

{tag} International Journal of Computer Applications
Foundation of Computer Science (FCS), NY, USA

[Volume 169](#)

-
[Number 1](#)

Year of Publication: 2017

Authors:

O. Srinivasa Rao, N. V. Ganapathi Raju, V. Vijaya Kumar

10.5120/ijca2017914587

{bibtex}2017914587.bib{/bibtex}

Abstract

Authorship attribution is one of the important problem, with many applications of practical use in the real-world. Authorship identification determines the likelihood of a piece of writing produced by a particular author by examining the other writings of that author. Every author has a unique style of writing pattern. This paper identifies the unique style of an author(s) using lexical stylometric features including function words using balanced training corpus. The present paper calculates the frequencies of the lexical based stylometric features by balancing training and test corpus on English editorial documents. The present paper compares various machine learning algorithms for the authorship attribution and achieved highest average accuracy 95.58 using Random Forest classifier and 92.59 using Multilayer Perceptron algorithms.

References

1. Zheng, Rong, et al. "A framework for authorship identification of online messages: Writing-style features and classification techniques." Journal of the American Society for

Information Science and Technology 57.3 (2006): 378-393.

2. Stamatatos, Efstathios. "Author identification: Using text sampling to handle the class imbalance problem." *Information Processing & Management* 44.2 (2008): 790-799.
3. Grieve, Jack. "Quantitative authorship attribution: An evaluation of techniques." *Literary and linguistic computing* 22.3 (2007): 251-270.
4. López-Monroy, Adrián Pastor, et al. "A new document author representation for authorship attribution." *Mexican Conference on Pattern Recognition*. Springer Berlin Heidelberg, 2012.
5. Luyckx, Kim, and Walter Daelemans. "The effect of author set size and data size in authorship attribution." *Literary and linguistic Computing* 26.1 (2011): 35-55.
6. Stamatatos, Efstathios. "A survey of modern authorship attribution methods." *Journal of the American Society for information Science and Technology* 60.3 (2009): 538-556.
7. Cheng, Na, Rajarathnam Chandramouli, and K. P. Subbalakshmi. "Author gender identification from text." *Digital Investigation* 8.1 (2011): 78-88.
8. Layton, Robert. "A Simple Local n-gram Ensemble for Authorship Verification." *CLEF*. 2014.
9. Wei, Qiong, and Roland L. Dunbrack Jr. "The role of balanced training and testing data sets for binary classifiers in bioinformatics." *PloS one* 8.7 (2013): e67863.
10. Van Halteren, Hans, et al. "New machine learning methods demonstrate the existence of a human stylome." *Journal of Quantitative Linguistics* 12.1 (2005): 65-77.
11. V. Vijaya Kumar, N V Ganapathi Raju, O Srinivasa Rao, "Histograms of Term Weight Feature (HTWF) model for Authorship attribution", *International Journal of Applied Engineering Research (IJAER)*, vol10, number 16 , pp 36622-36628, ISSN 0973-4562, 2015
12. N V Ganapathi Raju, V. Vijaya Kumar, O Srinivasa Rao, "Authorship attribution of Telugu texts based on Syntactic features and Machine learning techniques", *Journal of Theoretical and Applied Information Technology (JATIT)*, volume 85, No.1, ISSN: 1992-8645, march 2016
13. N V Ganapathi Raju, V. Vijaya Kumar, OSrinivasa Rao, "Author based Rank Vector Coordinates (ARVC) model for Authorship attribution", *International Journal of Image, Graphics and Image Processing (IJIGSP)*, Vol. 8, No. 5, May 2016.
14. McMenamin, Gerald R. "Style markers in authorship studies." *International Journal of Speech Language and the Law* 8.2 (2007): 93-97.
15. Stamatatos, Efstathios. "Text Sampling and Re-Sampling for Imbalanced Authorship Identification Cases." *Frontiers in Artificial Intelligence and Applications* 141 (2006): 813.
16. Zhao, Ying, and Justin Zobel. "Searching with style: Authorship attribution in classic literature." *Proceedings of the thirtieth Australasian conference on Computer science-Volume 62*. Australian Computer Society, Inc., 2007.
17. Eder, Maciej. "Style-markers in authorship attribution a cross-language study of the authorial fingerprint." *Studies in Polish Linguistics* 6.1 (2011): 99-114.
18. Sanderson, C., & Guenter, S. "Short text authorship attribution via sequence kernels, Markov chains and author unmasking: An investigation", In *Proceedings of the International Conference on Empirical Methods in Natural Language Engineering*, Pages 482-491, 2006.

Index Terms

Computer Science

Artificial Intelligence

Keywords

Authorship Clustering; Stylometry; Supervised Classification