

Bài thực hành PEFT: Phân loại phản hồi khách hàng (Feedback)

Dưới đây là **một bài thực hành PEFT gắn liền với thực tế**, cụ thể là bài toán **phân loại phản hồi khách hàng (Feedback)** của người dùng tiếng Việt thành các nhóm: **Tích cực, Tiêu cực, Trung lập**. Bài này áp dụng PEFT (LoRA) để fine-tune một mô hình nhỏ và dễ triển khai thực tế trong doanh nghiệp hoặc startup.

"Phân loại phản hồi khách hàng tiếng Việt bằng PEFT + LoRA"

Mục tiêu:

- Hiểu cách áp dụng PEFT với LoRA để fine-tune mô hình pre-trained cho phân loại văn bản.
- Làm việc với dữ liệu phản hồi khách hàng (giả lập từ thực tế).
- Tối ưu hoá chi phí và thời gian train bằng cách chỉ fine-tune một phần nhỏ mô hình.

Bối cảnh thực tế

Một cửa hàng TMĐT muốn tự động **phân loại phản hồi khách hàng** sau mỗi đơn hàng để:

- Gửi báo cáo chất lượng dịch vụ theo tuần/tháng.
- Phân nhóm khiếu nại để ưu tiên xử lý.
- Tự động gắn thẻ sentiment vào feedback để dễ lọc.

Cài đặt thư viện

```
pip install transformers datasets accelerate peft bitsandbytes
```

Dataset giả lập

```
from datasets import Dataset

data = {
    "text": [
        "Giao hàng nhanh, sản phẩm đẹp", # tích cực
```

```

    "Tôi rất thất vọng vì sản phẩm bị hỏng", # tiêu cực
    "Bình thường, không có gì đặc biệt", # trung lập
    "Đóng gói kỹ, tôi sẽ quay lại mua", # tích cực
    "Hàng giao trễ, đóng gói cầu thả", # tiêu cực
    "Ổn, chưa biết có dùng lâu được không" # trung lập
],
"label": [2, 0, 1, 2, 0, 1] # 0: tiêu cực, 1: trung lập, 2: tích cực
}

dataset = Dataset.from_dict(data).train_test_split(test_size=0.3)

```

Tiền xử lý

```

from underthesea import word_tokenize

def tokenize_vi(example):
    example["text"] = word_tokenize(example["text"], format="text")
    return example

dataset = dataset.map(tokenize_vi)

```

Dùng PhoBERT + LoRA để fine-tune

```

from transformers import AutoTokenizer, AutoModelForSequenceClassification
from peft import get_peft_model, LoraConfig, TaskType

model_id = "vinai/phobert-base"
tokenizer = AutoTokenizer.from_pretrained(model_id, use_fast=False)
base_model = AutoModelForSequenceClassification.from_pretrained(model_id, num_labels=3)

lora_config = LoraConfig(
    r=8,
    lora_alpha=32,
    target_modules=["encoder.layer.11.output.dense"],
    lora_dropout=0.1,
    bias="none",
    task_type=TaskType.SEQ_CLS
)

peft_model = get_peft_model(base_model, lora_config)

```

```
peft_model.print_trainable_parameters()
```

Huấn luyện mô hình

```
def encode(example):
    return tokenizer(example["text"], truncation=True, padding="max_length", max_length=128)

encoded = dataset.map(encode)

from transformers import TrainingArguments, Trainer
from sklearn.metrics import classification_report
import numpy as np

def compute_metrics(eval_pred):
    logits, labels = eval_pred
    preds = np.argmax(logits, axis=1)
    return {"accuracy": (preds == labels).mean()}

training_args = TrainingArguments(
    output_dir="./sentiment_model",
    per_device_train_batch_size=4,
    num_train_epochs=3,
    evaluation_strategy="epoch",
    logging_dir="./logs"
)

trainer = Trainer(
    model=peft_model,
    args=training_args,
    train_dataset=encoded["train"],
    eval_dataset=encoded["test"],
    tokenizer=tokenizer,
    compute_metrics=compute_metrics
)

trainer.train()
```

Dự đoán phản hồi mới (ví dụ thực tế)

```
import torch

label_map = {0: "Tiêu cực", 1: "Trung lập", 2: "Tích cực"}

def predict(text):
    text = word_tokenize(text, format="text")
    inputs = tokenizer(text, return_tensors="pt", truncation=True, padding="max_length", max_length=128)
    with torch.no_grad():
        logits = peft_model(**inputs).logits
    pred = torch.argmax(logits, axis=1).item()
    return label_map[pred]

# Ví dụ thực tế
feedback = "Mình không hài lòng với chất lượng vải"
print("Phân loại:", predict(feedback))
```

Ứng dụng thực tế

- Tích hợp vào backend của trang web TMĐT, nhận phản hồi → phân loại tự động.
- Lưu label sentiment vào cơ sở dữ liệu để thống kê cảm xúc khách hàng theo thời gian.
- Gửi cảnh báo nội bộ nếu số phản hồi tiêu cực tăng vọt.

Phiên bản #1

Được tạo 8 tháng 5 2025 04:21:55 bởi Đỗ Ngọc Tú

Được cập nhật 8 tháng 5 2025 04:28:24 bởi Đỗ Ngọc Tú