Mitschrift: Sohail Malih

Referat Student Dayyan Smith Advisor Katharina Kann (Montag, den 29.05.2017)

Exploring the effect of regularization of a neural network for stance classification

Fake News. Is a made up story, with an intention to deceive to NY Times.
So how do we detect them?
Assessing the veracity of a news story is a complex and cumbersome task (even for trained experts)
Automatic FN detection can be broken into different stages.

FN challenge, is exploring how AI technologies could be leveraged to combat fake news.
It would be helpful to know which news organizations agree with given topic.

First question we need to talk about is, which news organisation agree with given claim?
Is the first FN chalange stage.

Next what is Stance?
Text A Mango smoothies are a good summer drink
Text B Mango smoothies are so refreshing especially in summer
Text C
2 Statements agree the third disagress with A and B.

2 Statements, Possible stances:
there is agree, disagree, discuss and unrelated
in the shown case it's agree but it's hard to determine the relation.
Annotation is often tricky, even more for fake news.
Several examples are shown for how Headlines and bodies are related.

Next he was talking about Enconding.
w2vec word embeddings to get vectors for each word. Use GRU to get sentence representations
Use sequence of word vectors to produce sentence embeddings and then concatenate them

Put them through 2 hidden layers. Concatenate the embeddings with outputs of both hidden layers for both headline and body. Then put it through the Cassification Layer.
“With four parameters i can fit an elephant... “ Quote by John von Neumann

Meaning one should not be impressed when a complex model fits a data set well. With enough parameters you can fit any data set.

At the end he was talking about different Regularizations.
We have the L2 Regularization Big weights are pushed down more than small weights because the square of weights is penalized.

Next is the L1. Both big weiggts and small weights are pushed down a little because absolut value is penalized.

Then we have the Dropout Regularization:
You don't change the output of the model but the model itself.
You drop out a certain number of neurons from the neural network and train your model, then you backpropagate the network. That way you try to cut through the noise and try to find out which neurons really matter. You find probabilities for neuron to be dropped.

Regularization causes big fluctuations right now, hasn't had the time to train it the way he wanted to yet.

Right now Results with and without Regularization are still about the same with L1 and L2.