Paper: *On the difficulty of training Recurrent Neural Networks*

Razwan Ahmed Tanvir

Date: 02/13/2022

Quote *"The vanishing gradients problem refers to the opposite behaviour, when long term components go exponentially fast to norm 0, making it impossible for the model to learn correlation between temporally distant events"*

Overview Vanishing Gradient and Exploding Gradient are two known issues with Recurrent Neural Network while training. Bengio et al. (1994) defined the exploding gradients problem as the large increment of the gradient at the time of training and it occurs for exponentially growing long term components. On the other hand, vanishing gradients is when long term components go to norm 0 exponentially. For this reason the correlation between the farther events cannot be learned by the model. Authors of this paper took an attempt to address these mentioned problems by proposing a gradient clipping strategy for exploding gradient and soft constraint for vanishing gradients. Moreover, the authors also explored the dynamical systems perspective to better understand exploding and vanishing gradients. Also, they considered geometrical interpretation. They used a regularization technique by a regularization term which ensures that the vanishing and exploding do not increase or decrease in huge magnitude.

This study has contributed to the better understanding of the gradient and vanishing gradient for the Recurrent Neural Networks and it produced effective output which can be integrated to eliminate the mentioned problems from the model. Authors evaluated their results empirically and validated their hypothesis and outlined their experiments to summarize the research results.

Intellectual Merit *Exploding and Vanishing gradient are important concerns of the Recurrent Neural Networks. To eliminate the possibility of these problems occurring, this paper provided insights and solutions which advanced the natural text processing field. Authors introduced creative ideas to address the issues with RNN and the work is novel in this field in terms of experiment results. Results of this research is validated empirically. One of the mentionable outputs of this research is that their trained model can generalize of twice as long text sequence compared to the sequence provided during training. The authors had their required resources like datasets available to them to carry their research.*

Broader Impact This research has impacts on the field of natural language processing, more specifically in Recurrent Neural Networks. It has addressed the major issues with RNN models and overcoming these concerns advanced this field of research. Furthermore, the authors of this paper explored the previous studies in this area and compared to their results to show that their approach is better in terms of solving the exploding and vanishing gradients in Recurrent Neural Networks during training. The authors are from Montreal University and Brno University.

Keywords Natural Language Processing, Vanishing Gradient, Exploding Gradient, Recurrent Neural Network

Discussion Questions

- Authors have mentioned that their clipping algorith works similar to the Mikolov's algorithm for gradient clipping. Is their any approach of selecting clipping algorithms, we can think of which could potentially improve the RNN model?
- How to identify the threshold for scalling down the gradients? Can we optimally choose the threshold for a given problem?

Table 1: Grade deductions by section

Overview	Intellectual M.	B. Impact	Keywords	Questions	Is Online?