

HACETTEPE UNIVERSITY

DOCTORAL THESIS

Evaluating Bilingual Embeddings in Bilingual Dictionary Alignment

Author:

Yiğit SEVER

Supervisor:

Dr. Gönenc ERCAN

*A thesis submitted in fulfillment of the requirements
for the degree of Master of Science*

in the

Computer Engineering Department

May 19, 2019

Declaration of Authorship

I, Yiğit SEVER, declare that this thesis titled, “Evaluating Bilingual Embeddings in Bilingual Dictionary Alignment” and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

“Thanks to my solid academic training, today I can write hundreds of words on virtually any topic without possessing a shred of information, which is how I got a good job in journalism.”

Dave Barry

HACETTEPE UNIVERSITY

Abstract

Faculty Of Science and Engineering
Computer Engineering Department

Master of Science

Evaluating Bilingual Embeddings in Bilingual Dictionary Alignment

by Yiğit SEVER

Dictionaries catalog and describe the semantic information of a lexicon. WordNet provides an edge by presenting distinct concepts with the hierarchy information among them. Research in computer science has been using this hand crafted tool in natural language applications such as text summarization and machine translation. Original WordNet has been compiled for English yet counterparts for other languages are not as readily available nor as comprehensive. In order for research on languages other than English to benefit from the power of a WordNet, machine assisted creation and evaluation methods are essential.

Word embeddings can provide a mapping between words and points in a real valued vector space. Using these vectors, representing documents as well as forming geometric relationships between them is a well studied area of research. In this thesis we start by hypothesizing that a dictionary definition captures the semantic basis of the described word. We used word embeddings as building blocks to map dictionary definitions into a multidimensional space. These spaces can be aligned to accommodate two languages, allowing the transfer of information from one language to another. We investigate the success of retrieving and matching discrete senses across languages by employing supervised and unsupervised methods. Our experiments show that dictionary alignment can be evaluated successfully by using both unsupervised and supervised methods but corpora sizes should be taken into consideration. We further argue that some methods are not viable considering their poor performance.

Acknowledgements

The acknowledgments and the people to thank go here, don't forget to include your project advisor...

Contents

Declaration of Authorship	iii
Abstract	vii
Acknowledgements	ix
1 Introduction	1
A Frequently Asked Questions	3
A.1 How do I change the colors of links?	3

List of Figures

List of Tables

List of Abbreviations

LAH List Abbreviations **Here**
WSF What (it) Stands For

Physical Constants

Speed of Light $c_0 = 2.997\,924\,58 \times 10^8 \text{ m s}^{-1}$ (exact)

List of Symbols

a	distance	m
P	power	W (J s ⁻¹)
ω	angular frequency	rad

For/Dedicated to/To my...

Chapter 1

Introduction

To our best knowledge, a succinct set of words to define a dictionary where words are presented alongside their translations and one where words from a single language are presented with their definitions with groupings for synyonyms and hypernyms are not available. Thus, we will refer the first type as a bilingual dictionary and the second type as a monolingual dictionary. James Somers puts down the currently available modern dictionaries by saying “The definitions are these desiccated little husks of technocratic meaningese, as if a word were no more than its coordinates in semantic space.” which is a bother for a writer but gives us further encouragement with our intuition.

Appendix A

Frequently Asked Questions

A.1 How do I change the colors of links?

The color of links can be changed to your liking using:

```
\hypersetup{urlcolor=red}, or  
\hypersetup{citecolor=green}, or  
\hypersetup{allcolor=blue}.
```

If you want to completely hide the links, you can use:

```
\hypersetup{allcolors=.}, or even better:  
\hypersetup{hidelinks}.
```

If you want to have obvious links in the PDF but not the printed text, use:

```
\hypersetup{colorlinks=false}.
```