

# Atribuição autoral de textos digitais

José Eleandro Custódio

Programa de Mestrado em Sistemas de Informação - PPgSI  
Universidade de São Paulo - USP

Novembro de 2018

# Informações gerais

- **Orientador:** Prof. Dr. Ivandré Paraboni
- **Semestre no curso:** 4o.
- **Qualificação:** 29/10/2018
- **Defesa:** realização planejada para 30/06/2019
- **Linha de pesquisa:** Inteligência de sistemas
- **Área de pesquisa:** Inteligência artificial
- **Área de aplicação:** Linguística computacional / Língua natural

# Agenda

1 Introdução

2 Referências

## Contexto

- A atribuição autoral de textos digitais (AA) (do inglês, *Authorship Attribution*) visa identificar quem é o autor de um determinado texto a partir de um conjunto de autores possíveis (POTTHAST et al., 2017).
- A premissa principal da AA é que o autor deixa rastros de seu estilo, sendo que esses rastros podem ser a preferência por certas palavras, o tamanho do vocabulário, a utilização de pontuação e a repetição de certos elementos gramaticais.
- Quantificar essas informações é uma tarefa conhecida por estilometria (STAMATATOS, 2009).

# Aplicações da Atribuição Autoral

## Aplicações da AA

Sua aplicação pode ajudar:

- em casos de escândalos de corrupção, como no caso Enron (KLIMT; YANG, 2004; CHEN et al., 2011).
- na identificação de abusos na utilização da internet (VARTAPETIANCE; GILLAM, 2012).
- na detecção de notícias falsas (PENG; CHOO; ASHMAN, 2016).
- na detecção de casos onde uma pessoa tenta se passar por outra (KOPPEL; SEIDMAN, 2018).
- na atribuição autoral de código-fonte (ALSULAMI et al., 2017)
- na detecção de pseudônimos (JUOLA, 2015)


## Área de interesse

- Humanidades Digitais
- Análise Forense
- Linguística computacional

# Métodos para atribuição autoral


Os métodos computacionais para atribuição autoral utilizam:

- Análise estatística multivariada (SAVOY, 2016; EVERT et al., 2017).
- Métodos baseados em vizinho mais próximo (KOCHER; SAVOY, 2017; KOPPEL; SEIDMAN, 2018; VARELA et al., 2016).
- Aprendizado de máquina com SVM (SCHWARTZ et al., 2013; STAMATATOS, 2017).
- Redes neurais recorrentes (BAGNALL, 2016).
- Redes neurais de convolução (SHRESTHA et al., 2017; SARI; STEVENSON, 2016). modelos de compressão (HALVANI; GRANER, 2018)


 ALSULAMI, B. et al. Source code authorship attribution using long short-term memory based networks. In: *Computer Security - ESORICS 2017 - 22nd European Symposium on Research in Computer Security, Oslo, Norway, September 11-15, 2017, Proceedings, Part I*. [s.n.], 2017. p. 65–82. Disponível em: [⟨https://doi.org/10.1007/978-3-319-66402-6\\\_6⟩](https://doi.org/10.1007/978-3-319-66402-6\_6).

 BAGNALL, D. Authorship clustering using multi-headed recurrent neural networks. In: Cappellato L. Ferro N., M. C. B. K. (Ed.). *CEUR Workshop Proceedings*. [S.l.]: CEUR-WS, 2016. v. 1609, p. 791–804. ISSN 16130073.

# Referências II


 CHEN, X. et al. Authorship similarity detection from email messages. In: *Machine Learning and Data Mining in Pattern Recognition - 7th International Conference, MLDM 2011, New York, NY, USA, August 30 - September 3, 2011. Proceedings.* [S.l.: s.n.], 2011. p. 375–386.


 EVERT, S. et al. Understanding and explaining delta measures for authorship attribution. *Digital Scholarship in the Humanities*, v. 32, n. suppl\_2, p. ii4–ii16, 2017.


 HALVANI, O.; GRANER, L. Cross-Domain Authorship Attribution Based on Compression: Notebook for PAN at CLEF 2018. In: CAPPELLATO, L. et al. (Ed.). *Working Notes Papers of the CLEF 2018 Evaluation Labs.* [S.l.]: CLEF and CEUR-WS.org, 2018. (CEUR Workshop Proceedings). ISSN 1613-0073.




# Referências III


 JUOLA, P. The rowling case: A proposed standard analytic protocol for authorship questions. *Digital Scholarship in the Humanitie*, v. 30, n. Suppl-1, p. i100–i113, 2015.


 KLIMT, B.; YANG, Y. The enron corpus: A new dataset for email classification research. In: BOULICAUT, J. et al. (Ed.). *Machine Learning: ECML 2004, 15th European Conference on Machine Learning, Pisa, Italy, September 20-24, 2004, Proceedings*. [S.l.]: Springer, 2004. (Lecture Notes in Computer Science, v. 3201), p. 217–226.

 KOCHER, M.; SAVOY, J. A simple and efficient algorithm for authorship verification. *Journal of the Association for Information Science and Technology*, v. 68, n. 1, p. 259–269, 2017.


 KOPPEL, M.; SEIDMAN, S. Detecting pseudepigraphic texts using novel similarity measures. *Digital Scholarship in the Humanities*, v. 33, n. 1, p. 72–81, 2018.

# Referências IV


 PENG, J.; CHOO, K. kwang R.; ASHMAN, H. Astroturfing detection in social media: Using binary n-gram analysis for authorship attribution. In: *2016 IEEE Trustcom/BigDataSE/ISPA*. [S.l.: s.n.], 2016. p. 121–128. ISBN 9781509032051.

 POTTHAST, M. et al. Overview of PAN'17: Author identification, author profiling, and author obfuscation. *Lecture Notes in Computer Science*, v. 10456 LNCS, p. 275–290, 2017. ISSN 16113349.


 SARI, Y.; STEVENSON, M. Exploring Word Embeddings and Character N -Grams for Author Clustering Notebook for PAN at CLEF 2016. *CEUR Workshop Proceedings*, 2016. ISSN 16130073.


 SAVOY, J. Estimating the probability of an authorship attribution. *Journal of the Association for Information Science and Technology*, v. 67, n. 6, p. 1462–1472, 2016. ISSN 23301643.

# Referências V


 SCHWARTZ, R. et al. Authorship Attribution of Micro-Messages. In: *Empirical Methods in Natural Language Processing*. [S.l.: s.n.], 2013. p. 1880–1891. ISBN 9781937284978.


 SHRESTHA, P. et al. Convolutional Neural Networks for Authorship Attribution of Short Texts. In: *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*. [S.l.]: Association for Computational Linguistics (ACL), 2017. v. 2, p. 669–674. ISBN 9781510838604.

 STAMATATOS, E. A survey of modern authorship attribution methods. *Journal of the American Society for Information Science and Technology*, v. 60, n. 3, p. 538–556, 2009. ISSN 15322882.

 STAMATATOS, E. Authorship attribution using text distortion. *Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017 - Proceedings of Conference*, v. 1, 2017.

# Referências VI

 VARELA, P. J. et al. A computational approach based on syntactic levels of language in authorship attribution. *IEEE Latin America Transactions*, v. 14, n. 1, p. 259–266, 2016. ISSN 15480992.

 VARTAPETIANCE, A.; GILLAM, L. Quite simple approaches for authorship attribution, intrinsic plagiarism detection and sexual predator identification. In: *CLEF 2012 Evaluation Labs and Workshop, Online Working Notes, Rome, Italy, September 17-20, 2012*. [S.l.: s.n.], 2012.