References

[1] f

[2] f

[3] f

[4] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," 2022

[5] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, A. Rodriguez, A. Joulin, E. Grave, and G. Lample, "Llama: Open and efficient foundation language models," 2023.

[6] OpenAI, "GPT-4 Technical Report," 2023.

[7] L. Bernstein, A. Sludds, R. Hamerly, V. Sze, J. Emer, and D. Englund, "Freely scalable and reconfigurable optical hardware for deep learning," *Scientific Reports*, vol. 11, 02 2021.

[8] R. Taori, I. Gulrajani, T. Zhang, Y. Dubois, X. Li, C. Guestrin, P. Liang, and T. B. Hashimoto, "Alpaca: A strong, replicable instruction-following model," 2023. [Online]. Available: https://crfm.stanford.edu/2023/03/13/alpaca.html

[9] ——, "Stanford alpaca: An instruction-following llama model," https://github.com/tatsu-lab/stanfordalpaca, 2023.

[10] S. Zhang, S. Roller, N. Goyal, M. Artetxe, M. Chen, S. Chen, C. Dewan, M. Diab, X. Li, X. V. Lin, T. Mihaylov, M. Ott, S. Shleifer, K. Shuster, D. Simig, P. S. Koura, A. Sridhar, T. Wang, and L. Zettlemoyer, "Opt: Open pre-trained transformer language models," 2022.

[11] A. Kroshchanka and V. Golovko, "The reduction of fully connected neural network parameters using the pre-training technique," in *2021 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, vol. 2, 2021, pp. 937–941.

[12] H. Mostafa and X. Wang, "Parameter efficient training of deep convolutional neural networks by dynamic sparse reparameterization," 2019.

[13] S. Han, J. Pool, J. Tran, and W. J. Dally, "Learning both weights and connections for efficient neural networks," 2015.

[14] P. Smolensky, "Information processing in dynamical systems: Foundations of harmony theory," *Parallel Distributed Process*, vol. 1, 01 1986.

[15] V. Golovko and V. Krasnoproshin, *Neural network data processing technologies*. BSU, 2017. [Online]. Available: https://elib.bsu.by/handle/123456789/193558

[16] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, no. 521 (7553), pp. 436–444, 2015.

[17] G. E. Hinton, P. Dayan, B. J. Frey, and R. M. Neal, "The "wake-sleep" algorithm for unsupervised neural networks," *Science*, vol. 268, no. 5214, pp. 1158–1161, 1995. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.7761831

[18] G. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural computation* , vol. 18, pp. 1527–54, 08 2006.

[19] Y. Bengio, "Learning deep architectures for AI," *Foundations and Trends in Machine Learning*, no. 2(1), pp. 1–127, 2009.

[20] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations," in *Proceedings of the 26th Annual International Conference on Machine Learning*, ser. ICML '09. New York, NY, USA: Association for Computing Machinery, 2009, p. 609–616. [Online]. Available: https: //doi.org/10.1145/1553374.1553453

[21] V. Golovko, A. Kroshchanka, and E. Mikhno, "Deep Neural Networks: Selected Aspects of Learning and Application," in *Pattern Recognition and Image Analysis*. Cham: Springer International Publishing, 2021, pp. 132–143.

[22] Y. LeCun and C. Cortes, "MNIST handwritten digit database." [Online]. Available: http://yann.lecun.com/exdb/mnist/

[23] A. Krizhevsky, "Learning multiple layers of features from tiny images," pp. 32–33, 2009. [Online]. Available: https: //www.cs.toronto.edu/ kriz/learning-features-2009-TR.pdf

Редуцирование нейросетевых моделей в интеллектуальных компьютерных системах нового поколения

Крощенко А. А.

Статья посвящена разработке метода редуцирования глубоких нейронных сетей в контексте интеграции подобных моделей в ostis-системы. Предлагается альтернативный подход к обучению глубоких нейронных сетей, базирующийся на использовании RBM и CRBM. Предлагается метод для снижения размерности "тяжелых" моделей. Полученные теоретические результаты подтверждаются вычислительными экспериментами, демонстрирующими эффективность предложенного подхода к редуцированию.

Received 18.03.2023