

Implementing a friendship classifier module for a virtual agent peer tutor

Holly Tancredi, Alexandros Papangelis, Justine Cassell
TH0084@bcmail.brooklyn.cuny.edu, {apapa,justine}@cs.cmu.edu

Motivation

Our goal is to build a fully automatic embodied conversational agent peer tutor that can build, maintain, and break rapport (Spencer-Oatey) in real time and over multiple interactions.

Based on prior work (Ogan et al) that proposes that friends and strangers build rapport using different approaches, we created a theoretical framework for rapport management between the user and our virtual agent.

The **Friendship Classifier** uses extracted low-level audio features, low-level visual features, and dyadic features, collected from peer tutoring sessions in which dyads of teenagers taught each other linear algebra (Yu et al), to determine whether or not a dyad are friends.

Based on the output of the Friendship Classifier, our system is able to pick an appropriate rapport strategy to execute.



Method

Feature Extraction

- OpenSMILE for audio
- OKAO for visual
- Human annotation for dyadic features

Training & Testing

- Leave-one-out cross validation

Classification

- Linear Support Vector Machine
- Implemented in C++ with libsvm library

Feature Selection

- Calculate f-score



Results and Analysis

- Initial tests of the classifier trained on data from 12 dyads (half friends and half strangers) yielded low accuracy results, so the data was split into 24 individuals.
- We performed feature selection by finding the f-scores for each set of features.
- Accuracy results show that multimodal data (audio, visual, and dyadic) is a more accurate indication of friendship status than only verbal or nonverbal data.
- The final data set of most important features has 4,518 data points across 24 individuals, and 75 multimodal features. It predicts friendship status with 82% accuracy.

The Data Set

Audio

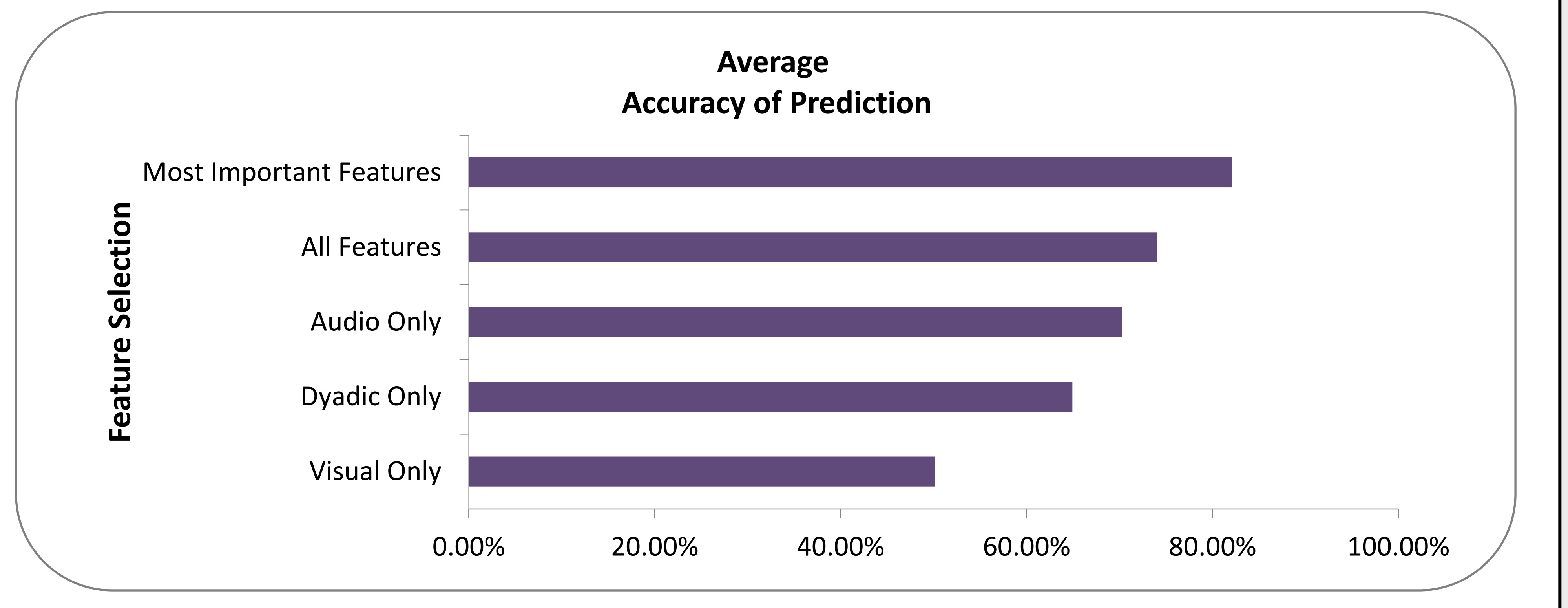
- Prosodic features
- Voice quality features
- Spectral features

Visual

- Facial position
- Gaze
- Smile

Dyadic

- Number & average length of turns
- Mutual/non-mutual smile
- Mutual/non-mutual gaze
- Mutual/non-mutual lean forward
- Mutual gaze with mutual smile



Conclusions

We created a module for a virtual agent peer tutor that bases its rapport strategies on whether or not the agent and human are perceived as friends.

We modeled this module on human-human interaction in a dyadic peer tutoring setting.

Extensive training and testing of the Friendship Classifier indicates that multimodal data produces the most accurate prediction of friendship.

Future Work

Currently, the Friendship Classifier is implemented as a module of the Virtual Human Toolkit. Eventually, the Friendship Classifier module will be connected with MultiSense, a module that extracts visual and audio features from a user in real time. The Friendship Classifier will also gather input from the virtual agent's natural language understanding module.

When implemented with the rest of the modules that make up our embodied conversational agent peer tutor, the Friendship Classifier will predict, in real time, the friendship status of the human user and virtual agent.

Future work on the Friendship Classifier could also incorporate a scale of friendship, rather than a binary classification.

Literature cited

A. Ogan, S. Finkelstein, E. Walker, R. Carlson, and J. Cassell. Rudeness and rapport: Insults and learning gains in peer tutoring, 2012.

Chih-Chung Chang and Chih-Jen Lin, LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1—27:27, 2011. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.

Hartholt, Arno, et al. All together now. *Intelligent Virtual Agents*. Springer Berlin Heidelberg, 2013.

H. Spencer-Oatey. (im) politeness, face and perceptions of rapport: unpackaging their bases and interrelationships, 2005.

Z. Yu, D. Gerritsen, A. Ogan, A. Black, and J. Cassell. Automatic prediction of friendship via multi-model dyadic features, August 2013.

Acknowledgments

This work was generously funded by the DREU program, a collaboration between the CRA-W and CDC, and has been supported by a grant from the NSF Broadening Participation in Computing program (NSF CNS-0540631).



ArticuLab