$\mathcal{Y}$-Tuning: An Efficient Tuning Paradigm for Large-Scale Pre-Trained Models via Label Representation Learning

Yitao LIU, Chenxin AN, Xipeng QIU

Frontiers of Computer Science, DOI: 10.1007/s11704-023-3131-8

Problems & Ideas

- Problems of conventional lightweight-tuning paradigms:
  - Need to save the computational graph for gradient descent, which is not training-efficient.
  - Many very-large-scale PTMs only supply inference API, which hinders the application of invasive tuning paradigms such as adapter-tuning.
- Ideas: Tune additional modules via label representation learning while no backpropagation is required on large models.

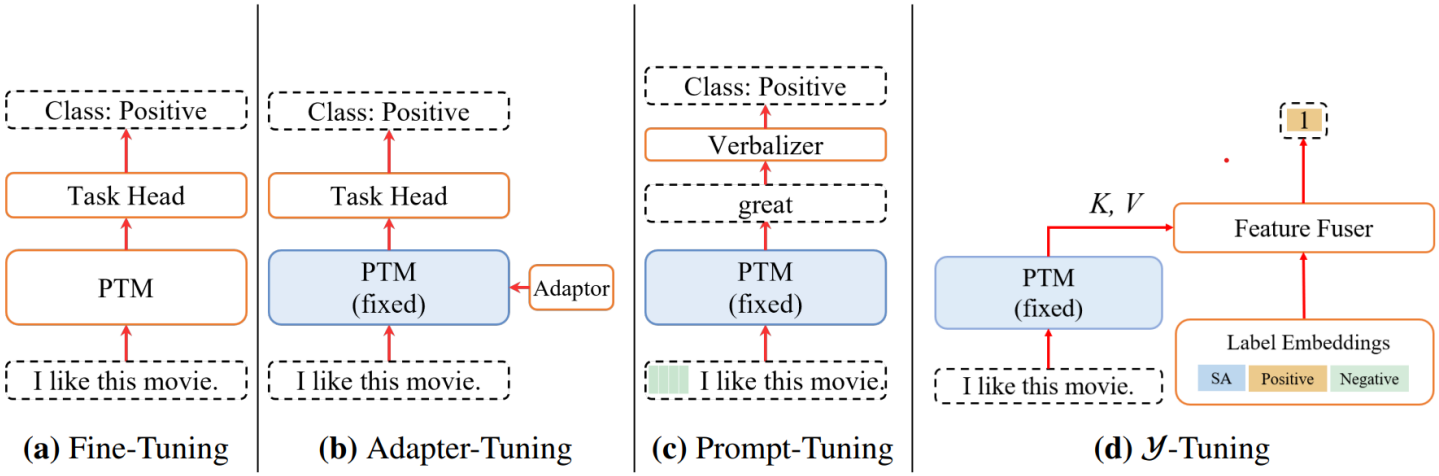
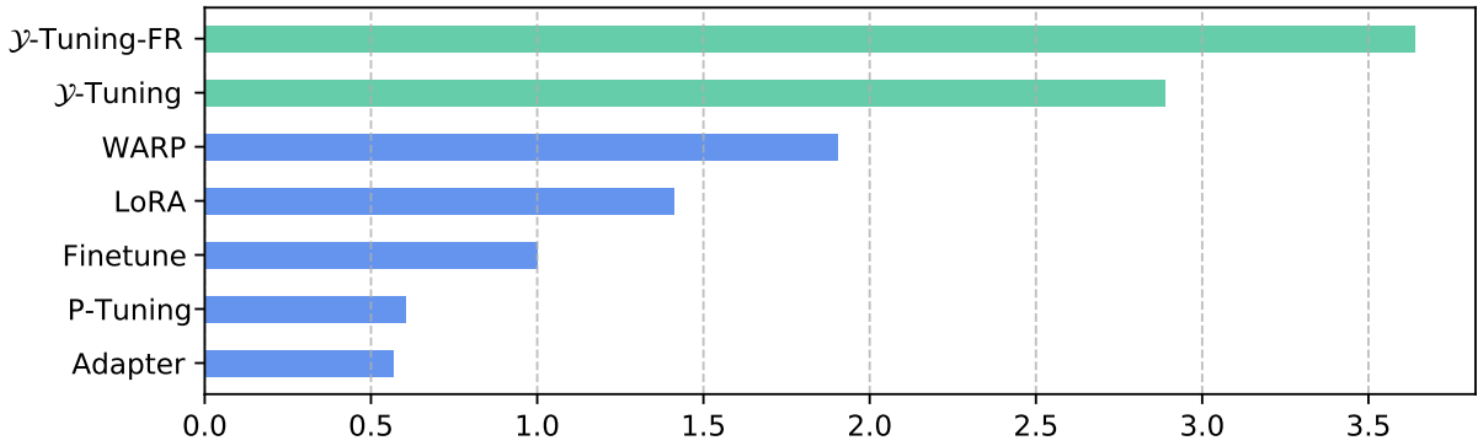


Illustration of four tuning paradigms. The proposed method, $\mathcal{Y}$-Tuning, aims to build a more powerful label space and adapt it to feature space. With this tuning paradigm, we can train PTMs efficiently with less time and memory consuming than other tuning paradigm.

Main Contributions

- Contributions:
  - A parameter-efficient and training-efficient tuning paradigm which doesn't require gradients of large PTMs in the training phase while achieves close experimental performance with full fine-tuning.
  - Label representation learning on text feature leads to better robustness and difficult to text-based attack.
  - Generalize proposed method to broader scenarios like sequence-labeling and span-based question answering.



Training speed-up ratio over fine-tuning on various methods. $\mathcal{Y}$-Tuning is much faster than directly fine-tuning. For $\mathcal{Y}$-Tuning the feature representations generated by the frozen PTM can be saved on disks during the first epoch and reused for subsequent epochs (denoted as $\mathcal{Y}$-Tuning-FR), which further reduces the overhead of computation tremendously.