

# A quick look at semi-supervised learning for NLP

Danul K. Gunatilleka

# Semi-supervised learning

Supervised learning algorithms are extremely successful.

We are now producing immense amount of data.

How do we label them?

Other approaches are needed!

# Semi-supervised learning

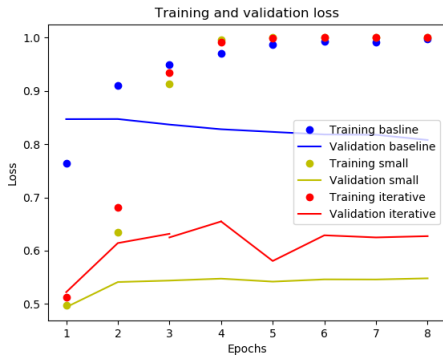
Use some labeled data, some unlabeled data.

Many approaches.

We will train the model then let it label some of the data.

We train again.

We use the IMDB data set.

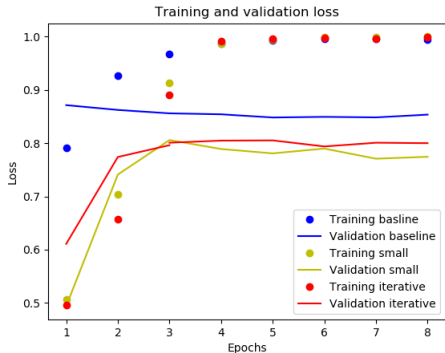


Iterative test accuracy: 0.63

Small model test accuracy: 0.54

Benchmark test accuracy: 0.49

Benchmark uses 20000 training example while the others use 2000 samples.



Iterative test accuracy: 0.79

Small model test accuracy: 0.78

Benchmark test accuracy: 0.5