Understanding the Perception of Courteous and Humorous Behavior using Prosodic and Lexical Features

Sidd Jagadish*, Ranjay Krishna‡, Gabriele Carotti-Sha∓

siddj@stanford.edu, rak248@stanford.edu, gcarotti@stanford.edu
*Department of Statistics, Stanford University
‡Department of Computer Science, Stanford University
∓ Symbolic Systems, Stanford University

Abstract

Computationally understanding the perception of a speaker's characteristics is an important task speech processing, behavioral outcomes and dialogue systems. Our work focuses on the classification and understanding the perception of two such characteristics: courtesy and humor. We analyze ratings surveyed from the SpeedDate corpus to build models using prosodic and lexical features of the speakers. We find that lexical features like *Hedge-words* and prosodic features like *Intensity Min SD* are important for detecting courteous while *Exclamations* and listener's *Intensity* are important for funny. Using an Adaboost Classifier, we detect the funny with an accuracy of SHIT GOES HERE and courteous at SHIT GOES HERE.

Index Terms: Prosody, Speed Date, AdaBoost, Gender differences in prosody, factor analysis, courteous, funny

1. Introduction

2. Related Works

3. The SpeedDate Dataset

We utilize the SpeedDate dataset introduced by [1] to build our model. The dataset contains approximately 1100 heterosexual 4-minute speed dates. Each date is stored as a wav file recording of the speed date along with text files of the dates annotations. On average, each date contains 812 words, with an average of 406 words per speaker. These worded are divided into an average of 93 turns per date where the speaker changes to the other.

3.1. Feature Extraction

3.1.1. Prosodic Feature

We use OpenSMILE to extract Prosodic features from audio files of the speed date. Lexical features are extracted from the transcripts with the help of the LIWC dictionary.

3.1.2. Lexical Feature

3.1.3. LIWC Feature

3.1.4. Accomodation Feature

[2], [3] and [4] demonstrate through their work the convergence of vocal intensities between interlocutors. To capture this convergence, we extract features from a speaker that accommodate the previous speaker's speech. These features include the *Rate*
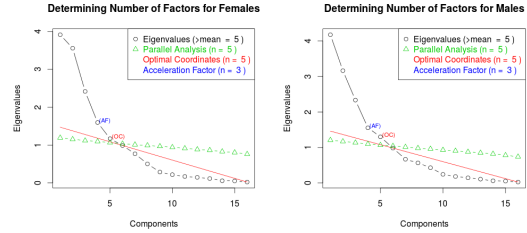

Figure 1: *Determining Number of Factors*

of Speech of the two speakers over time, the number of *functional* and *content* words also used in the other's previous turn and *laughter* that immediately proceeds the other's laugh.

4. Exploratory Analysis

Before attempting to classify speakers as funny or not, we examine the underlying dimensions along which the data varies. To do, so we use exploratory factor analysis. To determine the number of factors, we use a scree test and various non-graphical measures, including parallel analysis and an optimal coordinates test, as described in (CITE). We find an optimal number of factors $k = 5$ for both males and females, conducted separately. Figure 1 shows the two scree plots.

Interestingly, although we find that the various factors for male and female speech are similar, they explain different amounts of variation in the data. For males, the first factor reflects high intensity values and low intensity variation, explaining

5. Evaluation Approach

5.1. Methodology

We divide the dataset into two separate groups based on gender. For each gender, we train a classifier for detecting the perception of funny and courteous characteristics.

We create a training set for a given characteristic for a given gender by splitting the dataset into deciles. We use the top decile as positive and the bottom decile as negative examples.

5.2. Classifiers

We start by using an SVM with a linear kernel. To improve our classifications, we next enforce an L1 norm on the SVM to compensate for the large number of lexical and prosodic features we

extract. In case our decision boundaries are non-linear, we also train a RBF kernelized SVM. However, we lose interpretability of the features by using an RBF. So, we train a fourth Adaboost Classifier with decision stumps of unary depth to maintain interpretability while also allowing for a non-linear decision boundary.

6. Classification Results

7. Analysis

8. Future Work

9. Conclusions

10. Acknowledgements

We thank Dan Jurafsky for giving us access to the Speed Dating Dataset and for his continued guidance during this project.

11. References

[1] Ranganatha, R., Jurafsky, D., McFarland, D.A., "Detecting friendly, flirtatious, awkward, and assertive speech in speed-dates", in Computer Speech & Language Volume 27, Issue 1, Pages 89-115, 2013

[2] Natale, M., "Convergence of mean vocal intensity in dyadic communication as a function of social desirability", in Journal of Personality and Social Psychology 32 (5), 790-804, 1975

[3] Ireland, M.E., Slatcher, R.B., Eastwick, P.W., Scissors, L.E., Finkel, E.J., Pennebaker, J.W., "Language style matching predicts relationship initiation and stability", in Psychological Science 22 (1), 39, 2011

[4] Levitan, R., Hirschberg, J., "Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions", in ACL 2011