SIGIR'24 (author)

Docs / Log out

New Submission | Submission 2613 | Help | Conference | News | EasyChair

SIGIR'24 Submission 2613

Submission information updates are disabled. If you need to update your paper, contact SIGIR'24 chairs.

For all questions related to processing your submission you should contact the conference organizers. Click here to see information about this conference.

All **reviews sent to you** can be found at the bottom of this page.

Submission 2613	
Title	Investigating the Synergistic Effects of Dropout and Residual Connections on Language Model Training
Paper:	(Feb 09, 02:07 GMT)
Track	Short paper
Author keywords	Large Language Model Generative Pretrained Transformer (GPT) Transformer Networks Deep Learning
Topics	Machine Learning and Natural Language Processing for Search and Recommendation
Abstract	This paper examines the pivotal role of dropout techniques in mitigating overfitting in language model training. It conducts a comprehensive investigation into the influence of variable dropout rates on both individual layers and residual connections within the context of language modeling. Our study conducts training of a decoder implementation on the classic Tiny Shakespeare data to examine the effects of the adjustments on training efficiency and validation error. Results not only confirm the benefits of dropout for regularization and residuals for convergence, but also reveal their interesting interactions. There exists an important trade-off between the depth of residual connections and the dropout on these connections for optimal deep neural network convergence and generalization.
Submitted	Jan 30, 04:25 GMT
Last update	Feb 09, 02:11 GMT
Author roles	Qingyang Li: Investigation, Data curation, Writing – Original Draft Preparation Weimao Ke: Conceptualization, Methodology, Supervision
Student Paper	Yes
Official publication date	Yes
Presentation at the conference	Yes
Intention about Archival version	No intention to do so
Author conflicts	none

Authors

first name	last name	email	country	affiliation	Web page	corresponding?
Qingyang	Li	jl4533@drexel.edu	United States	Drexel University		
Weimao	Ke	wk77@drexel.edu	United States	Drexel University		✓

Reviews

Review 1	
Relevance to SIGIR	**4**: (good)

Technical soundness	**3**: (fair)
Quality of presentation	**3**: (fair)
Adequacy of citations	**3**: (fair)
Strengths	1. This work concentrates on the effects of dropout and residual connections in LLM. The writing is clear.
Weaknesses	1. The novelty is limited and should be clarified. 2. The evaluation should be conducted on different settings (e.g., different types of LLMs/sizes)
Overall recommendation	**-2**: (reject)
Detailed comments to authors	1. Both Dropout and Residual connections have been widely and thoroughly studied in previous works. The authors should clearly explain their unique contribution and insights compared to existing findings. For example, [1][2][3][4] have explored different types of dropout strategies in language models. However, the authors do not give detailed discussions on the correlations between the proposed model and existing efforts, and also do not compare with competitive models. The conclusions also seem to be intuitive and should be clarified (especially explaining what are the NEW findings in this work). 2. The evaluation should be conducted on different settings (e.g., different types of LLMs/sizes) to verify the universality of the conclusions. This work only conducts on the basic Transformer structure with relatively few tokens (nearly 200k words), which is far from the practical large language model setting. References: [1] Fan A, Grave E, Joulin A. Reducing transformer depth on demand with structured dropout[J]. arXiv preprint arXiv:1909.11556, 2019. [2] Wu Z, Wu L, Meng Q, et al. Unidrop: A simple yet effective technique to improve transformer without extra cost[J]. arXiv preprint arXiv:2104.049 [3] Wu L, Li J, Wang Y, et al. R-drop: Regularized dropout for neural networks[J]. Advances in Neural Information Processing Systems, 2021, 34: 10890-10905. [4] Pham H, Le Q. Autodropout: Learning dropout patterns to regularize deep networks[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2021, 35(11): 9351-9359.

Review 2

Relevance to SIGIR	**2**: (poor)
Technical soundness	**5**: (excellent)
Quality of presentation	**3**: (fair)
Adequacy of citations	**3**: (fair)
Strengths	- This paper presents an empirical study on how dropout affects the performance of the language models, which is an important research direction.
Weaknesses	- Lack of novelty, previous works have discussed the impact of using dropout [1,2,3,4] - Limited experimental results. This paper explores a small pertaining text corpus and language models. These findings may not be able to be generalised to other settings. The authors merely test a very toy case with a small pertaining corpus and a small language model. This raises my concerns whether this paper's findings can offer a generalizable conclusion for the commnuity. - Irrelevant to SIGIR. [1] Suman Adhya, Avishek Lahiri, and Debarshi Kumar Sanyal. 2023. Do Neural Topic Models Really Need Dropout? Analysis of the Effect of Dropout in Topic Modeling. In Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics, pages 2220–2229, Dubrovnik, Croatia. Association for Computational Linguistics. [2] A theoretically grounded application of dropout in recurrent neural networks. NeurIPS 2016. [3] Pushing the bounds of dropout. [4] Analysing Dropout and Compounding Errors in Neural Language Models.
Overall recommendation	**-2**: (reject)
Detailed comments to authors	- How could the Number of iterations and Training time be Evaluation Metrics?

Review 3	
Relevance to SIGIR	**5**: (excellent)
Technical soundness	**3**: (fair)
Quality of presentation	**3**: (fair)
Adequacy of citations	**4**: (good)
Strengths	1. The article empirically validates the importance of Dropout and Residual Connections in language model training, especially in preventing overfitting and improving model convergence. 2. It reveals the interaction between Dropout rates and the depth of Residual Connections, providing new insights for optimizing deep neural networks. 3. Experiment results suggest that appropriate configurations of Dropout rates and Residual Connections can enhance the model's generalization capabilities.
Weaknesses	1. The experiments mentioned in the article are conducted on relatively small models and datasets, which may not fully represent the performance on larger networks and datasets. 2. The differences observed in the article are minor, which may imply that the impact of these optimization techniques might not be as significant in practical applications as they are in the experiments. 3. The article does not provide specific guidance on how to precisely adjust these parameters in real-world applications, which could make it challenging to apply these findings to other tasks.
Overall recommendation	**-1**: (weak reject)
Detailed comments to authors	The paper explores how dropout and residual connections interact to improve language model training. It uses the Tiny Shakespeare dataset for training to study the effects of different dropout rates on model layers and residual connections. The results highlight the trade-off between regularization (dropout) and model convergence (residual connections) for optimal neural network performance. However, training tokens and the number of model parameters are so important that these experiment results are necessary. It is the key evidence for scaling law.

Metareview

Metareview for paper 2613	
Title	Investigating the Synergistic Effects of Dropout and Residual Connections on Language Model Training
Authors	Qingyang Li and Weimao Ke
Text	The paper explores how dropout and residual connections interact to improve language model training. As strengths, the reviewers mention that the writing is clear and that the research direction is important. The main weaknesses mentioned are: limited novelty, small models/datasets in experiments and minor differences findings that lack sufficient explanation. Ultimately, the reviewers have decided that the paper is not ready for publication.