User-Level Race and Ethnicity Predictors from Twitter Text

Daniel Preotiuc-Pietro, Lyle Ungar

Penn
University of Pennsylvania

Task

User-level Race/Ethnicity* prediction from tweets

* as defined by the US census
Four major classes:
White/Caucasian, African-American, Hispanic/Latino, Asian

Race affects:
- health
- voting patterns
- preferences

Previous data collection

Use perceived race labels (annotation)
- annotation is inaccurate using limited data about the user
- annotation reflects human stereotypes about user traits (Flekova et al 2016)

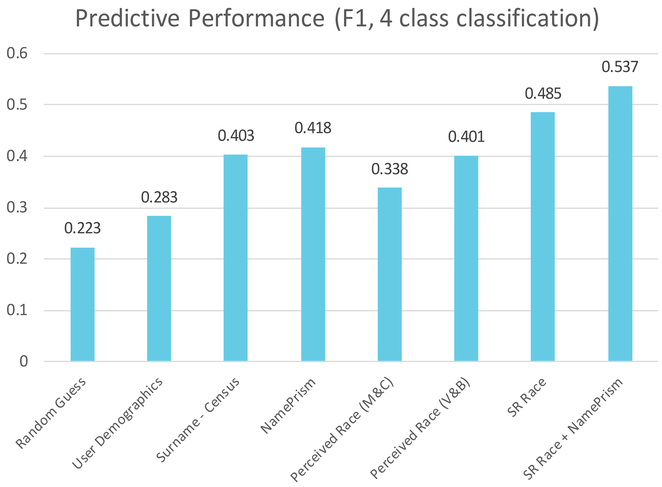
Examples:
(Mohammady & Culotta 2014) - M&C 2014
(Volkova & Bachrach 2016) - V&B 2016

Map geo-located tweets to census statistics (distant supervision)
- census statistics are outdated
- Twitter population is not representative
- users who geolocate tweets are not representative

Example:
(Blodgett et al 2016)

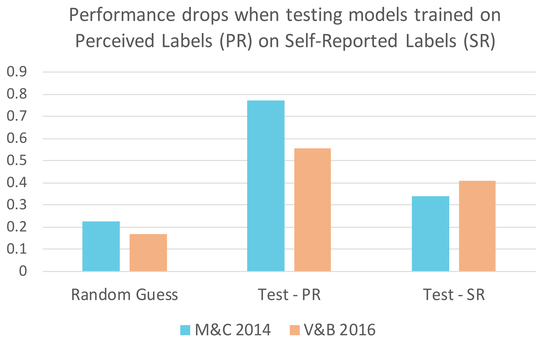
Our data collection

Use self-reported race labels
- Self-reported by users through surveys
- 4132 Twitter accounts
- 5.4 Million tweets
- only U.S. users
- several other demographic traits:
  age, gender, education, income



Experimental setup

- Logistic Regression classifier with L1 regularization
- Trained with unigram features
- Tested on users with self-reported race labels



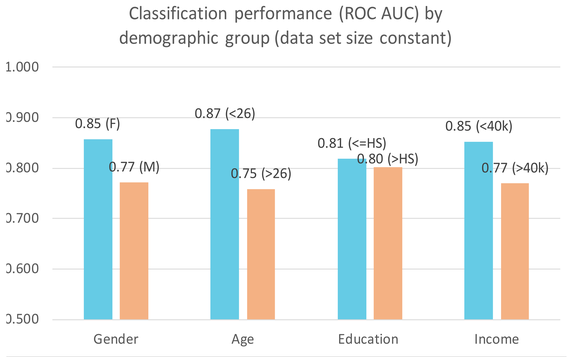
Feature analysis

White/Caucasian
- temporal references
- superlatives and adverbs of degree
- modals
- lower anger

African-American
- AAVE slang
- g-dropping
- verbs
- higher anger, disgust, fear, sadness

Hispanic/Latino
- Spanish words and S. American place-names
- polite words
- vove
- frustration
- lower positive and negative emotion

Asian
- Asian language words and placenames
- elongated 'yes'
- phonetic spellings
- interjections



Applications

- Covariate in Twitter-based community-level analyses
- Identify racial biases
- Passive polling

Predicting race of individual users in commercial applications is against Twitter TOS

Data availability

- models and aggregate data available
- user traits available only upon written request for academic purposes (sensitive user information)

The paper also covers

- One-vs-all classification results
- Cross data set analyses
- Evaluation of distant supervised collection methods
- Full feature analysis
  - LIWC
  - POS tags
  - Sentiment
  - Emotions
  - Topics (word2vec)

Take aways

- Models for predicting user race from tweets for the four most important US racial/ethnic groups
- Adding name information improves text-based prediction
- Label collection methods can
  - oversimplify prediction
  - bias predictive results
- Predicitive accuracy differs across demographic groups