Memory in ANN

Author: Jatin Aggarwal
Degree: M. Sc. RCSE
Date : 01/12/22

Recurrent neural network

To address the time-series data that contain sequential dependencies we use recurrent neural networks. Recurrent neural network can run any algorithm as it is turing complete. The basis architecture of RNN is discussed as: Input at time t is X, the hidden state H and output a predicted probabilities is Y. Here hidden state is contain a function of input F(H,X). This function is a weight metrics and activation function.The recurrecy exist in the output error with learn the weights through back-propagation.

There are problems with RNN like vanishing and exploding gradient problem. To understand this let assume a neural network with weight of the edges w1,w2..and function applied on each network layer. With back-propagation the magnitude of the gradient reduce this problem is vanishing gradient. To solve this we add large gradient to activation function which can lead to explodes the gradient. With the design variation in architecture on RNN we overcome with these challenges.

Bidirectional associative memory

The architecture of Bidirectional RNN is use to improve the knowledge about future state into the past state. Here we have separate hidden states for both forward and backward directions. Each state can interact with other state. The only condition is first computation of forward and backward hidden state are independent. Application of this is speech recognition, handwriting.

LSTM

To solve the above mentioned problems LSTM (long short term memory) used. It is a design to control gradient. To do this we introduce a new cell state act as long term memory. that contain a small part for information for previous state. This is like update the cell information over a time this help to avoid instability that can lead vanishing and exploding gradient problem. Through this small modification mechanism makes LSTM a good memory that remember and forget the information. The architecture is based on three gates namely forget, input and output gates. There are alot of application for the LSTM stock prediction, text generation are the promising one. Although the LSTM is good for sequence and time series problems there are hardware constraints. Like, It take a lot of time and resources to train the model.

GRU

To simplify LSTM in a more general way we introduce the GRU gated recurrent units. instance of using 3 different gates in LSTM, GRU is update and reset gate to perform same task. Reset gate is decisive for how much hidden state to carry over from previous state and update gate use for relative strength of matrix based update.

Refferences

Charu C. Aggarwal, Neural networks and deep learning, springer international publishing (2018)