

Words Embedding and News Topic Classification

Jean Lucien Randrianantenaina
Applied Mathematics, Machine Learning and Artificial Intelligence
Stellenbosch University

Abstract—abstract

I. INTRODUCTION

II. DATA AND DATA PREPARATION

III. WORDS EMBEDDING

Computers works principlly with numbers, to work with string it require to have a good representations.

A. Word2vec

B. Improved Word2vec with subsampling and negative sampling

IV. TEXT CLASSIFICATION

V. RESULTS AND DISCUSSION

VI. CONCLUSION

REFERENCES

- [1] Herman Kamper. *NLP817*. <https://www.kamperh.com/nlp817/>, 2022–2024.
- [2] Daniel Jurafsky and James H. Martin. *Speech and Language Processing*. 3rd edition, Stanford University and University of Colorado at Boulder, 2024.
- [3] *Compare languages - online genetic proximity calculator*. http://www.elinguistics.net/Compare_Languages.aspx, 2023.