Selectional Preferences for Word Sense Disambiguation – error detection

Hamdiye Arslan* , Martin F. Simmeth *, Oumar M. Dia*

*Departement of Computerlinguistik, Universität Heidelberg

1 Introduction/Motivation

✓Selectional Preferences:

Resnik s association strength:

$$\frac{p(class\,|\,predicate)}{p(class\,|\,predicate)} \times \log \frac{p(class\,|\,predicate)}{p(class)}$$

To determine how closely the verb selects his arguments (object / subject), calculate the Selectional preferences for this verb. This means the association between the predicate and the semantic categories with reference to the arguments

✓WSD(word sense disambiguation):

Is the process of identifying which sense of a word is used in a sentence, when the word has multiple meanings

WSD is a difficult task for every WSD-System

✓**Goal** : compute the Selectional Preferences of verb-object-pairs in order to compare the gold standard senses with the given Ukb_WSD-senses
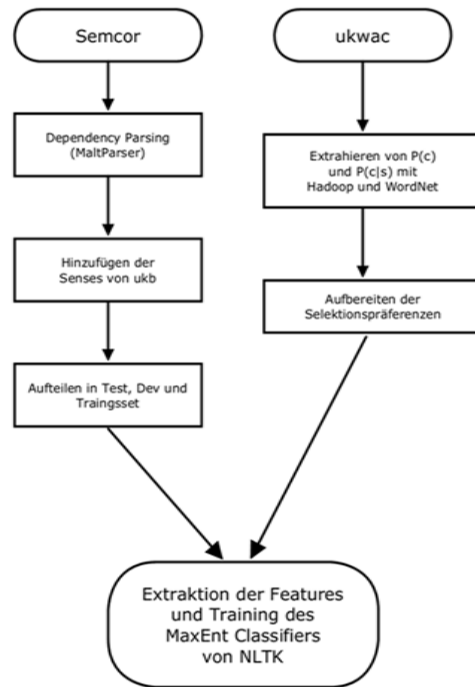
2 Approach (Figure 1)

1.Compute selectional association

-Hadoop (Figure 2)

  -Framework for distributed computers

  -Base on MapReduce Algorithm

  -Implemented in Java

-**Dumbo**

  -Python interface for Hadoop

  -Facilitate the writing and running of -hadoop programs in Python

  -could not get up to now a complete access to Hadoop API

  -poor documented

- **Hadoop/ Dumbo**

  -enable the use of UKWAC corpus as Database

2. Semcor for UKB_WSD-System

   -Convert Semcor-xml-corpus into .conll Format

   -Parse Semcor -conll-corpus with Malt Parser to get a dependent parsed basedata in order to identify verb-object-pair.
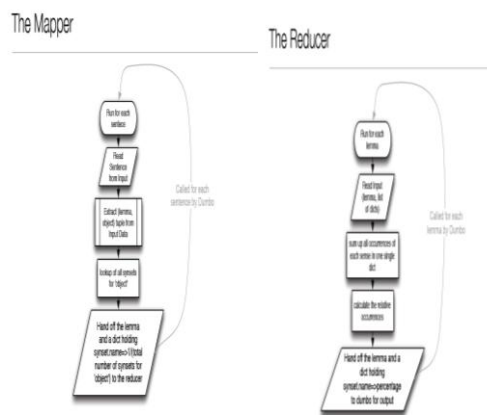
   -run UKB_WSD to disambiguate

Figure 1: Procedure

Figure 2: Mapper and Reducer

Example: ukb_output

ctx_02 w9 07190941-n !! charge
ctx_02 w12 06693198-n !! praise
ctx_03 w9 06681551-n !! report
ctx_04 w12 05682950-n !! interest

3 Classify

✓MaxEntClassifier with NLTK

-takes FeatureSet with Label as Input

  Example of a featureset:
 ({'context: record ': True, 'obj: action': True, 'verb: see': True, 'context: be ': True, 'selPrefInTopTen': 'False', 'context: start ': True, 'SenseNumber: 01': True, 'context: week ': True}, False)

- Data has been splitted into Train-set, Development-set and Test-Set

4 Evaluation

Evaluation was done on the Test-Set.

Precison:	0.712
Recall:	0.712
F1-Measure:	0.712

5 References

-Agirre and Soroa 2009. UKB: Graph Based Word Sense Disambiguation and Similarity

-Resnik, P. (1996). "Selectional constraints: an information-theoretic model and its computational realization . Cognition,61:127-159.

-Resnik, Philip. 1997. Selectional preference and sense disambiguation. In Proceedings of the SIGLEX Workshop on Tagging Text with Lexical Semantics: Why, What, and How? pages 52–57, Washington, DC.

•www.maltparser.org Malt Parser

•http://nltk.googlecode.com/svn/trunk/doc/api/nltk.classify.maxent.MaxentClassifier-class.html MaxEntClassifier