

SemEval 2015 Task 1

Mike Meding & Hoanh Nguyen

March 4, 2015

1 Introduction

Our choice for the project this semester is SemEval 2015 Task 1. This task has to do with paraphrasing tweets to divulge similarity between them. In bold below is the problem as stated directly from the SemEval website. **Given two sentences, the participants are asked to determine whether they express the same or very similar meaning and optionally a degree score between 0 and 1.**

2 Relevant Text

Andrew, Goldberg (2007). *Automatic Summarization*

3 Approach

The great part about doing a project from SemEval is that all baseline models and data are given. For our baseline we have the option of choosing between two models. One is a logistic regression model with an f-score of 0.6, the other model is a weighted factorization matrix.

Given these baseline models we will likely improve them using state-of-the-art semantic analysis tools such as word2vec.

4 Experiments & Evaluation

5 Data