

NLP Project: Yarowsky Word Sense Disambiguation Algorithm

Juho Kallio and Otto Wallenius

4. joulukuuta 2014

1 Implementation

We implemented Yarowsky's unsupervised word sense disambiguation algorithm with Python. We used test-driven development.

2 Results

Pattern	Seeds	k	Threshold	Epsilon	Accuracy
plant	growth, car	19	4.5	0.0001	94%
light	wind, bright	19	7.5	0.0001	60%

3 Conclusions

The algorithm did not perform very well, worse than we expected. With low values of k , we started to experience problems with common words. Words like *since* started to show up as good rules: we believe this was because of the smallish data set. When the lexicon is too big compared to the corpus, some common ones start to appear only with the more common sense of the original word. With a bigger value of k this effect diminishes.