NLANGP: Supervised Machine Learning System for Aspect Category

Classification and Opinion Target Extraction

1 Introduction

For Slot 1, we model the problem as a multi-class classification problem where binary classifiers are trained to predict the aspect categories. For Slot 2, we model the problem as a sequential labeling task, using Conditional Random Fields (CRF) as the training algorithm.

2 System Description

2.1 Features

2.1.1 Word

The current word is used as a feature. For opinion target extraction, the previous word and next word are also used as features.

2.1.2 Bigram

All word bigrams found in a sentence are used as features.

2.1.3 Name List

2.1.4 Head Word

From the sentence parse tree, we extract the head word of each word and use it as a feature.

2.1.5 Word Cluster

We induce Brown clusters and K-means clusters。

2.1.6 Name List Generated using Double Propagation

2.2 Approaches

2.2.1 Aspect Category Classification (Slot 1)

Aspect category classification is based on a set of

one-vs-all binary classifiers, one classifier for each category found in the training set.

2.2.2 Opinion Target Extraction (Slot 2)

Opinion target extraction is modeled as a sequential labeling task, where each word in the sentence is assigned a label using the IOB2 scheme.

The classifier is trained using Conditional Random Fields (CRF).

3 Results

4 Feature Ablation

5 Conclusion

In this paper, we report our work on aspect category classification and opinion target extraction using supervised machine learning approaches.