INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

Garden of Knowledge and Virtue

KULLIYYAH OF INFORMATION & COMMUNICATION TECHNOLOGY

CSCI 4342 NATURAL LANGUAGE PROCESSING
SEMESTER 1, 2020/2021

Paper Review
(Universal Language Model Fine-tuning for Text Classification)

PREPARED BY: GROUP Alpha

NAME	MATRIC NO.
Abdella Mame Abdo	1714883
Islam MD Shariful	1720601
Mohamed Moubarak Mohamed Misbahou Mkouboi	1820705

LECTURER

DR. SURIANI BT. SULAIMAN

Review of: Universal Language Model Fine-tuning for Text Classification

Summary of the paper

- What is the paper all about
- The objective/goals of the research
- Research problem addressed by the authors
- Methodology used

Contribution of the paper

- Short introduction to the research
- The main question addressed by the research
- The research work relevancy
- Originality / Unique work
- Successful aspects of the paper

Analytical and Critical review of the paper

- Describe the algorithm/technique used
- Experimental design and methods
- Research Process
- The results achieved
- Data processing steps
- Tables or figures adding

Conclusion

Fine-tuning the Universal Language Model for Text Classification ULMFiT, or Universal Language Model Fine-tuning, is a transfer learning and architectural approach that may be applied to NLP tasks. For its representations, it employs a three-layer AWD-LSTM architecture.

Deep learning models benefit from fine-tuning the Universal Language Model for Text Classification while training new models. It makes the procedure easier by saving time. A massive quantity of data from prior models is loaded. As a result, it can assist save a significant amount of time.

These recurrent neural networks, however, have their own set of issues. One significant difficulty is that RNNs cannot be parallelized since they only accept one input at a time. An RNN or LSTM would take one token at a time as input in the instance of a text sequence. As a result, it will go through the sequence token by token. As a result, training such a model on a large dataset will take a long time.

The Universal Language Model Fine-Tuning (ULMFiT) technique is a recent strategy that suggests training a language model and then transferring its information to a final classifier. ULMFiT selects meaningful information from an embedding sequence using max and average pooling layers during the classification stage.

The authors in this paper introduced approaches for fine-tuning a language model and presented a novel method, Universal Language Model Fine-tuning (ULMFiT), an effective transfer learning method that can be used for any task in NLP.

The study clearly suggested Universal Language Model Fine-tuning (ULMFiT), a mechanism for achieving CV-like transfer learning in NLP for any task. The authors, on the other hand, made a contribution by suggesting discriminative fine-tuning, tilted triangular learning rates, and progressive unfreezing, all of which are unique strategies for retaining prior information and avoiding catastrophic forgetting during fine-tuning.

On six example text classification datasets, the author summarizes the study by greatly beating the state-of-the-art, with an error reduction of 18-24 percent on the majority of datasets. Furthermore, they also demonstrated that their technique allows for exceptionally sample-efficient transfer learning as well as a comprehensive ablation analysis. Additionally, they also made the pre-trained models and their source code publicly accessible in order to facilitate broader adoption.

On six well-researched text classification tasks, the searches by the authors were focused on obtaining the same 3-layer LSTM architecture with the same hyper-parameters and no changes other than tweaked dropout hyper-parameters beating highly developed models and trans- 329 for learning techniques.

Finally, the study technique surpasses the current state-of-the-art on six text classification tasks, lowering error by 18% to 24% on the majority of datasets. Furthermore, it equals the performance of training from scratch on 100 times more data with just 100 tagged instances.

Universal Language Model Fine-tuning, is an architecture and transfer learning method that can be applied to NLP tasks. It involves a 3-layer AWD-LSTM architecture for its representations. The training consists of three steps: 1) general language model pre-training on a Wikipedia-based text, 2) fine-tuning the language model on a target task, and 3) fine-tuning the classifier on the target task.

As different layers capture different types of information, they are fine-tuned to different extents using discriminative fine-tuning. Training is performed using Slanted triangular learning rates, a learning rate scheduling strategy that first linearly increases the learning rate and then linearly decays it.

Fine-tuning the target classifier is achieved in ULMFiT using gradual unfreezing. Rather than fine-tuning all layers at once, which risks catastrophic forgetting, ULMFiT gradually unfreezes the model starting from the last layer, such as closest to the output, as this contains the least general knowledge. First, the last layer is unfrozen and all unfrozen layers are fine-tuned for one epoch. Then the next group of frozen layers is unfrozen and fine-tuned and repeated until all layers are fine-tuned until convergence at the last iteration.

For data processing, the study uses a pre-processing that has been used in other studies. Also, in order to make the algorithm able to recognize other aspects of the language model that could be considered relevant when it comes to classification, the study added special tokens for the upper-case words, elongation as well as repetition.

Lastly, when it comes to the results, this method outperforms both CoVe, a state-of-the-art transfer learning method based on hypercolumns, as well as the state-of-the-art on both datasets. On IMDb, errors have been reduced dramatically by 43.9% and 22% with regard to CoVe and the state-of-the-art respectively. There are promising aspects to these results since the existing state-of-the-art requires complex architectures, multiple forms of attention, and sophisticated embedding schemes. On IMDb, with 100 labeled examples, ULMFiT matches the performance of training from scratch with 10× and is given 50k unlabeled examples with 100× more data.

The author has successfully proposed ULMFiT, an effective and extremely sample-efficient transfer learning method that can be applied to any NLP task. And also novel fine-tuning techniques in conjunction prevent catastrophic forgetting.