Identifying Emotions, Intentions, and Attitudes in Text
Using a Game with a Purpose
Lisa Pearl & Mark Steyvers, University of California, Irvine

How useful is the database?
To demonstrate the utility of the created database for developing computational approaches to social information identification in text, we applied a Sparse Multinomial Logistic Regression (SMLR) classifier (Krishnapuram et al. 2005) to the subset of the Crowd-Labeled messages where the intended social information was perceived (624 messages).

First-pass measure: use shallow features
• unigrams, bigrams, and trigrams
• number of word types, word tokens, and sentences
• number of exclamation marks, question marks, and punctuation marks
• average sentence and word length
• word type to word token ratio
• average word log frequency for words appearing more than once in the database


The SMLR classifier model was trained to produce the label (one of eight) corresponding to the generated social information using all the features as input. Using a 10-fold cross-validation procedure, the model was trained on 90% of the messages and tested on the remaining 10%.

Confusion matrix for the SMLR classifier:
The rows represent the intended social information for a message while the columns represent the labeled social information, averaged over messages and participants.

GWAP Results

How well do our non-experts do individually?
How well do they express the intended social information?
56% correctly expressed on average

How well do they perceive the intended social information?
58% correctly labeled on average

How reliable are the messages?
Confusion matrix for the human participants. The rows represent the intended social information for a message while the columns represent the labeled social information, averaged over messages and participants.

Diagonal represents correctly perceived messages.
Average correct: 57%.

Why so low? Some messages are better than others at expressing social information in a way obvious to humans.

Main Questions
1. Do we observe a “wisdom of the crowds” effect for identifying social information in text? That is, can pooling non-expert opinion yield something more reliable?
2. Can we construct a useful database for social information in text using a GWAP?

GWAP Design

Participants:
Game players

Motivation:
Getting points in the game for both generating messages that are easy to label and correctly labeling previously generated messages.

Social information types explored:
- politeness
- rudeness
- embarrassment
- formality
- persuading
- deception
- confidence
- disbelief

Asynchronous game play

Context pictures:
Generic context pictures randomized, could be used for any prompt

Labeling:
Messages labeled by multiple participants, only one label per message, participants asked to label more often than generate messages.

Taboo words:
Morphological variants of the social information type

Social Information Generated Message

deception
dehception
emabrrassment
embarrasement
disbelief
disbelieve
rudenes
persuasing
politeness
decception
dehception
emabrrassment
embarrasement
disbelief
disbelieve
rudenes
persuasing
politeness

GWAP-created databases are useful
• When the classifier makes the same mistakes as humans do, this suggests humans might be using similar shallow linguistic features to make their decisions. A classifier trained on these shallow features may be able to predict which messages are likely to be confusing for humans and what linguistic features are diagnostic.
• Since the social information types we used in our GWAP can be identified automatically with some success, this suggests that these social information types are useful to pursue with automatic classification techniques.
• Future Work: The GWAP methodology is easy to extend to create large-scale databases in both English and other languages, via online versions of the GWAP that run as web applications.