Celebrity Profiling

What this is about

Celebrity profiling is author profiling applied to celebrities.

- They are prolific social media users, supplying lots of writing samples.
- Lots of personal details are public knowledge.
- They build a consistent public persona, either themselves or with the help of agents.
- A number of demographics apply only to this population.

Celebrities are a great population to study!

Contributions

1. We built a large corpus of celebrity profiles by matching Twitter usernames with Wikidata items.
2. We compared profiling celebrities with the SOTA on the general population and held a competition on profiling. With this, we showed where celebrity profiling works and where it does not.
3. We obtained some insights into celebrities on Twitter by analyzing our corpus.

Corpus Construction

1. Find authors
   We started with a list of 297,876 verified Twitter accounts.

2. Link to Wikidata
   We then created several candidate names, resolved them via Wikipedia in different languages and acquired the corresponding Wikidata items.

3. Verify matches
   We removed matches of non-human and memorial accounts, detected errors, and ambiguous accounts.

4. Evaluate matching
   For evaluation we reversed the procedure: acquired verified tweeters from Wikidata and counted misses and errors.

Experiment Results

Can we profile celebrities like other authors?

We profiled the benchmark demographic gender on four general population datasets, the respective SOTA models, and our own data and model. Results are comparable, no matter which data we trained on.

We held a competition at PAN to predict four demographics of celebrities. The performances of the eight submitted algorithms show:

What works?

- Binary gender, as usual
- Distinguishing the most from the least famous celebrities
- Predicting the occupations sports, politics, and performers
- Age in the range of ca. 20–40 years

What does not work?

- Rare demographics: non-binary gender and religious occupations
- Cross-topic occupations: i.e. managers and scholars
- Ages outside the range of ca. 20–40 years

Corpus Analysis

Some insights into the population of celebrities on Twitter:

- Young men are most often famous for doing sports.
- Most young women are actors or musicians.
- Politics and management are an old man’s business.
- The occurrence ratio of occupations is more balanced for celebrities on Twitter older than 40.

Matti Wiegmann, Benno Stein, Martin Potthast

Web Technology and Information Systems