



Leveraging Hidden Dialogue State to Select Tutorial Moves

Kristy Elizabeth Boyer, Robert Phillips,
Eun Young Ha, Michael Wallis,
Mladen Vouk, & James Lester

Department of Computer Science
North Carolina State University

Introduction



Tutorial Dialogue Systems

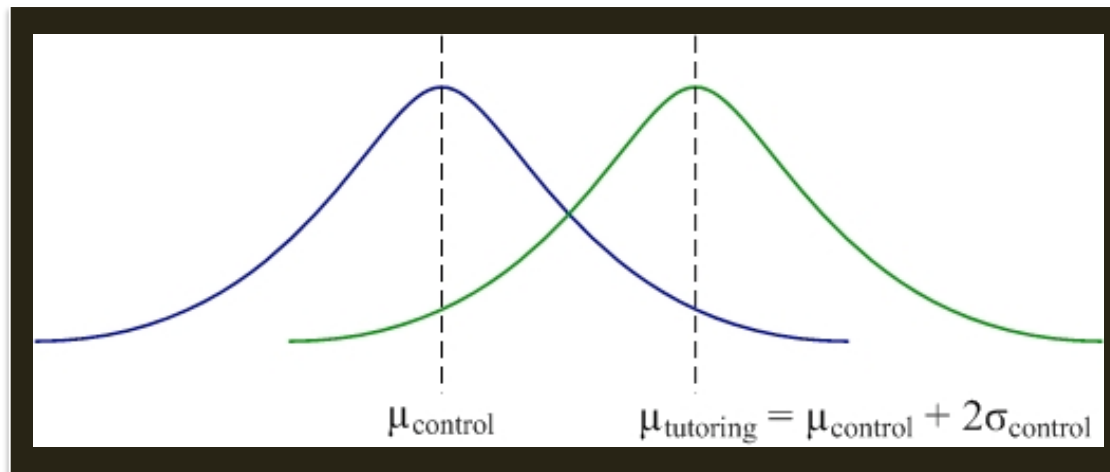


- Geometry (Alevan et al., 2004)
- Circulatory System (Evens & Michael, 2006)
- Physics (Graesser et al., 2005; Jordan et al., 2006)
- Data Structures (Fossati et al., 2008)
- And many more...

Challenge



Never achieved effectiveness equal to the most effective human tutors



(Bloom, 1984)

Reasons?



- Need more sophisticated natural language dialogue (Graesser et al., 1994)

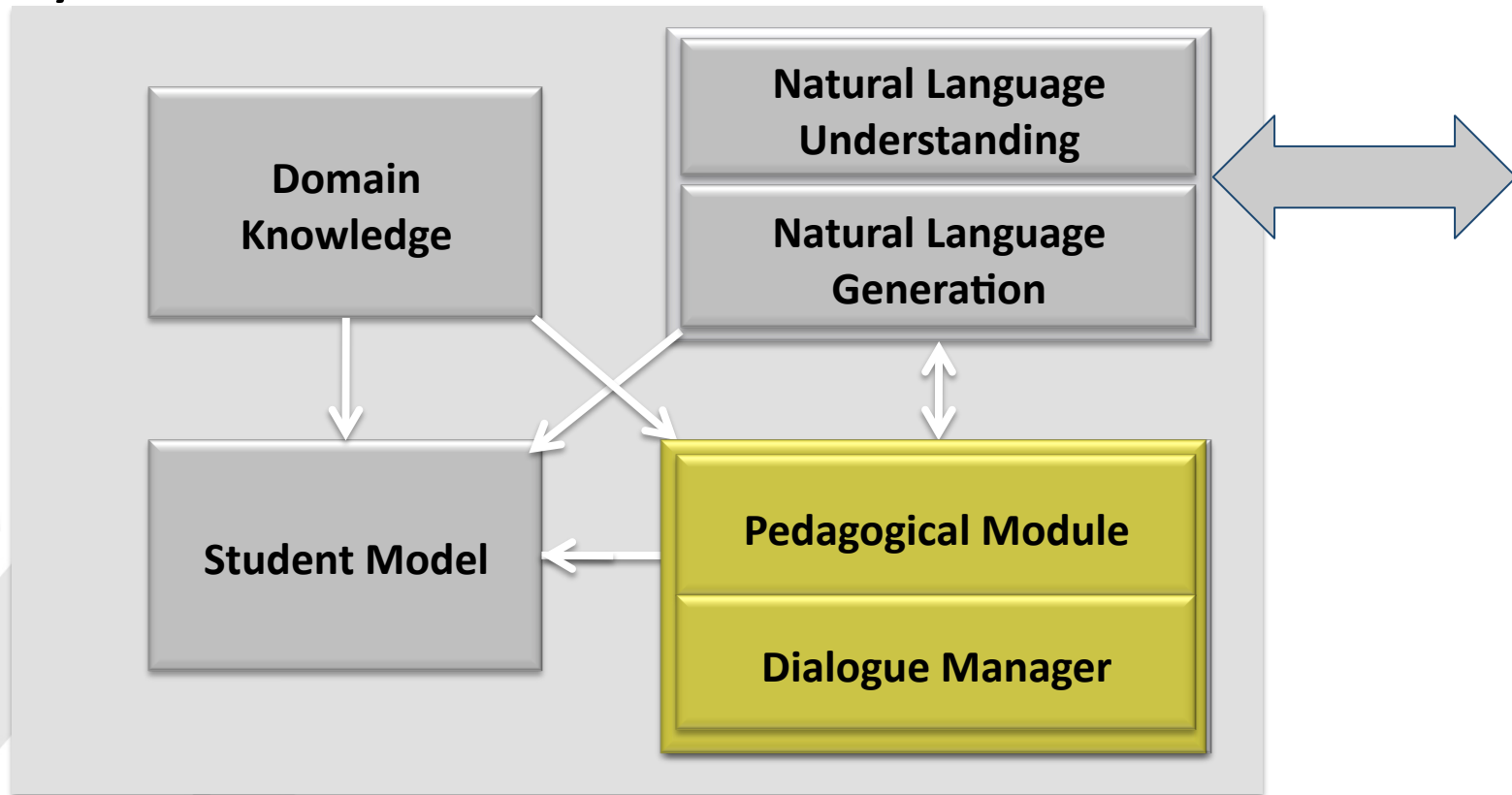
AND/OR

- Need to facilitate mastery learning (Van Lehn, 2008)

Possible Solution in Either Case



Data-driven authoring of tutorial dialogue system behavior



Hypothesized Benefits



Data-driven tutorial dialogue systems:

- Reduced development time
- Increased number of hours of instruction for students
- Flexible dialogue strategies
- Reflect approaches of effective human tutors

Goal of This Work



Learn

a data-driven tutorial dialogue strategy

as evidenced by

predicting human tutors' dialogue acts

within a corpus

Approach



Perform

sequence prediction

with models that leverage

hidden dialogue state

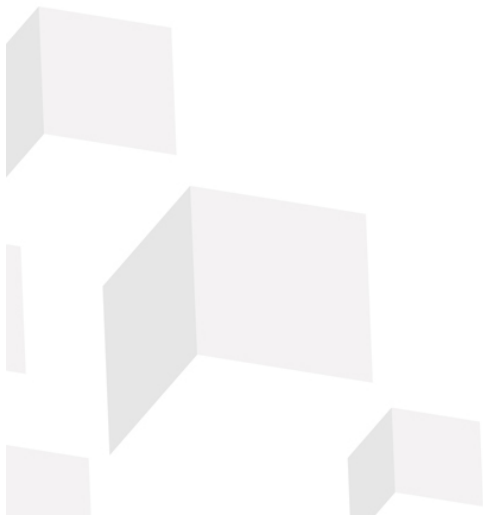


Related Work



Data-driven dialogue policy creation through:

- Reinforcement learning
- Direct corpus-based extraction



Reinforcement Learning



- Markov assumption (Levin et al. 2000)
- Reinforcement learning (Frampton & Lemon, 2009)
- Challenges include sparse data and large state spaces (Ai et al., 2007; Tetreault & Litman, 2008; Henderson et al., 2008; Heeman, 2007; Young et al., 2009)
- Comparing specific tutorial dialogue tactic choices (Chi et al., 2008)

Corpus-Based Extraction



Assumes that a good dialogue policy is realized in successful human-human dialogues

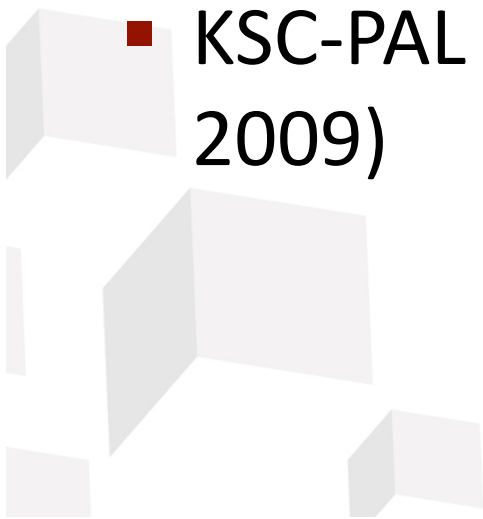
- Financial domain (Hardy et al., 2006)
- Catalogue ordering (Bangalore et al., 2008)
- Maptask conversational game (Poesio & Mikheev, 1998)

Strategies in Tutorial Dialogue



Data-driven exploration

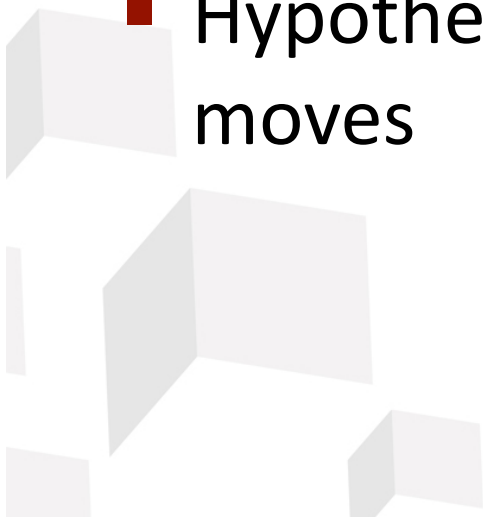
- CIRCSIM-TUTOR (Evens & Michael, 2006)
- ITSPOKE (Forbes-Riley, Rotaru, Litman, & Tetreault, 2007; Forbes-Riley & Litman, 2009)
- KSC-PAL (Kersey, Di Eugenio, Jordan, & Katz, 2009)



Hidden Dialogue State



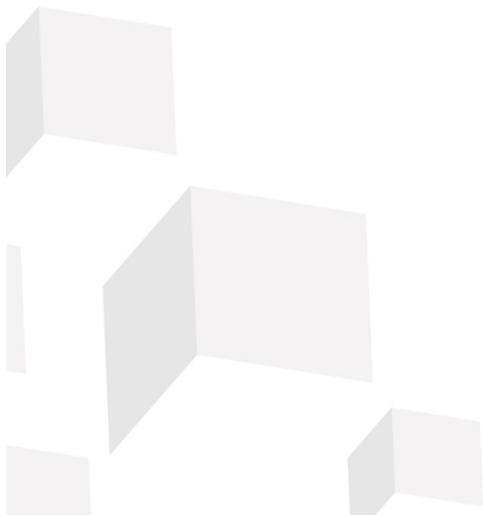
- Learned by hidden Markov models (Boyer et al., 2009a)
- Qualitative resemblance to tutorial dialogue strategies (Boyer et al., 2009b)
- Hypothesized to boost prediction of tutor moves



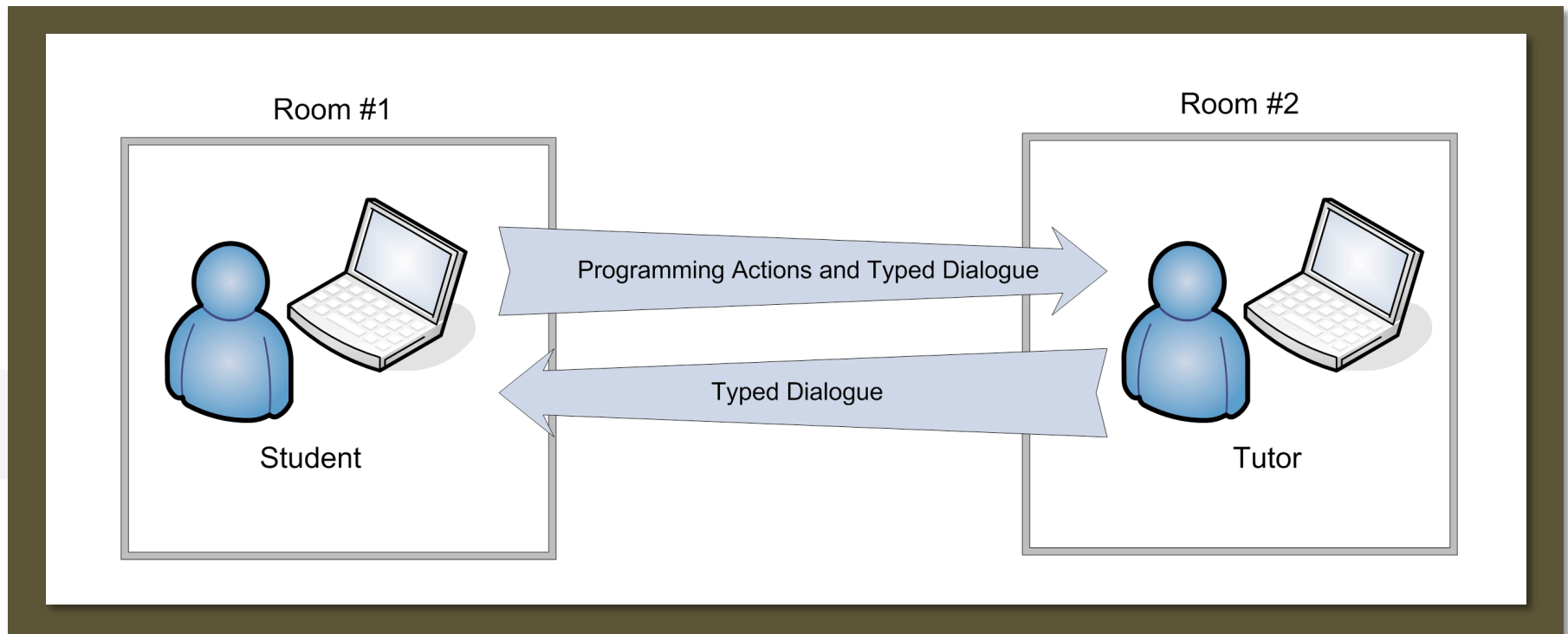
Corpus Study



- 48 Human-human tutoring dialogues
- Domain: introductory computer programming
- Task-oriented, separate, parallel task event stream



Corpus Collection



Remote Tutoring Environment



The screenshot displays the Eclipse IDE interface. The main editor window shows the following Java code:

```
71  /**
72   * Extract the individual digits stored in the ZIP code
73   * and store their values as private data
74   */
75  private void extractDigits() {
76
77      for(int i = 0; i < 5; i++)
78      {
79          digits[i] = (int)(zipCode / Math.pow(10, 5 - i))
80      }
81
82      barCode.clearCode();
83  }
84
85  /**
86   * Calculate correction digit and draw it
87   * (used by actionPerformed())
88   */
```

Overlaid on the left is a chat window titled "RIPPLE Chat - Student". The chat history includes:

- Tutor (Navigator 09:10:01): No, we've already declared it. We can just use it.
- Student (Driver 09:11:10): I'm sorry. I'm still stuck on that last loop we wrote because I feel like that's where my digits should come from.
- Tutor (Navigator 09:11:18): That's fine.
- Tutor (Navigator 09:11:25): That is where the digits are coming from.
- Tutor (Navigator 09:11:32): They're stored in the digits array.
- Tutor (Navigator 09:11:41): Does that make sense?
- Student (Driver 09:12:25): Sort of, but then how do I add them together? Is it something like $digits[1] + digits[2] + digits[3] + \dots$ or is there a different way to approach this since we're using an array?
- Student (Driver 09:12:43): That's exactly what we're going to be doing, but we can do it with

The chat window also shows "Nobody is typing." and a "Send Message" button at the bottom.

Corpus

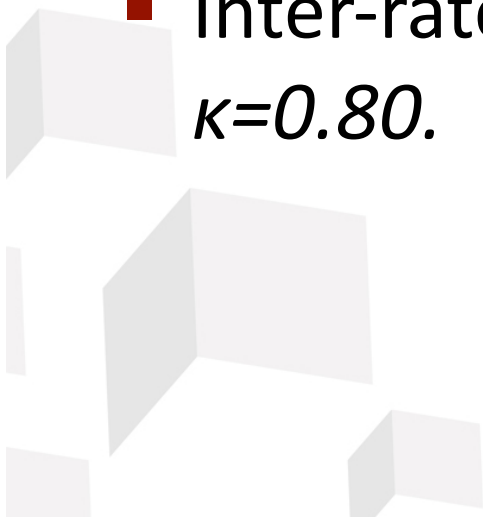


- 1,468 student utterances and 3,338 tutor utterances
- 3,793 semantic student task actions
- Significant learning gain from pretest to posttest (7% average, $p < 0.0001$)
- Annotated with dialogue act tags and task/subtask structure

Dialogue Act Annotation



- Scheme inspired by tags for
 - Conversational speech (Stolcke et al., 2000)
 - Task-oriented dialogue (Core & Allen, 1997)
 - Tutoring (Litman & Forbes-Riley, 2006)
- Inter-rater reliability on 10% of corpus was $\kappa=0.80$.



Dialogue Act Tags



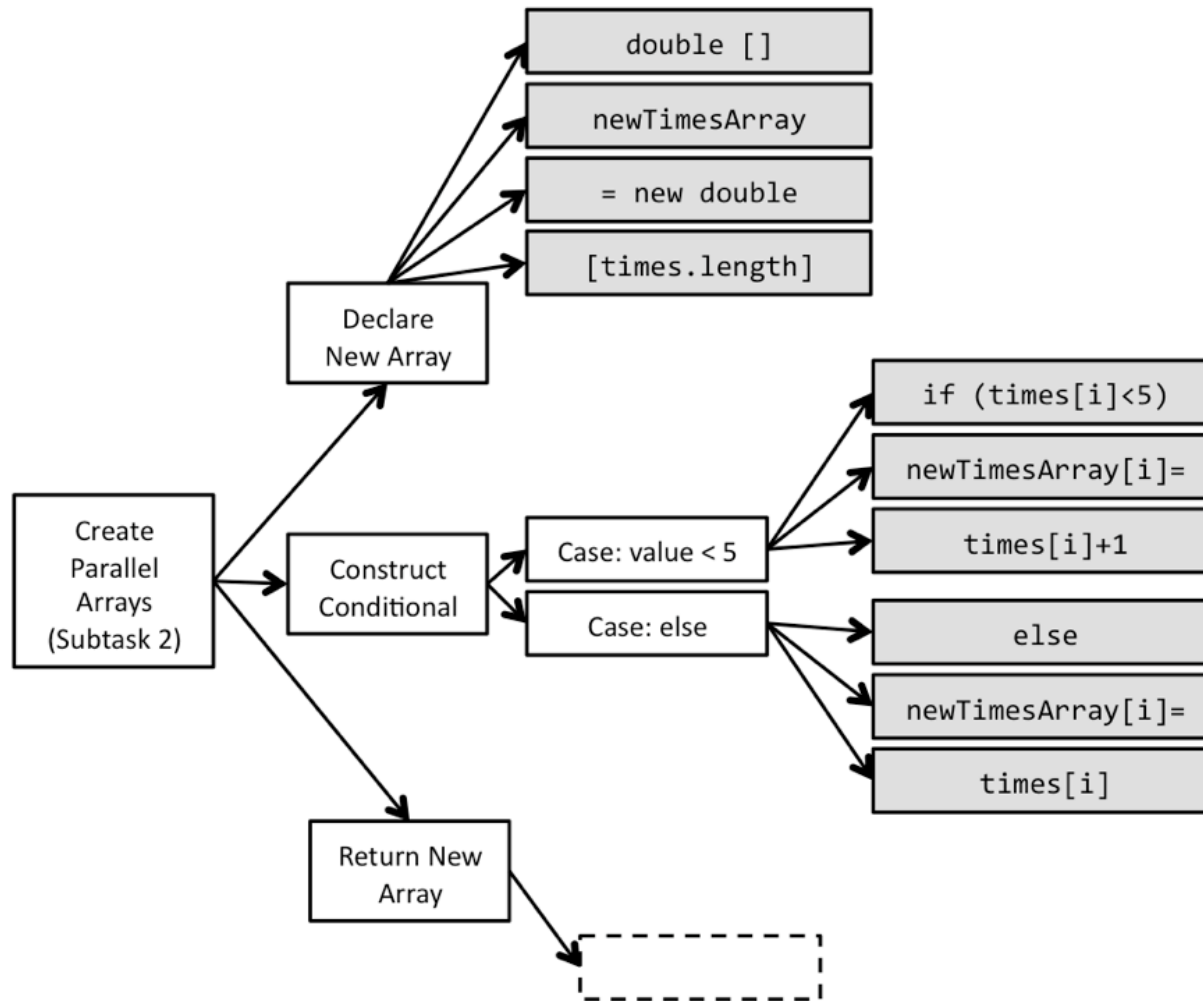
Dialogue Act Tag	Description
ASSESSING QUESTION (AQ)	Request for feedback on task or conceptual utterance.
EXTRA-DOMAIN (EX)	Asides not relevant to the tutoring task.
GROUNDING (G)	Acknowledgement/thanks.
LUKEWARM CONTENT FEEDBACK (LCF)	Negative assessment with explanation.
LUKEWARM FEEDBACK (LF)	Lukewarm 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.

Task Annotation



- Corpus includes 97,509 keystroke-level student task events
- Manually aggregated into 3,793 subtask actions
- Annotated with task/subtask structure
- Marked for correctness

Subtask Structure Annotation



Task Annotation Results



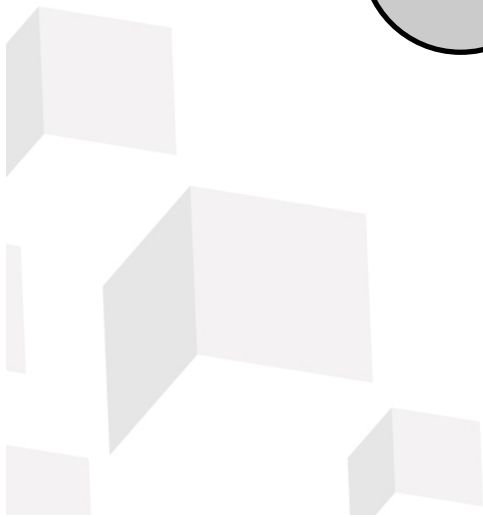
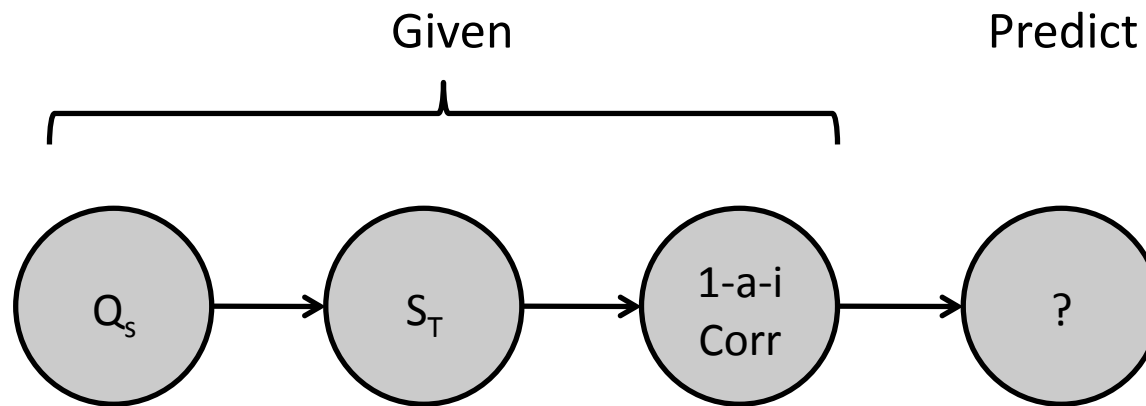
- Subtask annotation (66 leaves)
 - Inter-rater reliability study on 20% of corpus
 - Simple kappa = 0.58
 - Weighted kappa = 0.86
- Correctness annotation (4 tags)
 - CORRECT, BUGGY, INCOMPLETE, and DISPREFERRED
 - Simple kappa = 0.80

Tagged Excerpt



Time Stamp	Dialogue Stream	Task Stream	
2008-04-11 18:23:45	Student: so do i have to manipulate the array this time? [Q]		
2008-04-11 18:23:53	Tutor: this time, we need to do two things [S]		
2008-04-11 18:24:02	Tutor: first, we need to create a new array to hold the changed values [S]		
2008-04-11 18:24:28		i	} 1-a-l Buggy
2008-04-11 18:24:28		n	
2008-04-11 18:24:28		t	
2008-04-11 18:24:28		\sp	
2008-04-11 18:24:35		\del	
2008-04-11 18:24:36		\sp	} 1-a-i Correct
2008-04-11 18:24:36		d	
2008-04-11 18:24:36		o	
2008-04-11 18:24:36		u	
2008-04-11 18:24:36		b	
2008-04-11 18:24:37		l	
2008-04-11 18:24:37		e	
2008-04-11 18:24:37		\sp	
2008-04-11 18:24:39		[]	

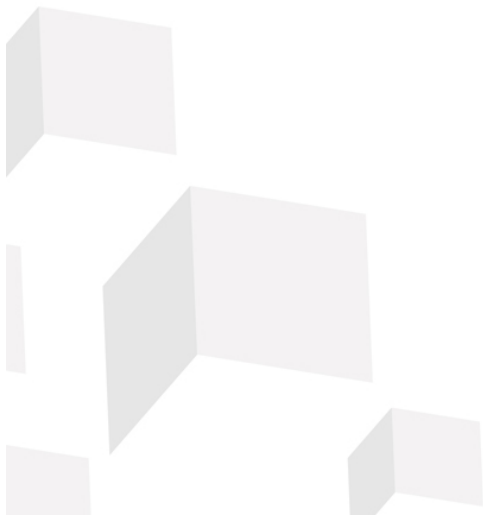
Sequence Modeling Task



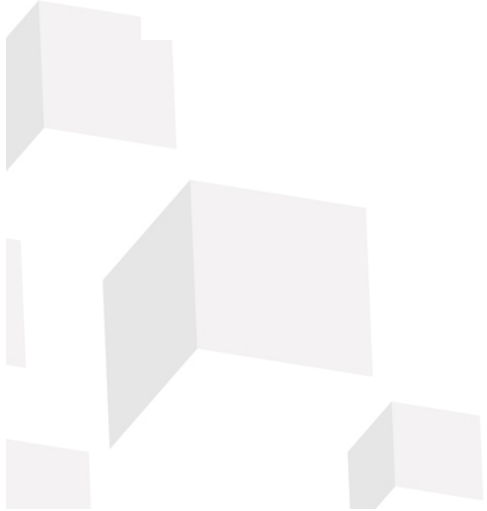
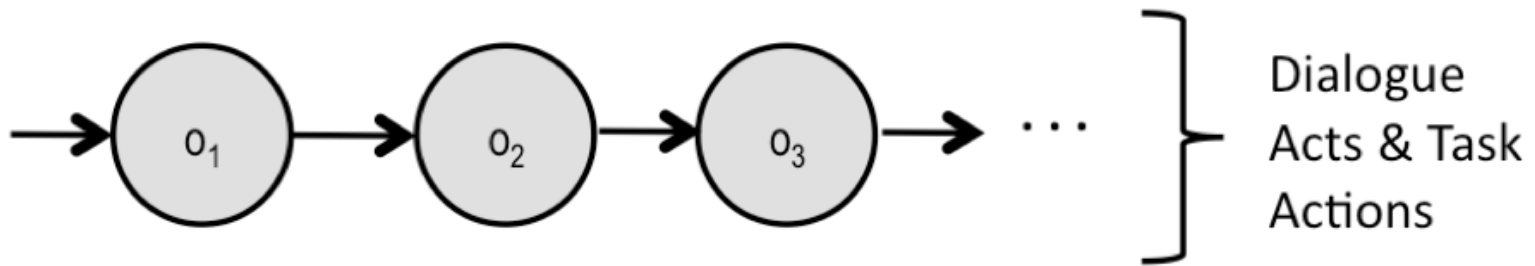
Three models



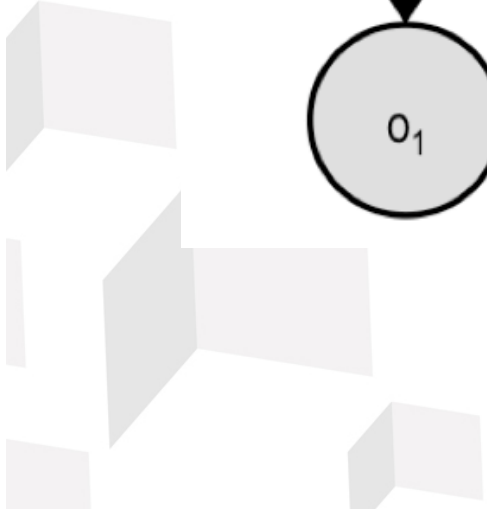
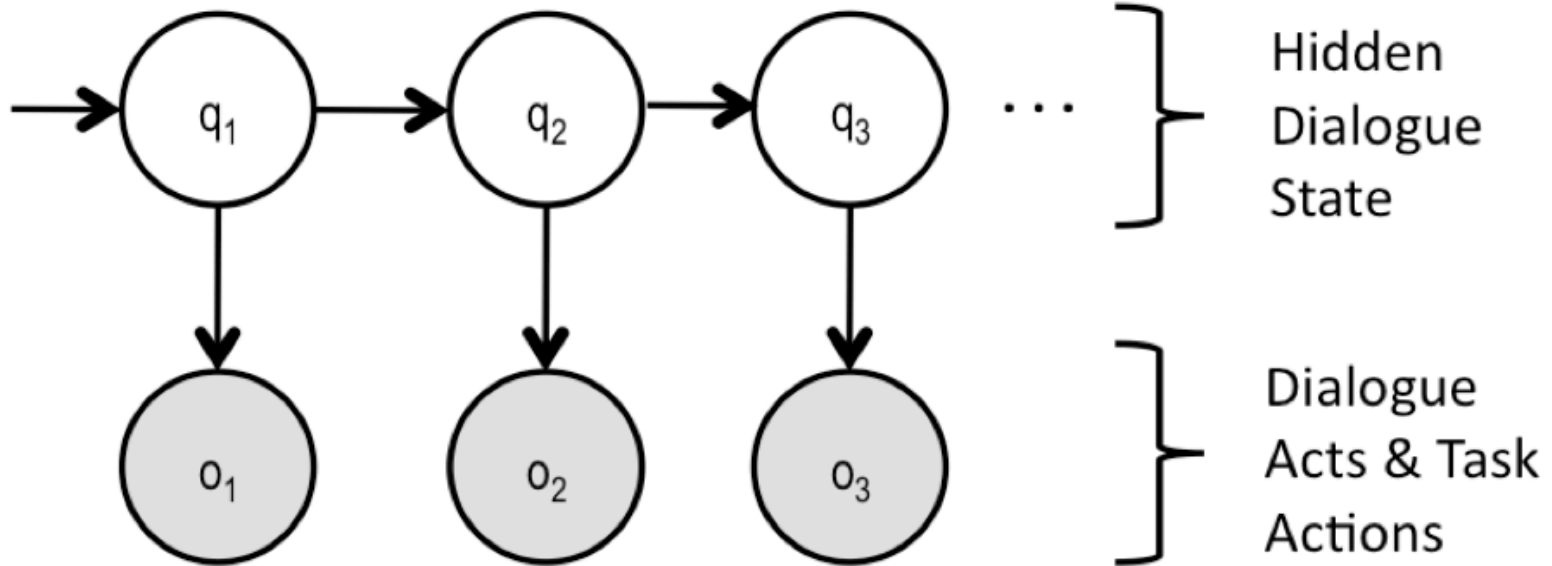
- First-order Markov models (MMs)
- Hidden Markov models (HMMs)
- Hierarchical Hidden Markov models (HHMMs)



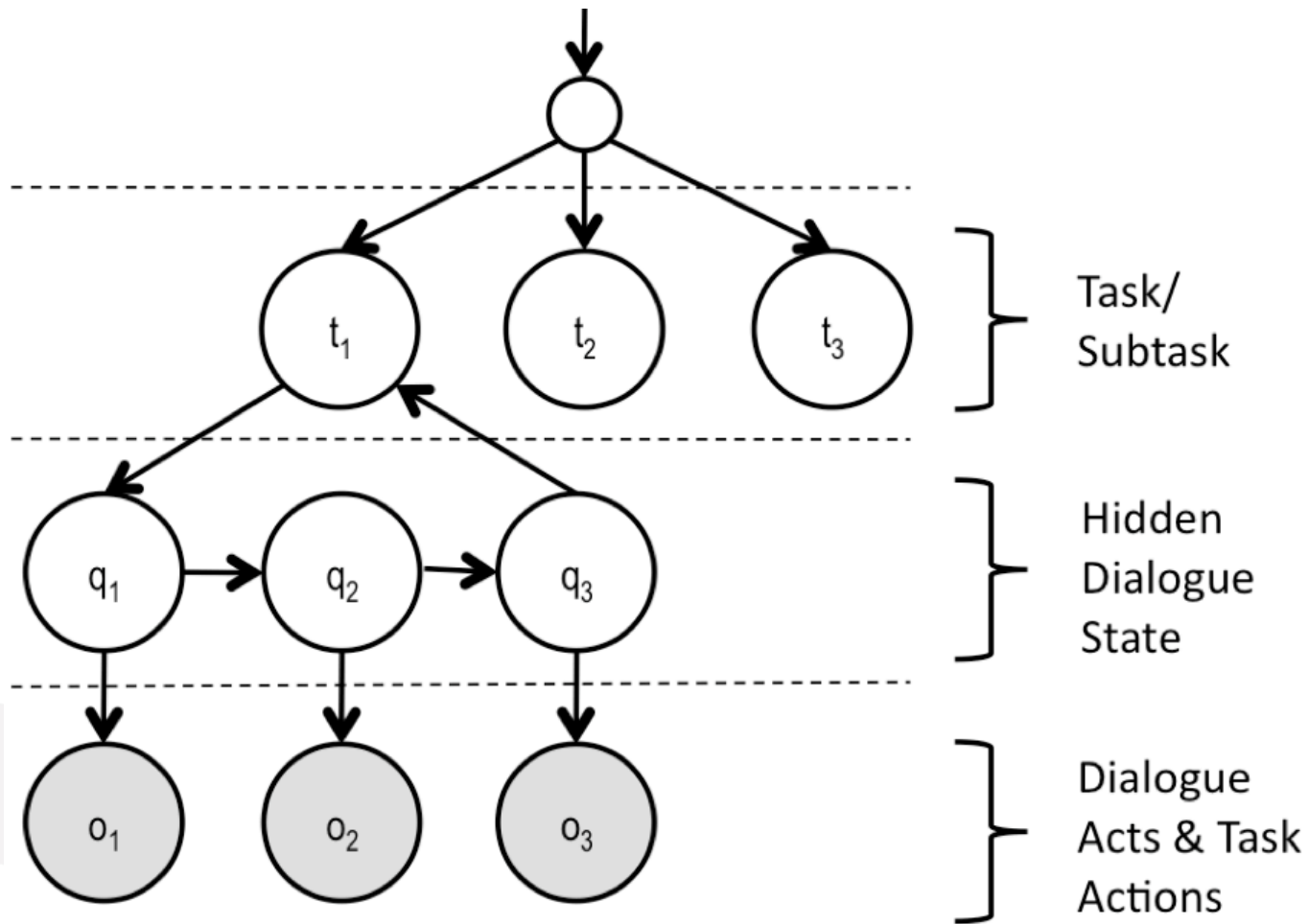
Markov model



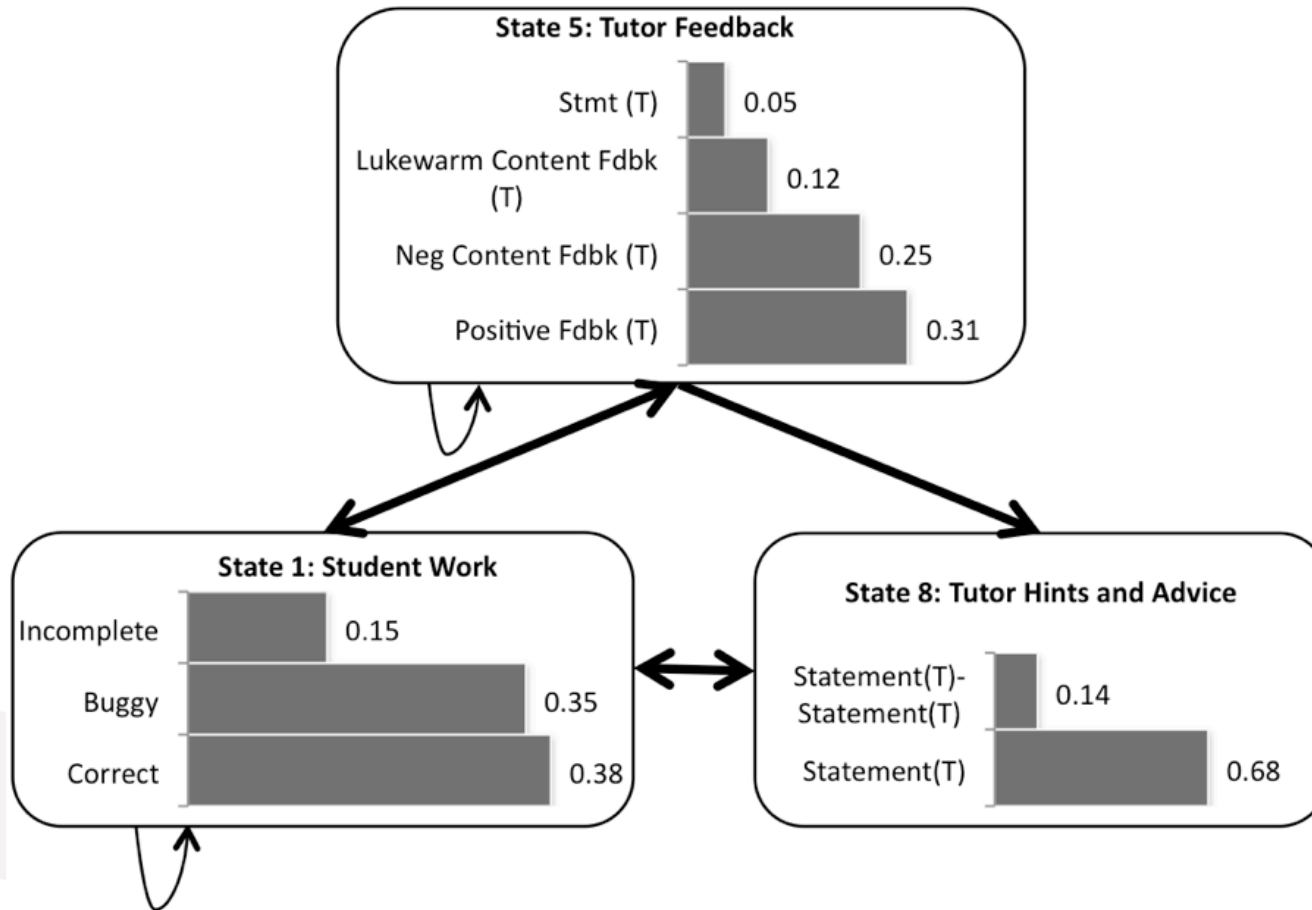
HMM



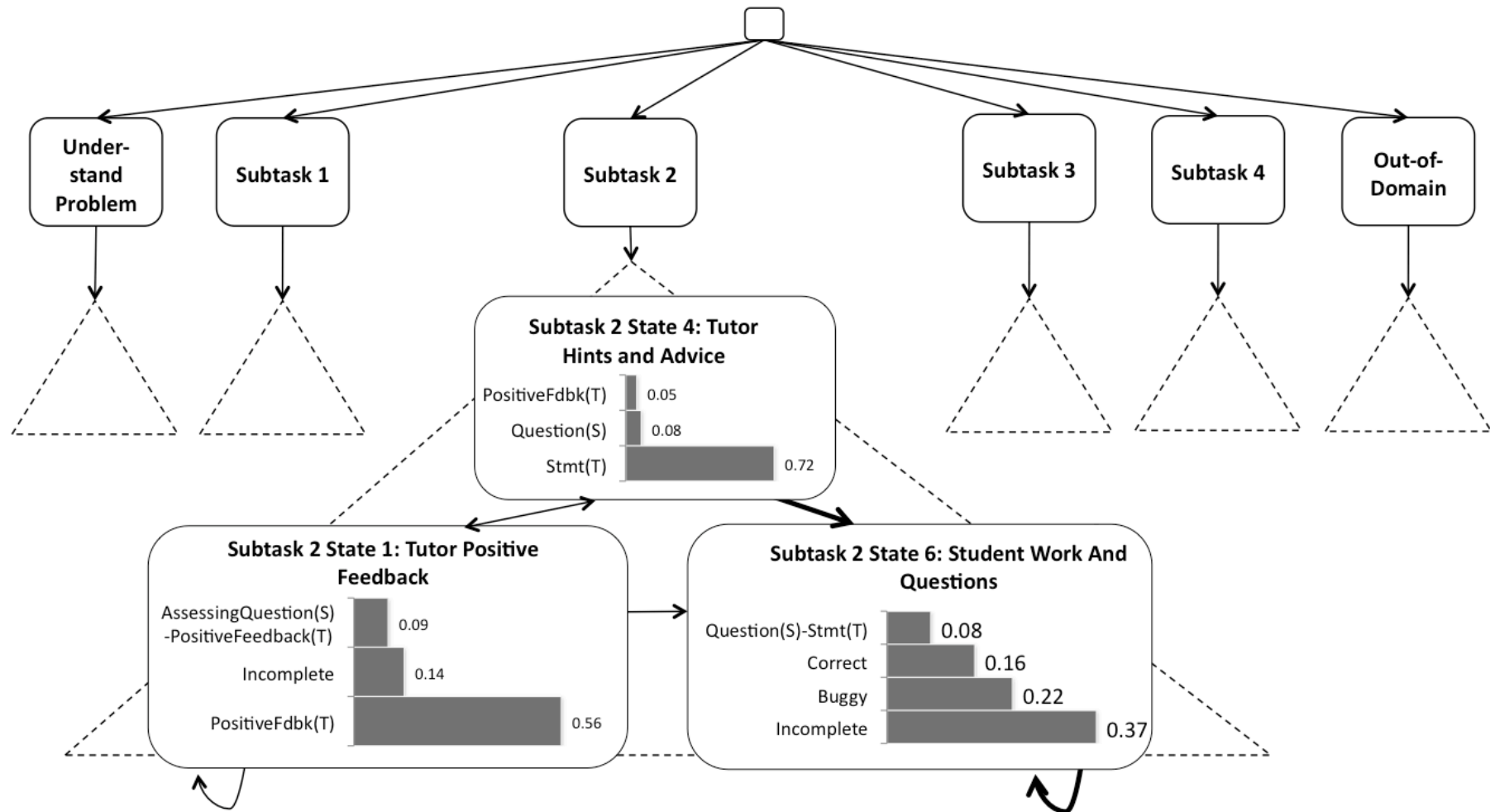
HHMM



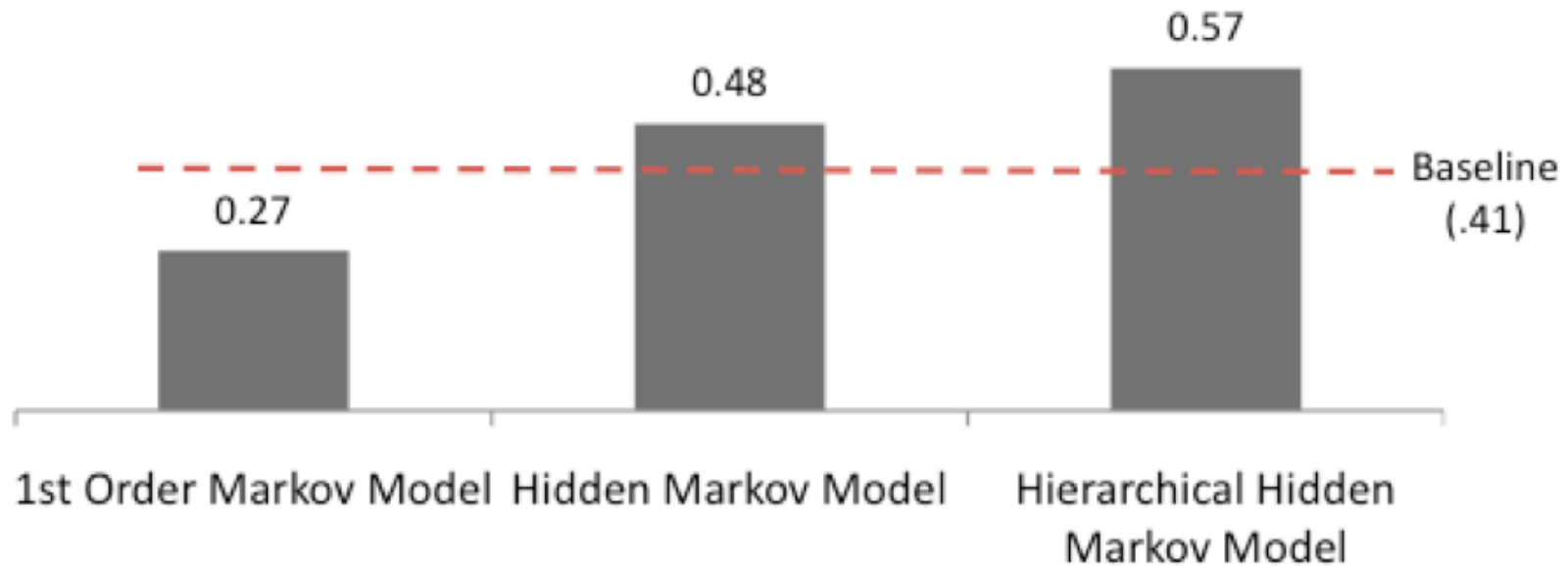
Portion of Learned HMM



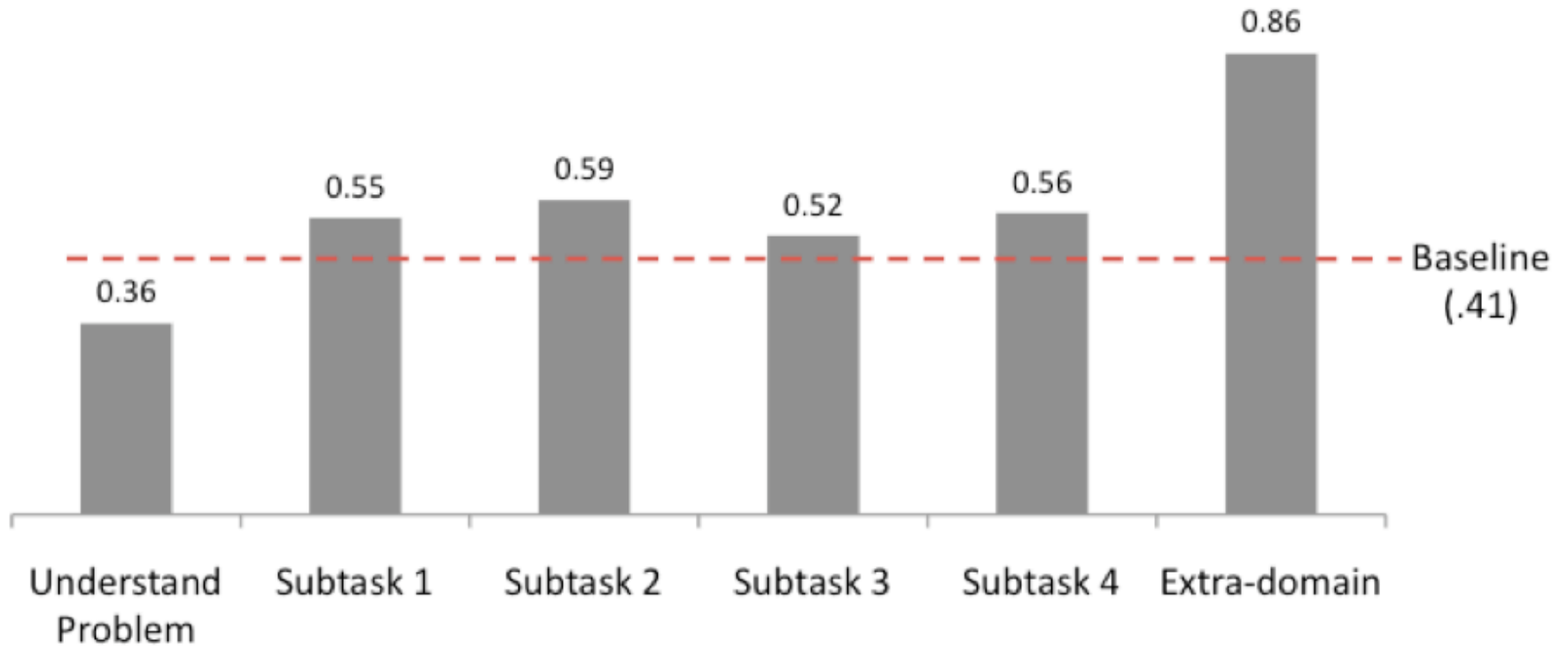
Portion of Learned HHMM



Prediction Results by Model



HHMM Results by Subtask



Results in Context



- Catalogue ordering domain
 - Flat model - 55% accuracy (Bangalore et al., 2008)
 - Hierarchical model - 35.6% accuracy (Bangalore & Stent, 2009)

- Our domain, tutoring for introductory computer programming
 - Flat model – 48% accuracy
 - Hierarchical model – 57% accuracy

Conclusions



- Hidden dialogue state useful; HMMs outperform MMs
- Task/subtask structure useful; HHMMs outperform HMMs
- *Understand Problem* phase very challenging to model
- *Extra-domain* conversation most straightforward to predict

Future Work



- Maintain multiple hypotheses for tutorial moves
- Leverage learner characteristics
- Develop fully unsupervised dialogue models
- Ultimate goal: create highly effective data-driven tutorial dialogue systems

A Shameless Plug



The Third Workshop on
Question Generation

QG 2010



www.questiongeneration.org/QG2010

June 18, 2010, Pittsburgh

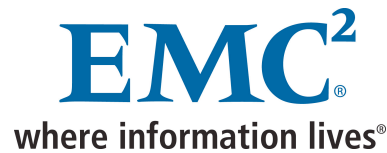
Acknowledgments



BlueCross BlueShield
of North Carolina



at&t





Contact: keboyer@ncsu.edu

