DARE: Deep Anaphora Resolution in Dialogue based Intelligent Tutoring Systems

Nobal B. Niraula
Institute for Intelligent Systems
University of Memphis
nbnraula@memphis.edu

Vasile Rus
Institute for Intelligent Systems
University of Memphis
vrus@memphis.edu

Dan Stefanescu
Institute for Intelligent Systems
University of Memphis
dstfnscu@memphis.edu

ABSTRACT

Anaphora resolution is a central topic in dialogue and discourse processing that deals with finding the referents of pronouns. There are no studies, to the best of our knowledge, that focus on anaphora resolution in the context of tutorial dialogues. In this paper, we present the first version of DARE (Deep Anaphora Resolution Engine), an anaphora resolution engine for dialogue-based Intelligent Tutoring Systems. The development of DARE was guided by dialogues obtained from two dialogue-based computer tutors: DeepTutor and AutoTutor.

Keywords

Anaphora Resolution, Tutoring System, Dialogue Systems

1. INTRODUCTION

Anaphora resolution is the task of resolving what a pronoun or a referential noun phrase refers to. As an example consider the dialogue segment in Table 1 (a). Here, the pronoun "it" in the STUDENT turn refers to the first mention of the word "force" earlier in the turn.

In dialogue based Intelligent Tutoring Systems (ITSs), pronouns are quite frequent in students' natural language responses. Examples of student responses that contain anaphoric pronouns are shown in Table 1. These examples are from DeepTutor, a conversational ITS[1]. DeepTutor mimics the dialogue between a computer tutor and tutee and is based on constructivist theories of learning according to which students construct their knowledge themselves and only get help when floundering. The help consists of hints in the form of questions - see the DeepTutor dialogue turns in Table 1. Students responses are assessed for accuracy and appropriate feedback is provided by DeepTutor. Students can ask questions themselves as well.

Solving anaphors in student responses in dialogue-based ITSs is very important as it has a direct impact on assessing the correctness of student responses.

Table 1: Use of pronouns in students' responses

(a) Intra-turn : **DEEPTUTOR**:What does Newton's second law say? **STUDENT**:for every force, there is another equal force to counteract it
(b) Inter-turn immediate: **DEEPTUTOR**:What can you say about the acceleration of the piano based on Newton's second law and the fact that the force of gravity acts on the piano? **STUDENT**: It remains constant.
(c) Inter-turn history: **DEEPTUTOR**: Since the ball's velocity is upward and its acceleration is downward, what is happening to the ball's velocity? **STUDENT**: increasing **DEEPTUTOR**: Can you please elaborate? **STUDENT**: it is increasing

While anaphora resolution is a well-studied problem in Natural Language Processing [1, 2], there is no previously reported work, to the best of our knowledge, which addresses the problem of anaphora resolution in dialogue based ITSs. As already mentioned, resolution of pronouns in ITSs is a key step towards understanding students' responses which impacts the accuracy of the student model. Failing to resolve pronouns can make the computer tutor assess incorrectly a student response and react ineffectively which would lead to suboptimal learning. Incorrect feedback from the system could frustrate students sometimes to the point of quitting interacting with the system.

Given the importance of accurate assessment of student responses in ITSs, a highly accurate anaphora resolution mechanism is needed. To this end, we have been developing Deep Anaphora Resolution Engine (DARE) for conversational ITSs. The design of DARE is guided by an analysis of actual interactions between students and two dialogue based ITSs : AutoTutor[2] and DeepTutor. DeepTutor is a fully online tutoring systems that has been used by close to a thousand students who can access DeepTutor anytime, anywhere. In its first version, DARE distinguishes two categories of pronouns in dialogue-based ITSs based on the lo-

[1] www.deeptutor.org

[2] www.autotutor.org

cation of the referents: *intra-turn* anaphors and *inter-turn* anaphors. The *intra-turn* anaphors refer to an entity located in the current student dialogue turn. The *inter-turn* anaphors refer to entities in previous dialogue turns (or dialogue history) or entities present in the problem description or other contextual elements, e.g. even world knowledge. Examples of intra-turn and inter-turn anaphoric pronouns are shown in Table 1.

2. THE METHODOLOGY

As a starting point for developing DARE, we analyzed pronoun use in 24,945 student responses from DeepTutor log files and 1,978 student responses from AutoTutor logs. The results are shown in Figure 1. Both the students and computer tutor use first-person pronouns (i.e. "i","me", "we", "us") during the interaction. It should be noted that these pronouns do not need be resolved for assessment purposes (for space reasons we do not elaborate). Of the remaining pronouns, the top two most frequent pronouns, one of which is "it", account for more than 80% of the anaphors. Thus, considering a very few, very frequent anaphors may be a good start for developing DARE. Moreover, it was observed (see Section 3) that most pronouns used in student responses can be resolved withing the same responses or just looking at the previous system turn, i.e. the previous hint from the system. Although these aspects of anaphors in dialoge-based ITSs simplify the problem of pronoun resolution, the task is still challenging because "it" is often an pleonastic, i.e. "it" is not always an anaphoric pronoun [1].

Another important aspect of anaphors in tutorial dialogues is the location of pronouns in students' responses. In our data, we observed that most pronouns at the beginning of a student response refer to an antecedent present in the most recent system response, i.e. the previous dialogue turn. On the other hand, pronouns that occur in the middle or last part of a student response most likely refer to an entity in the current/same student response. Thus, we added in DARE a classifier that relies on the position of the anaphors to identify the text where the antecedent should be searched for. Given this fact, input to the DARE's resolution engine is the concatenation of previous tutor turn and current student response if the pronoun to be resolved occurs at the beginning of the student response and current student response only if the pronoun occurs in the middle or near the end of the student response. DARE then uses the coreference resolution module in the Stanford CoreNLP package in order to perform anaphora resolution. It should be noted that sometimes a pronoun in the student response may refer to entity in the problem description, e.g. the current Physics problems the student is working on. Also, it may be possible that a pronoun refers to something mentioned much earlier in the dialogue than the previous system turn. In the current version of DARE, we do not handle these latter cases. Another case we do not directly handle in DARE currently is the use of eliptic anaphors - see the first student response in (c) in Table 1 where instead of saying "it is increasing" the student simply says "increasing" (a typical exaple of ellipsis).

3. EXPERIMENTS AND RESULTS

In order to evaluate the performance of the DARE system, we extracted student-tutor interactions from the DeepTutor's log files. We only considered student-tutor interactions which contain at least one pronoun in the student responses. Since the pronouns in students' responses could refer to entities that were mentioned at any moment during the dialogue, i.e. the whole dialogue history, or the problem descriptions, we retain all this information for each instance in the data set. In total, we extracted 5,589 instances out of which 112 were annotated manually.

Figure 1: Pronouns used by (left) DeepTutor students in 24,945 dialogue turns (right) AutoTutor students in 1,978 dialogue turns

Pronoun	Count
it	3249
they	1152
i	526
you	207
her	163
she	158
he	90
them	89
we	81
her	62
us	56
me	48

Pronoun	Count
it	285
they	227
he	39
them	25
i	18
you	11
him	6
me	3
we	3
one	1
us	1
she	1

Antecedents in	Count(%)
H_0	85 (75.89%)
A	22 (19.64%)
H_1	5 (4.46%)
H_2	0 (0.00%)

Table 2: Location of antecedents for anaphors

The analysis of the annotated instances is shown in Table 2. Almost 76% of the time, students' anaphors refer to entities in the most recent tutor turn (hint/question H_0). They also use pronouns to refer to entities in their response (intra-turn anaphors) which accounts for almost 19% of the time. Next, they refer 4% of the time to entities in the hint that immediately precedes the most recent hint (i.e. H_1). Interestingly, there were no references to entities beyond H_1. Currently, DARE does not look back for referents beyond H_0. The accuracy of the current version of DARE is 46.36 %.

4. CONCLUSION AND FUTURE WORK

The current version of DARE presented here offers a good baseline which we plan to improve in future iterations of development. We also intend to evaluate the performance on a larger data set which we will make publicly available.

5. ACKNOWLEDGMENTS

This research was supported in part by Institute for Education Sciences under awards R305A100875.

6. REFERENCES

[1] R. Mitkov. Anaphora resolution: The state of the art. Technical report, University of Wolverhampton, 1999.

[2] A. Rahman and V. Ng. Supervised models for coreference resolution. In *Proceedings of EMNLP*, pages 968–977. ACL, 2009.