Identifying Reasoning in Speech

Laura Willson

Language Technologies Institute, Carnegie Mellon University, Pittsburgh Pennsylvania
Laura.A.Willson@gmail.com

Abstract

Transactive reasoning has been shown to increase learning. We would like to be able to determine when Transactive reasoning is occurring not only in written transcripts but also in audio files. Thus, we are developing tools to train a model to be able to identify instances of Transactive reasoning with only audio files. In order to be able to use our corpus of data, first we needed to develop coding manuals to identify reasoning and transactivity in speech, as an important first step in training the automatic detection of collaborative learning.

1. Introduction

We would like to be able to evaluate students, but a lot of time can be spent in the process. It requires teachers to write tests that test the student on the relevant knowledge, time for the students to take the tests, and then the teachers have to grade them. We would like to train a computer to be able to automatically recognize when students are learning without the students actually having to take tests. This would have many practical applications, but would be especially useful for spoken dialogue tutoring systems.

2. Previous Work

It has been demonstrated, in Psychology and Computer Science, that there are certain types of reasoning that have been highly correlated with learning (Weinberger & Fischer '06). When students are engaging in a dialogue, one of them builds on the other's argument, either by acknowledging their partner's point and offering evidence for or against it, or by critiquing the logic they used in their reasoning. This is known as Transactive Reasoning, and it has been shown to result in a restructuring of a person's knowledge base (Azmitia & Montgomery '93).

The idea of Transactive Reasoning is based in the work of Swiss developmental psychologist Jean Piaget, who established the theory of constructivism, which emphasized that knowledge acquisition in social and peer learning enhances outcomes (O'Donnell et al '99). A transactive discussion is one where the reasoning of one participant builds on the other's, and is a marker of a dialogue where the participants are engaging with one another in a dynamic way.

3. Hypothesis

We believe that if you could develop a method to identify when transactive reasoning is occurring, then you would be able to predict when students are learning with reasonable accuracy by counting the occurrences of transactive dialogues. We attempted to identify transactive reasoning by taking advantage of social accommodation theory, which states that when two people speak, their speaking styles will become either more or less similar depending on the outcome of their discourse. When a dialogue is constructive and the participants are engaging each other, they will start to speak more like each other. Each individual has a specific way of pronouncing different sounds, and the quantitative description of how the vowels are pronounced is known as a vowel space. By comparing the change in vowel spaces of individuals over the course of a dialogue and looking for

convergence we can estimate the amount of influence they had on each other.

4. The Corpus

The ottoman corpus is a series of debates between pairs of University of Pittsburgh undergraduate students. Each side is given a list of points as to why the Ottoman Empire fell, either due to external or internal causes. Each debate is 8 minutes long, and consists of turns of two people talking. The speech is natural free form speech, that has been transcribed. The students were assessed at the beginning and end, so we know when they are learning. They were given some basic background information. The students were incentivized to do well because they were given a prize if they won.

5. Segmenting Speech Transcripts

Because we will ultimately determine how much a student is learning by amount of transactive discussions that occur, it is important to have a well-defined unit of measurement. Although written text is easy enough to break up, as sentences are an intuitive unit, naturally occurring speech is much harder. When people produce spontaneous speech the boundaries between different ideas are often not clear. Sometimes people switch thoughts mid-sentence, have extended run-on sentences, or interrupt each other. Although this might sound normal in a conversation, there is no way to form a grammatically correct sentence from it. The subject of the debates was the fall of Ottoman Empire, which is something all but a few of the students knew very little about. Because of this the data was full of repairs, repeats, and other linguistic disfluencies, (such as um, uh, okay, mmhm), that make spontaneous speech different than written text or prepared speech. Examples follow:

False Starts
It had been a, but World War 1, it's been a long time since the Ottomans attacked Vienna.

Repeats and Repairs
wh, what…what what how was their technology not built up?

We ultimately chose an independent clause as our smallest unit of analysis. We devised a coding manual that took the idiosyncratic behaviors of speech, and created rules around them to develop a consistent and clear method for getting meaningful units of speech.

6. Identifying Reasoning

In order to determine where transactive reasoning was occurring, first we needed to determine when reasoning was being used. We defined reasoning as an argument that explicitly shows part of speaker's thought process, and does not just state facts. It must involve a component of explanation. In order to prevent double counting occurrences of reasoning, a single segment on its own does not contain reasoning, but when paired with a later segment does contain reasoning, then we would mark the later segment as reasoning.

We divided reasoning into three categories: causal relationships, comparing and contrasting, and value judgments. We empirically derived a manual to code for reasoning that was specific to the debating style of our corpus. Sometimes we could use certain key words, for instance "because" always indicates a causal relationship, and therefore reasoning is being used.

7. Identifying Transactivity

For transactivity we had identified four different categories that overlapped with reasoning-- reasoning critique, counter consideration, competitive juxtaposition, and comparative critique (Berkowitz '79). We then only coded for instances of transactivity if they had already been coded for reasoning. For instance, if a student corrected a factual statement of another's, they would be

engaging with the other student, but if there was no element of reasoning in their correction, then we did not consider it relevant.

8. Analyzing Vowel Spaces

Formants are amplitude peaks of the frequency spectrums of sound that measure acoustic resonance. The first one represents the backness of a vowel, and the second one represents the openness of a vowel (Eckert & Rickford '01). Knowing the frequency of these two formants is enough to identify a vowel. Every person has vowel space that is slightly different, so we can measure the convergence of vowel spaces with a Euclidian distance metric of the distance between formants for the two speakers. If the speakers' vowels begin to converge, then social accommodation theory would suggest they are influencing each other. We choose to analyze the vowel spaces of the speakers, rather than their consonants because vowels are held for longer, and are often easier to distinguish.

9. Results

My contribution to the project was writing the various coding manuals to properly segment and then label the corpus for reasoning and transactivity. The success of the coding manuals was measured by Cohen's Kappa, which measures inter-rater agreement between two raters. Using Cohen's Kappa allowed us to determine whether two people labeling the data were agreeing at rates greater than by chance.

I revised the manuals until we successfully achieved Kappa scores that were consistently greater than .7, which is considered very good agreement. Once the manuals were complete we had the entire corpus coded using them, which allowed to data to be reliably hand-labeled.

10. Conclusion and Further Work

We were able to consistently and reliably identify reasoning and transactivity so that we could label the corpus and use the resulting data to train models to determine if learning had occurred.

After my work on the coding manuals the results allowed for further work analyzing the vowel spaces to be completed. The results were submitted to two journals, EACL and ICLS.

11. References

[1] Azmitia, M., and Montgomery, R.. "Friendship, Transactive Dialogues, and the Development of Scientific Reasoning." *Social Development* 2.3 (1993): 202-21. Print.

[2] Berkowitz, M., & Gibbs, J. (1979). A Preliminary Manual for Coding Transactive Features of Dyadic Discussion. Unpublished manual for coding transactivity.

[3] Eckert, Penelope, and John R. Rickford. *Style and Sociolinguistic Variation*. Cambridge, UK: Cambridge UP, 2001. Print.

[4] Gweon, G., Bhiksha, R., Rose, C., Pulkit, A., Udani, M., 2011. The Automatic Assessment of Knowledge Integration Processes in Project Teams. Proceedings of the International Journal of Computer-Supported Collaborative Learning.

[5] Kane, A., Argote, L., Levine, K., 2005. Knowledge transfer between groups via personnel rotation: Effects of social identity and knowledge quality. *Organizational Behavior and Human Decision Processes*.

[6] O'Donnell, Angela M., and Alison King. *Cognitive Perspectives on Peer Learning*. Mahwah, NJ: L. Erlbaum, 1999. Print.

[7] UCLA Phonetics website. http://www.phonetics.ucla.edu/course/chapter1/vowels.html
[8] Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. Computers & Education, 46, 71-95.