Multitask Fine-Tuning for Enhanced Model Performance

Introduction

Multitask fine-tuning extends single-task fine-tuning by training the model on a dataset comprising examples for multiple tasks. By simultaneously improving the model's performance across various tasks, multitask fine-tuning mitigates the risk of catastrophic forgetting.

Key Concepts

Training Dataset: Contains examples instructing the model to perform diverse tasks like summarization, review rating, code translation, and entity recognition.
Simultaneous Improvement: Through multitask fine-tuning, the model learns to excel at multiple tasks concurrently, enhancing its versatility and utility.
Data Requirement: Multitask fine-tuning necessitates a substantial amount of data, typically 50,000 to 100,000 examples, across various tasks.
FLAN Models: FLAN (Fine-tuned Language Net) models, such as FLAN-T5 and FLAN-PALM, are products of multitask fine-tuning. These models undergo fine-tuning across numerous datasets and task categories to achieve broad capabilities.

FLAN-T5: A Case Study

Training Datasets: FLAN-T5 is fine-tuned on 473 datasets across 146 task categories, covering a wide range of tasks.
Example Dataset: SAMSum, a dataset for dialogue summarization, is utilized to train FLAN-T5 on summarizing conversational dialogues.
Prompt Templates: Templates like "Summarize a dialogue" are applied to the SAMSum dataset to create instruction prompts for fine-tuning FLAN-T5.

Customization for Specific Use Cases

Scenario: Consider a data scientist developing a customer service app requiring dialogue summarization.
Fine-Tuning with Domain-Specific Data: Additional fine-tuning using a dataset like dialogsum, consisting of support chat dialogues, further enhances FLAN-T5's summarization capabilities for customer service contexts.

Evaluation and Quality Assessment

Pre-Fine-Tuning Evaluation: Assessing FLAN-T5's initial response to prompts reveals its performance gaps.
Post-Fine-Tuning Improvement: After fine-tuning on dialogsum, FLAN-T5 produces more accurate and contextually relevant summaries, aligning closely with human-produced summaries.