Finding Gender Differences within Problem-Solving Dialogue Interactions

Angelica Willis Zachery Thomas
Advisors: Dr. Jung Hee Kim, Dr. Kelvin S. Bryant, Dr. Taehee Kim
Department of Computer Science
North Carolina A&T State University

Abstract

Project Aims

The COMPS (Computer-Mediated Problem-Solving) project analyzes conversations of students working together in class via computer chat. An eventual goal is to have a computer monitor the students while they work. The research problem here is to have the computer identify the gender of the students as they type.

Method

From log files of the sessions, transcripts are extracted. Each turn of the dialogue is annotated with features. These features are input to machine learning algorithms to derive a classifier. Software recognizes in each dialogue turn features such as the number of words a student typed, the presence of question marks, smiley faces, and words that often indicate reasoning ("so", "therefore"). Unlike the strict turn-taking of spoken conversation, COMPS affords students the possibility to all type at the same time, so we have the computer annotate typing overlaps and various timing characteristics. Human coders manually annotate each term with additional features from the literature of gender-related linguistic differences.

Data Collection

- Our work has focused on transcripts from a 1st year Java programming class.
- Students work in 3 or 4 person groups solving a problem.
- Chat via NC A&T's COMPS web page, which keeps logs.
- Transcripts are extracted from logs: contain student text plus timestamps.

Feature Selection Examples

Emotional Attributes: Below are the extracted emotional attributes, including examples we decided to look for when parsing the dialogues. Each of these give insight as to how each student is thinking as well as how the group interacts with each other as a whole.

Apologetic: refers to a user expressing regret for previous action this type of message is usually aimed towards another user or towards the group as a whole.

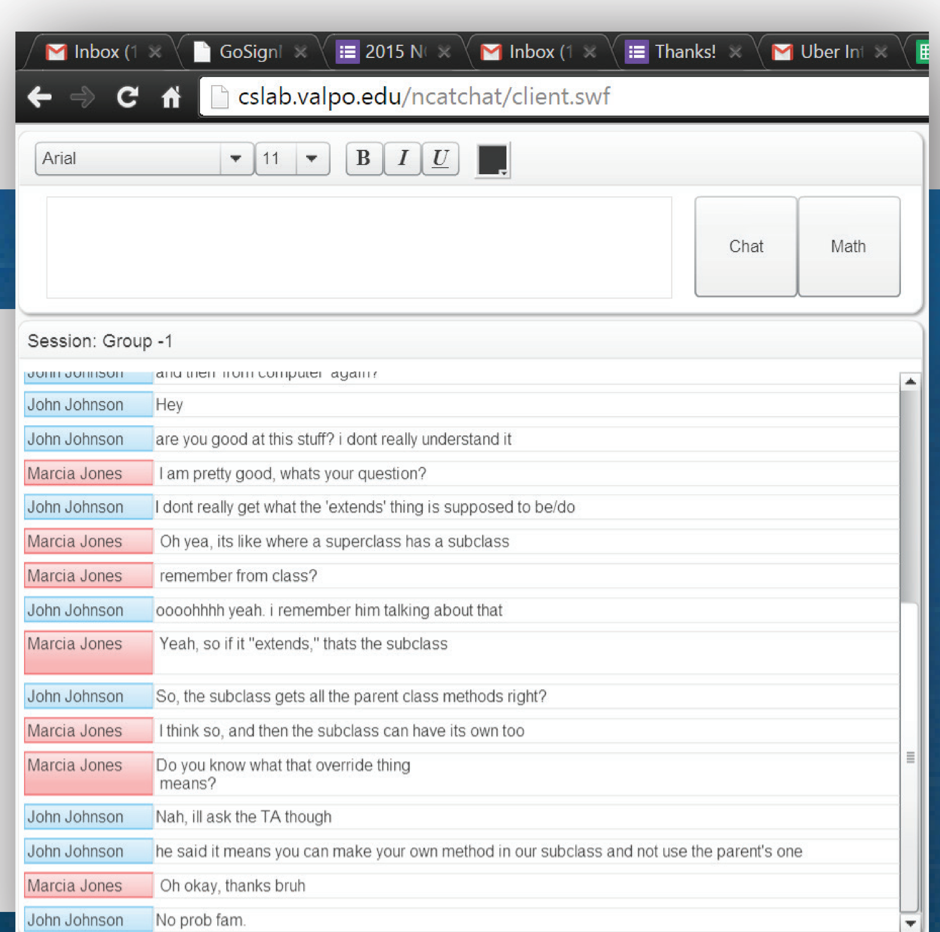
Confused: determined by user explicitly expressing confusion, or by user not appearing to be with the rest of the group (ie asking a lot of questions).

Sad: a negative emotion determined by keywords and sad emoticons that are usually directed at self.

Example feature words, symbols and phrases

Excited	Apologetic	Confused
:D yay yes! !!! cool!	sorry my bad nvm whoops i messed up	i'm confused how why what is I don't understand
Frustrated	**Sad**	**Humor**
D:<):< this is hard	:( ): I feel stupid	Context Joke haha lol

Number of turns: We also considered the number of turns, or number of independent messages sent by each student.



Example Dialogue

The analysis of collaborative learning transcripts include the classification of keywords and phrases that may be used as factors in determining the gender of a student.

Sample Transcript

User	Timestamp	Message
Student A	06:44.2	f and foo are the refernece variables
Student A	07:05.2	so those together make 16? for the refrence types
Student B	07:11.9	yup yup
Student A	07:27.9	16 bytes
Student C	07:30.2	2a = 20
Student C	07:36.0	:D
Student B	07:39.7	there ya go lol
Student D	07:54.86	Wait where did you get 16?
Student D	08:05.8	wouldnt it be 48 at least for the main method
Student D	08:18.3	because the array creates 5 object
Student A	08:26.1	oh yeah i looked over that was just countingm f and foo
Student C	08:28.7	those are on the heap not the stack
Student D	08:48.0	So the objects created by an array are on the heap
Student A	09:13.8	yeah run time stack = 48

Excited Apologetic Confused Humor

Misspelling: In order to still understand the meaning behind misspelled words, we decided to use word stems as context clues. Using word stems allows us to deduce the true meaning of that dialogue turn.

Acknowledgement

Partial support for this work was provided by the National Science Foundation's Improving Undergraduate STEM Education (IUSE) program under Award No. 1504917. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Results

Statistical Differences

Knowledge post-test. 43 students took a post-test in the topic of the exercise, with a score range of 0 to 3 points. Female students (Mean=1.50, SD=1.557) did better than male students (Mean=1.03, SD=1.426). However an independent samples t-test indicated the difference was not significant: p=.30.

Affective States. Among the affective states that we annotated in the dialogues, the one most significantly different between the genders was being apologetic. Men expressed an apology more often (0.48 apologies per dialogue) than women (0.19). This was close to significant: p = 0.06

Amount of participation. Men typed many more turns per dialogue (Mean=46 turns) than women (Mean=36), which was not significant: p=0.34. They also typed more words in total (291 vs. 233).

Machine Classifiers

Using about 1800 dialogue turns, we tried training Weka J48 decision trees to classify the gender of individual dialogue turns. Each turn was tagged a variety of hand- and machine-annotated features. Although the training algorithm discovered the associations noted above, apologies and wordiness, none of the classifiers stood up to cross-validation.

Future Work

The immediate plan is to develop and test more features, with an eye towards the socio-linguistics literature on the gender differences in language and dialogue. An example is perception phrases, e.g. "I think that" and "it seems that." We will try to rely on features that can be machine-annotated.