Tutorial

How Encoder Decoder Work Stacked LSTM

Course:

Natural Language Processing

By:

Dr. Zohair Ahmed

Great! Let's walk through **how the sentence "How are you?" is passed through the model during training**, using:

- **2 LSTM layers in the Encoder**

- **3 LSTM layers in the Decoder**

Step-by-Step Training: "How are you?" → "آپ کیسے ہیں؟"

◆ **1. Preprocessing**

Input Sentence:

- "How are you?" → ["How", "are", "you", "<eos>"]

Target Sentence:

- "آپ کیسے ہیں؟" → [<start> آپ کیسے ہیں]

---

2. Encoder Forward Pass (2 LSTM layers)

At each time step **t = 1 to 4**:

t = 1:

- Word: "How"

- Embedding → LSTM Layer 1 → hidden state h_1^1

- h_1^1 → LSTM Layer 2 → hidden state h_2^1

t = 2:

- Word: "are"

- Embedding → LSTM Layer 1 → h_1^2

- h_1^2 → LSTM Layer 2 → h_2^2

t = 3:

- Word: "you"

- Embedding → LSTM Layer 1 → h_1^3

- h_1^3 → LSTM Layer 2 → h_2^3

t = 4:

- Word: "<eos>"

- Embedding → LSTM Layer 1 → $h1^4$

- h_1^4 → LSTM Layer 2 → h_2^4

<center>**Final encoder hidden and cell states from Layer 2 are passed to the decoder.**</center>

DR. ZOHAIR AHMED
▶ www.youtube.com/@zohairai
🌐 www.begindiscovery.com

PAGE **2** OF **4**

3. Decoder Forward Pass (3 LSTM layers)

At each time step **t = 1 to 5**:

t = 1:

- Input: <start>

- Embedding → LSTM Layer 1 (initialized with encoder state) → h_1^1

- h_1^1 → LSTM Layer 2 → h_2^1

- h_2^1 → LSTM Layer 3 → h_3^1

- Output: Predict "آپ"

t = 2:

- Input: "آپ"

- Embedding → LSTM Layer 1 → h_1^2

- h_1^2 → LSTM Layer 2 → h_2^2

- h_2^2 → LSTM Layer 3 → h_3^2

- Output: Predict "کیسے"

t = 3:

- Input: "کیسے"

- Embedding → LSTM Layer 1 → h_1^3

- h_1^3 → LSTM Layer 2 → h_2^3

- h_2^3 → LSTM Layer 3 → h_3^3

- Output: Predict "ہیں؟"

t = 4:

- Input: "ہیں؟"

- Embedding → LSTM Layer 1 → h_1^4

- h_1^4 → LSTM Layer 2 → h_2^4

- h_2^4 → LSTM Layer 3 → h_3^4

- Output: Predict <eos>

◆ **4. Loss Calculation**

DR. ZOHAIR AHMED
▶ www.youtube.com/@zohairai
🌐 www.begindiscovery.com

- Compare each predicted word to the actual target word.

- Use **cross-entropy loss** at each time step.

- Accumulate total loss.

◆ **5. Backpropagation**

- Loss is backpropagated:

    - Through decoder layers (3 LSTMs)

    - Into encoder layers (2 LSTMs)

- All weights are updated.

DR. ZOHAIR AHMED   PAGE **4** OF **4**
▶ www.youtube.com/@zohairai
🌐 www.begindiscovery.com