ML1819 Research Assignment 1

Team 47

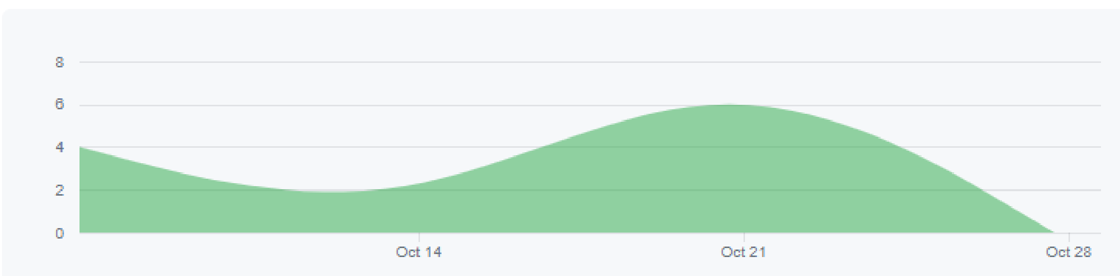
107. How well can the gender of Twitter users be predicted?

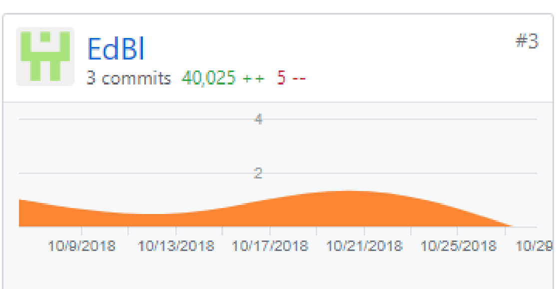
15326006 Chia (Jason) Lau	Research, classifier design, writing
15325303 Rory Hughes	Research, data download and enrichment, writing
15323612 Edvardas Blazevicius	Research, data collection, writing, report editing

Word count: 936

Source code repository:
https://github.com/RoryDH/ML1819--task-107--team-47

Source code repository activity:
https://github.com/RoryDH/ML1819--task-107--team-47/graphs/contributors





Predicting a Twitter User's Gender Using Machine Learning

Chia J. Lau
lauch@tcd.ie

Rory Hughes
hughesr8@tcd.ie

Edvardas Blazevicius
blazevie@tcd.ie

1 INTRODUCTION

The vast amounts of user generated data available on social networks is multiplying year on year at an ever increasing rate. With many social networks being over a decade old, this makes them bottomless wells of user generated data.

Twitter does not ask its users for their gender. Armed with megabytes of other data on each member, it would be interesting to see how well their gender could be predicted. On top of that it would be interesting to investigate what subsets of that data are the best predictors of their gender.

2 RELATED WORK

A lot of work has already gone into trying to classify a user's gender based on their Twitter data.
You et al. [1] attempted predicting a social media user's gender using the images they post.
Zamal et al. [2] and Burger et al. [3] tried predicting a Twitter user's gender by applying an n-gram feature model to the user's data such as their real name, username, bio and tweets.
Tenuto of Figure Eight [4] tried predicting whether a Twitter user is male, female or a brand by analyzing key words and emoji's in their tweets.
Liu and Ruths [5] attempted predicting a Twitter user's gender using their first name only.
Rao et al. [6] tried predicting a Twitter user's gender by analyzing the sentiment of their tweets.
Alowidbu et al. [7] attempted predicting a Twitter user's gender by analyzing their use of colors as well as the phonemes found in their names.

3 METHODOLOGY

We found a dataset [9] that included 14,000 Twitter users and their already-classified genders. We stripped the rest of the user data and only kept the username and the gender. Using the Twitter API, we collected the data of 9,500 of these users and stored it in a JSON. About 7,000 of these were male or female and the rest were brands or unknowns.

We later found another list of 40,000 twitter usernames [10]. After downloading all of their profiles, we combined the datasets and used a gender guesser module [8] that takes in a first name and gets the gender for that name, supplying also how confident it is. From that list, we only used names which it believed were *definitely* male or *definitely* female.

We tried a few different approaches to building a classifier:

1. In our first attempt we tried to see how well using only the user's first name could predict their gender. We broke the name strings into a list of ASCII ordinals. We fed the names labeled with known genders into a sequential model consisting of an embedding layer, a pooling layer and one hidden dense layer.
2. Our second attempt added more features including users' last name, full name and screen name. We also added the colours they use for their profile page. We pre-processed the data so that the name characters were a range between -1 and 1 and translated the hex string to three RGB channels each between 0 and 1. A dropout layer was also added to prevent overfitting.
3. In our third iteration we used the same data points but translated the name strings into phonetic tokens (known as phonemes).
4. Our final attempt added the bio of users, storing each word in a common dictionary of occurrences. The layout of the pipeline was also updated so that names were fed into LSTMs. The outputs were merged with the user bio and colours data similar to our third attempt.

For each attempt we used TensorFlow's Sequential Keras with a learning rate of 0.001 and the sparse categorical cross entropy loss function.

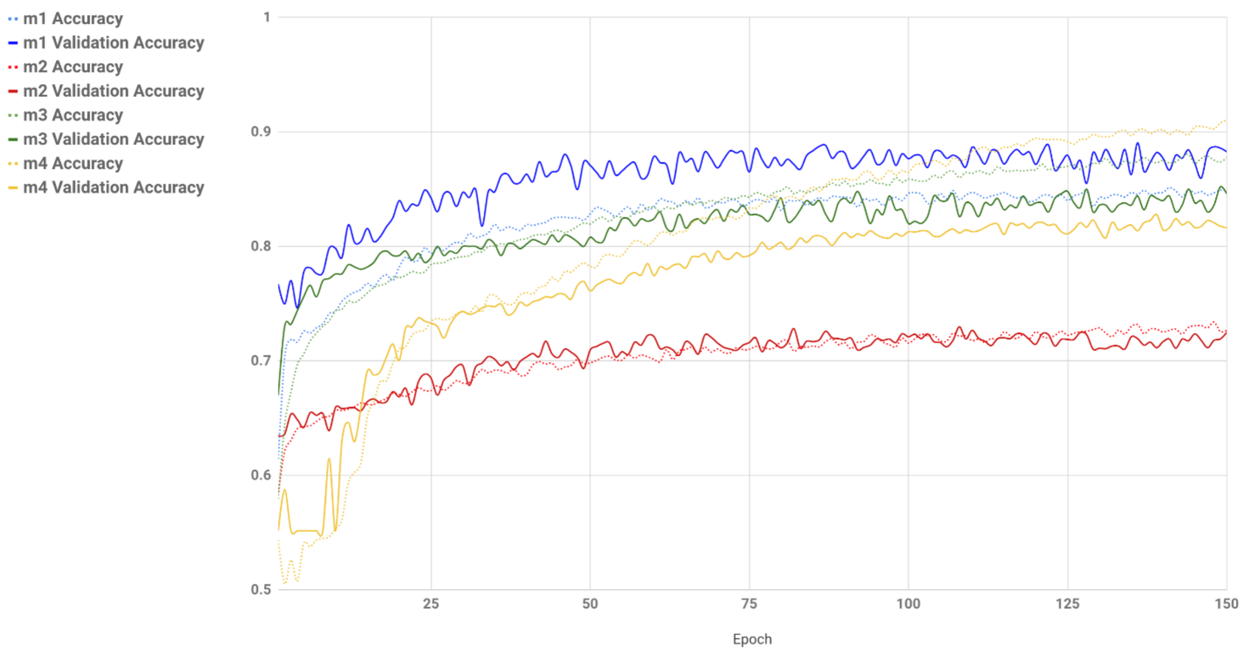


Figure 1: Training the Four Methods

4 RESULTS AND DISCUSSION

Table 1: Prediction accuracy of various methods used

Method	Accuracy	Loss
1	87.7%	.308
2	71.2%	.579
3	83.6%	.370
4	82.0%	.503

Table 2: Results from referenced research papers

Research Paper	Best Method's Accuracy
Zamal et al. [2]	80.2%
Burger et al. [3]	92.0%
Liu et al. [5]	87.1%
Rao et al. [6]	72.3%
Alowibdi et al. [7]	82.5%

A Twitter user's gender can be predicted with a high enough amount of confidence using a small subset of data from their user profiles. Using only a user's name it can be predicted with 87.7% confidence using a dataset of 40,000 users' information as training data.

Initially method 4 was most successful but its accuracy plateaus around 82% and it appears to overfit very early. Our first method worked better for the larger data set in the end. We believe that this is because it is harder to find patterns unique to genders amongst the other features we considered (bio text, profile colour, screen name).

Method 3 was the second most effective of the methods used. This is consistent with the findings of Alowidbi et Al. [7] who in their paper used a similar method with n-grams applied to phonemes. They used a dataset of a much smaller size compared to other researchers, but achieved comparable results to the other researcher's methods.

5 LIMITATIONS AND OUTLOOK

For the first part of the project we were using a small data set of 40,000 users - the next smallest dataset that we saw in the referenced research papers had 150,000 users. We would like to increase our dataset by at least a factor of three.
We would also like to apply the same phoneme analysis to the user's most popular tweets to see if that helps the accuracy in any way. We would like try to analyzing the sentiment of users' bios and tweets to see if that would help with the accuracy. On top of that we would like to analyze each user's profile picture for color, objects, maybe facial features and see if that would help with the accuracy.

REFERENCES

[1] Q. You, S. Bhatia, T. Sun and J. Luo, "The Eyes of the Beholder: Gender Prediction Using Images Posted in Online Social Networks," 2014 IEEE International Conference on Data Mining Workshop, Shenzhen, 2014, pp. 1026-1030.
DOI: 10.1109/ICDMW.2014.93

[2] Al Zamal, Faiyaz & Liu, Wendy & Ruths, Derek. (2012). Homophily and Latent Attribute Inference: Inferring Latent Attributes of Twitter Users from Neighbors.

[3] L. Burger, J. Henderson, G. Kim, and G. Zarrella, "Discriminating
gender on Twitter". In Proceedings of EMNLP'11 1301–1309, 2011.

[4] Justin Tenuto. Figure Eight. *Using machine learning to predict gender.* https://www.figure-eight.com/using-machine-learning-to-predict-gender [Accessed October 2018]

[5] Liu, Wendy and Derek Ruths. "What's in a Name? Using First Names as Features for Gender Inference in Twitter." AAAI Spring Symposium: Analyzing Microtext (2013).

[6] Delip Rao, David Yarowsky, Abhishek Shreevats, and Manaswi Gupta. 2010. Classifying latent user attributes in twitter. In Proceedings of the 2nd international workshop on Search and mining user-generated contents (SMUC '10). ACM, New York, NY, USA, 37-44.
DOI: https://doi.org/10.1145/1871985.1871993

[7] Alowibdi, Jalal & Buy, Ugo & Yu, Philip. (2013). Empirical Evaluation of Profile Characteristics for Gender Classification on Twitter. 1. 365-369. 10.1109/ICMLA.2013.74.

[8] gender-guesser, pip module, https://pypi.org/project/gender-guesser/

[9] Dataset with users with pre-classified genders, https://www.kaggle.com/crowdflower/twitter-user-gender-classification

[10] Large dataset containing Twitter usernames, https://www.kaggle.com/hwassner/TwitterFriends