Neural Self-Training through Spaced Repetition

Problem:

Paper discusses an approach of spaced repetition of training instances to achieve better performance of a neural network model in a semi-supervised learning task. Semi-supervised learning process broadly performs four steps starting with learning on labeled instances, then sampling a subset from unlabeled instances, and obtaining their labels from the trained model, followed by learning on labeled and sampled instances. Authors mainly target the subset sampling step and highlight the importance of accounting characteristics of the unlabeled instances and the model while sampling. The problem can approximately be written as follows,

Repeat for episodes

Return

Method:

The spaced repetition method presented in the paper traces origins from the field of psychology and a fundamental observation that humans and machines can learn better if more learning time is spent on difficult instances than on easy instances. Algorithmically, spaced repetition method takes unlabeled data as an input and segregates it in number of strata/queues based on the difficulty of the unlabeled instances and the model performance.

A previously published spaced repetition system, Leitner system, is used in the paper to achieve partitioning of the unlabeled dataset. Basic principle of Leitner system is promotion (into higher strata/sets/queues) of easy instances and demotion (into lower strata/sets/queues) of difficult instances over training period. Predicted labels of the unlabeled data (with previous model version) are taken as ground truth by the Leitner system (STEP 1 in ‘Problem’ sections provides the labels for unlabeled dataset). A new model is trained from scratch for epochs in the Leitner system subroutine and the model being trained is tested on the unlabeled data. The notion of ‘difficulty’ of an unlabeled instance is based on prediction accuracy of the model being trained i.e., correctly guessed unlabeled instances are promoted and incorrect predictions are demoted. After performing promotions and demotions -times, partitions of unlabeled dataset are finalized (STEP 2 in ‘Problem’ section).

The partitions are then combined with the labeled dataset, one by one, and another models are trained on these datasets (STEP 3 in ‘Problem’ section). At the end, the best model is selected (out of and one model trained initially on the labeled data only, STEP 1, 3 and 4 in ‘Problem’ section) maximizing performance on validation set.

If a model trained on combination of labeled and unlabeled partition set is found to be the best performing, then the unlabeled partition set (called designated queue), and corresponding label predictions are moved from the unlabeled set to the labeled set. This entire process the repeated for episodes.

Strengths:

Authors compared performance of the presented method with five other semi-supervised learning methods/approaches and achieved an improvement of at least 1% to maximum of 5% in Macro-F1 score. These results also surpass the state-of-the-art benchmarks on two text classification problems. Authors justify their results with additional analyses that investigates the size of the Leitner queues (partitions), similarity between the queues and the training set, diversity of the queues and movement (number of promotions or demotions) of the instances in queues of the Leitner system. Interestingly, the presented spaced repetition systems adds that Leitner queue (subset of the unlabeled data) which is less similar to the training set than other queues and relatively more diverse than most of the other queues (non-designated). Additionally, authors also explain that the number of promotions and demotions in the lowest and the highest queue are the least i.e., the simplest and toughest instances in the unlabeled data are efficiently distilled. Lastly, the presented method is generalizable for semi-supervised learning tasks and hence can achieve better performance on broad spectrum of tasks.

Limitations:

Every episode contains k-iterations of training on labeled dataset and evaluation on unlabeled dataset. As the size of the training dataset is not significant (generally for self-supervised tasks), the k-training iterations might not increase the computational expense as much but together with evaluations on large unlabeled set might lead to significant increase in the computational expense. The presented method does not handle the issue associated with high deviation in the movement patterns of the instances in the same queue. In such case more micro level analysis and better sampling strategies will be required.

Contribution:

Unlike previously published methods the presented method not only considers the model’s prediction confidence but also considers the instance characteristics to effectively sample instances from the unlabeled dataset. This key contribution deals with the issue of over sampling those unlabeled data instances which are the most similar to the training dataset. The introspective analysis shows that the samples of the unlabeled dataset added to the labeled data are less similar to the labeled data than non-sampled instances. The spaced repetition framework presented explores the unlabeled dataset dynamically and overcomes the limitations of the predefined sampling policies used for unlabeled data sampling in self-learning tasks.