



Fig. 3 A walkthrough of recent developments in NLP

Neural Networks those are recurrent in nature due to performing the same function for every data, also known as Recurrent Neural Networks (RNNs), have also been used in NLP, and found ideal for sequential data such as text, time series, financial data, speech, audio, video among others, see article by Thomas (2019) [137]. One of the modified versions of RNNs is Long Short-Term Memory (LSTM) which is also very useful in the cases where only the desired important information needs to be retained for a much longer time discarding the irrelevant information, see [52, 58]. Further development in the LSTM has also led to a slightly simpler variant, called the gated recurrent unit (GRU), which has shown better results than standard LSTMs in many tasks [22, 26]. *Attention mechanisms* [7] which suggest a network to learn what to pay attention to in accordance with the current hidden state and annotation together with the use of transformers have also made a significant development in NLP, see [141]. It is to be noticed that Transformers have a potential of learning longer-term dependency but are limited by a fixed-length context in the setting of language modeling. In this direction recently Dai et al. [30] proposed a novel neural architecture Transformer-XL (XL as extra-long) which enables learning dependencies beyond a fixed length of words. Further the work of Rae et al. [104] on the Compressive Transformer, an attentive sequence model which compresses memories for long-range sequence learning, may be helpful for the readers. One may also refer to the recent work by Otter et al. [98] on uses of Deep Learning for NLP, and