

這段程式碼的目的是介紹如何在循環神經網絡（RNN）中加入注意力機制，並探討自注意力機制及其在變壓器（Transformer）架構中的應用。以下是程式碼的主要步驟和意義：

1. **自注意力機制基礎**：

- **句子嵌入**：將單詞映射到實數向量空間中，每個單詞被表示為一個 16 維的向量。
- **計算注意力權重**：首先計算每對詞嵌入之間的點積，然後通過 softmax 函數將這些點積轉換為注意力權重，確保每行的權重和為 1。
- **計算上下文向量**：利用注意力權重對所有詞的嵌入進行加權平均，生成上下文向量。這可以通過矩陣乘法來高效計算。

2. **自注意力機制的參數化**：

- **縮放點積注意力**：使用查詢（query）、鍵（key）和值（value）矩陣來計算注意力分數，這些矩陣通過學習得到。將點積結果除以向量維度的平方根進行縮放，再用 softmax 計算注意力權重，最後加權平均得到上下文向量。

3. **多頭自注意力**：

- **多頭注意力**：將注意力機制擴展到多個注意力頭，每個頭學習不同的注意力模式。多頭注意力允許模型同時關注輸入的不同部分，並從中提取多樣化的信息。
- **線性層映射**：將多頭注意力的輸出通過線性變換整合，最終生成一個上下文向量。

這段程式碼的重點是展示如何在模型中實現和應用自注意力機制，這是變壓器架構的核心組成部分。自注意力機制允許模型在處理序列數據時有效地捕捉長期依賴關係，從而提升模型的性能。