CS-502 Project proposal:
Extending Few-shot benchmark on biomedical datasets with Relation Network algorithm

Marija Zelic
Section of Life Sciences Engineering

Elena Mrdja
Section of Life Sciences Engineering

November 2023

1 Introduction

Machine learning has experienced immense growth in recent years. A large part of its success lies in the available data - a sufficient amount of it will likely yield more accurate results. More often than not, it is challenging and unrealistic to acquire an adequate quantity of labeled data. This is especially the case with biomedical datasets since its labeling requires the effort and agreement of many medical professionals. This is where the few-shot learning approach comes in handy, allowing us to make algorithms that will be able to generalize well given only a few labeled examples per class.

2 Motivation

Even though the importance of few-shot learning algorithms for biomedical problems has been recognized, there is a modicum of related publications and they are usually medical imaging-oriented [1]. Therefore, we want to establish a comprehensive few-shot learning benchmark that will incorporate various biomedical datasets and evaluate different algorithms. The existing benchmark is based on two datasets: *Tabula Muris*, which proposes a cell-type annotation task across tissues, and *SwissProt*, which proposes a gene function prediction task from the sequence information. The benchmark currently incorporates five different algorithms: Standard NN, baseline-finetune, MAML, ProtoNet, and MatchingNet. We would like to contribute to the expansion of the present benchmark by implementing an additional few-shot algorithm, particularly the Relation Network [2].

3 Project Details

Relation Network (ResNet) proposes a few-shot learning algorithm that performs few-shot recognition by learning to compare *query* images against few-shot labeled *sample* images. First, an *embedding module* generates a representation of the query and training images. Then these embeddings are compared by *relation module* that determines if they are from matching categories or not. Defining an episode-based strategy, the embedding and relation module are meta-learned to support few-shot learning [2]. Having this in mind, the methodology we are going to adopt for incorporating the ResNet in the existing benchmark is:

- First, try to accomplish the results from the paper to make sure everything is working as expected.
- Adapt the ResNet algorithm to specific requirements of the datasets already included in the benchmark (original ResNet architecture is intended for the image classification task, therefore uses convolutional blocks) by using a fully connected backbone. Ensure compatibility and seamless integration in the current benchmark.
- Evaluate and compare the performance of the expanded benchmark against the original benchmark using various metrics, such as accuracy. Asses whether ResNet keeps outperforming other few-shot algorithms, as it is proposed in the original paper.
- Provide insights into the effectiveness of ResNet architecture in biomedical non-image classification tasks and few-shot learning algorithms in general.

4 Conclusion

The anticipated final deliverables of this project are a fully integrated ResNet algorithm in the existing benchmark and a report that provides insights into the implementation and evaluation of the extended benchmark on the existing biomedical datasets. The outcomes are expected to contribute to the improvement in the assessment of few-shot learning algorithms and their application in biomedical field.

References

[1] Y. Ge, Y. Guo, S. Das, M. A. Al-Garadi, and A. Sarker, "Few-shot learning for medical text: A review of advances, trends, and opportunities," *Journal of Biomedical Informatics*, vol. 144, p. 104458, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S153204642300179X

[2] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. S. Torr, and T. M. Hospedales, "Learning to compare: Relation network for few-shot learning," 2018.