

Evaluation

Adam Poliak — apoliak1 David Russell - drusse19

March 22, 2016

1 METEOR

1.1 METEOR Implementation

As instructed in the assignment, we implemented the METEOR in the script `./evaluation` by calculating and using precision and recall as described in the assignment. The implementation is contained in the method `meteor`

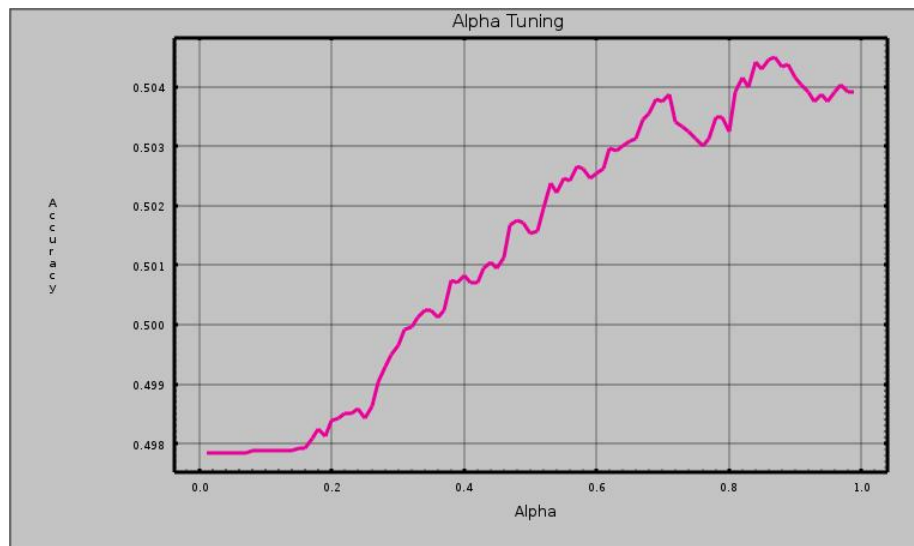
1.2 Running METEOR

To run our implementation of METEOR, `./evaluation` must be called with the α flag. To add an α parameter, use the `-a` flag followed by a decimal number to represent the α . If the `-a` flag is not provided, then the baseline implementation of `./evaluation` provided will be ran instead.

1.3 α tuning

After running `./evaluation` 100 times, we determined that the best alpha on the dev data is $\alpha = .87$ resulted in an accuracy of 0.5045

The graph below shows the results from tuning alpha. The x-axis represents the alphas and the y-axis represents the accuracy for corresponding alphas used in METEOR.



The data in the graph was generated by running the script `test.alpha` to determine the best α . This script takes about 5 minutes to run.

2 WordNet Synonyms