

# Supervised Word Sense Disambiguation

Supervised WSD is commonly used whenever we have sufficient data that has been hand-labeled with correct word senses.

J+M\_C WordNet

郑娜

## Components

❖ Datasets: sense-tagged corpus

- Each open-class word is labeled with its word sense from a dictionary/thesaurus

❖ Algorithm: standard classification algorithm

- Features including word identity, POS tags, embeddings of surrounding words, etc.

# Datasets

## ❖ Lexical Sample datasets

- Line-hard-serve, interest, etc.

## ❖ Project: SENSEVAL

- Produce sense-labeled lexical corpora

## ❖ Corpus: SemCor

- A subset of the Brown Corpus
- Consisting of 234,000 words
- Tagged with WordNet senses

Fig.SemCor

```
<wd>side</wd><sn>[noun.location.0]</sn><tag>NN</tag>
<wd>of</wd><tag>IN</tag>
<wd>the</wd><tag>DT</tag>
<wd>tub</wd><sn>[noun.artifact.1]</sn><tag>NN</tag>
<wd>he</wd><tag>PP</tag>
<wd>saw</wd><mwd>see</mwd><msn>[verb.perception.0]
  </msn><tag>VBD</tag>
<wd>a</wd><tag>DT</tag>
<wd>pair</wd><sn>[noun.quantity.0]</sn><tag>NN</tag>
<wd>of</wd><tag>IN</tag>
<wd>blue</wd><sn>[adj.all.0.col.3]</sn><tag>JJ</tag>
<wd>wool</wd><sn>[noun.artifact.0]</sn><tag>NN</tag>
<wd>swimming_trunks</wd><sn>[noun.artifact.0]</sn>
  <tag>NN</tag>
<wd>.</wd><tag>.</tag>
</s>
```

## Features

❖ Features: POS tags & surrounding word information

➤ Bag-of-words, POS tags, unigram/bigram collocation features

The linguistic and English tutor walked into the class.

Feature: [ $W_{i-2}$ ,  $POS_{i-2}$ ,  $W_{i-1}$ ,  $POS_{i-1}$ ,  $W_{i+1}$ ,  $POS_{i+1}$ ,  $W_{i+2}$ ,  $POS_{i+2}$ ,  $W_{i-2}^{i-1}$ ,

$W_{i+1}^{i+2}$ ,  $E(W_{i-2}, W_{i-1}, W_{i+1}, W_{i+2})$ ,  $bag()$ ]

Result: [linguistic, NN, and, CC, tutor, NN, walked, VB, linguistics and,

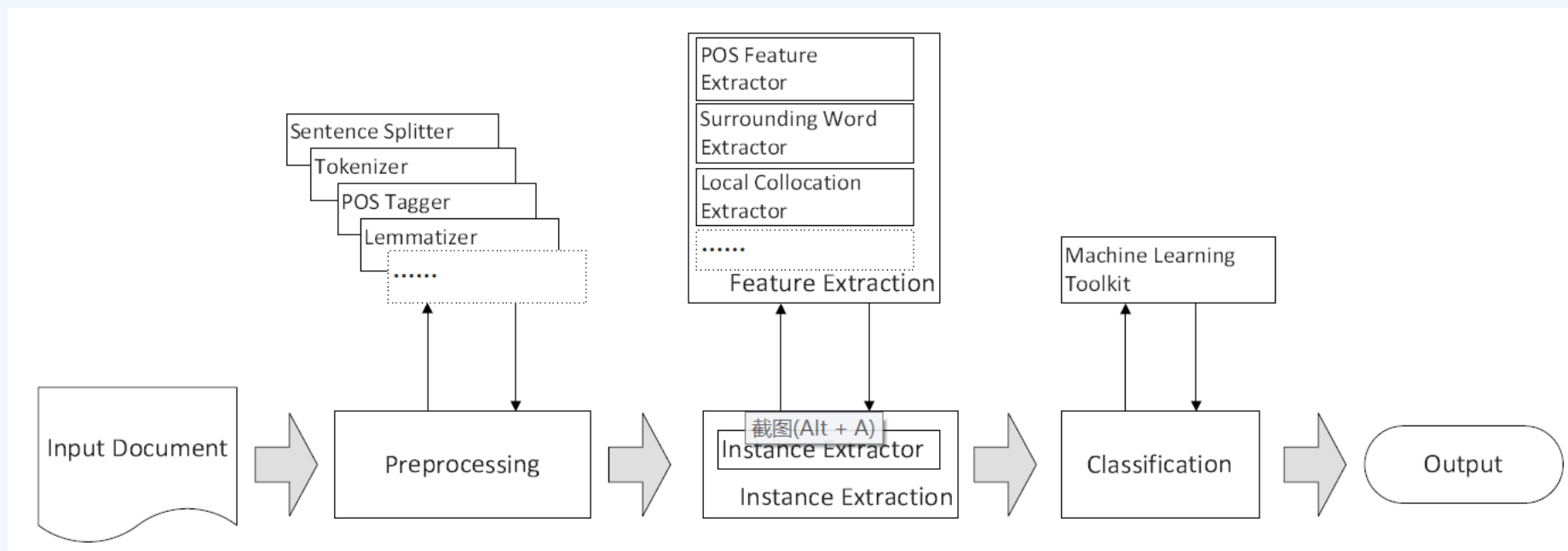
tutor walked,  $E(\text{linguistic, and, tutor, walked})$ ,  $bag(\text{linguistic,$

tutor, walked)]

# Algorithm

❖ IMS: a supervised English all-words WSD system (open-source)

- Classifier: linear support vector machines (SVM)
- Knowledge sources: POS tags, surroundings, collocations, etc.
- Results: the state-of-art results



# Algorithm

❖ Preprocessing – OpenNLP toolkits

❖ Feature & Instance Extraction

❖ Classification – Liblinear Method

Algorithm 1 The coordinate descent method for (13)

- Given  $\alpha$  and the corresponding  $w_m$
- While  $\alpha$  is not optimal, (outer iteration)
  1. Randomly permute  $\{1, \dots, l\}$  to  $\{\pi(1), \dots, \pi(l)\}$
  2. For  $i = \pi(1), \dots, \pi(l)$ , (inner iteration)
    - If  $\bar{\alpha}_i$  is active and  $x_i^T x_i \neq 0$  (i.e.,  $A \neq 0$ )
      - Solve a  $|U_i|$ -variable sub-problem (17)
      - Maintain  $w_m$  for all  $m$  by (18)

Algorithm 2 Solving the sub-problem

- Given  $A, B$
- Compute  $D$  by (25)
- Sort  $D$  in decreasing order; assume  $D$  has elements  $D_1, D_2, \dots, D_{|U_i|}$
- $r \leftarrow 2, \beta \leftarrow D_1 - AC$
- While  $r \leq |U_i|$  and  $\beta/(r-1) < D_r$ 
  1.  $\beta \leftarrow \beta + D_r$
  2.  $r \leftarrow r + 1$
- $\beta \leftarrow \beta/(r-1)$
- $\alpha_i^m \leftarrow \min(C_{y_i}^m, (\beta - B_m)/A), \forall m$

❖ Train

## REFERENCES

### ❖ State-of-art Unsupervised WSD Model:

Zhong, Z. , & Ng, H. T. . (2010). It Makes Sense: A Wide-Coverage Word Sense Disambiguation System for Free Text. Acl, Meeting of the Association for Computational Linguistics, July, Uppsala, Sweden, System Demonstrations. DBLP.

### ❖ Liblinear for Classification:

Fan, R. E., Chang, K. W., Hsieh, C. J., Wang, X. R., & Lin, C. J. (2008). Liblinear: a library for large linear classification. *Journal of Machine Learning Research*, 9(9), 1871-1874.