

Master's Thesis Essay

Bjørn-Andreas Lamo

Spring 2021

1 Introduction

In this thesis we will discuss the viability and applicability of state of the art machine learning models inferring gender from a Reddit users comment history. BERT has been the state of the art model for NLP tasks for a few years,[1] and has already been used for author profiling.[2][3][4]

However this paper will take a different approach than previous work and instead of making *one model* for all the classification, we will explore making models per interest or topic. Potentially with many models we could avoid the model learning that certain keywords are incredibly gendered, and force it to learn more subtle linguistic features.

Reddit is an ideal source for text because the content is already sorted by topic, and it is easy to retrieve all or many of a users comments site wide. The users are in principle anonymous, but voluntarily identify their gender in various ways. We will take advantage of this to extract a large number of gendered users, and then fetch their comment history.

2 Author Profiling

Gender classification based on posted text on the internet has been in academic interest for

decades[5].

3 Reddit

Reddit is a popular site for sharing content and discussions. In 2020 alone there were 2 billion user comments on the site[6]. A user can make a post, and other users can upvote/downvote or comment and discuss the post. The votes a post receive is based on how relevant it is to the specific subreddit it is submitted to. Reddit is divided into many subreddits, this is how they subdivide content for different interests. For maximum upvotes the key is to submit quality content to the relevant subreddit. Reddit users are very good at policing which post belong in the subreddit by means of the voting system.

In a similar manner, the comment section is acute in discussion relevant to the post.

4 BERT

5 Previous works

6 Methodology

References

- [1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805, 2018.
- [2] Roberto López-Santillán, Manuel Montes-Y-Gómez, Luis Carlos González-Gurrola, Graciela Ramírez-Alonso, and Olanda Prieto-Ordaz. Richer Document Embeddings for Author Profiling tasks based on a heuristic search. *Information Processing & Management*, 57(4):102227, July 2020.
- [3] Marco Polignano, Marco de Gemmis, and Giovanni Semeraro. Contextualized bert sentence embeddings for author profiling: The cost of performances. In Osvaldo Gervasi, Beniamino Murgante, Sanjay Misra, Chiara Garau, Ivan Blečić, David Taniar, Bernady O. Apduhan, Ana Maria A. C. Rocha, Eufemia Tarantino, Carmelo Maria Torre, and Yeliz Karaca, editors, *Computational Science and Its Applications – ICCSA 2020*, pages 135–149, Cham, 2020. Springer International Publishing.
- [4] Rosa María Ortega Mendoza, Anilú Franco Árcega, and Manuel Montes y Gómez. Author Profiling on Social Media using New Weighting Schemes that Emphasize Personal Information. *Computación y Sistemas*, 23(2):501–510, June 2019.
- [5] M. Koppel. Automatically Categorizing Written Texts by Author Gender. *Literary and Linguistic Computing*, 17(4):401–412, November 2002.
- [6] Reddit’s 2020 year in review. <https://redditblog.com/2020/12/08/reddits-2020-year-in-review/>.