

神经网络 Final Paper

191240060 吴雨欣 匡亚明学院

本次实验要求是通过链长为 6、8、10、12 的高分子链的能量和位置分布数据进行训练，对链长为 20 的高分子链的能量和位置分布进行预测。

1 方法介绍

整个训练代码分为五个部分：网络构建、训练代码、加载 FU 数据，训练 FU，加载 Structure 数据，训练 Structure。均在代码中有注释。

采用 transformer 模型进行训练，即将序列 seq 和 z 坐标经过 embedding 后，传入 transformer 模块，最后将 transformer 模块的输出经过平均后与经过 embedding 的 z 相乘，结果传入全连接输出最后的预测。具体网络结构如下：

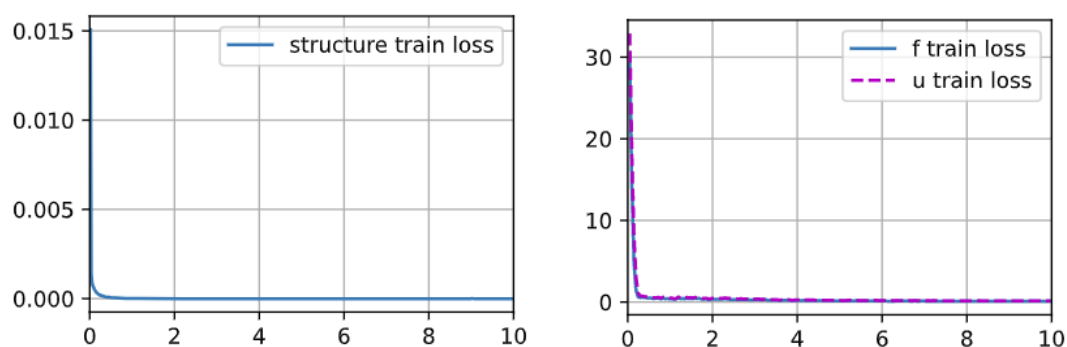
```
# construct model
class model(nn.Module):
    def __init__(self, zr_dimension, d_model=16):
        super(model, self).__init__()
        self.embedding = nn.Embedding(3, d_model)
        self.z_embedding = nn.Embedding(zr_dimension, d_model)
        self.encoder_layer = nn.TransformerEncoderLayer(d_model=d_model, nhead=4)
        self.transformer_encoder = nn.TransformerEncoder(self.encoder_layer, num_layers=2)
        self.fc = nn.Sequential(
            nn.Linear(d_model, 32),
            nn.BatchNorm1d(32),
            nn.ReLU(),
            nn.Linear(32, 16),
            nn.BatchNorm1d(16),
            nn.ReLU(),
            nn.Linear(16, 1)
        )

    def forward(self, seq, z):
        embedded_seq = self.embedding(seq)
        embedded_z = self.z_embedding(z).squeeze(1)
        h = self.transformer_encoder(embedded_seq.permute(1, 0, 2)).permute(1, 0, 2)
        h = torch.mean(h, dim=1)
        y = self.fc(torch.mul(h, embedded_z))
        return y
```

模型中 zr_dimension 可以为 48 或 480，48 对应了预测 FU，480 对应了预测 Structure。d_model 为 embedding 大小，默认为 16。预测 FU 需要传入两个参数，seq 为序列，zr 为 z 坐标；预测 Structure 也需要传入两个参数，seq 为序列，zr 为 r 坐标。

加载 FU 数据后，得到 train_loader_fu，包含四个部分：seq, z, f, u，分别表示序列（长度统一为 20，不足用 2 补齐），z 坐标，F 能量，U 能量。加载 Structure 数据后，得到 train_loader_structure，包含三个部分：seq, r, structure，分别表示序列，r 坐标，structure 的平均位置，这里对原 structure.txt 的每一行取平均作为平均位置。

训练 10 轮，优化器为 Adam，学习率为 1e-3，batch size 设置为 128，损失函数设置为 MSEloss，其他超参数详见代码。训练过程如下所示，可以看到，整个模型的收敛速度是很快的。



2 结果分析

模型训练完成后，选取 5 个有代表性的序列，分别是：

11111100000000111111

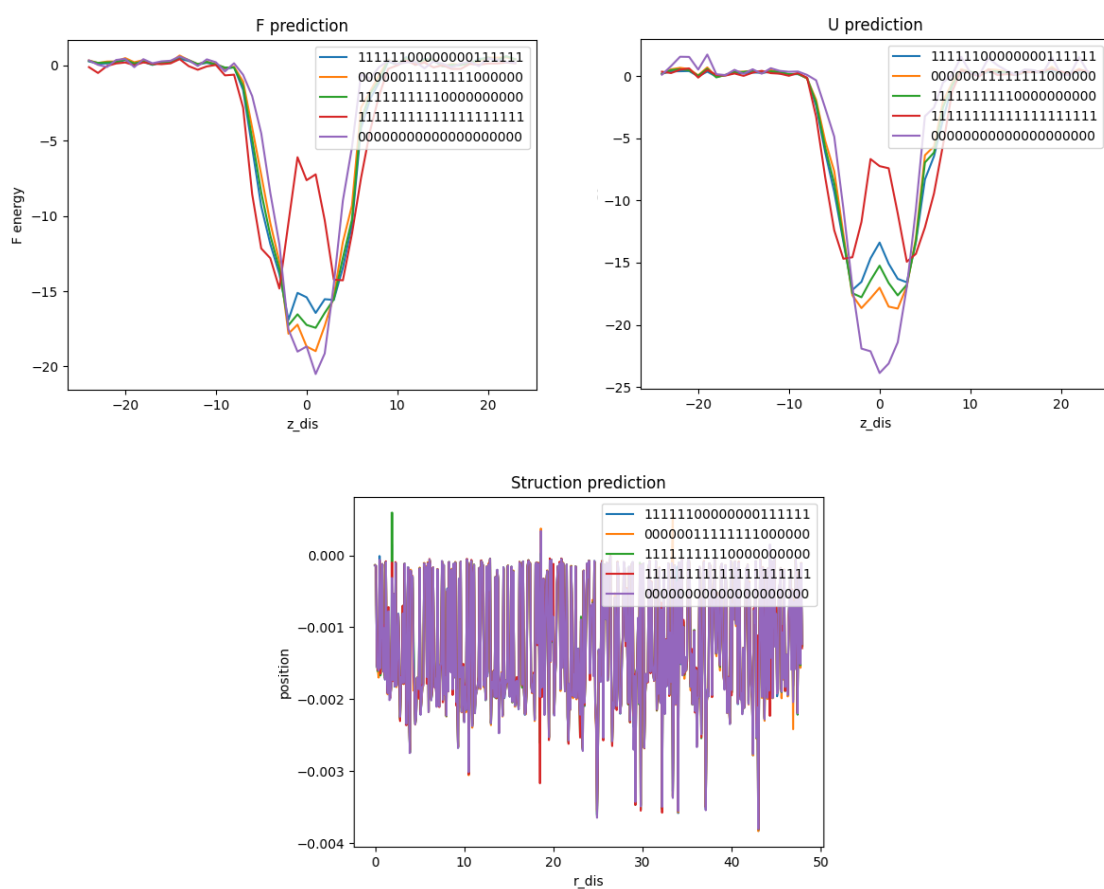
00000011111111000000

11111111110000000000

11111111111111111111

00000000000000000000

对其 F，U，结构进行预测，结果如下：



可以看到，F 和 U 的预测结果与作业中的示例基本一致，而结构的预测则存在较大波动，可能存在一定问题。

3 总结

本次实验，通过 transformer 模型对高分子链的能量和位置分布数据进行训练，取得了相对不错的训练结果，同时对一些序列的能量和平均位置进行了预测。通过这次实验，我不仅对高分子链的若干属性有了更多的了解，对 transformer 和序列数据的神经网络训练方式也有了更深刻的理解。

最后，感谢张子谦同学在网络架构上给我提出的建议，相比于一开始我尝试的 rnn 结构，transformer 确实有明显的性能提升。