Master Project/ Intern Project Proposal: Pre-training Deep Learning Architectures for Online Time-Series Classification

Viktor Losing

1 Background

In the RT-Phume project, we focus on the task of online time-series classification and regression using different modalities such as IMU sensors, foot-sole pressure sensors and gaze data. As we target real-time performance, we plan to process the data with a high frequency of at least 30 Hz. We will apply standard Deep Neural Network (DNN) models such as Recurrent Neural Networks (RNN), Long-short Term Memories (LSTM) and Gated Recrrent Units (GRU) [1]. However, there have been various extensions that significantly improve the performance of recurrent DL architectures. In particular, pre-trainining the network with unsupervised data has been been successfully applied in various domains [2, 3, 4]. The relation of this proposal to the RT-Phume targets is given by means of applying state-of-the-art approaches on the recorded data. Specifically, pre-training is one of the most promising techniques to get the benefits of Deep Learning on a comparably small labeled dataset (as it is in our case).

2 Research /Technical Target

We are interested in applying techniques of pre-training and subsequent fine-tuning on time-series data. Specifically, we want to investigate their value in relation to IMU-based time-series data. Pre-training has been sucessfully applied to improve model generalization and is usually justified based on the two following perspectives [5]. First, it can be seen as a robust way of network parameter initizalization, generating starting conditions and subsequent gradient trajectories that are intrinsically different from those generated by random initialization, leading at least to a quicker convergence [6, 7]. Another perspective is the one of shared task-knowledge (transfer learning), where training the network on one task facilitates the learning of another, as long both tasks depend on a similar data representation in the latent space.
On the one hand, pre-training has been done with unsupervised data, where self-prediction is usually used as proxy learning task, e.g. Natural Language Processing (NLP) with BERT / GPT [3, 4]. On the other, pre-training is performed using supervised data of a related task, e.g. in object recognition pre-training on Imagenet [8] is very common. We already have a collection of publicly available classification datasets of labeled Xsens IMU data with similar motion classes

[10, 11] and the work will be mainly based on those[1]. Therefore, both options are possible and of interest. It is already clarified with Taizo Yoshikawa that the student can work on the data recorded at HGRX as long as he/she does not keep them after the project. A corresponding agreement with the student will be made.

Recently, pre-training in domains apart from NLP is losing in relevance due to the availability of larger supervised datasets and the application of modern regularization methods such as dropout, weight-sharing and batch-normalization. For instance, it has been recently shown that competetive results could be achieved without any form of pre-training on the COCO dataset [12]. Nonetheless, it was noted that pre-training enabled a quicker model convergence.
In the context of IMU-based data and motion classification, pre-training has rarely been used and available datasets are rather small, encouraging the application of pre-training. Therefore, the focus of this master thesis is the exploration and evaluation of pre-training methods in the domain of IMU-based motion classification / prediction.

3 Estimated Timeline

1. month: The student conducts a literature research and gets familiar with the available data.

2. month: Programming of the required evaluation pipelines.

3.-5. month: Application and evaluation of state-of-the-art methods using at least two different XSens IMU datasets. If applicable exploration of ideas to adapt methods for further performance improvement, i.e. targeting specifically IMU-data.

6. month: Writing the thesis. The advisors will encourage to write the related work beforehand to reduce the pressure in the last month. Potentially preparing a conference submission with the help of advisors (paper, poster or workshop).

4 Live Demonstration

The resulting model could be easily transferred into the live demonstrator of RT-PHUME to showcase the real-time classification / prediction.

References

[1] Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and Pierre-Alain Muller. Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*, 33(4):917–963, 2019.

[2] Alaa Sagheer and Mostafa Kotb. Unsupervised pre-training of a deep lstm-based stacked autoencoder for multivariate time series forecasting problems. *Scientific Reports*, 9(1):1–16, 2019.

[1] We will shortly record futher data mainly consisting of walking motions which poentially could be used as well

[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019.

[4] Tom B Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*, 2020.

[5] Yoshua Bengio, Pascal Lamblin, Dan Popovici, Hugo Larochelle, et al. Greedy layer-wise training of deep networks. *Advances in neural information processing systems*, 19:153, 2007.

[6] Dumitru Erhan, Pierre-Antoine Manzagol, Yoshua Bengio, Samy Bengio, and Pascal Vincent. The difficulty of training deep architectures and the effect of unsupervised pre-training. In David van Dyk and Max Welling, editors, *Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics*, volume 5 of *Proceedings of Machine Learning Research*, pages 153–160, Hilton Clearwater Beach Resort, Clearwater Beach, Florida USA, 16–18 Apr 2009. PMLR.

[7] Dumitru Erhan, Aaron Courville, Yoshua Bengio, and Pascal Vincent. Why does unsupervised pre-training help deep learning? In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 201–208. JMLR Workshop and Conference Proceedings, 2010.

[8] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[9] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.

[10] Pauline Maurice, Adrien Malaisé, Clélie Amiot, Nicolas Paris, Guy-Junior Richard, Olivier Rochel, and Serena Ivaldi. Human movement and ergonomics: An industry-oriented dataset for collaborative robotics. *The International Journal of Robotics Research*, 38(14):1529–1537, 2019.

[11] Viktor Losing, Barbara Hammer, and Heiko Wersing. Personalized online learning of whole body motions using multiple inertial measurement units. In *IEEE International Conference on Robotics and Automation*, 2019.

[12] Kaiming He, Ross Girshick, and Piotr Dollár. Rethinking imagenet pre-training. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4918–4927, 2019.