Authorship Attribution with Document Embeddings and Neural Networks

1st Miles Baer
Dept. of Computer Science
Middle Tennessee State University
Murfreesboro, United States
mtb3x@mtmail.mtsu.edu

2nd Nathaniel Boyer
Dept. of Computer Science
Middle Tennessee State University
Murfreesboro, United States
njb3n@mtmail.mtsu.edu

3rd William Cope
Dept. of Computer Science
Middle Tennessee State University
Murfreesboro, United States
wrc2t@mtmail.mtsu.edu

4th Charles Johnson
Dep. of Computer Science
Middle Tennessee State University
Murfreesboro, United States
cwj2z@mtmail.mtsu.edu

5th Robert Smith
Dep. of Computer Science
Middle Tennessee State University
Murfreesboro, United States
rws2p@mtmail.mtsu.edu

***Abstract*—Our goal is to determine authorship of short texts using neural networks and document level encodings. Neural networks (NNs) are capable of finding correlations between inputs and their expected outputs. This allows them to perform quite well at classification tasks such as authorship attribution where the inputs would be documents and the classifications their respective authors. Currently, Word2Vec and Doc2Vec are popular methods used to build vocablary models. The models can then be used to encode words and documents respectively into fixed sized vectors containing floating point values. We propose various methods that utilize Word2Vec and Doc2Vec as a means of building document level encodings that retain the stylistic elements of the author. Primarily, we used three different NN architectures inspired by previous work in authorship attribution and sentiment analysis research. With each network we tested multiple methods of encoding amazon reviews, treating each review as a document whose classification was the author. Testing shows that document embeddings for short texts are not able to outperform previous methods.**

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

Start typing here [1].

II. Background

Start typing here [2].

III. Methods

Start typing here [3].

IV. Results

Start typing here [4].

V. Discussion

Start typing here [5].

References

[1] L. D. Trang and Z. Mebkhout, "Variétés caractéristiques et variétés polaires," *C. R. Acad. Sc. Paris*, vol. 296, pp. 129–132, 1983.
[2] J. Birançon, P. Maisonobe, and M. Merle, "Localisation de systèmes différentiels, stratifications de Whitney et condition de Thom," *Invent. Math.*, vol. 117, pp. 531–550, 1994.
[3] A. Parusiński and P. Pragacz, "A formula for the Euler characteristic of singular hypersurfaces," *J. Alg. Geom.*, vol. 4, pp. 337–351, 1995.
[4] K. Elissa, "Title of paper if known."
[5] R. Nicole, "Title of paper with only first word capitalized," *J. Name Stand. Abbrev.*