Cerebrium Follow

Oct 24 · 5 min read · Listen

SetFit outperforms GPT-3 while being 1600x smaller

Everyone is very familiar with the current hype around Large Language Models (LLM) such as GPT-3 and Image Generation models such as DALL-E 2 and Stable diffusion. However, the results of these models come at a price.

- GPT-3: ~$12 Million
- DALL-E: ~$500k- $1 million
- Stable Diffusion: ~$600k

This is due to the large number of GPU's required to process and train these models. Besides the cost, the amount of labelled data required to achieve these results is difficult to source. Previously it was not possible for startups to train models of this calibre because of these two factors until now— introducing Sentence Transformer Fine-tuning (SetFit) a simple and efficient alternative for few-shot text classification unveiled by the teams at Intel Labs, UKP Labs and Hugging Face.

> *Few shot classification is a NLP task in which a model aims to classify text into a large number of categories, given only a few training examples per category*

Compared to other few-shot learning methods, SetFit has several unique features:

- **No prompts**: Most techniques around few-shot learning require handcrafted prompts to convert examples into a format thats suitable for the underlying model which is also known as 'Prompt Engineering'. SetFit does not require this but instead generates rich embeddings directly from a small number of labeled examples. This radically reduces the labelled data requirement and does not require extensive prompt engineering which was often one of the biggest drawbacks

> *Prompt engineering is when the description of the task is embedded in the input, e.g., as a question instead of it being implicitly given.*

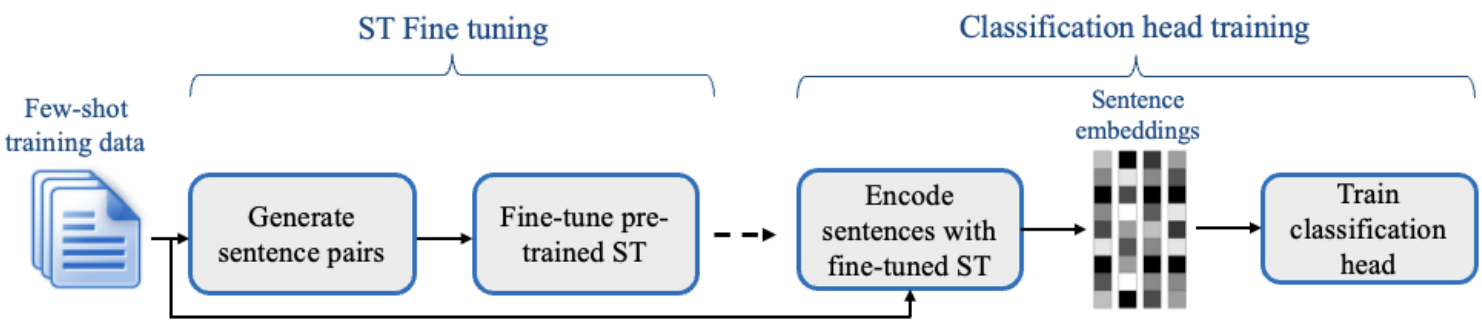
- **Fast to train:** Due to the smaller data requirement, SetFit does not require large scale in order to achieve similar (and in some cases better) accuracy to GPT-3 (175b Parameters). This reduces the training time required by orders of magnitude.
- **Multilingual support:** SetFit can be used with any Sentence Transformer on Hugging Face Hub, which means you can classify text in multiple languages by simply fine-tuning a multilingual checkpoint.

How does it work

To start, you have to understand what a sentence transformer is. It's a popular approach for text and image embeddings that encode a vector representation based on its semantic signature. The representation is built during contrastive training which is a form of self-suprvised learning that augments views of the same input. For example, "What time is it?" is semantically the same as "How late is it". The aim in contrastive training is to minimise the distance between semantically similar sentences and maximise distances between sentences that are semantically distant. Model Hub on Hugging Face contains over 100 pre-trained sentence transformers based on a variety of datasets.

The first step of SetFit is choosing a Sentence Transformer (ST) from from the model hub. The ability to select any ST from model hub is what enables SetFit to have multilingual support since there are ST models for over 100 languages. SetFit then fine-tunes the Sentence Transformer based on a small number of labelled examples using contrastive learning — where the positive pairs are two sentences chosen randomly from the same class and the negative pairs are two sentences chosen randomly from different classes. An adapted ST is now produced.

The sentences in the training data are now encoded using the adapted ST which creates sentence embeddings. These sentence embeddings are then utilised to create a simple logistic regression model for simplicity. At inference time, the data is encoded at using the adapted ST and classified using the trained Logistic Regression model.



SetFit's two-stage training process

Benchmarking

RAFT is a few-shot classification benchmark designed to match real world scenarios by restricting the number of training samples to 50 labeled examples per task and not providing validation sets.

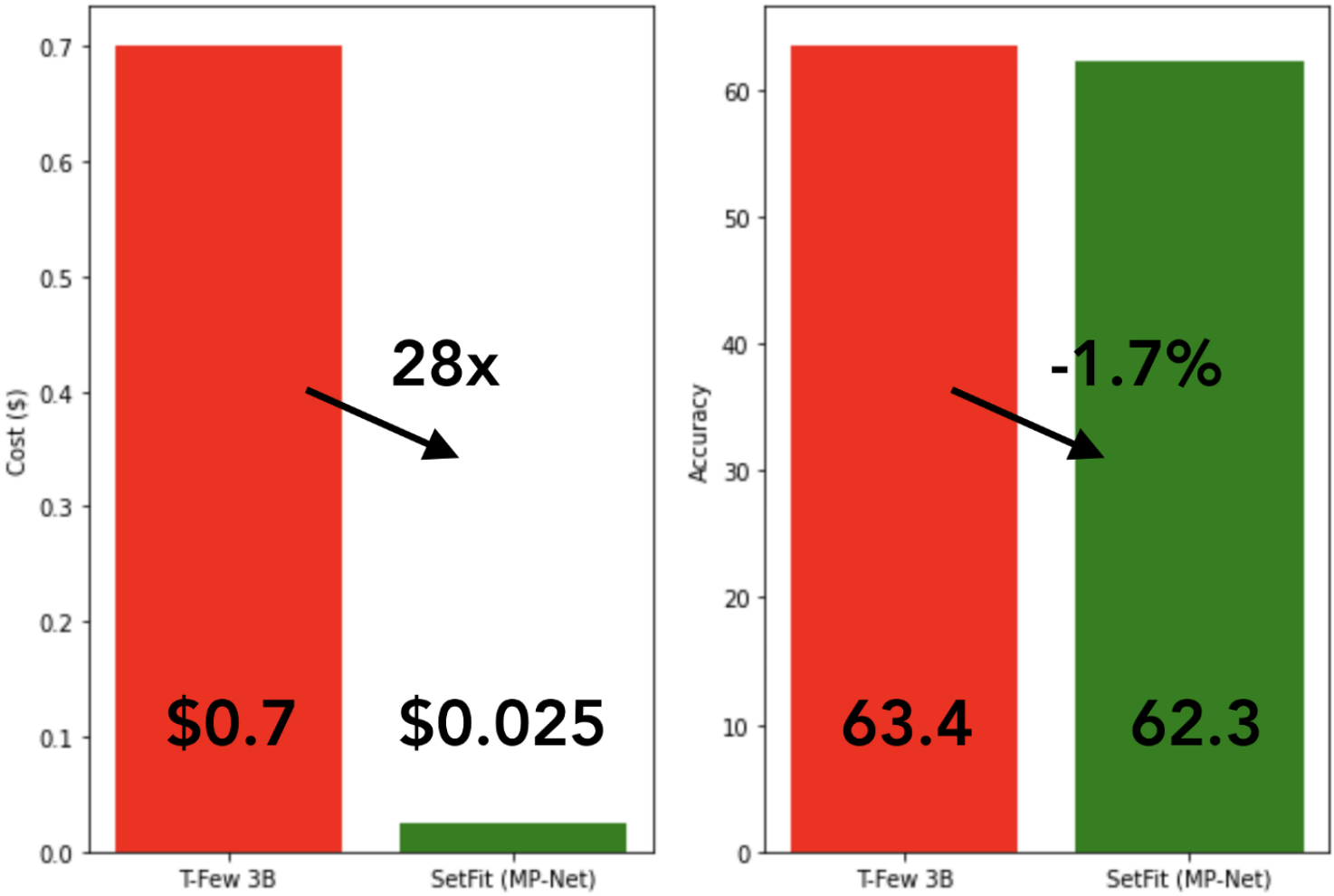
Rank	Method	Accuracy	Model Size
2	T-Few	75.8	11B
4	Human Baseline	73.5	N/A
6	SetFit (Roberta Large)	71.3	355M
9	PET	69.6	235M
11	SetFit (MP-Net)	66.9	110M
12	GPT-3	62.7	175 B

Prominent methods on the RAFT Leaderboard

The prominent characteristics of SetFit that impact its performance include:

1. The type of ST: there are many types of STs in the sentence-transformers model hub, and the question is how to choose the best ST for a given task.
2. Input data selection: in some of the RAFT tasks, several data fields are provided as input, for example title, abstract, author name, ID, data, etc. This gives rise to the question: which data fields should be used as input and how to combine them?
3. Choice of hyperparameters: how does one choose the best fine-tuning hyperparameter set (e.g. #epochs, number of sentence-pairs generation iterations)?

Training and inference



Comparing training cost and average performance for T-Few 3B and SetFit (MPNet), with 8 labeled examples per class

Due to the small data requirement of SetFit, its extremely fast to train. Hugging Face reported the following statistics on training:

> *Training SetFit on an NVIDIA V100 with 8 labeled examples takes just 30 seconds, at a cost of $0.025. By comparison, training T-Few 3B requires an NVIDIA A100 and takes 11 minutes, at a cost of around $0.7 for the same experiment — a factor of 28x more. In fact, SetFit can run on a single GPU like the ones found on Google Colab and you can even train SetFit on CPU in just a few minutes! As shown in the figure above, SetFit's speed-up comes with comparable model performance.*

SetFit is an extremely effective method for few-shot classification tasks as shown by the evaluation on RAFT. It allows businesses with a small amount of data to build powerful text classifiers at a cheap cost — two criteria that often prohibited startups from implementing these types of solutions. If you want to learn how to fine-tune your own model you can go through the short tutorial on Hugging Face here. Otherwise get in touch with us and we will help you!

Thanks to Jonathan Irwin