

# Nemesyst: A Hybrid Parallelism Deep Learning Framework

George Onoufriou

**Abstract**—Nemesyst is a deep learning framework that is built above and agnostic of any single node deep learning framework such as tensorflow, keras, pytorch. Our Nemesyst implementation uses databases along with model sequentialisation in python to allow models to be fed unique and transformed data at the point of need, along with facilitating near real-time application, and making those models available for further training or use at any node which currently has access to the database simultaneously. Further we show how Nemesyst implementation can achieve compelling performance in several different scenarios and real world problems, while still being malleable for on demand adjustments.

**Index Terms**—Deep Learning, Framework, Recurrent Neural Networks, Long Short Term Memory, Generative Adversarial Nets, MongoDB, Distributed Computing, Parallel Computing

## 1 INTRODUCTION

DEEP LEARNING is widely known as the sub-field of machine learning that broadly seeks to learn data representations usually using some form or variant of artificial neural networks, which are intuited by their biological counterparts. Deep learning has two distinct tasks:

- Training; where the deep neural network (DNN) learns the data representation through iterative methods, such as stochastic gradient descent. This is a very time complex operation, taking hours, days, weeks depending on the task, implementation and computational power available. This usually occurs at most on one node or a few nodes if we consider a single computer with several CPUs/ GPUs to be a multi-node system, which here we shall not.
- Inference; where the DNN model now with its learned weights can be used in one form or another to make predictions. This is usually significantly faster than training as it does not require as many operations, but can still suffer from a similar bottleneck [61]

Both training and inference can be improved by use of 'better' hardware, which here shall mean faster and/or higher quantities of processing, and also distributed/parallel hardware. The issue with current deep learning approaches, besides being a time consuming process overall, is that from data handling to evaluation and deployment there is a limit to how much hardware a single computer can support, which is where distributed training and inference is required. Here parallel shall mean any single system that according to flyns taxonomy is using either:

- single instruction, multiple data (SIMD)
- multiple instruction, single data (MISD)
- multiple instruction, multiple data (MIMD)

This is separate from a distributed system, as a distributed system is one where each computing node is a different networked computer, as opposed to a single computer with multiple CPUs/GPUs for computation in parallelisation.

Currently the majority of deep learning frameworks are primarily focused and have matured in parallelisation, but they lack or are still in their infancy with regards to distribution, albeit there are other efforts currently emerging to solve this problem with varying success, such as TensorFlow-distribute, NVIDIA GPU cloud, or IBMs distributed deep learning. [61]

Lastly the ability to share data between distributed nodes using these deep learning frameworks is also in its infancy despite the prevalence of databases providing very mature and advanced functionality such as MongoDBs aggregate pipeline. The unification of deep learning interfaces to databases will also aid in reproducibility as it will standardise the message protocol interface (MPI) and describe what data in what manner is used for training, and inference, in MongoDBs case, by using aggregate pipeline json files. Clearly the integration of deep learning into database models is critical for any (near) real-time application such as Netflix, Amazon, Google, as 'rolling' a bespoke data management system or MPI is likely too costly, and would result in inferior performance and capabilities.

We seek to contribute a new hybrid parallelisation deep learning framework that facilitates practical application of deep learning on a large scale, while also standardising deep learning such that it can be more easily reproducible and consistent. We briefly also show examples of Nemesystin use for national grids demand side response (DSR) using data from both a large chain retailer, and our mock store facilities, which seek to simulate the normal operation of an expected retailer store, and its refrigeration units.

## REFERENCES

- [1] Fabio De Sousa Ribeiro, Francesco Caliva, Mark Swainson, Kjartan Gudmundsson, Georgios Leontidis, and Stefanos Kollias. Deep bayesian uncertainty estimation for adaptation and self-annotation of food packaging images. *arXiv preprint arXiv:1812.01681*, 2018.
- [2] Ramon Granell, Colin J Axon, David CH Wallom, and Russell L Layberry. Power-use profile analysis of non-domestic consumers for electricity tariff switching. *Energy Efficiency*, 9(3):825–841, 2016.
- [3] Fei Teng, Marko Aunedi, Danny Pudjianto, and Goran Strbac. Benefits of demand-side response in providing frequency response service in the future gb power system. *Frontiers in Energy Research*, 3:36, 2015.
- [4] Philipp Grünewald and Jacopo Torriti. Demand response from the non-domestic sector: Early uk experiences and future opportunities. *Energy Policy*, 61:423–429, 2013.
- [5] Andrey Postnikov, Argyrios Zolotas, Chris Bingham, Ibrahim M Saleh, Corneliu Arsene, Simon Pearson, and Ronald Bickerton. Modelling of thermostatically controlled loads to analyse the potential of delivering ffr dsr with a large network of compressor packs. In *2017 European Modelling Symposium (EMS)*, pages 163–167. IEEE, 2017.
- [6] Peter Bradley, Alexia Coke, and Matthew Leach. Financial incentive approaches for reducing peak electricity demand, experience from pilot trials with a uk energy provider. *Energy Policy*, 98:108–120, 2016.
- [7] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- [8] Qingchen Zhang, Laurence T Yang, Zhikui Chen, and Peng Li. A tensor-train deep computation model for industry informatics big data feature learning. *IEEE Transactions on Industrial Informatics*, 14(7):3197–3204, 2018.
- [9] John Ringland, Martha Bohm, and So-Ra Baek. Characterization of food cultivation along roadside transects with google street view imagery and deep learning. *Computers and Electronics in Agriculture*, 158:36–50, 2019.
- [10] Jayme Garcia Arnal Barbedo. Plant disease identification from individual lesions and spots using deep learning. *Biosystems Engineering*, 180:96–107, 2019.
- [11] Andreas Kamilaris and Francesc X Prenafeta-Boldu. Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, 147:70–90, 2018.
- [12] Liangzhi Li, Kaoru Ota, and Mianxiong Dong. Deep learning for smart industry: efficient manufacture inspection system with fog computing. *IEEE Transactions on Industrial Informatics*, 14(10):4665–4673, 2018.
- [13] Qingchen Zhang, Laurence T Yang, Zheng Yan, Zhikui Chen, and Peng Li. An efficient deep learning model to predict cloud workload for industry informatics. *IEEE Transactions on Industrial Informatics*, 14(7):3170–3178, 2018.
- [14] Qingchen Zhang, Laurence T Yang, Zhikui Chen, and Peng Li. A survey on deep learning for big data. *Information Fusion*, 42:146–157, 2018.
- [15] Hao Ye, Geoffrey Ye Li, and Biing-Hwang Juang. Power of deep learning for channel estimation and signal detection in ofdm systems. *IEEE Wireless Communications Letters*, 7(1):114–117, 2018.
- [16] Tom Young, Devamanyu Hazarika, Soujanya Poria, and Erik Cambria. Recent trends in deep learning based natural language processing. *IEEE Computational Intelligence Magazine*, 13(3):55–75, 2018.
- [17] Fabio De Sousa Ribeiro, Francesco Caliva, Dionysios Chionis, Abdelhamid Dokhane, Antonios Mylonakis, Christophe Demaziere, Georgios Leontidis, and Stefanos Kollias. Towards a deep unified framework for nuclear reactor perturbation analysis. In *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 120–127. IEEE, 2018.
- [18] Fabio De Sousa Ribeiro, Liyun Gong, Francesco Caliva, Mark Swainson, Kjartan Gudmundsson, Miao Yu, Georgios Leontidis, Xujiang Ye, and Stefanos Kollias. An end-to-end deep neural architecture for optical character verification and recognition in retail food packaging. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 2376–2380. IEEE, 2018.
- [19] Siqi Nie, Meng Zheng, and Qiang Ji. The deep regression bayesian network and its applications: Probabilistic deep learning for computer vision. *IEEE Signal Processing Magazine*, 35(1):101–111, 2018.
- [20] Dimitrios Kollias, Miao Yu, Athanasios Tagaris, Georgios Leontidis, Andreas Stafylopatis, and Stefanos Kollias. Adaptation and contextualization of deep neural network models. In *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1–8. IEEE, 2017.
- [21] Andre Esteva, Alexandre Robicquet, Bharath Ramsundar, Volodymyr Kuleshov, Mark DePristo, Katherine Chou, Claire Cui, Greg Corrado, Sebastian Thrun, and Jeff Dean. A guide to deep learning in healthcare. *Nature medicine*, 25(1):24, 2019.
- [22] Sean D Holcomb, William K Porter, Shaun V Ault, Guifen Mao, and Jin Wang. Overview on deepmind and its alphago zero ai. In *Proceedings of the 2018 International Conference on Big Data and Education*, pages 67–71. ACM, 2018.
- [23] Elif Dede, Madhusudhan Govindaraju, Daniel Gunter, Richard Shane Canon, and Lavanya Ramakrishnan. Performance evaluation of a mongodb and hadoop platform for scientific data analysis. In *Proceedings of the 4th ACM workshop on Scientific cloud computing*, pages 13–20. ACM, 2013.
- [24] François Chollet et al. Keras: The python deep learning library. *Astrophysics Source Code Library*, 2018.
- [25] Ibrahim Saleh, Andrey Postnikov, Chris Bingham, Ronald Bickerton, Argyrios Zolotas, Simon Pearson, et al. Aggregated power profile of a large network of refrigeration compressors following ffr dsr events. *International Conference on Energy Engineering*, 2018.
- [26] Sara Sabour, Nicholas Frosst, and Geoffrey E Hinton. Dynamic routing between capsules. In *Advances in Neural Information Processing Systems*, pages 3856–3866, 2017.
- [27] Evgeny Frolov and Ivan Oseledets. Tensor methods and recommender systems. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 7(3), 2017.
- [28] Steffen Rendle. Factorization machines. In *Data Mining (ICDM)*, 2010 IEEE 10th International Conference on, pages 995–1000. IEEE, 2010.
- [29] Gediminas Adomavicius and Alexander Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering*, 17(6):734–749, 2005.
- [30] Le Hoang Son. Dealing with the new user cold-start problem in recommender systems: A comparative review. *Information Systems*, 58:87–104, 2016.
- [31] Carlos A Gomez-Urbe and Neil Hunt. The netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems (TMIS)*, 6(4):13, 2016.
- [32] Brent Smith and Greg Linden. Two decades of recommender systems at amazon. com. *IEEE Internet Computing*, 21(3):12–18, 2017.
- [33] John Riedl, Joe Konstan, Loren Terveen, Lana Yarosh, and Haiyi Zhu. Movielens, 2017.
- [34] Xavier Amatriain and Josep M Pujol. Data mining methods for recommender systems. In *Recommender Systems Handbook*, pages 227–262. Springer, 2015.
- [35] Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8), 2009.
- [36] 2017.
- [37] G. Linden, B. Smith, and J. York. Amazon.com recommendations: item-to-item collaborative filtering. *IEEE Internet Computing*, 7(1):76–80, 2003.
- [38] Shuo Chang, F. Maxwell Harper, and Loren Terveen. Using groups of items for preference elicitation in recommender systems. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & #38; Social Computing, CSCW '15*, pages 1258–1269, New York, NY, USA, 2015. ACM.
- [39] Qian Zhao, Gediminas Adomavicius, F Maxwell Harper, Martijn C Willemsen, and Joseph A Konstan. Toward better interactions in recommender systems: Cycling and serpentine approaches for top-n item lists. In *CSCW*, pages 1444–1453, 2017.
- [40] Qian Zhao. Re-thinking top-n recommendation lists: should recommenders always show the best content?, 2017.
- [41] Najdt Mustafa, Ashraf Osman Ibrahim, Ali Ahmed, and Afnizan-faizal Abdullah. Collaborative filtering: Techniques and applications. In *Communication, Control, Computing and Electronics Engineering (ICCCCEE)*, 2017 International Conference on, pages 1–6. IEEE, 2017.
- [42] Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B Kantor. *Recommender systems handbook*. Springer, 2015.

- [43] Andriy Mnih and Ruslan R Salakhutdinov. Probabilistic matrix factorization. In *Advances in neural information processing systems*, pages 1257–1264, 2008.
- [44] Jianxun Lian, Fuzheng Zhang, Xing Xie, and Guangzhong Sun. Cccfnet: A content-boosted collaborative filtering neural network for cross domain recommender systems. In *Proceedings of the 26th International Conference on World Wide Web Companion*, pages 817–818. International World Wide Web Conferences Steering Committee, 2017.
- [45] Chao-Yuan Wu, Amr Ahmed, Alex Beutel, Alexander J Smola, and How Jing. Recurrent recommender networks. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pages 495–503. ACM, 2017.
- [46] Oleksii Kuchaiev and Boris Ginsburg. Training deep autoencoders for collaborative filtering. *arXiv preprint arXiv:1708.01715*, 2017.
- [47] Lifang Ren and Wenjian Wang. An svm-based collaborative filtering approach for top-n web services recommendation. *Future Generation Computer Systems*, 78:531–543, 2018.
- [48] Vipul Narayan, RK Mehta, Mahendra Rai, Ashish Gupta, Madhu Singh, Swati Verma, Ankita Patel, and Shristhi Yadav. E-commerce recommendation method based on collaborative filtering technology. *International Journal of Current Engineering and Technology*, 2017.
- [49] Liang Xiang and Qing Yang. Time-dependent models in collaborative filtering based recommender system. In *Web Intelligence and Intelligent Agent Technologies, 2009. WI-IAT'09. IEEE/WIC/ACM International Joint Conferences on*, volume 1, pages 450–457. IEEE, 2009.
- [50] Pavlos Kefalas and Yannis Manolopoulos. A time-aware spatio-textual recommender system. *Expert Systems with Applications*, 78:396–406, 2017.
- [51] Tulasi K Paradarami, Nathaniel D Bastian, and Jennifer L Wightman. A hybrid recommender system using artificial neural networks. *Expert Systems with Applications*, 83:300–313, 2017.
- [52] Xiaopeng Li and James She. Collaborative variational autoencoder for recommender systems. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 305–314. ACM, 2017.
- [53] Mehdi Elahi, Francesco Ricci, and Neil Rubens. A survey of active learning in collaborative filtering recommender systems. *Computer Science Review*, 20:29–50, 2016.
- [54] Paul Covington, Jay Adams, and Emre Sargin. Deep neural networks for youtube recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems*, New York, NY, USA, 2016.
- [55] Shuai Zhang, Lina Yao, and Aixin Sun. Deep learning based recommender system: A survey and new perspectives. *arXiv preprint arXiv:1707.07435*, 2017.
- [56] Shuai Zhang, Lina Yao, Aixin Sun, and Yi Tay. Deep learning based recommender system: A survey and new perspectives. *ACM Computing Surveys (CSUR)*, 52(1):5, 2019.
- [57] Ryan R. Curtin, James R. Cline, Neil P. Slagle, William B. March, P. Ram, Nishant A. Mehta, and Alexander G. Gray. mlpack: A scalable C++ machine learning library. *Journal of Machine Learning Research*, 14:801–805, 2013.
- [58] Zhilin Yang, Zihang Dai, Ruslan Salakhutdinov, and William W Cohen. Breaking the softmax bottleneck: A high-rank rnn language model. *arXiv preprint arXiv:1711.03953*, 2017.
- [59] Lantao Yu, Weinan Zhang, Jun Wang, and Yong Yu. Seqgan: Sequence generative adversarial nets with policy gradient. *CoRR*, abs/1609.05473, 2016.
- [60] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.
- [61] Dhableswar K DK Panda, Hari Subramoni, and Ammar Ahmad Awan. High performance distributed deep learning: A beginners guide. In *Proceedings of the 24th Symposium on Principles and Practice of Parallel Programming*, pages 452–454. ACM, 2019.
- [62] Takuya Akiba, Keisuke Fukuda, and Shuji Suzuki. Chainermn: Scalable distributed deep learning framework. In *Proceedings of Workshop on ML Systems in The Thirty-first Annual Conference on Neural Information Processing Systems (NIPS)*, 2017.
- [63] Patrick L McDermott and Christopher K Wikle. Bayesian recurrent neural network models for forecasting and quantifying uncertainty in spatial-temporal data. *Entropy*, 21(2):184, 2019.
- [64] Kaihao Zhang, Wenhan Luo, Yiran Zhong, Lin Ma, Wei Liu, and Hongdong Li. Adversarial spatio-temporal learning for video deblurring. *IEEE Transactions on Image Processing*, 28(1):291–301, 2019.
- [65] David Mascharka, Philip Tran, Ryan Soklaski, and Arjun Majumdar. Transparency by design: Closing the gap between performance and interpretability in visual reasoning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4942–4950, 2018.
- [66] Wenqi Fan, Yao Ma, Qing Li, Yuan He, Eric Zhao, Jiliang Tang, and Dawei Yin. Graph neural networks for social recommendation. In *Proceedings of the 2019 World Wide Web Conference (WWW)*, 2019.
- [67] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.
- [68] Paraskevas Panagiotidis, Andrew Effraimis, and George A Xydis. An r-based forecasting approach for efficient demand response strategies in autonomous micro-grids. *Energy & Environment*, 30(1):63–80, 2019.
- [69] Fabiano Pallonetto, Mattia De Rosa, Federico Milano, and Donal P Finn. Demand response algorithms for smart-grid ready residential buildings using machine learning models. *Applied Energy*, 239:1265–1282, 2019.
- [70] Eklas Hossain, Imtiaj Khan, Fuad Un-Noor, Sarder Shazali Sikander, and Md Samiul Haque Sunny. Application of big data and machine learning in smart grid, and associated security concerns: A review. *IEEE Access*, 7:13960–13988, 2019.
- [71] Peiyang Guo, Jacqueline CK Lam, and Victor OK Li. Drivers of domestic electricity users price responsiveness: A novel machine learning approach. *Applied energy*, 235:900–913, 2019.
- [72] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*, 2014.
- [73] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. *NIPS*, 2017.
- [74] Frank Seide and Amit Agarwal. Cntk: Microsoft’s open-source deep-learning toolkit. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 2135–2135. ACM, 2016.
- [75] George Onoufriou. Nemesyst lstm. <https://github.com/DreamingRaven/Nemesyst/blob/master/examples/lstm.py>, 2019.
- [76] Hoang Dung Vu, Kok-Soon Chai, Bryan Keating, Nurislam Tursynbek, Boyan Xu, Kaige Yang, Xiaoyan Yang, and Zhenjie Zhang. Data driven chiller plant energy optimization with domain knowledge. *CoRR*, abs/1812.00679, 2018.
- [77] Weicong Kong, Zhao Yang Dong, David J Hill, Fengji Luo, and Yan Xu. Short-term residential load forecasting based on resident behaviour learning. *IEEE Transactions on Power Systems*, 33(1):1087–1088, 2018.
- [78] Golriz Amooee, Behrouz Minaei-Bidgoli, and Malihe Bagheri-Dehnavi. A comparison between data mining prediction algorithms for fault detection(case study: Ahanpishegan co.). *CoRR*, abs/1201.6053, 2012.
- [79] Ivan Sekulic, Matej Gjurkovic, and Jan Snajder. Not just depressed: Bipolar disorder prediction on reddit. *CoRR*, abs/1811.04655, 2018.
- [80] Sanjeev Raja and Ernest Fokoué. Multi-stage fault warning for large electric grids using anomaly detection and machine learning. *CoRR*, abs/1903.06700, 2019.
- [81] H Han, B Gu, T Wang, and ZR Li. Important sensors for chiller fault detection and diagnosis (fdd) from the perspective of feature selection and machine learning. *International Journal of Refrigeration*, 34(2):586–599, 2011.
- [82] Sepp Hochreiter, Yoshua Bengio, Paolo Frasconi, Jürgen Schmid-

huber, et al. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies, 2001.