1 Introduction

This thesis investigates the efficacy of ranking scalar adjectives from corpus data using monolingual corpus and data derived from bilingual sources. We make three major contributions:

1. successfully reproduce the state of the art in adjective ranking using monolingual corpus

2. offer two alternative formulations of ranking adjectives using monolingual corpus, each one achieves parity with the state of the art.

3. successfully incorporate bilingual data with monolingual data, and outperforms the state of the art by a non-trivial amount.

Additionally, all code and data needed to reproduce the successful experiments in this thesis are distributed and can be found at: https://github.com/lingxiao/adj-relation.

This thesis is organized as follows: in the rest of this chapter we will motivate the problem of adjective ranking and give an overview of prior work. In chapter 2 we will discuss the two sources of data we consider in this thesis, as well as how the gold standards are procured. We will also discuss how the rankings are typically evaluated. Chapter 3 gives thorough review of the state of the art method, and implementation details that arose while we reproduced the method. In chapter 4 we offer a very different formulation of the problem at a high level. Chapter 5 will present a simple baseline in the spirit of this reformulation that performs surprisingly well. Chapter 6 refines the baseline using a variety of models, and finally in chapter 7 we combine the models and outperforms the state of the art by a significant amount. This chapter will conclude with several motivations for

future work. Chapters 8 and 9 are optional reading for those who wish to consider other points of view on the problem. Chapters 8 will gives a formulation that is on parity with the previous state of the art, although this model uses monolingual data only. Chapter 9 gives a detailed account of a failed attempt at ranking adjectives, readers should note that this chapter is meant to amuse rather than to inform.