

# LM-BFF: Making Pre-trained Language Models Better Few-shot Learners

---

*By the DeepSingers:*

Xinhu Li, Stephen Iota, Kai Chen, Shuhong Lu, Zhi Tu

# Introduction

*Task:* few-shot learning in NLP models

*Motivation:*

- Previous models (GPT-3) have achieved great performance
  - but with a too **big** model so that it is too **time-consuming**.
- **practical (smaller)** language models e.g., BERT

*Proposal:* **LM-BFF** (Better Few-shot Fine-tuning of Language Models)

- A suite of **fine-tuning techniques** on a **small number** of annotated examples.

# Prompt-based Fine-tuning

Motivation:

1. Lack of training data (fewer shot learning)

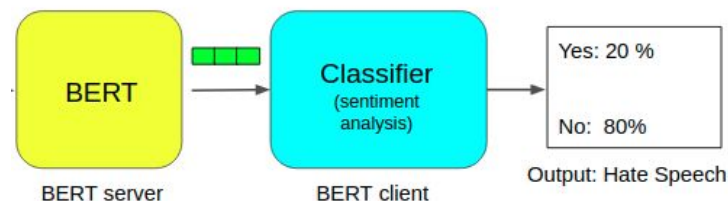
Design:

1. Pick a prompt and view the task as a masked language model (MLM).

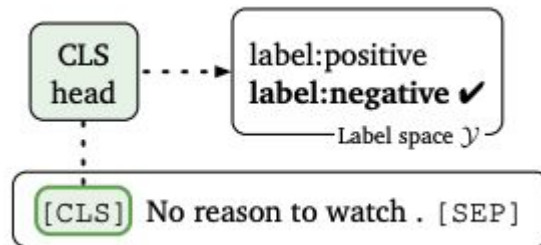
[CLS]  $x_1$  It was [MASK] . [SEP]

Advantage::

1. No additional parameters.
2. Make better use of the pre-trained model.



Traditional Approach



Prompt-based Approach

$$\begin{aligned} p(y \mid x_{\text{in}}) &= p([\text{MASK}] = \mathcal{M}(y) \mid x_{\text{prompt}}) \\ &= \frac{\exp(\mathbf{w}_{\mathcal{M}(y)} \cdot \mathbf{h}_{[\text{MASK}]})}{\sum_{y' \in \mathcal{Y}} \exp(\mathbf{w}_{\mathcal{M}(y')} \cdot \mathbf{h}_{[\text{MASK}]})} \end{aligned}$$

# Automatic-prompt Generation

## Motivation:

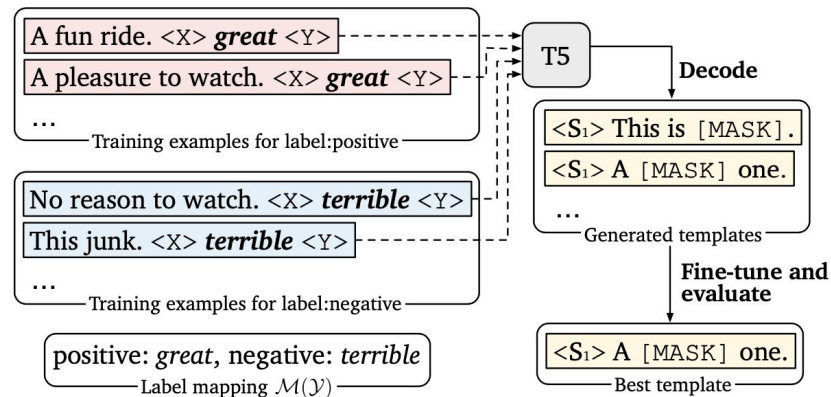
1. **Find more optimal settings** than those that we manually choose.
2. **Reduce the human involvement** required to design prompts

## Design:

1. Automatic selection of label words

$$\text{Top-}k_{v \in \mathcal{V}} \left\{ \sum_{x_{\text{in}} \in \mathcal{D}_{\text{train}}^c} \log P_{\mathcal{L}} \left( [\text{MASK}] = v \mid \mathcal{T}(x_{\text{in}}) \right) \right\}, \quad (3)$$

2. Automatic generation of templates



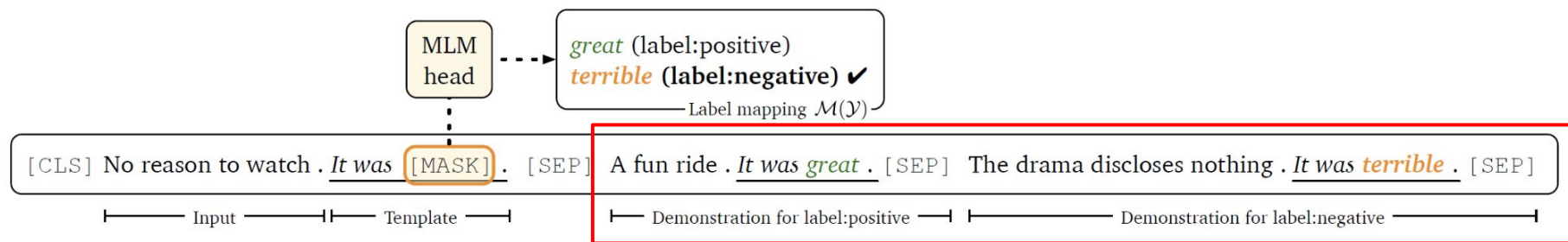
# Fine-tuning with Demonstration

## Problem:

- Limitation of maximum input length
- Extremely long contexts

## Solution:

- Sample one example from each class
- Obtain embedding from pretrained SBERT model, Calculate the cosine similarity, sample from top 50%



(c) Prompt-based fine-tuning with demonstrations (our approach)

# Final thoughts on LM-BFF

*Prompting* is a very promising direction for few-shot learning tasks in NLP.

- **Leverages** existing smaller LMs.
- **Accelerates** research in typical academic settings.
- **Reduces** NLP's carbon footprint.

*However...*

- Prompting has limited use cases.
- More data is still better than less data.



*Potential follow-ups...*

- *Linguistic insights*: What can LLMs teach us about language?
  - e.g., use BERT to mine knowledge, facts, reasoning.