**Orientações Gerais sobre a estrutura do código**

**Gabriel Bicalho Ferreira 18.1.4024**

O código foi dividido em algumas partes

**Pasta CompressPyTorch:**  Esta pasta contém o arquivo compression.py, que basicamente pega uma rede neural devidamente treinada, realiza a compressão seguindo alguns parâmetros definidos pelas variáveis de projeto e realiza a inferência, devolvendo os valores dos objetivos, acurácia e taxa de não nulos. São considerados dois modelos treinados, VGG16 (modelo.pt) e Resnet50 (Modelo2.pt).

Outro arquivo desta pasta, é o problem\_compress.py, que utilizando o pymoo define o problema de compressão com 7 variáveis e 2 objetivos. Esse problema definido, chama um método do arquivo compression.py para obter os valores dos dois objetivos.

**Arquivo Run\_benchmarks.py:** Arquivo principal do trabalho, que chama o procedimento de otimização baseada em modelos substitutos. Como são vários testes, foram considerados alguns argumentos que precisam ser passados. O primeiro argumento, é a rede a ser comprimida (‘vgg16’ ou ‘resnet50’). O segundo parâmetro é um número que vai ditar as métricas a serem utilizadas, sendo as possibilidades:

1 -> MSE para objetivos combinados com as 4 possibilidades para restrições;

2 -> MAPE para objetivos combinados com as 4 possibilidades para restrições;

3 -> SPEARMAN para objetivos combinados com as 4 possibilidades para restrições;

4 -> RAND para objetivos combinados com as 4 possibilidades para restrições;

O terceiro argumento, é apenas um número para orientar na quantidade de testes, foram realizados 5 testes para cada combinação, que foram salvos em um arquivo pickle na pasta resultados. Cada arquivo de resultados contém 4 testes, sendo a escolha de uma métrica para os objetivos com todas as combinações de métricas para restrições.

Por exemplo: arquivo de resultados **vgg162T2,** significa que esse arquivo contém resultados da seleção de métricas conforme descrito na tabela abaixo. Além disso, T2 significa que foi o experimento 2 para essas 4 configurações. Isso levando em consideração a rede neural VGG16.

|  |  |
| --- | --- |
| **Objetivos** | **Restrições** |
| MAPE | MSE |
| MAPE | MAPE |
| MAPE | SPEARMAN |
| MAPE | RAND |

Logo, o arquivo Run\_benchmarks.py gera todos os resultados do trabalho com otimização em modelos substitutos.

Arquivo **cta.py:** Arquivo responsável por otimizar sem a utilização de modelos substitutos, otimizando diretamente no problema original, definido em problem\_compress.py. A finalidade de realizar esses testes, é para comparar com os resultados obtidos com a otimização em modelos substitutos. Os testes foram salvos na Pasta Original.

O projeto necessita da versão 0.5.0 do pymoo e da versão 0.7.0 do PyTorch.