Transfer Learning in NLP: A Survey

The limitations of deep learning models, such as requiring a large amount of data to train models, and also demand for huge computing resources, forces research for the knowledge transfer possibilities. Nowadays many large DL models are emerging that demand the need for transfer learning. This survey aims to discuss the recent advances in using transfer learning in Natural Language Processing (NLP). In this survey, we discuss various models that are used for NLP based on the three main architecture that is Recurrent-Based Models, Attention-Based Models, and CNN-Based Models. We briefly discuss the popular datasets (such as SQuAD 2.0, GLUE, RACE, etc.), language models, and tasks that have been used in researches related to transfer learning.

Further, we provide a taxonomy for transfer learning that divides it into two categories: Transductive Transfer Learning and Inductive Transfer Learning. Transductive Transfer Learning is when for the same task, the target domain or task doesn't have labeled data or has very few labeled samples. It can further be divided into sub-categories such as Domain Adaptation (involves learning about a different data distribution in the target domain) and Cross-lingual learning (involves adapting to a different language in the target domain). Inductive Transfer Learning is when for different tasks in source and target domain we have labeled data in the target domain only. It is further divided into Sequential Transfer Learning (involves learning multiple tasks in a sequential fashion and Multi-Tasks Learning (involves learning multiple tasks at the same time).

In comparison to RNN-based and CNN-based language models, we see that attention-based models are more powerful in the literature. Also, we see BERT is the default architecture used for language modeling in many of the tasks because of its bidirectional architecture that makes it useful for many downstream tasks. For transfer learning in NLP, sequential fine-tuning appears to be more successful in the studies in comparison to other approaches such as zero-shot learning. In recent years, multi-task fine-tuning is gaining more popularity as researches suggest training a model to learn multiple tasks at the same time gives better results than learning them independently. Also, we see that text classification datasets are more popular than other NLP tasks as text-classification models are easier to fine-tune.

Nowadays, as NLP is drawing more attention and can solve several new problems, it has become important to leverage data from the domains, languages, and tasks. We cannot go without transfer learning.