DistilBERT (Distilled BERT) is a smaller, faster, and lighter version of BERT (Bidirectional Encoder Representations from Transformers). It was introduced by Hugging Face and leverages knowledge distillation to achieve performance similar to BERT with significantly fewer parameters.

Key Concepts of DistilBERT

1. Knowledge Distillation:
   - A teacher model (like BERT) trains a student model (DistilBERT) to mimic its behavior.
     - During training, the student model learns not only from the true labels (hard targets) but also from the "soft labels" (output probabilities) produced by the teacher model.
   - This helps the smaller model retain much of the performance of the larger one.

2. Size Reduction:
   - DistilBERT has about 40% fewer parameters than BERT.
   - This reduction is achieved by:
     - Removing the token-type embeddings (used for tasks like question-answering).
     - Reducing the number of transformer layers by half (BERT-base has 12 layers; DistilBERT has 6).

3. Speed and Efficiency:
   - DistilBERT is about 60% faster in inference while retaining 97% of BERT's performance on various natural language understanding tasks.

4. Training Objectives:
   - Masked Language Modeling (MLM): Predicts randomly masked words in a sentence (like BERT).
   - Distillation Loss: Matches the soft probabilities of the teacher and student models.

Applications of DistilBERT

DistilBERT can perform various NLP tasks, similar to BERT:

1. Text classification (e.g., sentiment analysis)
2. Named Entity Recognition (NER)
3. Question Answering
4. Text Summarization
5. Language Translation

⌄ Advantages of DistilBERT

1. Compact: Requires less memory and computational power.
2. Fast: Ideal for real-time applications like chatbots or mobile devices.
3. Generalizable: Retains the versatility of BERT for various NLP tasks.

Here's how you can use DistilBERT for text classification with the Hugging Face library

```
!pip install transformers datasets
```

```
Requirement already satisfied: transformers in /usr/local/lib/python3.10/dist-packages (4.47.1)
Requirement already satisfied: datasets in /usr/local/lib/python3.10/dist-packages (3.2.0)
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-packages (from transformers) (3.16.1)
Requirement already satisfied: huggingface-hub<1.0,>=0.24.0 in /usr/local/lib/python3.10/dist-packages (from transformers) (
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.10/dist-packages (from transformers) (1.26.4)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from transformers) (24.2)
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.10/dist-packages (from transformers) (6.0.2)
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.10/dist-packages (from transformers) (2024.11.6)
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-packages (from transformers) (2.32.3)
Requirement already satisfied: tokenizers<0.22,>=0.21 in /usr/local/lib/python3.10/dist-packages (from transformers) (0.21.0
Requirement already satisfied: safetensors>=0.4.1 in /usr/local/lib/python3.10/dist-packages (from transformers) (0.5.0)
Requirement already satisfied: tqdm>=4.27 in /usr/local/lib/python3.10/dist-packages (from transformers) (4.67.1)
Requirement already satisfied: pyarrow>=15.0.0 in /usr/local/lib/python3.10/dist-packages (from datasets) (17.0.0)
Requirement already satisfied: dill<0.3.9,>=0.3.0 in /usr/local/lib/python3.10/dist-packages (from datasets) (0.3.8)
Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages (from datasets) (2.2.2)
```

```
Requirement already satisfied: xxhash in /usr/local/lib/python3.10/dist-packages (from datasets) (3.5.0)
Requirement already satisfied: multiprocess<0.70.17 in /usr/local/lib/python3.10/dist-packages (from datasets) (0.70.16)
Requirement already satisfied: fsspec<=2024.9.0,>=2023.1.0 in /usr/local/lib/python3.10/dist-packages (from fsspec[http]<=20
Requirement already satisfied: aiohttp in /usr/local/lib/python3.10/dist-packages (from datasets) (3.11.11)
Requirement already satisfied: aiohappyeyeballs>=2.3.0 in /usr/local/lib/python3.10/dist-packages (from aiohttp->datasets) (
Requirement already satisfied: aiosignal>=1.1.2 in /usr/local/lib/python3.10/dist-packages (from aiohttp->datasets) (1.3.2)
Requirement already satisfied: async-timeout<6.0,>=4.0 in /usr/local/lib/python3.10/dist-packages (from aiohttp->datasets) (
Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.10/dist-packages (from aiohttp->datasets) (24.3.0)
Requirement already satisfied: frozenlist>=1.1.1 in /usr/local/lib/python3.10/dist-packages (from aiohttp->datasets) (1.5.0)
Requirement already satisfied: multidict<7.0,>=4.5 in /usr/local/lib/python3.10/dist-packages (from aiohttp->datasets) (6.1.
Requirement already satisfied: propcache>=0.2.0 in /usr/local/lib/python3.10/dist-packages (from aiohttp->datasets) (0.2.1)
Requirement already satisfied: yarl<2.0,>=1.17.0 in /usr/local/lib/python3.10/dist-packages (from aiohttp->datasets) (1.18.3
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.10/dist-packages (from huggingface-hub<1
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests->transform
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests->transformers) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests->transformers) (
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests->transformers) (
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.10/dist-packages (from pandas->datasets) (2.
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas->datasets) (2024.2)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.10/dist-packages (from pandas->datasets) (2024.2)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.8.2->pandas->dat
```

```python
from transformers import DistilBertTokenizer, DistilBertForSequenceClassification
from transformers import Trainer, TrainingArguments
from datasets import load_dataset
```

```python
import os
os.environ["WANDB_MODE"] = "disabled"
```

```python
# Load dataset (e.g., IMDb for sentiment analysis)
dataset = load_dataset("imdb")
tokenizer = DistilBertTokenizer.from_pretrained("distilbert-base-uncased")
```

```
/usr/local/lib/python3.10/dist-packages/huggingface_hub/utils/_auth.py:94: UserWarning:
The secret `HF_TOKEN` does not exist in your Colab secrets.
To authenticate with the Hugging Face Hub, create a token in your settings tab (https://huggingface.co/settings/tokens), set
You will be able to reuse this secret in all of your notebooks.
Please note that authentication is recommended but still optional to access public models or datasets.
  warnings.warn(
```

```python
# Tokenize data
def preprocess_data(example):
    return tokenizer(example['text'], truncation=True, padding='max_length', max_length=128)

tokenized_datasets = dataset.map(preprocess_data, batched=True)
tokenized_datasets = tokenized_datasets.rename_column("label", "labels")
tokenized_datasets.set_format(type="torch", columns=["input_ids", "attention_mask", "labels"])
```

```python
# Load model
model = DistilBertForSequenceClassification.from_pretrained("distilbert-base-uncased", num_labels=2)

# Training arguments
training_args = TrainingArguments(
    output_dir="./results",
    evaluation_strategy="epoch",
    learning_rate=2e-5,
    per_device_train_batch_size=16,
    num_train_epochs=3,
    weight_decay=0.01,
)
```

```
Some weights of DistilBertForSequenceClassification were not initialized from the model checkpoint at distilbert-base-uncase
You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.
/usr/local/lib/python3.10/dist-packages/transformers/training_args.py:1575: FutureWarning: `evaluation_strategy` is deprecat
  warnings.warn(
```

```python
# Trainer
trainer = Trainer(
```

```
    model=model,
    args=training_args,
    train_dataset=tokenized_datasets["train"],
    eval_dataset=tokenized_datasets["test"],
)

# Train model
trainer.train()
```

⋯ **wandb**: WARNING The `run_name` is currently set to the same value as `TrainingArguments.output_dir`. If this was not intended

[9/4689 01:23 < 15:35:46, 0.08 it/s, Epoch 0.01/3]

Epoch Training Loss Validation Loss

```
# Evaluate
trainer.evaluate()
```

Summary

- DistilBERT is a smaller, faster alternative to BERT that uses knowledge distillation to retain most of BERT's performance.
- It is ideal for real-world applications where computational resources are limited.
- You can use libraries like Hugging Face to quickly fine-tune DistilBERT for your specific tasks.