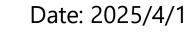


Can Textual Gradient Work in Federated Learning?

Author: Minghui Chen^{1,2,3}, Ruinan Jin^{1,2}, Wenlong Deng^{1,2}, Yuanyuan Chen³, Zhi Huang⁴, Han Yu³, Xiaoxiao Li^{1,2}

Affiliation: ¹The University of British Columbia, ²Vector Institute, ³Nanyan Technological University, ⁴University of Pennsylvania

Speaker: Minghui Chen







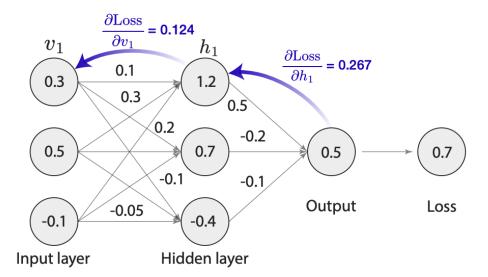


0. Overview: TextGrad Background

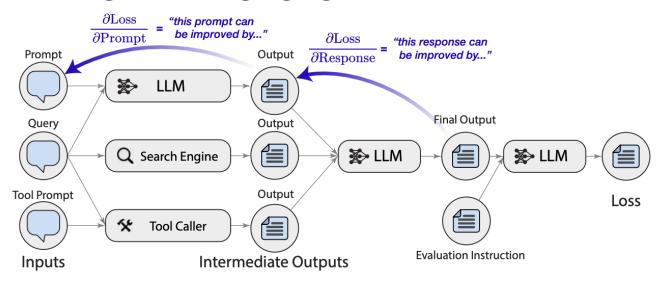
TextGrad: An automated framework that **optimizes compound AI systems** by **backpropagating** LLM feedback through natural language suggestions. It improves tasks like question answering, molecule design, and treatment planning **without manual tuning**, following PyTorch-like syntax for flexibility.

Applications: (i) Non-differentiable loss and inaccessible gradients; (ii) End-to-end optimization: coding, medical report generation, complex reasoning, and molecule design.

a Neural network and backpropagation using numerical gradients

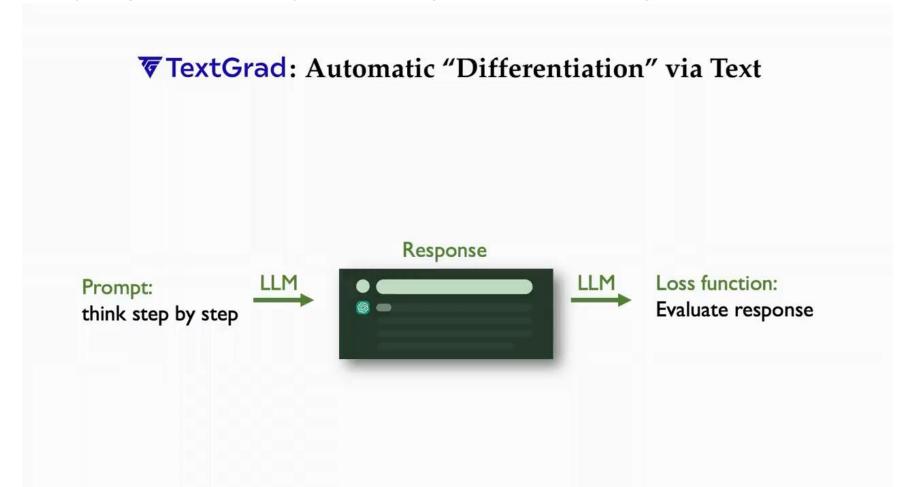


b Blackbox AI systems and backpropagation using natural language 'gradients'



0. Overview: TextGrad Background

TextGrad Applications: (i) Non-differentiable loss and inaccessible gradients; (ii) End-to-end optimization: coding, medical report generation, complex reasoning, and molecule design.



0. Overview of FedTextGrad Motivations A Practical Federated **Learning New Setting:** Federated TextGrad Framework

Challenges: Effective and **Efficient** Prompt Aggregation

How to aggregate?



• Concatenation?

Aggregated Prompt: You will answer a reasoning question by thinking step-by-step and providing a clear and concise response that accurately addresses the question being asked. Consider the context of the question and provide a response that is relevant and accurate. To achieve this, break down the problem into smaller steps and provide a clear explanation of each

Updated Prompt: You will answer a reasoning question by thinking step-by-step and providing a clear and concise response that accurately addresses the question being asked. Consider the context of the question and provide a response that is relevant and accurate. Ensure that your response includes a clear and concise answer in the format: 'Answer: \$VALUE' where VALUE is a numerical value. ...

5. Prompt

Aggregation

Server

Query: I have three oranges, a pig, a frog, a cow, three bananas, a nectarine, and a snail. How many animals do I have?

Prompt: You will answer a reasoning question. Think step by step. The last line of your response should be of the following format: 'Answer: \$VALUE' where VALUE is a numerical value.

Response: To find the total number of animals, we need to identify and count the animals in the given list.

- 1. The list includes a pig, a frog, and a cow. These are all animals.
- 2. We don't need to count the fruits and snail as they are not animals.

So, the total number of animals is

Ground Truth: 4 Answer: 3

Evaluation: 0

Multi-Epoch

∂Prompt

 $\frac{1}{1}$ = "improve the prompt by ..."

= "improve the response by ..."

4. Textural Grad

1.Prompt

2.Respons

3. Evaluation Client 1

= "improve the prompt by ... ∂Prompt

LLM

LLM

= "improve the response by ..."

LLM

LLM

4. Textural Grad

Challenges



step, ...

Heterogeneity



Robustness



Privacy



Efficiency

Updated Prompt: You will answer a reasoning question. Think step by step and provide a clear and concise explanation of your reasoning and calculations. Make sure to include any intermediate steps or calculations and clearly indicate the final answer within your explanation. Do not simply state the final answer without providing any explanation or justification...

1.Prompt

Query: I have a chicken, a frog, a mouse, a cat, two pigs, and two rabbits. How many animals do I have?

Prompt: You will answer a reasoning question. Think step by step. The last line of your response should be of the following format: 'Answer: \$VALUE' where VALUE is a numerical value.

2.Respons

Response: To find the total number of animals, we need to count each type of animal individually and then add them together.

- 1. You have 1 chicken.
- 2. You have 1 frog.

Now, let's add all the animals together:

- 1 (chicken) + 1 (frog) + 1 (mouse)
- + 1 (cat) + 2 (pigs) + 2 (rabbits) = 8

3. Evaluation Client 2

Answer: 8

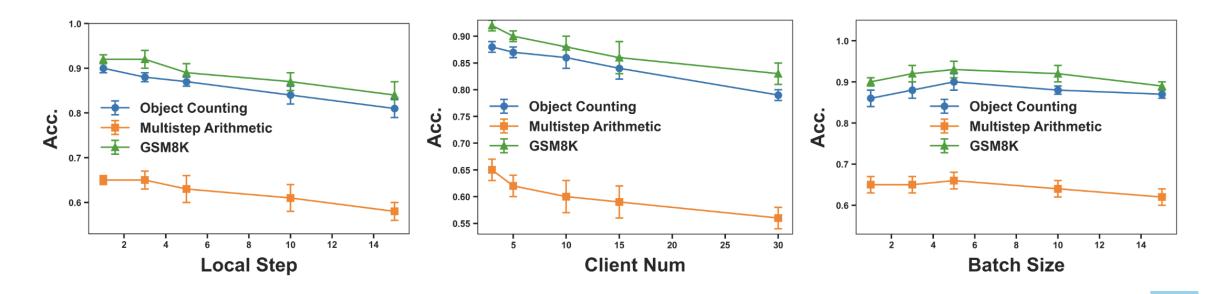
Ground Truth: 8

Evaluation: 1

3. Experiments

Key Results on Evaluations

Trend Analysis: Increasing the number of local steps and clients results in performance drops, likely due to added heterogeneity and overfitting. Conversely, increasing batch size initially boosts performance but eventually causes a slight decline, indicating an optimal batch size range for maintaining effective prompt optimization.



4. Conclusion

Takeaways:

- 1. **FedTextGrad Introduction**: A novel FL framework leveraging LLM-generated textual gradients for prompt optimization, enabling privacy-preserving collaboration on tasks without numerical loss functions.
- 2. **Prompt Aggregation Challenges**: Effective aggregation requires balancing length and content; UID-based summarization outperforms concatenation and traditional summarization.
- 3. **Key Insights**: Optimal local steps and batch sizes enhance performance, while excessive updates or client heterogeneity reduce global alignment. Validated on tasks like BBH and GSM8K.