Tutor Dialogue Planning with Contextual Information and Discourse Structure

Robert Fisher and Reid Simmons

Carnegie Melon University
5000 Forbes Ave
Pittsburgh PA,15213

Abstract. In this paper, we present two techniques to improve the effectiveness of a conversational tutoring system when interacting with groups of users. First, we propose the usage of linguistic features derived from *discourse parsing* to better understand the structure of the ongoing dialogue between the students and the tutoring system. Discourse parsing gives us the relational structure of a dialogue, and can be used to improve the tutoring system's understanding of conversation structure. Second, we discuss dialogue planning using contextual information stored in a Contextual Knowledge Base (CKB). A conversational agent with access to a information about linguistic and environmental context will be able to respond to changing conditions in ways that a simple question answering agent is not capable of. We present previous empirical results for these two tasks and discuss how they can be incorporated into an Intelligent Tutoring System.

Keywords: Intelligent Tutoring Systems, Machine Learning, Natural Language Processing, Discourse Parsing, Spectral Learning

1 Introduction

Modern conversational intelligent systems have become quite adept at question answering tasks, but are much less effective when faced with complex, structured dialogue and debate that can arise when interacting with groups of users. In a group setting, students may reference or follow up on statements made by other users of the system. As such, it is paramount to understand the structure of the conversation, particularly the relationship between different user utterances. A tutoring system instructing a group of students should also be able to react to other contextual factors, such as the number of students, their relative aptitude for the material, and their proclivity towards active participation.

In this work, we will focus on two keys areas that we believe can improve social intelligent tutors, and conversational agents in general. First, we propose the usage of linguistic features derived from *discourse parsing* to better understand the structure of a dialogue with a group of users. Discourse parsing gives us the inter-sentential structure of a whole dialogue, and helps us understand the relationship between utterances given by the users. To utilize the dialogue structure, and other contextual information when planning a sequence of interactions with the students, the tutor require a rich, expressive state space representation. Therefore, we propose the use of contextual features stored in a *Contextual Knowledge Base* (CKB) [9]. A Contextual Knowledge Base represents information in logical form by composing predicates, relations, and concepts. Unlike a traditional knowledge base such as NELL [3] and FreeBase [1], a Contextual Knowledge Base includes information relative to the agent's environment and conversational partners. The contextual knowledge base can be populated using information gained through discourse parsing, as well as through semantic parsing, and environmental sensor processing algorithms. By planning the agent's dialogue using the CKB as input, the system is better able to demonstrate environmental and linguistic contextual awareness, as well as personalized behavior for the current set of users. If we were to use a knowledge tracing algorithm [4] to estimate the users' aptitude towards the material being taught, the knowledge component estimates could also be stored in the CKB.

2 Spectral Discourse Structure

Discourse parsing is a natural language parsing task in which logical relations are discovered between pairs of text units. Unlike sentence level syntactic or semantic parsing, discourse parsing discovers inter-sentential structure within a larger discourse or piece of text. Examples include causation, attribution, and paraphrase relations. For instance, the following text contains two sentences of texts coupled by an elaboration relation.

(1) "Lactose is milk sugar. The enzyme lactase breaks it down."

In the above example, the relation is *implicit* and not instantiated by a coordinating connective word such as *and* or *because*. Discovering implicit relations in text is a difficult task that requires the use of syntactic and linguistic features, as well as entity recognition capability, for instance recognizing that the word *it* in example (1) refers to lactose.

A common framework for discourse parsing is Rhetorical Structure Theory (RST), an example of which is shown in Figure 1. RST defines tree-like structures of discourse relations over atomic discourse units. While RST provides a solid foundation, a different set of discourse relations may be appropriate for the tutoring domain. In particular, we may wish to define relations such as agreement, clarification, elaboration, and rebuttal. An example of how discourse parsing might be applied to a tutoring dialogue is shown in Figure 2.

While state of the art discourse parsing systems incorporate the necessary linguistic features to make relation learning tractable, they often rely on relatively simplistic classifiers such as linear support vector machines and maximum entropy classifiers [7, 12]. To address the shortcomings of current discourse parsing systems, we have developed a discourse parser that utilizes a spectral, latent variable model to efficiently and optimally estimate the parameters of a hidden state classifier similar to an HMM. Spectral methods, the underlying optimization procedure of our framework, are a class of algorithms that utilize matrix

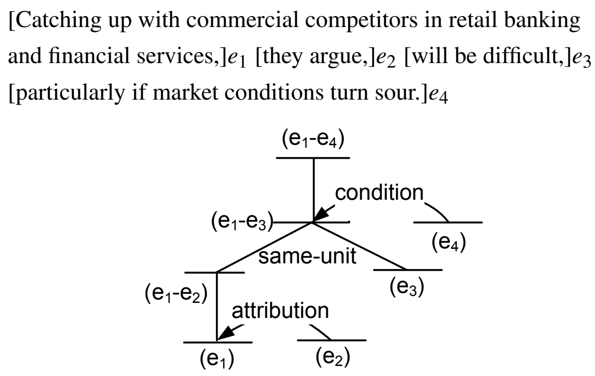

Fig. 1. An example of a Rhetorical Structure Theory discourse structure

decomposition of empirical data matrices to discover parameter values. When using latent variable models, these matrix decompositions are often low rank. As a result, the optimization will not suffer from local optima and can be orders of magnitude faster than comparable methods such as EM or gradient descent. Spectral methods have been successfully applied to a variety of tasks, including training HMMs [10], dependency parsing [5], and knowledge tracing [6]. Details of a similar latent variable spectral model can be found in [10].

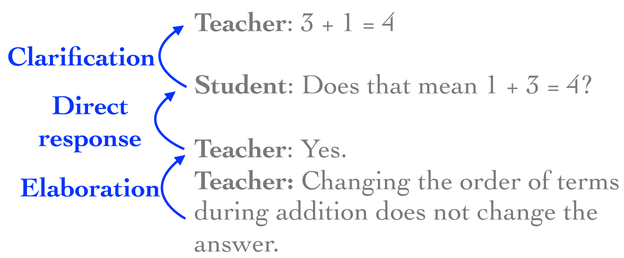

Fig. 2. An example of tutoring dialogue with discourse structure

3 Planning with Contextual Knowledge Bases

A Contextual Knowledge Base is a set of facts about a user and the environment [9]. In dialogue systems, examples of entries in a CKB may include the semantic representation of the user's most recent utterances, the current time of day, the user's age and gender, or an estimate of the user's aptitude for various deliberative tasks.

The information is stored as logical forms created by combining concepts, relations and predicates. This method of representing information allows us to incorporate the output of semantic and discourse parsing algorithms into the CKB very easily. Information from other sources, such as cameras, microphones, or GPS units, can also be incorporated into the knowledge base.

By allowing the agent to use the CKB when planning dialogue interactions, we can respond to critical information that would be ignored by a traditional conversational agent. The difficulty in using the CKB for planning is the innately sparse, high-dimensional nature of the feature space, so we utilize subspace identification algorithms such as Predictive State Representations [2]. If a measurable outcome metric is defined, the system can be trained using a traditional supervised dynamical system such as a Conditional Random Field (CRF) [11].

4 Results

In this section we present prior empirical results with spectral discourse parsing as well as planning using a contextual knowledge base.

For discourse parsing we used the Penn Discourse Treebank (PDTB) [13]. This is the largest discourse parsing treebank available; it includes eighteen thousand labeled relations from over one million words of text taken from the Wall Street Journal. There are 16 implicit relation types in this dataset, such as cause, restatement, and concession. The empirical results of our spectral discourse parser are shown in table 1, as well as results from two comparable methods.

Method	**Precision**	**Recall**	**F_1 Score**
Spectral Discourse Parser	0.51	0.42	0.46
SVM [7]	0.48	0.39	0.43
Max Entropy [12]	0.40	0.40	0.40

Table 1. Empirical Results for Spectral Discourse Parsing

For the CKB planning task, we created a simulated environment in which the system can be trained in an online fashion using a synthetic CKB. Each training instance represented one sequential interaction between the system and the simulated user. The goal is to minimize the cost of each interaction, where the cost is defined as a function that maps a state-interaction pair (S, i) to a real-valued, non-negative integer. The system behavior was governed using a Conditional Random Field. Figure 3 shows the average cost of each interaction as a function of the number of training instances. For comparison, a system that always greedily selects the interaction with the lowest average cost (and does

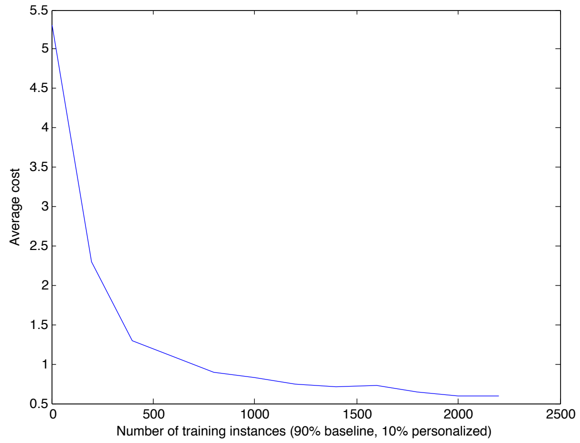

Fig. 3. Interaction costs with a Contextual Knowledge Base

not have access to the CKB) will have an average interaction cost of 5.4. More details are available in [9].

5 Conclusions and Future Work

Discourse parsing and Contextual Knowledge Bases give us tools to improve the social intelligence of future conversational agents. The largest challenge for incorporating these techniques into an ITS is the current need for annotated training data, particularly with discourse parsing. Going forward, we will be investigating active, semi-supervised training to reduce the need for human supervision [8]. This will allow us to leverage unlabeled data, which is readily available, and only request annotations for the most informative data points.

Dialogue structure and contextual information are critical components in human conversation, and they must be addressed to improve the effectiveness of conversational tutoring systems. When coupled with domain knowledge, semantic parsing, and sound principles of human-computer interaction, the tools discussed in this paper could help to close the gap between automated and human tutors.

6 Acknowledgements

This material is based upon work supported by the National Science Foundation under Cooperative Agreement EEC-0540865 as well as by a National Science Foundation Graduate Research Fellowship. We also acknowledge the Pittsburgh chapter of the American Rewards for College Scientists (ARCS) program for their generous support.

References

1. Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, pages 1247–1250. ACM, 2008.
2. Byron Boots, Sajid M Siddiqi, and Geoffrey J Gordon. Closing the learning-planning loop with predictive state representations. *The International Journal of Robotics Research*, 30(7):954–966, 2011.
3. Andrew Carlson, Justin Betteridge, Bryan Kisiel, Burr Settles, Estevam R Hruschka Jr, and Tom M Mitchell. Toward an architecture for never-ending language learning. In *AAAI*, 2010.
4. Albert T Corbett and John R Anderson. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction*, 4(4):253–278, 1994.
5. Paramveer S Dhillon, Jordan Rodu, Michael Collins, Dean P Foster, and Lyle H Ungar. Spectral dependency parsing with latent variables. In *Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning*, pages 205–213. Association for Computational Linguistics, 2012.
6. Mohammad H Falakmasir, Zachary A Pardos, Geoffrey J Gordon, and Peter Brusilovsky. A spectral learning approach to knowledge tracing. 2010.
7. Vanessa Wei Feng and Graeme Hirst. Text-level discourse parsing with rich linguistic features. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*, pages 60–68. Association for Computational Linguistics, 2012.
8. Robert Fisher. Active, semi-supervised learning to utilize human oracles. Master's thesis, Carnegie Mellon University, 2012.
9. Robert Fisher, Tom Kollar, and Reid Simmons. Building and learning from a contextual knowledge base for a personalized physical therapy coach. In *ICML Workshop on Robot Learning*, Atlanta GA, June 2013.
10. Daniel Hsu, Sham M Kakade, and Tong Zhang. A spectral algorithm for learning hidden markov models. *Journal of Computer and System Sciences*, 78(5):1460–1480, 2012.
11. John Lafferty, Andrew McCallum, and Fernando CN Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. 2001.
12. Ziheng Lin, Hwee Tou Ng, and Min-Yen Kan. A pdtb-styled end-to-end discourse parser. *Natural Language Engineering*, pages 1–34, 2012.
13. Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind K Joshi, and Bonnie L Webber. The penn discourse treebank 2.0. In *LREC*. Citeseer, 2008.