

這段程式碼是用於情感分析的模型訓練，主要包括以下幾個部分：

1. ****數據集下載與預處理****：

- 下載並解壓 IMDB 資料集。
- 定義分詞函數來處理文本數據，移除 HTML 標籤和處理表情符號。
- 建立自定義的 `IMDbDataset` 類來加載和處理數據，並構建詞彙表。

2. ****數據處理****：

- `text_pipeline` 函數將文本轉換為數字索引。
- `collate_batch` 函數將批次數據進行填充和轉換，以便可以餵給模型進行訓練。

3. ****模型定義****：

- 定義了一個簡單的 LSTM 模型，包含嵌入層、LSTM 層、全連接層和激活函數。
- 使用 `nn.Embedding` 進行詞嵌入，`nn.LSTM` 進行序列建模，最後通過全連接層進行分類。

4. ****訓練與評估****：

- 定義了訓練和評估函數，訓練模型並計算訓練和驗證集的準確度。
- 訓練過程中保存最佳模型，並在測試集上進行最終評估。

整體來說，這段程式碼展示了如何使用 PyTorch 來構建和訓練一個處理 IMDB 影評數據的情感分析模型，包括數據預處理、模型定義、訓練、評估等步驟。