Name: Asfahan Shah (aks7824@psu.edu)

Homework 1

Paper 1: Deep Active Learning with Contrastive Learning Under Realistic Data Pool Assumption

1. What problem does this paper try to solve, i.e., its motivation

In the active learning paradigm, it is seen that most of the methods used assume that unlabeled data and labeled data have the same distribution as in both the labeled and unlabelled set of data have useful information for the given task. But it is often seen in real world data (unlabelled), consists of out of distribution and sometimes even useless samples (not useful for a particular task). So if we consider this type of data and use it for training of models, the performance of the model will be affected.

2. How does it solve the problem?

To solve these problems authors first created two benchmark datasets (MixMNIST and MixCIFAR60) that simulate realistic unlabeled data pools i.e (datasets within distribution, ambiguous/useless , and out of distribution samples). Authors have also created a novel active learning method that combines contrastive learning. The details of algo are mentioned as under:
First use both labeled and unlabeled data and train a contrastive learning model. Now using this model, we get feature representation of each data point (both labeled and non labeled). Now run k means clustering algorithm on these feature representations. Now check all the clusters and remove the cluster that contains the largest number of labeled non-in-distribution samples as it is a cluster of non-in-distribution samples. For the remaining cluster we do the following.For each cluster we calculate distance between the cluster centroid and the centroid-closest non-in distribution features. This distance can be considered a radius and using this radius and cluster centroid as the center we get a circle. Now for each circle we do as follows (sequentially):
(i) take samples close to centroid but inside the circle
(ii) samples inside the circle regardless of distance from center
(iii) samples close to the centroid but outside the circle.

3. A list of novelties/contributions

1) Created two new benchmark datasets (MixMNIST and MixCIFAR60) that simulate realistic unlabeled data pools with in-distribution, ambiguous, and out of distribution samples.
2) Proposed an acquisition strategy for active learning that is based on contrastive learning which outperforms existing methods.

4. What do you think are the downsides of the work?

One potential downside is the additional overhead that is being used for contrastive learning and for calculating distance. If we compare it with other methods then the acquisition method based on random sampling on average has 1 to 2% less accuracy as compared to the proposed algorithm, but the random method does not require the above overhead and also is much simpler to implement. Another downside is that this method is only tested for image datasets (the datasets proposed by authors are image based) hence this algorithm's performance on language datasets is not yet known.

Paper 2: ActiveGLAE: A Benchmark for Deep Active Learning with Transformers

1. What problem does this paper try to solve, i.e., its motivation

For transformer based language models, there is a lack of standardized evaluation protocol of deep active learning. Data plays a crucial role in active learning so evaluation wildly varies from dataset to dataset. So this lack of evaluation standard makes research of active learning methods on transformer based language methods difficult.

2. How does it solve the problem?

To solve this authors have proposed Active General Language Adaption Evaluation (ActiveGLAE) benchmark. ActiveGLAE comprises a wide variety of NLP classification tasks and dataset along with evaluation guidelines that helps in providing a uniformity. Besides this they also provide a baseline which further helps in providing a reference for future research efforts.

3. A list of novelties/contributions

1) Authors propose ActiveGLAE which helps in providing a uniform evaluation setting for active learning methods in context of transformer based language methods. ActiveGLAE consists of ten NLP classification tasks that cover a wide range of text genres, data set sizes, degrees of difficulty etc.
2) Authors also propose a guildins that help in developing a standardized evaluation for transformer based luggage models

4. What do you think are the downsides of the work?

The Author's study mainly focuses on NLP but it doesn't cover other fields like computer vision. Also experiments conducted by authors are not exhaustive, so more models and different strategies could also be tried on ActiveGLAE

Paper 3: Deep Active Learning with Noisy Oracle in Object Detection

1. What problem does this paper try to solve, i.e., its motivation

For complex tasks such as Object detection, data annotation is an expensive and time consuming task. Besides that oracle (usually human) which is used to annotate unlabeled data , is not noise/ error free. So we further need an additional review task which is again as costly as the original annotation task. Specifically for object detection tasks we have two common types of errors: missed bounding box labels and bounding box labels with incorrect labels.

2. How does it solve the problem?

To reduce this reviewing cost, authors have proposed an automated label review method. First we use the oracle to get the predictions of unlabeled data. Then we use both this unlabeled data and labeled data with their respective noisy predictions for the review module. Review module has two components one for misses (missed bounding box labels) and one for flips (bounding box labels with incorrect labels).

1) For the misses, using the noisy labels and combined data, we feed it into the model to get background loss. Then data points are sorted in descending order on the bases of background loss. After which ets say top k points are sent for review.
2) For the flips using the noisy labels and combined data, we feed it into the model to get classification loss. Then data points are sorted in descending order on the bases of classification loss. After which ets say top k points are sent for review.

3. A list of novelties/contributions

1) The authors method helps in improving the performance metric while keeping the total annotations cost same
2) AUthors method also highlight that partitioning a part of annotation budget for correction after review, for each active learning round is a valid strategy that can improve the performance significantly

4. What do you think are the downsides of the work?

The method still relies on humans to review the suspected points and correct them. This is again time consuming and error prone. Other drawback is that authors have not compared their method with other state of art object detection methods,