APPLYING LSTM TO TEXT GENERATION

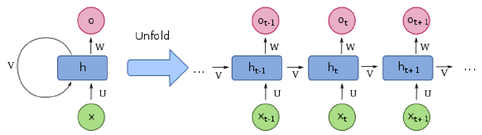
J. LEROUX, N. LALIBERTE, S. FREDERIC BOILEAU

MILA

Université de Montréal

RECURRENT NEURAL NETWORKS

Recurrent neural networks (RNN) are a family of neural networks specialized for processing a sequence of values. They are feedforward neural networks with the addition of time dependency in the model by introducing edges that span the adjacent time steps in the network. At a given time, the nodes with recurrent edges receive input from the current data and from the output of the hidden layer in the previous state, see figure below. Thus, an input at time t can influence the output at time $t + \delta$ by way of recurrent connections.



Each time step are computed as follows:

$$h_t = \sigma(Ux_t + Vh_{(t-1)} + b_h),$$
$$o_t = \text{softmax}(Wh_t + b_o)$$

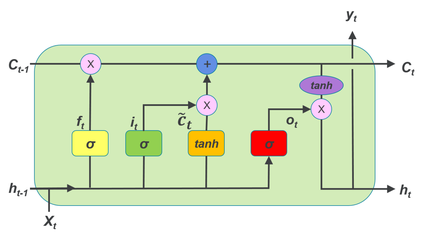
Here, U, V and W are weights matrix. The vectors b are bias parameters. Learning with RNN is challenging due to dependencies between long time steps. Consider the gradient with respect to h_t of $o_{t+\delta}$. How does it vary with δ? Following the graph above and applying the chain rule we can see that

$$\nabla_{h_t} o_{t+\delta} = \left(\prod_{k=t+1}^{t+\delta} V^T \text{diag}(1 - h_k^2) \right) \nabla_{h_{t+\delta}} o_{t+\delta}.$$

Thus, as δ grows, the gradient grows exponentially with V. If V is small or large than the gradient will either vanish or explode. This problem is well known. Solutions exist, which brings us to present the Long Short-Term Memory network (LSTM).

LSTM

The LSTM model has been introduced primarily to solve the vanishing and exploding gradients problem. This model is a RNN in which we replaced every hidden nodes by a *memory cell*.



Intuitively, RNN have *long-term memory* in the form of matrix weights, they change during the training encoding general knowledge about the data. RNN also have *short-term memory* in the form of activation passing from each node to successive ones. The memory cell introduced in the LSTM model provides storage for those memories. We now describe components of the cell.

- Gates (f_t, i_t, c_t): They are sigmoidal units that takes activation from the input x_t and the output of the hidden layer from previous state h_{t-1}. Note that f_t multiply the value of the previous cell c_{t-1}. The term *gate* stands for the literal meaning in the sense that if f_t is close to 0, then the gate is *closed* and the flow from the previous cell is cut off. If f_t is closed to 1 then all flow is passed through.
- Cell state ($c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t$): Cell state maintains information on the input. Also refered as the internal state, c_t has a self-connected edges with a fixed unit weight. This constant weight implies that the error can flow across time without vanishing or exploding.

RESULTS

Implemented network

For the implementation section of this projet, we decided to use Recurent Neural Network, more specificaly LSTM, for the generation and the classification of sequential data. More concretely, we gathered text data that we found around the internet to train a sequence classifier. The training of the classifier consist of identifying from which corpus between Harry Potter, Lord of the rings, some random quotes and Shakespeare, the sequence corresponds to. Further to this, we trained one model per sequence type for the text generation. Finally, we verified that our generated sequences were well classified by our classifier.

TextGen and Classification

The architecture we used for the text generation is of the form of a many-to-many LSTM. For the training, the corpus was separated in sequences of 50 tokens

- " well , we ' ll do it with a wand , " said hermione . " really ? " said harry , looking at each other .

Parameters	Classification	TextGen
Input dim	60k	60k
Embedding dim	256	256
Hidden/cell dim	256	512
Output dim	4	60k

Preprocessing and model details

Before doing any sort or training, we had to do a bit of preprocessing on the data. First, we tokenized each corpus in sequences of 50 tokens.

- what looked about this way , the black citadel , was far from the darkness , the ring was heard , but the sea big was big , and a great ring was in his battle .
- failure is a beginning of love and a family which comes from god .
- " that now my mind shall screens his music , " " and i to give thee my vulgar heaven , " " i am your brother .

REFERENCES

1. Zachary Chase Lipton, 2015. A Critical Review of Recurrent Neural Networks for Sequence Learning, CoRR.