

CS221 Prroject Report

Melvin Johnson Premkumar

School of Electrical and
Computer Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332-0250

Email: <http://www.michaelsell.org/contact.html>

Isaac Caswell

Twentieth Century Fox
Springfield, USA
Email: homer@thesimpsons.com

Zane Silver

Starfleet Academy
San Francisco, California 96678-2391
Telephone: (800) 555-1212
Fax: (888) 555-1212

Abstract—We are working with Xiao Cheng, a PhD candidate in the Computer Science Dept.

In which we discuss various things

I. INTRODUCTION

Our project is to detect paraphrases in Twitter, as part of the SemEval challenge for 2015. Given two sentences (Tweets), we determine whether they express the same or very similar meaning and optionally a degree score between 0 and 1. The training dataset is provided by semEval and contains about 17,790 annotated sentence pairs, and comes with tokenization, part-of-speech and named entity tags. The testing dataset consists of a further 1k examples from a different time period, annotated by an expert. SemEval also provides several baselines against which to compare our algorithm's performance. In this progress report we report the result from using Google Wordvec and Dynamic Pooling and Unfolding Recursive Autoencoders for Paraphrase Detection for this task; and discuss how to improve upon them.

A. Initial Steps: Wordvec

Subsection text here.

1) *Subsubsection Heading Here:* Subsubsection text here.

II. CONCLUSION

The conclusion goes here.

ACKNOWLEDGMENT

The authors would like to thank...

REFERENCES

- [1] H. Kopka and P. W. Daly, *A Guide to L^AT_EX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.