Can language help in the characterization of user behavior?  
Feature engineering experiments with Word2Vec

Eduardo Lopez*  
McMaster University  
Hamilton, Ontario  
lopeze1@mcmaster.ca

Kamran Sartipi  
East Carolina University  
Greenville, NC  
sartipik16@ecu.edu

ABSTRACT
Among the many significant advances in the area of deep learning−use of deep neural networks −, the Natural Language Processing (NLP) space holds a special place. The availability of very large datasets, along with the existence of powerful computing environments have created a fascinating environment for researchers. One of the algorithms recently created is Word2Vec − it enables the creation of embeddings: low-dimensional, meaningful representation of language that can be used for machine learning tasks such as prediction or classification. In this study, we experiment with Word2Vec and apply it to a different domain: representation of user behavior in information systems. We demonstrate that several feature engineering tasks for user behavior characterization can be enriched by the use of NLP concepts.

CCS CONCEPTS
• Security and privacy → Formal methods and theory of security; Domain-specific security and privacy architectures; Human and societal aspects of security and privacy.

KEYWORDS
feature engineering with word2vec; natural language processing in user behavior; feature transformation in information systems use

Unpublished working draft. Not for distribution.

*Eduardo Lopez. DeGroote School of Business, McMaster University. Correspondence concerning this article should be addressed to: Eduardo Lopez 1280 Main St W, Hamilton, ON L8S 4L8. E-mail: lopeze1@mcmaster.ca

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

© 2020 Copyright held by the owner/author(s).

CASCON'20, Canada  
https://doi.org/10.475/123_4