Fall 10-27-2017

Identifying Student Discussion in Computer-Mediated Problem Solving Chat

Lindsey K. Arndt
Valparaiso University, lindsey.arndt@valpo.edu

Stamatina E. Kalafatis
Valparaiso University, stamatina.kalafatis@valpo.edu

Emily Graham
Valparaiso University, emily.graham1@valpo.edu

Identifying Student Discussion in Computer-Mediated Problem-Solving Chat

Lindsey Arndt, Emily Graham, Tina Kalafatis

Dept. of Mathematics and Statistics, Valparaiso University

Research Questions

1. Can we write machine classifiers that can recognize productive student discussion?
2. Can we achieve this using only a common English vocabulary?

Abstract

The COMPS project employs computer chat for students working in small groups solving classroom problems. This summer's project aims to build computer classifiers that could effectively "look over the shoulders" of the students while working, to approximately recognize whether the students are engaging in productive discussion.

Several thousand lines of COMPS transcripts were manually annotated. A topic modelling program determined 10 main topics which appeared in the transcripts and the words in those topics. A Linear Classifier and a Support Vector Machine Classifier used the topic model to predict the annotation of each line of dialogue.

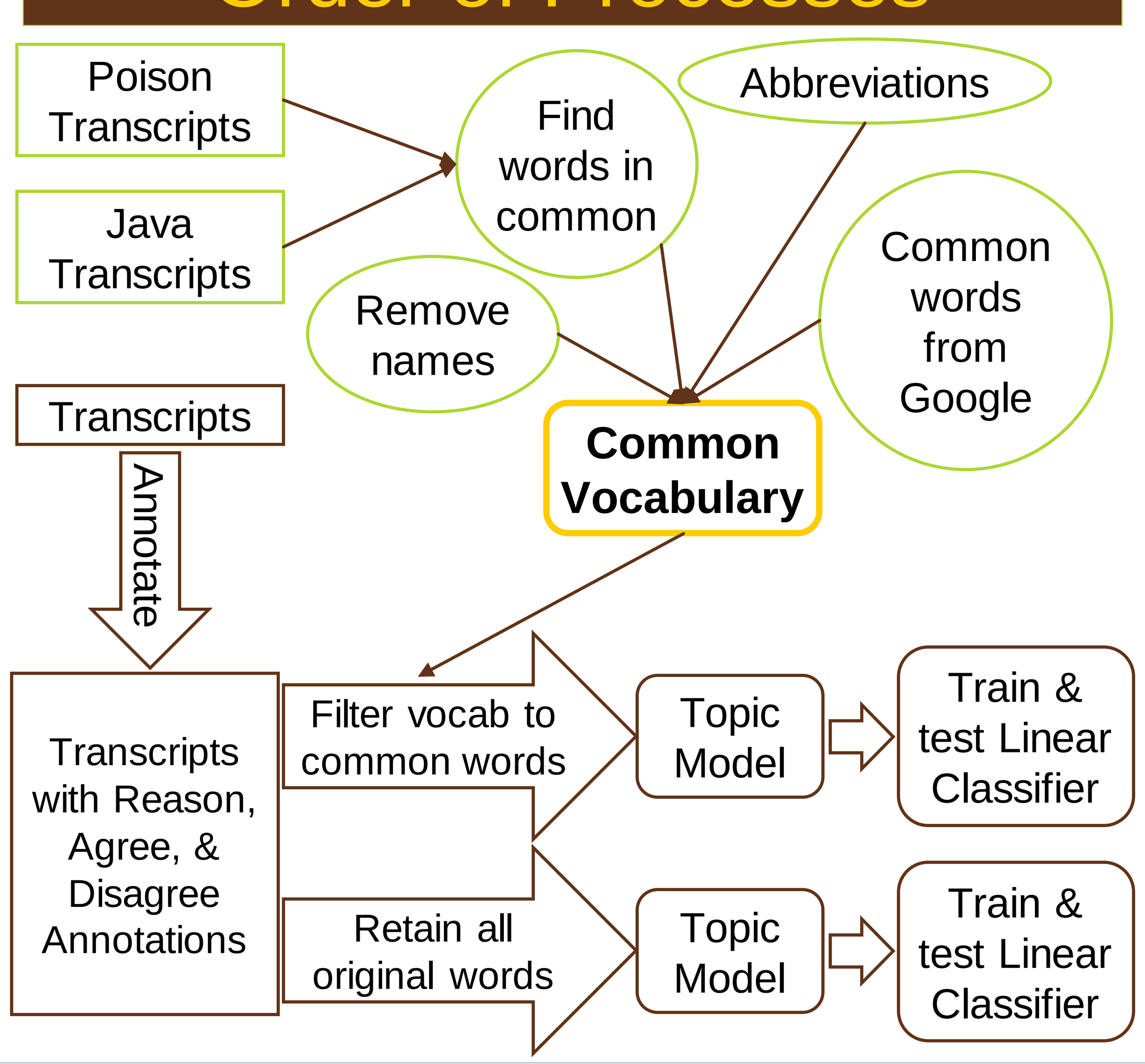
To address the common English vocabulary research question, an intersection of many transcripts from various sources was combined with Google word lists and modified to accommodate text-chat conventions.

Productive Discussion

In order to determine whether or not the students are engaging in a productive discussion, we need to see evidence that the students are:

1. Talking about the problem
2. Using key agreement and disagreement terms

Order of Processes



Dialogue Transcript Data

Student	Text	Reason	Agree	Disagree
A	now you took opposite, but it ended up being even, is that still fine?	1	0	0
B	I don't think so. We want it to be odd. So that didn't work. What do you guys think?	1	0	1
C	im confused	0	0	0
A	I wonder if it comes back to multiples somehow. I'm sorry to bring it back up but I feel like that might somehow apply?	1	0	0
B	I think you're right too!	0	1	0
C	no i think youre right i just dont know the pattern youre thinking multiples of 4 still?	1	1	1
A	No. Because that isn't always possible.	1	0	1
B	sooooooooooo what multiples? if you dont mind me asking	1	0	0
A	*shrugs shoulders* haha	0	0	0

We manually annotated several thousand lines of chat dialogue with a binary system (1=y, 0=n). A coding manual was written which details how we decided on the annotations.

For Reason turns, we looked to see if the students were taking any steps towards the solution of the problem, such as talking or asking a question about the problem. For marking Agree, we looked for key affirmative words such as "yes", "okay", "sure", etcetera. Similarly for Disagree, we looked for key disagreement words such as "no", "not sure", "don't think so", etcetera. The key idea when annotating dialogue is to think completely literally because the computer cannot read the context of the dialogue, so neither could we.

Annotated transcripts are needed for training and testing our classifiers. The computer classifiers need to know some correct answers in order to know a lot of the correct answers.

Vocabulary

Student	Original Text	Filtered Text	Text with Common Vocabulary
A	works*	$g02 works	$g02 works
A	alright well i think *(Student C)* should go this time	alright well i think *(Student C)* should go this time	alright well i think $g01 should go this time
A	first i mean	first i mean	first i mean
B	Let's try it. Keep in mind you want these numbers as your goal.	let's try it keep in mind you want these numbers as your goal	let's try it keep in mind you want these numbers as your goal
A	alrigh	Alrigh	$g01
B	Go ahead *(Student C)*.	go ahead *(Student C)*	go ahead $g01
A	*(Student C)* will win either way	*(Student C)* will win either way	$g01 will win either way
B	I think you can say *(Student C)*has won.	i think you can say *(Student C)* has won	i think you can say $g01 has won

Our goal in finding a common vocabulary is for us to be able to determine if, in a chat, students are solving the problem at hand without needing the context of the problem or problem-specific words. In compiling a common vocabulary with which to apply to our transcripts, we used a list of 10,000 words from Google as our basis. Then we removed all common names that occurred in the chats as well as we added in abbreviations and slang that was used in our transcripts. We used transcripts of students working in two different problem areas from two different universities. We also added to the vocabulary words that occurred in both transcripts but did not appear in the Google list.

When converting dialogue turns from original vocabulary to our chosen common vocabulary, we keep all of the words in the common vocabulary. The uncommon words are converted to the token $g01. Other lexical phenomena used primarily for emphasis such as #, @, * and ^ were stripped from the words and other tokens were inserted.

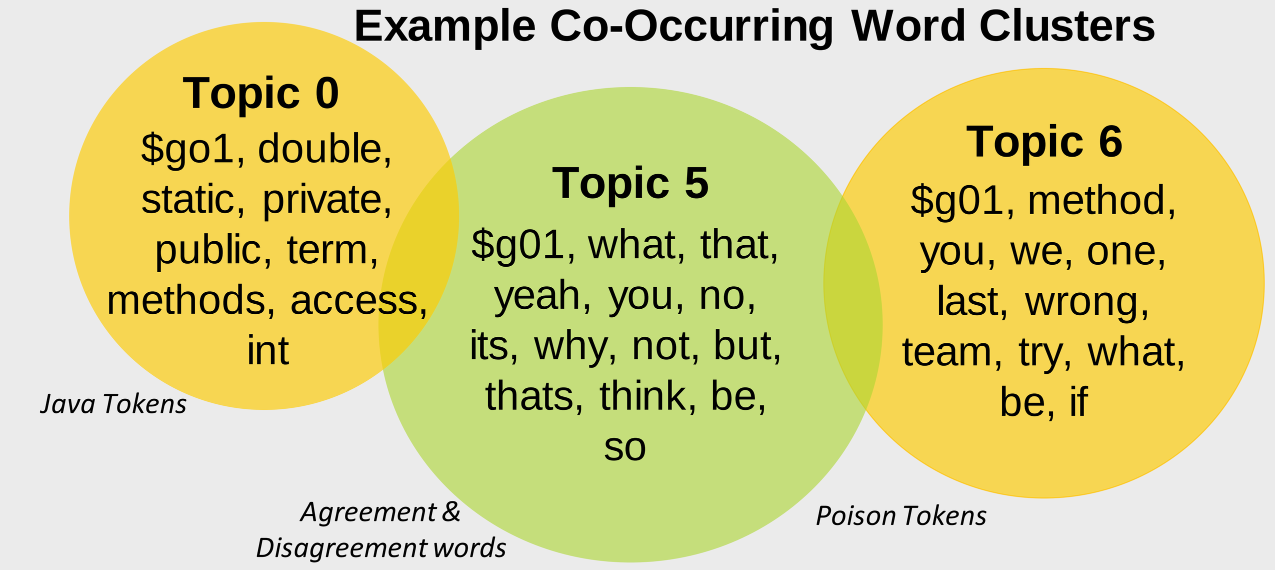
Filtering COMPS Chat Transcripts for Computer Modeling Using Common Vocabulary by Nathaniel Bouman has more details on the algorithms used to analyze and regularize the vocabulary.

Topic Modeling

Each topic is a cluster of words that tend to co-occur within dialogue turns in our transcripts. Ten topics are determined by the computer. The topic modeling program outputs the probability of a specific topic appearing in each dialogue turn.

We ran the topic model program on the same transcripts twice – once with the transcript containing original vocabulary and a second time with that same transcript filtered for common vocabulary.

Example Co-Occurring Word Clusters



Classifiers

We worked with two classifiers: our Linear Classifier and our Support Vector Machine (SVM) Classifier. We trained on 60% of our data and then tested on the remaining 40%. For the Linear Classifier, the computer is doing a linear regression to predict the annotation based on the topic values outputted by the topic modeling program and it fits the annotated lines with the following equation:

$$A_0 * T_0 + A_1 * T_1 + A_2 * T_2 + \cdots + A_9 * T_9 + C = Y$$

The A_i values are the coefficients of the linear regression and the T_i are the topic values that are outputted by our topic modeling program. Y is our annotation mark, so if $Y > \alpha$ where α is our cutoff value, then the annotation is 1. If $Y \leq \alpha$, then the annotation is 0.

To judge the success of our classifiers, we look at precision, recall, and the harmonic mean ($f1$), which is the balance between precision and recall. Ideally, $f1 > 0.6$.

The SVM Classifier fits a hyperplane that separates the 0's and 1's in a scatter plot.

Results

Linear Classifier Reasoning f1 Scores

	Original Vocab	Common Vocab
Poison	0.737	0.658
Java	0.592	0.578
Combined	0.695	0.701

Linear Classifier Agree f1 Scores

	Original Vocab	Common Vocab
Poison	0.509	0.455
Java	0.278	0.321
Combined	0.377	0.397

Our reasoning scores were by far the best, mostly meeting our target of 0.6 and above. Agree scores were less promising around 0.4. The disagree scores were much lower, ranging from 0.03 to 0.1. The SVM Classifier results were very similar to those of the Linear Classifier.

In general, our classifiers worked better on the Poison transcripts than on the Java transcripts.

Next Steps

- Create the dashboard program
- Explore other conversational behaviors
- Investigate other applications of the COMPS program

Acknowledgments

Thank you to Professors Melissa Desjarlais and Michael Glass and fellow students Nathaniel Bouman and Yesukhei Jagvaral at Valparaiso University for their assistance with this project.

Partial support for this work was provided by the National Science Foundation's Improving Undergraduate STEM Education (IUSE) program under Award No. 1504918 and by the MSEED program under Award No. 1068346. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.