B_h\}} \arrow[r] & \boxed{P(O|\lambda_h)} \arrow[dr] & \\
\boxed{\text{new sequence } O} \arrow[ur] \arrow[dr] & & & \boxed{\text{classification}} \\
& \boxed{\lambda_l = \{\pi_l, A_l, B_l\}} \arrow[r] & \boxed{P(O|\lambda_l)} \arrow[ur] &
\end{tikzcd}$$

Two training sets based on the tutee normalize gain from pre- to post-test, $g = (S_{post} - S_{pre})/(1 - S_{pre})$

high gain: $g > 10\%$ (20 dyads, $\bar{g} = 28\%$)

low gain: $g \leq 0\%$ (20 dyads, $\bar{g} = -3\%$)

Performance estimated using leave one out cross validation (slow).

ETS®

Confidential and Proprietary. Copyright © 2013 by Educational Testing Service. All rights reserved.

Listening. Learning. Leading.®

Performance comparison with logistic regression

Classifier	Predicted label	Actual high	Actual low
naive logistic	high	11	9
	low	9	11
best logistic	high	12	5
	low	8	15
HMM	high	16	5
	low	4	15

naive logistic model predicts class using all dyad aggregates as predictors (same observables as HMM)

best logistic model (step-wise AIC) uses only significant predictors

HMM classifier has 8 hidden states

ETS

Confidential and Proprietary. Copyright © 2013 by Educational Testing Service.
All rights reserved.

Listening. Learning. Leading.®

Making sense of the fitted parameters

Hidden state transition matrix (high-gain)

	1	2	3	4	5	6	7	8
1	0.39	0.01	0.16	0.06	0.00	0.00	0.09	0.30
2	0.03	0.06	0.14	0.19	0.07	0.14	0.34	0.01
3	0.00	0.07	0.44	0.01	0.11	0.32	0.05	0.00
4	0.10	0.21	0.06	0.08	0.00	0.00	0.53	0.02
5	0.01	0.04	0.01	0.00	0.85	0.04	0.00	0.04
6	0.03	0.05	0.52	0.00	0.06	0.01	0.02	0.31
7	0.04	0.11	0.05	0.68	0.00	0.04	0.02	0.06
8	0.26	0.00	0.09	0.01	0.00	0.00	0.06	0.58

Confidential and Proprietary. Copyright © 2013 by Educational Testing Service.
All rights reserved.

Making sense of the fitted parameters

Hidden state transition matrix (high-gain)

	1	2	3	4	5	6	7	8
1	0.39	0.01	0.16	0.06	0.00	0.00	0.09	0.30
2	0.03	0.06	0.14	0.19	0.07	0.14	0.34	0.01
3	0.00	0.07	0.44	0.01	0.11	0.32	0.05	0.00
4	0.10	0.21	0.06	0.08	0.00	0.00	0.53	0.02
5	0.01	0.04	0.01	0.00	0.85	0.04	0.00	0.04
6	0.03	0.05	0.52	0.00	0.06	0.01	0.02	0.31
7	0.04	0.11	0.05	0.68	0.00	0.04	0.02	0.06
8	0.26	0.00	0.09	0.01	0.00	0.00	0.06	0.58

Confidential and Proprietary. Copyright © 2013 by Educational Testing Service.
All rights reserved.

Listening. Learning. Leading.®

Making sense of the fitted parameters

Observation matrix (high-gain)

	1	2	3	4	5	6	7	8
incorrect	0.04	0.17	0.64	0.01	0.00	0.01	0.06	0.01
correct	0.43	0.04	0.05	0.01	0.00	0.02	0.06	0.65
undo	0.00	0.16	0.12	0.00	0.89	0.53	0.00	0.01
(tutee) chat	0.25	0.45	0.06	0.02	0.04	0.10	0.85	0.08
unelaborated.explanation	0.17	0.01	0.04	0.12	0.02	0.21	0.00	0.17
no codes (off topic)	0.05	0.13	0.01	0.40	0.02	0.03	0.00	0.01
Positivity	0.03	0.01	0.00	0.19	0.01	0.00	0.02	0.02
elaborated.hint	0.00	0.00	0.02	0.03	0.00	0.01	0.00	0.00
unelaborated.hint	0.00	0.00	0.02	0.04	0.01	0.07	0.00	0.01
Impoliteness	0.00	0.00	0.00	0.06	0.00	0.01	0.01	0.00
wrongstep.feedback	0.00	0.01	0.01	0.02	0.01	0.00	0.00	0.00
elaborated.explanation	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00
Impoliteness + Rudeness	0.01	0.00	0.01	0.02	0.00	0.00	0.00	0.00
Laughter	0.00	0.01	0.00	0.03	0.00	0.00	0.00	0.00
Laughter + Positivity	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.01
Rudeness	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01
unelab.expl + Positivity	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00

Confidential and Proprietary. Copyright © 2013 by Educational Testing Service.
All rights reserved.

Listening. Learning. Leading.®

Making sense of the fitted parameters

Observation matrix (high-gain)

	1	2	3	4	5	6	7	8
incorrect	0.04	0.17	0.64	0.01	0.00	0.01	0.06	0.01
correct	0.43	0.04	0.05	0.01	0.00	0.02	0.06	0.65
undo	0.00	0.16	0.12	0.00	0.89	0.53	0.00	0.01
(tutee) chat	0.25	0.45	0.06	0.02	0.04	0.10	0.85	0.08
unelaborated.explanation	0.17	0.01	0.04	0.12	0.02	0.21	0.00	0.17
no codes (off topic)	0.05	0.13	0.01	0.40	0.02	0.03	0.00	0.01
Positivity	0.03	0.01	0.00	0.19	0.01	0.00	0.02	0.02
elaborated.hint	0.00	0.00	0.02	0.03	0.00	0.01	0.00	0.00
unelaborated.hint	0.00	0.00	0.02	0.04	0.01	0.07	0.00	0.01
Impoliteness	0.00	0.00	0.00	0.06	0.00	0.01	0.01	0.00
wrongstep.feedback	0.00	0.01	0.01	0.02	0.01	0.00	0.00	0.00
elaborated.explanation	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00
Impoliteness + Rudeness	0.01	0.00	0.01	0.02	0.00	0.00	0.00	0.00
Laughter	0.00	0.01	0.00	0.03	0.00	0.00	0.00	0.00
Laughter + Positivity	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.01
Rudeness	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01
unelab.expl + Positivity	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00

Confidential and Proprietary. Copyright © 2013 by Educational Testing Service.
All rights reserved.

Summary and future work

Collaboration with Erin Walker (ASU) and Amy Ogan (CMU)

Two uses of DBNs in two quite different models

Model 1: an IOHMM to estimate assistance of tutor inputs, uses tutee "assisted capability" as (binary) hidden state and correctness as output state.

- Found differences between high and low pre-test groups
- Both the model and the data had flaws

Model 2: a discriminative HMM to classify high/low learning gain uses dyadic hidden state ($|S| = 8$) and 17 observed action categories including cognitive and affective labels.

- HMM classification accuracy (78%) exceeds best static logistic regression (68%)
- Dynamic (sequential) information has value!

ETS®

Confidential and Proprietary. Copyright © 2013 by Educational Testing Service.
All rights reserved.

Listening. Learning. Leading.®

Summary and future work

Still need more work to interpret HMMs in Model 2

Are there interactions between cognitive and affective factors?

54/60 dyads in study were friends. There is (small) evidence that dynamics are different for strangers.

A lot of "telling" is happening in these data. We might do better by incentivizing different behavior.

Need to collect more data!

Thank you!

ybergner@ets.org

ETS®

Confidential and Proprietary. Copyright © 2013 by Educational Testing Service. All rights reserved.

Listening. Learning. Leading.®