AL-DNN

parminder.bhatia243

September 2015

Abstract

Deep Neural Networks have shown great results in various fields. One of the major reasons has been abundance of data. However, major problem today is that we have lot of data but most of the data is unlabelled. Motivation is to explore Active Learning Techniques in context of Deep Neural Networks.

1 Background - Active Learning

If we pick the right examples to label, we can learn the concept from only a few labeled examples.

1.1 Active Learning Heuristics

- Start with the pool of unlabeled data.
- Pick a few points at random and get their labels
- Repeat the following Fit a classifier to labels seen so far
  Pick the BEST unlabeled point to get a label for

1.2 Query Selection Strategies

Expected Model Change - Pick the unlabeled instance that would cause the greatest change to the model, if we knew its label
Expected Error Reduction - Query the instances that would most reduce the error(most computationally expensive framework
Uncertainty Sampling

Attain good learning performance without demanding to many labeled examples.
- Semi supervised learning : use unlabeled data
- Active learning : choose labeled examples

Pool Based Active Learning - Uncertainty Sampling
Query the sample x that the learner is most uncertain about. - Maximum Entropy
- Smallest Margin between most likely and second most likely label
- Least Confidence

From Theories to Queries: Active Learning in Practice(Burr Settles)

Practical Challenges
- Querying in Batches - etc
Not that useful for now ...

Active Learning Literature Survey (Settles)

There are several different problem scenarios in which the learner may be able to ask queries. The three main settings that have been considered in the literature are (i) membership query synthesis, (ii) stream-based selective sampling, and (iii) pool-based sampling.
Uncertainty Sampling Query the sample x that the learner is most uncertain about.
- Maximum Entropy
- Smallest Margin between most likely and second most likely label
- Least Confidence

Expected Model Change
Instance that would cause greater change to current model if we knew its label. Query strategy in this framework is the "expected gradient length" (EGL) approach for discriminative probabilistic model classes.
In theory, the EGL strategy can be applied to any learning problem where gradient-based training is used.In other words, the learner should query the instance x which, if labeled and added to L, would result in the new training gradient of the largest magnitude.

Expected Error Reduction
Aims to measure not how much the model is likely to change, but how much its generalization error is likely to be reduced.

Variance Reduction
Minimizing the expectation of a loss function directly is expensive, and in general this cannot be done in closed form. However, we can still reduce generalization error indirectly by minimizing output variance, which sometimes does have a closed-form solution.

2 Equations

Entropy

$$Ent = min_x \sum_k p(x, y_k) \log p(x, y_k) \tag{1}$$

Least Confidence, (here y_k is the top predicted label)

$$lc = min_x p(x, y_k) \tag{2}$$

Marginal Sampling

$$msmpl = min_x p(x, y_1) - p(x, y_2) \tag{3}$$

3 CNN

CNN model used is similar to sentence modeling CNN model by Yoon Kim , where we had filters of various widths (3,4,5), each having 100 features maps which then passed through max pooling to get 300 vector representation of a sentence/tweet. This tweet was then based through a MLP(with softmax) to get the final prediction. It is a multiclass classification problem where labels are Pos, Neg or Neu .

4 Results

4.1 CNN - Entropy

1000	1064	1192	1448	1960	2984	4008
42.25	43.35	45.96	46.73	45.30	47.47	48.49
36.36	40.55	43.57	42.88	40.86	46.06	47.38
44.25	44.74	45.25	45.30	45.60	48.36	48.70
43.20	45.43	46.35	47.66	47.37	47.32	47.88

4.2 CNN - Marginal Sampling

1000	1064	1192	1448	1960	2984	4008
45.30	44.46	46.43	44.43	45.01	48.16	48.79
42.23	43.60	43.77	50.03	49.86	51.19	48.38
45.85	45.76	47.37	48.54	45.86	48.41	50.71

4.3 CNN - LC

1000	1064	1192	1448	1960	2984	4008
43.38	42.40	38.26	44.09	45.30	48.35	46.67
49.35	44.69	47.57	45.78	43.91	47.52	47.98
44.25	51.98	52.02	52.39	52.80	46.36	44.70
47.23	50.38	49.37	49.97	50.43	50.40	49.29

4.4 CNN - Note

I am getting different results in CPU and GPU version.(Maybe might be due to 32 bit. Need to find this.) Results in CPU are better by 5-7 percent in general.

5 Code and Data

Code is available at `https://github.com/parry2403/Active-Learning-in-DNN`
Data `sb/project/ycy-622-aa/ml_datasets/embeddings/data/`