Discriminating Gender on Twitter

David Arroyo Menéndez

March 8, 2020

Presentation

Discriminating Gender on Twitter John D. BurgerThe MITRE Corporation, Bedford, Massachusetts John HendersonThe MITRE Corporation, Bedford, Massachusetts George KimThe MITRE Corporation, Bedford, Massachusetts Guido ZarrellaThe MITRE Corporation, Bedford, Massachusetts
URL: https://dl.acm.org/ft_gateway.cfm?id=2145568&ftid=1146137&dwn=1&CFID=96718218&CFTOKEN=5ad99f87ffded6b-6BA0215F-CA4F-5037-02956322E689CADF Citations: 575 Published on: EMNLP '11 Proceedings of the Conference on Empirical Methods in Natural Language Processing

Abstract

Social Value

Accurate prediction of demographic attributes from social media and other informal online content is valuable for marketing, personalization, and legal investigation.

Brief paper description

This paper describes the construction of a large, multilingual dataset labeled with gender, and investigates statistical models for determining the gender of uncharacterized Twitter users.

Technical Solution and Advance in Science

We explore several different classifier types on this dataset. We show the degree to which classifier accuracy varies based on tweet volumes as well as when various kinds of profile metadata are included in the models. We also perform a large-scale human assessment using Amazon Mechanical Turk. Our methods significantly out-perform both baseline models and almost all

Introduction

Background:

The rapid growth of social media in recent years, exemplified by Facebook and Twitter, has led to a massive volume of user-generated informal text. This in turn has sparked a great deal of research interest in aspects of social media, including automatically identifying latent demographic features of online users.

Gap

Many latent features have been explored, but gender and age have generated great interest (Schler et al., 2006; Burger and Henderson, 2006; Argamon et al., 2007; Mukherjee and Liu, 2010; Rao et al., 2010). Accurate prediction of these features would be useful for marketing and personalization concerns, as well as for legal investigation.

Technical solution

In this work, we investigate the development of highperformance classifiers

Data

Twitter Description

Twitter is a social networking and micro-blogging platform whose users publish short messages or tweets. In late 2010, it was estimated that Twitter had 175 million registered users worldwide, producing 65 million tweets per day (Miller, 2010). Twitter is an attractive venue for research into social media because of its large volume, diverse and multilingual population, and the generous nature of its Terms of Service. This has led many researchers to build corpora of Twitter data (Petrovic et al., 2010; Eisenstein et al., 2010). In April 2009, we began sampling data from Twitter using their API at a rate of approximately 400,000 tweets per day. This represented approximately 2% of Twitter's daily volume at the time, but this fraction has steadily decreased to less than 1% by

1. This decrease is because we sample roughly the

same number of tweets every day while Twitter's overall volume has increased markedly. Our corpus thus far contains approximately 213 million tweets from 18.5 million users, in many different languages

Experiments

[. . .] The sheer volume of data presents a challenge for many of the available machine learning toolkits, e.g. WEKA (Hall et al., 2009) or MALLET (McCallum, 2002) [. . .]
[. . .] We performed initial feasibility experiments using a wide variety of different classifier types, including Support Vector Machines, Naive Bayes, and Balanced Winnow [. . .]

Field combinations

We performed a number of experiments with the Winnow algorithm described above. We trained it on the training set and evaluated on the development set for each of the four user fields in isolation, as well as various combinations, in order to simulate different use cases for systems that perform gender prediction from social media sources. [. . .]

Human performance

We wished to compare our classifier's efficacy to human performance on the same task. A number of researchers have recently experimented with

Results

Field combinations

Human performance

Self-training

Conclusion

What

In this paper, we have presented several configurations of a language-independent classifier for predicting the gender of Twitter users. The large dataset used for construction and evaluation of these classifiers was drawn from Twitter users who also completed blog profile pages.

Technical result (percentages)

These classifiers were tested on the largest set of gender-tagged tweets to date that we are aware of. The best classifier performed at 92% accuracy, and the classifier relying only on tweet texts performed at 76% accuracy. Human performance was assessed on this latter condition, and only 5% of 130 humans performed 100 or more classifications with higher accuracy than this machine.

Future Work

In future work, we will explore how well such models carry over to gender