Developing temporal word embeddings with Twitter

line 1: 1st Given Name Surname   
line 2: *dept. name of organization (of Affiliation)*  
line 3: *name of organization (of Affiliation)*line 4: City, Country  
line 5: email address or ORCID  
  
line 1: 2nd Given Name Surname  
line 2: *dept. name of organization (of Affiliation)*  
line 3: *name of organization (of Affiliation)*line 4: City, Country  
line 5: email address or ORCID  
  
line 1: 3rd Given Name Surname  
line 2: *dept. name of organization (of Affiliation)*  
line 3: *name of organization (of Affiliation)*line 4: City, Country  
line 5: email address or ORCID  
  
line 1: 4th Given Name Surname  
line 2: *dept. name of organization (of Affiliation)*  
line 3: *name of organization (of Affiliation)*line 4: City, Country  
line 5: email address or ORCID

line 1: 5th Given Name Surname  
line 2: *dept. name of organization (of Affiliation)*  
line 3: *name of organization (of Affiliation)*line 4: City, Country  
line 5: email address or ORCID

line 1: 6th Given Name Surname  
line 2: *dept. name of organization (of Affiliation)*  
line 3: *name of organization (of Affiliation)*line 4: City, Country  
line 5: email address or ORCID

*Abstract*— Human language is evolving, the meanings and associations between words is ever changing. Apple is a good example since once it was mostly known as a fruit, now it is also associated with the tech company. In this paper, we attempt to create dynamic word embeddings using data collected from twitter. By examining what words are closer to others we are able to infer semantic meaning. Then, creating a positive pointwise mutual information matrix, PPMI, we can attempt to align the time-separated matrices over time and examine how the semantic meaning can change over time. Then in the future, using data collected from twitter we may be able to detect changes in language associated with HIV/AID or the opioid epidemic and be potentially be able to predict outbreaks.

Keywords—natural language processing, big data, word embeddings, twitter

# Introduction

Word Embeddings are a helpful method when analyzing a large corpus of words and have many uses in Natural Language Processing. Using word embeddings, the distance between words is used to infer the similarity between words. Therefore, words that are more similar should have embeddings that are closer to each other. *Apple* and *Pear* being closer then *Apple* and *Chair*. One thing that word embeddings alone fail to do is take time into account is time. There is no way to determine whether a word used to be more similar to one word then it does now.

Creating Dynamic Word Embeddings can allow for a time aware analysis of semantic meaning between words. With Dynamic Word Embeddings the evlolving meanings between words can be discovered. Now how words like “hip” have shifted from being the body part to also being associated with fashionable. In this paper we developed Dynamic Word Embeddings to try to infer the shifting meaning of words on social media by using data collected from twitter.

We developed a model based on the paper “Developing Semantic Word Embedings for Semantic Discovery by Z. Yao et. al. using a Positive Pointwise Mutual Information matrix(PPMI), which they used on a NYTimes dataset. There are many differences between news articles and tweets though. Of those length and hashtags are the most prominent.

# Methodology

# Analysis

# IV. Future

##### Acknowledgment

##### References

1. Z. Yao, Y. Sun, W. Ding, N. Rao, and H. Xiiong, *“Dynamic Word Embeddings for Evolving Semantic Discovery”, 13 Feb 2018,* [arXiv:1703.00607](https://arxiv.org/abs/1703.00607)