Transformers and Transfer Learning - Exploring the limits of transfer learning.

With plethora of techniques and methods in the field of Machine Learning being emerged, transfer learning is a method/model which is pre-trained on a data-rich task before being finetuned on a downstream task, and has been emerged as a powerful technique in natural language processing (NLP). Unlike other computer vision tasks, where the pre-training is typically done via supervised learning on a large labeled data set like ImageNet, modern techniques for transfer learning in NLP often pre-train using unsupervised learning on unlabeled data. Beyond its empirical strength, unsupervised pre-training for NLP is particularly attractive because unlabeled text data is available en masse thanks to the Internet.

The basic idea is to treat every text processing task as a “text-to-text" problem. The systematic study compares pre-training objectives, architectures, unlabeled data sets, transfer approaches, and other factors on dozens of language understanding tasks. The goal is not to propose new methods but instead to provide a comprehensive perspective on where the NLP field stands. As such, the work primarily comprises a survey, exploration, and empirical comparison of existing techniques. The limits of current approaches by scaling up the insights from the systematic study (training models up to 11 billion parameters) to obtain state-of-the-art results in many of the tasks being considered is also explored.

The Transformer model architecture is being used for downstream tasks in NLP like, translation, question answering, and classification. Earlier for transfer learning in NLP, RNN were leveraged, however, with the increased ubiquity, all the tasks/models are based on Transformer architecture. E very task considered—including translation, question answering, and classification—is cast as feeding our model text as input and training it to generate some target text. This allows to use the same model, loss function, hyperparameters, etc. across our diverse set of tasks. It also provides a standard testbed for the methods included in our empirical survey. “T5” refers to model, which dub the “Text-to-Text Transfer Transformer”.