

Machine Learning 2023 Spring Homework 5 Report

生醫電資所 R11945071 莊聰賢

1. Visualize the similarity between different pairs of positional embedding and briefly explain the result. Additionally, attach the code that you used for visualization.

positional embedding 就是把位置嵌入向量 Visualize，例如繪製成 heatmap、散點圖等，以便清楚地觀察不同位置 embed 之間的 similarity、模式或其他特點。

下圖呈現出不同 pairs of positional embedding 之間的 similarity，而 neighboring words 之間的關係會比較緊密，因此 decoder 所得到的 similarity 也會比較高。

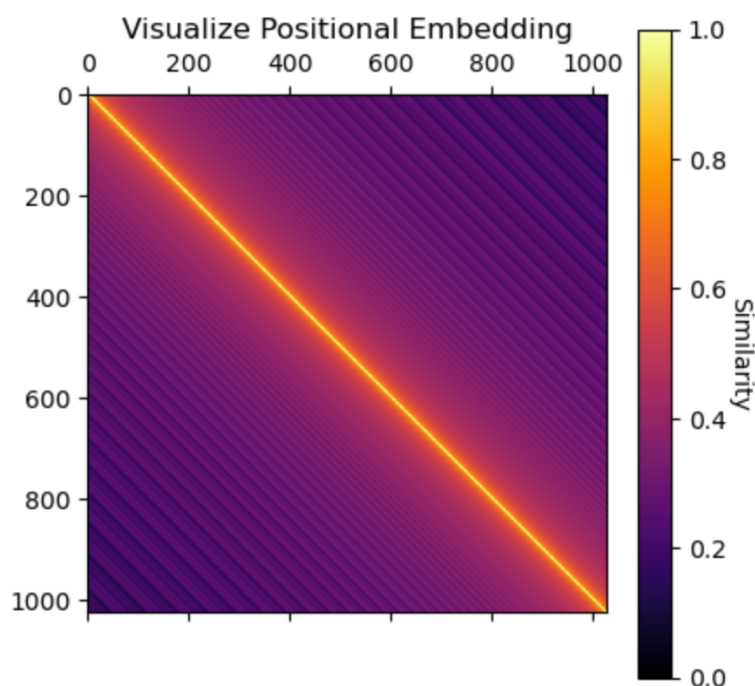


Fig1. Similarity between different pairs of positional embedding.

下圖為該次 visualize 使用的程式碼，主要參考 pytorch 寫出，figure size 定在(10,10)，matshow 參數的部分選用 cmap="inferno"，並加上 colorbar 協助理解顏色與 similarity 之間的關聯性。

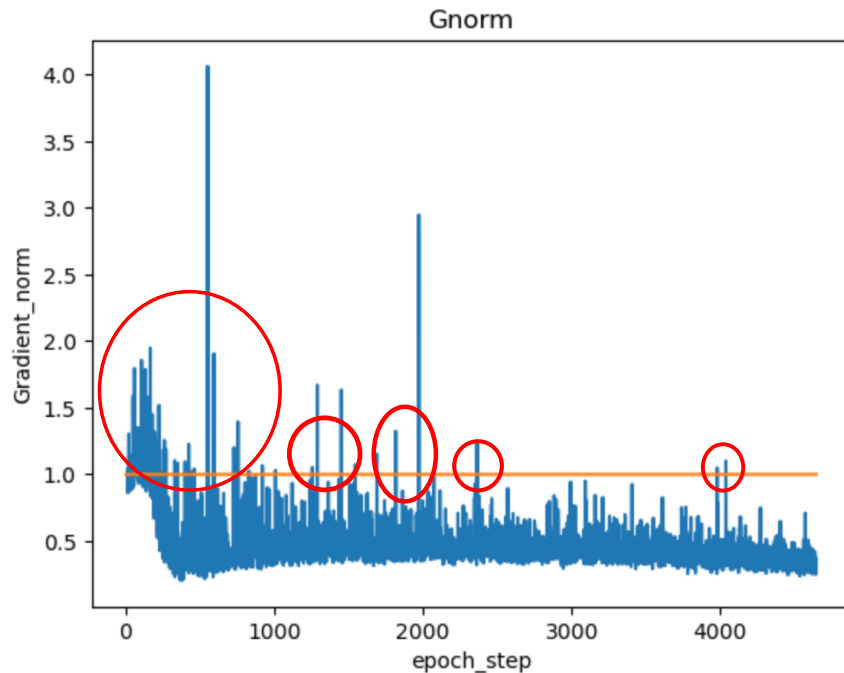
```
from torch.nn.functional import cosine_similarity as cos
pos_emb = model.decoder.embed_positions.weights.cpu().detach()
print("Size : ", pos_emb.size())
similarity = cos(pos_emb.unsqueeze(1), pos_emb, dim=2)
plt.figure(figsize=(10,10))
plt.matshow(similarity, cmap='inferno')
cbar = plt.colorbar()
cbar.set_label('Similarity', rotation=270, labelpad=10)
plt.title("Visualize Positional Embedding")
plt.show()
```

Fig2. code that I used for visualization.

2. Clip gradient norm and visualize the changes of gradient norm in different steps.
Circle two places with gradient explosion.

下圖為使用本次通過 strong 之 model(without back translation)所得到的 gradient norm 結果，通過 strong 的 epochs 為 30，而此圖為跑 5epochs 所得到的結果，可以由此圖得知多半在最先開始時發生所謂的 gradient explosions，後期較少發生。

Fig2. Gradient norm (5epochs)



此次模型參數:

```
arch_args = Namespace(
    encoder_embed_dim=256,
    encoder_ffn_embed_dim=1024,
    encoder_layers=6,
    decoder_embed_dim=256,
    decoder_ffn_embed_dim=1024,
    decoder_layers=6,
    share_decoder_input_output_embed=True,
    dropout=0.3,
)

# HINT: these patches on parameters for Transformer
def add_transformer_args(args):
    args.encoder_attention_heads=4
    args.encoder_normalize_before=True

    args.decoder_attention_heads=4
    args.decoder_normalize_before=True

    args.activation_fn="relu"
    args.max_source_positions=1024
    args.max_target_positions=1024

    # patches on default parameters for Transformer (those not set above)
    from fairseq.models.transformer import base_architecture
    base_architecture(arch_args)

add_transformer_args(arch_args)
```