LieCatcher: Game Framework for Collecting Human Judgments of Deceptive Speech

Sarah Ita Levitan, James Shin, Ivy Chen, Julia Hirschberg
Department of Computer Science
Columbia University

Games4NLP 2018
May 7, 2018

Outline

- Motivation
- Corpus
- Game design
- LieCatcher demo
- Pilot study
- Conclusions and future work

Motivation

- Goal: large-scale study of characteristics of trustworthy speech
- Problem: we need human annotation!
- Possible solutions:
  - Expert annotation
  - Crowdsourcing
  - Game With a Purpose

GWAP Advantages

- More engaging format than monotonous annotation tasks
- Built-in incentives
- Affordable/free annotation
- Easy to distribute, accessible

Columbia X-Cultural Deception (CXD) Corpus

- 340 dialogues between 170 interviewer/interviewee pairs
  - Interviewer asks 24 biographical questions
  - Interviewee lies for random half, tells truth for the rest

- Native speakers of SAE and MC, all speaking in English
- 7,141 question/answer pairs
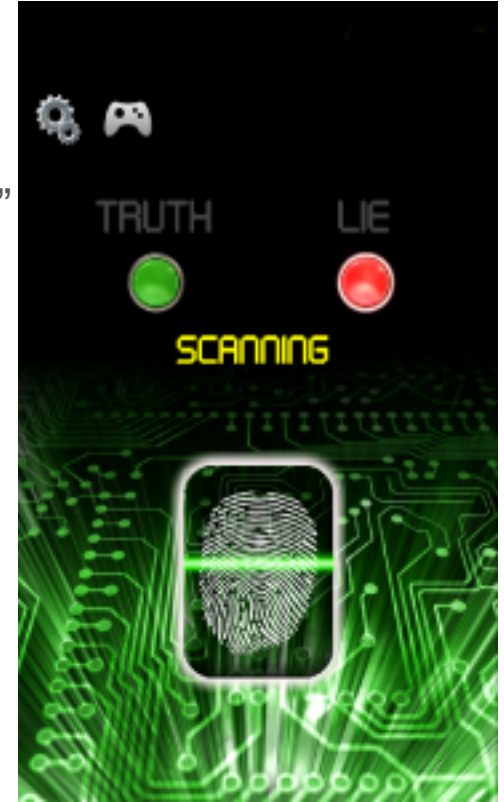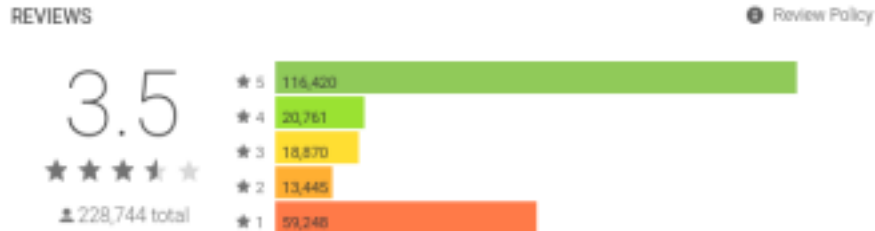- Orthographic transcription (Amazon Mechanical Turk)

"Have you ever tweeted?"

TRUE or FALSE?

"Have you ever tweeted?"

Lie Detection Games

- Amusing to assess lie detection ability
- Popular lie detection app: "Lie Detector Simulator Fun"
    - 10,000,000+ installs on Google Play alone
    - 3.5/5 rating, 228,744 reviews

Game Framework

- PhaserJS
- MongoDB backend
- JS fetch API as request handler to save player behavior information

Pilot Study

- Early feedback about game design
- 40 student participants
- Pre and post game surveys
- 2 levels – with and without instant feedback
- Quality control questions

Survey Responses

- Positive feedback!
  - 85% found game easy to use
  - 75% might or would definitely recommend to friend
  - 73% preferred level 2 – with instant feedback
  - 70% liked the premise of the game

Player Behavior

- Player accuracy: 49.86%
  - Level 1: 45.66%
  - Level 2: 54.44%
- 100% correct answer for quality control questions
- Some questions were "easier" than others
- Some samples were more "trusted" than others
  - But no clear consensus on "mistrusted" segments
- Gender differences

Conclusions

- LieCatcher GWAP
  - Assesses ability to detect deception
  - Collects human annotations of deception
- Pilot study with initial positive feedback

Next steps

- Incorporate feedback from pilot study
- Distribute game to wide audience (initial study with Amazon Mechanical Turk)
- Study acoustic-prosodic characteristics of trustworthy speech!

Thank you!

Questions?