

Abstract

The main purpose of this research is to compare the actual effects of different data augmentation methods applied in the field of WSA.

Introduction

In the case of a small amount of training data, machine learning applied to NLP may produce poor results. Data augmentation can be used to improve this problem. Data augmentation has relatively little research in the field of NLP, so we plan to analyze several data augmentation methods and compare their training results.

Research Plan

We will apply the following several data augmentation methods to the data set Semcor, and apply our own model.

1. Similar Word Replacement
2. Back Translation
3. Random Noise Injection
4. Pretrained model MLM
5. Conditional Pretrained model

In the end, we will compare the accuracy of the trained model to judge the quality of the data augmentation method.

Expected Conclusions

The results will be charted and compared. Finally, we will find the best augmentation method based on our results.