Bài 1: Mô tả kiến trúc mô hình Sequence to Sequence (dựa trên Transformer)

1. Kiến trúc tổng quát

- Thành phần của kiến trúc mô hình Seq2Seq bao gồm 2 thành phần Encoder và Decoder - Mục tiêu của mô hình là xây dựng một trình tóm tắt văn bản, với đầu vào là một chuỗi dài các từ (trong nội dung văn bản) và đầu ra là một bản tóm tắt ngắn (cũng là một chuỗi).

Mã hóa (Encoder):

- Nhận đầu vào là một chuỗi từ (có thể là câu, văn bản) và biến đổi nó thành một loạt các vector biểu diễn.

- Sử dụng nhiều lớp mã hóa Transformer để xử lý thông tin từ đầu vào. Giải mã (Decoder):

- Nhận các vector biểu diễn từ mã hóa và dùng chúng để dự đoán chuỗi đích. - Sử dụng nhiều lớp giải mã Transformer để tạo ra chuỗi đích.

2. Chi tiết tầng input mã hóa và xử lý token như thế nào

- Tầng embedding: Có, thường là tầng embedding ở phía mã hóa. Tầng này chuyển đổi các token thành các vector số có chiều thấp hơn một cách dày đặc, giúp mô hình hiểu được mối quan hệ giữa các từ trong không gian vector.

- Mã hóa token:

- Embedding Layer: Các token đầu vào được chuyển đổi thành các vector embedding dày đặc.

- Tầng mã hóa (Encoder Layer): Các vector embedding sau đó được đưa qua nhiều lớp mã hóa trong kiến trúc Transformer. Mỗi lớp mã hóa chứa các cơ chế tự chú ý và chú ý đa đầu để xử lý thông tin từ các token.

3. Output được sinh ra như thế nào? (tức là các token được sinh ra thế nào)

- Tầng Giải Mã (Decoder Layers): Gồm nhiều tầng attention và feedforward neural network xếp chồng lên nhau tương tự như tầng mã hóa. Tuy nhiên, ở đây, mô hình cũng sử dụng thông tin từ tầng mã hóa để dự đoán chuỗi đích.

- Tầng Output (Output Layer): Dự đoán xác suất của các token trong từ điển đích. Một phương pháp thông thường là sử dụng một tầng softmax để chuyển đổi các đầu ra từ các đặc trưng trước đó thành phân phối xác suất trên tất cả các từ có thể trong từ điển. Sau đó, từ có xác suất cao nhất được chọn làm đầu ra tiếp theo.

Bài 2:

1. Mô tả và ý nghĩa của pre trained models Pre-trained models là các mô hình máy học đã được đào tạo trước trên một lượng lớn dữ liệu và nhiều tác vụ khác nhau. Điều này cho phép chúng học được các biểu diễn tổng quát của ngôn ngữ hoặc thậm chí là kiến thức về thế giới từ dữ liệu lớn mà chúng được huấn luyện. Các pre-trained models thường là một phần quan trọng của tiến trình truyền thống trong học sâu và có thể được sử dụng như một điểm khởi đầu cho việc giải quyết các nhiệm vụ cụ thể hoặc để trích xuất các đặc trưng từ dữ liệu

2. Ý nghĩa của fine-tune Fine-tuning là quá trình tiếp tục huấn luyện một pre-trained model trên một tập dữ liệu cụ thể hoặc một nhiệm vụ cụ thể khác với một số lượng nhỏ các tham số của mô hình được cập nhật. Mục tiêu của việc fine-tuning là điều chỉnh các trọng số của mô hình đã được đào tạo trước để nó có thể hoạt động tốt trên nhiệm vụ mới mà không cần huấn luyện lại từ đầu. Việc này giúp tiết kiệm thời gian và tài nguyên đào tạo, đặc biệt là khi bạn có một lượng lớn dữ liệu và thời gian huấn luyện hạn chế.

3. Áp dụng fine-tune cho 1 pretrained models cho 1 bài toán nào đó ưu tiên cho bài toán Sequence to Sequence. Áp dụng fine-tuning cho một pre-trained model trong bài toán Sequence to Sequence có thể làm cho mô hình đã có kiến thức về ngôn ngữ chung và khả năng suy luận trở nên có hiệu suất cao trong việc dự đoán chuỗi đầu ra từ chuỗi đầu vào. Bằng cách fine-tuning, mô hình sẽ được điều chỉnh để tối ưu hóa cho nhiệm vụ cụ thể của Sequence to Sequence, như dịch máy hoặc tóm tắt văn bản, bằng cách cập nhật các trọng số để phản ánh dữ liệu huấn luyện mới và mục tiêu dự đoán. Điều này giúp mô hình hiệu quả hơn trong việc học các quy luật và mối quan hệ trong các chuỗi dữ liệu và cải thiện kết quả dự đoán.

Tầng Attention trong Seq2Seq: Trong mô hình Seq2Seq, cơ chế attention giúp decoder tập trung vào các phần quan trọng của input sequence khi dự đoán output sequence. Bằng cách này, mô hình có thể xử lý các chuỗi có độ dài khác nhau một cách hiệu quả.

Một số kiểu mô hình Seq2Seq khác: Ngoài Transformer, có các kiểu mô hình Seq2Seq khác như LSTM-based Seq2Seq, GRU-based Seq2Seq, hay CNN-based Seq2Seq. So sánh về hiệu suất và ưu nhược điểm của các kiểu mô hình này có thể giúp độc giả hiểu rõ hơn về lựa chọn mô hình cho bài toán cụ thể.

Các kỹ thuật Regularization và Optimization: Trong quá trình fine-tuning, việc áp dụng các kỹ thuật regularization như dropout hay weight decay cũng như tối ưu hóa thông qua các thuật toán như Adam hay SGD cũng rất quan trọng để tránh overfitting và tối ưu hóa hiệu suất của mô hình.

Thực hiện fine-tuning và đánh giá hiệu suất: Mô tả cụ thể quá trình thực hiện fine-tuning, bao gồm việc chia dữ liệu thành tập huấn luyện, tập kiểm tra và tập validation, cũng như các phương pháp đánh giá hiệu suất như BLEU score (đối với bài toán dịch máy) hoặc ROUGE score (đối với bài toán tóm tắt văn bản), sẽ giúp độc giả hiểu rõ quy trình thực hiện và kết quả đạt được.