Modeling Dialogue Structure with Adjacency Pair Analysis and Hidden Markov Models

Kristy Elizabeth Boyer[*1] **Robert Phillips**[1,2] **Eun Young Ha**[1] **Michael D. Wallis**[1,2] **Mladen A. Vouk**[1] **James C. Lester**[1]

[1]Department of Computer Science
North Carolina State University
Raleigh, NC, USA

[2]Applied Research Associates
Raleigh, NC, USA

[*]`keboyer@ncsu.edu`

Abstract

Automatically detecting dialogue structure within corpora of human-human dialogue is the subject of increasing attention. In the domain of tutorial dialogue, automatic discovery of dialogue structure is of particular interest because these structures inherently represent tutorial *strategies* or *modes*, the study of which is key to the design of intelligent tutoring systems that communicate with learners through natural language. We propose a methodology in which a corpus of human-human tutorial dialogue is first manually annotated with dialogue acts. Dependent adjacency pairs of these acts are then identified through χ^2 analysis, and hidden Markov modeling is applied to the observed sequences to induce a descriptive model of the dialogue structure.

1 Introduction

Automatically learning dialogue structure from corpora is an active area of research driven by a recognition of the value offered by data-driven approaches (*e.g.,* Bangalore *et al.*, 2006). Dialogue structure information is of particular importance when the interaction is centered around a learning task, such as in natural language tutoring, because techniques that support empirical identification of dialogue strategies can inform not only the design of intelligent tutoring systems (Forbes-Riley *et al.*, 2007), but also contribute to our understanding of the cognitive and affective processes involved in learning through tutoring (VanLehn *et al.*, 2007).

Although traditional top-down approaches (*e.g.,* Cade *et al.*, 2008) and some empirical work on analyzing the structure of tutorial dialogue (Forbes-Riley *et al.*, 2007) have yielded significant results, the field is limited by the lack of an automatic, data-driven approach to identifying dialogue structure. An empirical approach to identifying tutorial dialogue strategies, or *modes*, could address this limitation by providing a mechanism for describing in succinct probabilistic terms the tutorial strategies that actually occur in a corpus.

Just as early work on dialogue act interpretation utilized hidden Markov models (HMMs) to capture linguistic structure (Stolcke *et al.*, 2000), we propose a system that uses HMMs to capture the structure of tutorial dialogue implicit within sequences of already-tagged dialogue acts. This approach operates on the premise that at any given point in the tutorial dialogue, the collaborative interaction is in a dialogue *mode* that characterizes the nature of the exchanges between tutor and student. In our model, a dialogue mode is defined by a probability distribution over the observed symbols (*e.g.*, dialogue acts and adjacency pairs).

Our previous work has noted some limitations of first-order HMMs as applied to sequences of individual dialogue acts (Boyer *et al.*, in press). Chief among these is that HMMs allow arbitrarily frequent transitions between hidden states, which does not conform well to human intuition about how tutoring strategies are applied. Training an HMM on a sequence of adjacency pairs rather than

individual dialogue acts is one way to generate a more descriptive model without increasing model complexity more than is required to accommodate the expanded set of observation symbols. To this end, we apply the approach of Midgley *et al.* (2006) for empirically identifying significant adjacency pairs within dialogue, and proceed by treating adjacency pairs as atomic units for the purposes of training the HMM.

2 Corpus Analysis

This analysis uses a corpus of human-human tutorial dialogue collected in the domain of introductory computer science. Forty-three learners interacted remotely with a tutor through a keyboard-to-keyboard remote learning environment yielding 4,864 dialogue moves.

The tutoring corpus was manually tagged with dialogue acts designed to capture the salient characteristics of the tutoring process (Table 1).

Tag	Act	Example
Q	Question	*Where should I declare i?*
EQ	Evaluation Question	*How does that look?*
S	Statement	*You need a closing brace.*
G	Grounding	*Ok.*
EX	Extra-Domain	*You may use your book.*
PF	Positive Feedback	*Yes, that's right.*
LF	Lukewarm Feedback	*Sort of.*
NF	Negative Feedback	*No, that's not right.*

Table 1. Dialogue Act Tags

The correspondence between utterances and dialogue act tags is one-to-one. Compound utterances (*i.e.*, a single utterance comprising more than one dialogue act) were split by the primary annotator prior to the inter-rater reliability study.[1]

The importance of adjacency pairs is well-established in natural language dialogue (*e.g.*, Schlegoff & Sacks, 1973), and adjacency pair analysis has illuminated important phenomena in tutoring as well (Forbes-Riley *et al.*, 2007). For the current corpus, bigram analysis of dialogue acts yielded a set of commonly-occurring pairs. How-

---

[1] Details of the study procedure used to collect the corpus, as well as Kappa statistics for inter-rater reliability, are reported in (Boyer *et al.*, 2008).

ever, as noted in (Midgley *et al.,* 2006), in order to establish that two dialogue acts are truly related as an adjacency pair, it is important to determine whether the presence of the first member of the pair is associated with a significantly higher probability of the second member occurring. For this analysis we utilize a χ^2 test for independence of the categorical variables act_i and act_{i+1} for all two-way combinations of dialogue act tags. Only pairs in which speaker(act_i)≠speaker(act_{i+1}) were considered. Other dialogue acts were treated as atomic elements in subsequent analysis, as discussed in Section 3. Table 2 displays a list of the dependent pairs sorted by descending (unadjusted) statistical significance; the subscript indicates tutor (t) or student (s).

act_i	act_{i+1}	$P(act_{i+1}\|act_i)$	$P(act_{i+1}\|\neg act_i)$	χ^2 val	*p*-val
EQ_s	PF_t	0.48	0.07	654	<0.0001
G_s	G_t	0.27	0.03	380	<0.0001
EX_s	EX_t	0.34	0.03	378	<0.0001
EQ_t	PF_s	0.18	0.01	322	<0.0001
EQ_t	S_s	0.24	0.03	289	<0.0001
EQ_s	LF_t	0.13	0.01	265	<0.0001
Q_t	S_s	0.65	0.04	235	<0.0001
EQ_t	LF_s	0.07	0.00	219	<0.0001
Q_s	S_t	0.82	0.38	210	<0.0001
EQ_s	NF_t	0.08	0.01	207	<0.0001
EX_t	EX_s	0.19	0.02	177	<0.0001
NF_s	G_t	0.29	0.03	172	<0.0001
EQ_t	NF_s	0.11	0.01	133	<0.0001
S_s	G_t	0.16	0.03	95	<0.0001
S_s	PF_t	0.30	0.10	90	<0.0001
S_t	G_s	0.07	0.04	36	<0.0001
PF_s	G_t	0.14	0.04	34	<0.0001
LF_s	G_t	0.22	0.04	30	<0.0001
S_t	EQ_s	0.11	0.07	29	<0.0001
G_t	EX_s	0.07	0.03	14	0.002
S_t	Q_s	0.07	0.05	14	0.0002
G_t	G_s	0.10	0.05	9	0.0027
EQ_t	EQ_s	0.13	0.08	8	0.0042

Table 2. Dependent Adjacency Pairs

3 HMM on Adjacency Pair Sequences

The keyboard-to-keyboard tutorial interaction resulted in a sequence of utterances that were annotated with dialogue acts. We have hypothesized that a higher-level dialogue structure, namely the tutorial dialogue *mode*, overlays the observed dia-

logue acts. To build an HMM model of this structure we treat dialogue mode as a hidden variable and train a hidden Markov model to induce the dialogue modes and their associated dialogue act emission probability distributions.

An adjacency pair joining algorithm (Figure 1) was applied to each sequence of dialogue acts. This algorithm joins pairs of dialogue acts into atomic units according to a priority determined by the strength of the adjacency pair dependency.

Sort adjacency pair list L by descending statistical significance
For each adjacency pair (act_1, act_2) in L
    For each dialogue act sequence $(a_1, a_2, ..., a_n)$ in the corpus
        Replace all pairs $(a_i=act_1, a_{i+1}=act_2)$ with a new single act (act_1act_2)

Figure 1. Adjacency Pair Joining Algorithm

Figure 2 illustrates the application of the adjacency pair joining algorithm on a sequence of dialogue acts. Any dialogue acts that were not grouped into adjacency pairs at the completion of the algorithm are treated as atomic units in the HMM analysis.

Original Dialogue Act Sequence:

Q_s - S_t - LF_t - S_t - S_t - G_s - EQ_s - LF_t - S_t - S_t - Q_s - S_t

After Adjacency Pair Joining Algorithm:

$\mathbf{Q_sS_t}$ - LF_t - S_t - $\mathbf{S_tG_s}$ - $\mathbf{EQ_sLF_t}$ - S_t - S_t - $\mathbf{Q_sS_t}$

Figure 2. DA Sequence Before/After Joining

The final set of observed symbols consists of 39 tags: 23 adjacency pairs (Table 2) plus all individual dialogue acts augmented with a tag for the speaker (Table 1).

It was desirable to learn n, the best number of hidden states, during modeling rather than specifying this value *a priori*. To this end, we trained and ten-fold cross-validated seven models (each featuring randomly-initialized parameters) for each number of hidden states n from 2 to 15, inclusive.[2] The average log-likelihood was computed across

all seven models for each n, and this average log-likelihood l_n was used to compute the Akaike Information Criterion, a maximum-penalized likelihood estimator that penalizes more complex models (Scott, 2002). The best fit was obtained with $n=4$ (Figure 3). The transition probability distribution among hidden states is depicted in Figure 4, with the size of the nodes indicating relative frequency of each hidden state; specifically, State 0 accounts for 63% of the corpus, States 1 and 3 account for approximately 15% each, and State 2 accounts for 7%.
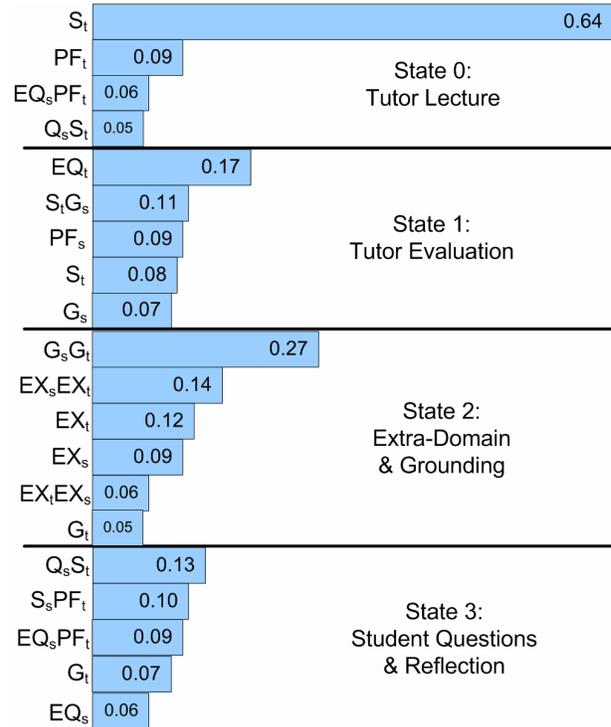

Figure 3. Dialogue Act Emission Probability Distribution by Dialogue Mode[3]

4 Discussion and Future Work

This exploratory application of hidden Markov models involves training an HMM on a mixed input sequence consisting of both individual dialogue acts and adjacency pairs. The best-fit HMM consists of four hidden states whose emission symbol probability distributions lend themselves to interpretation as tutorial dialogue modes. For example, State 0 consists primarily of tutor statements and positive feedback, two of the most common dialogue acts in our corpus. The transition probabili-
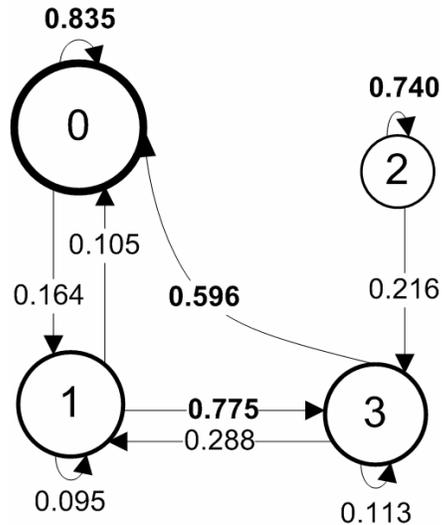
---

[2] $n=15$ was chosen as an initial maximum number of states because it comfortably exceeded our hypothesized range of 3 to 7 (informed by the tutoring literature). The Akaike Information Criterion measure steadily worsened above $n = 5$, confirming no need to train models with $n > 15$.

Figure 4. Transition Probability Distribution[4]

ties also reveal that State 0 is highly stable; a self-transition is most likely with probability 0.835. State 3 is an interactive state featuring student reflection in the form of questions, statements, and requests for feedback. The transition probabilities show that nearly 60% of the time the dialogue transitions from State 3 to State 0; this may indicate that after establishing what the student does or does not know in State 3, the tutoring switches to a less collaborative "teaching" mode represented by State 0.

Future evaluation of the HMM presented here will include comparison with other types of graphical models. Another important step is to correlate the dialogue profile of each tutoring session, as revealed by the HMM, to learning and affective outcomes of the tutoring session. This type of inquiry can lead directly to design recommendations for tutorial dialogue systems that aim to maximize particular learner outcomes. In addition, leveraging knowledge of the task state as well as surface-level utterance content below the dialogue act level are promising directions for refining the descriptive and predictive power of these models.

Acknowledgements

References

Boyer, K.E., Phillips, R., Wallis, M., Vouk, M., & Lester, J. (2008). Balancing cognitive and motivational scaffolding in tutorial dialogue. *Proceedings of the 9th International Conference on Intelligent Tutoring Systems,* Montreal, Canada, 239-249.

Boyer, K.E., Ha, E.Y., Wallis, M., Phillips, R., Vouk, M. & Lester, J. (in press). Discovering tutorial dialogue strategies with hidden Markov models. To appear in *Proceedings of the 14th International Conference on Artificial Intelligence in Education,* Brighton, U.K.

Bangalore, S., DiFabbrizio, G., Stent, A. (2006). Learning the structure of task-driven human-human dialogs. *Proceedings of ACL*, Sydney, Australia, 201-208.

Cade, W., Copeland, J., Person, N., & D'Mello, S. (2008). Dialog modes in expert tutoring. *Proceedings of the 9th International Conference on Intelligent Tutoring Systems,* Montreal, Canada, 470-479.

Forbes-Riley, K., Rotaru, M., Litman, D. J., & Tetreault, J. (2007). Exploring affect-context dependencies for adaptive system development. *Proceedings of NAACL HLT, Companion Volume,* 41-44.

Midgley, T. D., Harrison, S., & MacNish, C. (2007). Empirical verification of adjacency pairs using dialogue segmentation. *Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue,* 104-108.

Schlegoff, E.A., Sacks, H. (1973). Opening up closings. *Semiotica*, *8*(4), 289-327.

Scott, S. L. (2002). Bayesian methods for hidden Markov models: Recursive computing in the 21st century. *Journal of the American Statistical Association, 97*(457), 337-351.

Stolcke, A., Coccaro, N., Bates, R., Taylor, P., Van Ess-Dykema, C., Ries, K., Shirberg, E., Jurafsky, D., Martin, R., Meteer, M. (2000). Dialog act modeling for automatic tagging and recognition of conversational speech. *Computational Linguistics 26*(3), 339-373.

VanLehn, K., Graesser, A., Jackson, G.T., Jordan, P., Olney, A., Rose, C.P. (2007). When are tutorial dialogues more effective than reading? *Cognitive Science, 31*(1), 3-62.