921 U2620 HW4

TOTAL POINTS

4 / 4

QUESTION 1

1 Make a brief introduction about a variant of Transformer. 2 / 2

✓ **- 0 pts** *Correct*

- 0.5 pts slightly wrong

- 1 pts Not detailed enough.

- 1 pts incorrect

- 2 pts Wrong

- 2 pts Directly copy from the internet is not allowed.

- 2 pts Not a variant of Transformer.

QUESTION 2

2 Briefly explain why adding convolutional layers to Transformer can boost performance. 2 / 2

✓ **- 0 pts** *Correct*

- 0.5 pts Slightly incorrect.

- 0.5 pts Not detailed enough

- 1 pts This is not the main reason.

- 1 pts Should be more detailed.

- 1 pts There're some mistakes

- 2 pts Wrong

1. Make a brief introduction about a variant of Transformer.

Sparse Transformer 是將 vanilla transformer 稍作修改，原本 vanilla transformer 是將 attention score 做 softmax，而 sparse transformer 則是先將 attention score 做 sparsification 後再做 softmax.

但我發現好像有 2 篇論文都叫 sparse transformer.

我介紹的是: Explicit Sparse Transformer: Concentrated Attention Through Explicit Selection

另一篇是 OpenAI: Generating Long Sequences with Sparse Transformer (這個好像是利用 matrix factorization 的技術達到 sparse 的)

2. Briefly explain why adding convolutional layers to Transformer can boost performance.

從論文的 introduction 有介紹到說 transformer 和 convolution 在語音辨識都有不錯的成果，但是 transformer 比較擅長 long-range context 的特徵捕捉，而 convolution 比較擅長捕捉 receptive field 內的相關性(偏向 short-range)，所以 conformer 就是同時結合 2 個特性，同時利用 2 種優點。

1 Make a brief introduction about a variant of Transformer. 2 / 2

✓ **- 0 pts** *Correct*

- 0.5 pts slightly wrong

- 1 pts Not detailed enough.

- 1 pts incorrect

- 2 pts Wrong

- 2 pts Directly copy from the internet is not allowed.

- 2 pts Not a variant of Transformer.

1. Make a brief introduction about a variant of Transformer.

Sparse Transformer 是將 vanilla transformer 稍作修改，原本 vanilla transformer 是將 attention score 做 softmax，而 sparse transformer 則是先將 attention score 做 sparsification 後再做 softmax.

但我發現好像有 2 篇論文都叫 sparse transformer.

我介紹的是: Explicit Sparse Transformer: Concentrated Attention Through Explicit Selection

另一篇是 OpenAI: Generating Long Sequences with Sparse Transformer (這個好像是利用 matrix factorization 的技術達到 sparse 的)

2. Briefly explain why adding convolutional layers to Transformer can boost performance.

從論文的 introduction 有介紹到說 transformer 和 convolution 在語音辨識都有不錯的成果，但是 transformer 比較擅長 long-range context 的特徵捕捉，而 convolution 比較擅長捕捉 receptive field 內的相關性(偏向 short-range)，所以 conformer 就是同時結合 2 個特性，同時利用 2 種優點。

2 Briefly explain why adding convolutional layers to Transformer can boost performance. **2 / 2**

✓ **- 0 pts** *Correct*

- 0.5 pts Slightly incorrect.

- 0.5 pts Not detailed enough

- 1 pts This is not the main reason.

- 1 pts Should be more detailed.

- 1 pts There're some mistakes

- 2 pts Wrong