

ASSIGNMENT 5 - MACHINE LEARNING IN COMPUTATIONAL LINGUISTICS

The table shows results of WSD experiments on 32 words. The table shows the metric used such as MVDM, IB1, IGTREE etc., and the k neighbours. The scores are calculated as baseline, default and optimal settings for the training data set and fine-grained, medium and coarse for the evaluation data set. The experiment used 10-fold cross-validation on the training set and SENSEVAL coordinators for the evaluation set. The average scores have also been presented at the end of the table.

From the results, we deduce that optimization improves the results overall. It shows a better performance than default for every word irrespective of the used metric, the value of k or the type of the word. The word 'consume' had a score of 32.9 with default settings and 58.8 with optimal settings. Thus, it showed the highest difference of 25.9. After comparing all the values of default and optimal settings, we see that optimization doesn't have any negative effects. MVDM-IG with k=5 is the most robust metric as it gives more higher increase from baseline to optimal settings. We use an odd number for value of k to avoid the number of ties between the classification. If the number is even, each half will favour the different results. Thus, odd number of k is used. We see that 'sack-v' was the easiest as training (in default settings) and evaluation scores were exactly the same. Optimization did not help in the case of this word. The most difficult word is 'float-v.' It has a low baseline score of 21. In the evaluation set, coarse scores were 44.1. Optimization does help in this case as it increases the score to a whopping 44. The word 'giant-n' has the highest difference between of 18.7 between fine-grained and coarse-grained evaluation.

As a conclusion, we see that the default settings do not give the best results. To obtain better results we use optimization. Coarse-grained, on the other hand, gives a better performance in the evaluation set compared to fine-grained and medium-grained scores. Thus, we can say evaluation set was easier than the training set as we're getting better results on it.