Generative Properties of Universal Bidirectional Activation-Based Learning

Kristína Malinovská(✉) and Igor Farkaš

Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava, Bratislava, Slovakia
{malinovska,farkas}@fmph.uniba.sk
http://cogsci.fmph.uniba.sk/cnc/

Abstract. UBAL is a novel bidirectional neural network model with bio-inspired learning. It enhances contrastive Hebbian learning rule with an internal echo mechanism enabling self-supervised learning. UBAL approaches any problem as a bidirectional heteroassociation, which gives rise to emergent properties, such as generation of patterns while trained for classification. We briefly discuss and illustrate these properties using the MNIST dataset and conclude that with a slight trade-off in accuracy we can achieve feasible image generation without explicitly setting up the objective to do so.

Keywords: Biologically inspired learning · Bidirectional connectivity · Generative networks

1 Towards More Brain-Like Learning

Well-known error-backpropagation (BP) algorithm is traditionally argued to lack biological plausibility [2,7]. Learning in the brain is based on local interactions between presynaptic and postsynaptic neurons. The brain makes a great use of bidirectional flow of information in order to classify and reconstruct patterns [7], but the activation never propagates back via the same synaptic weights. Our Universal Bidirectional Activation-based Learning (UBAL) model [6] is mainly inspired by the canonical recirculation algorithm proposed by Hinton [3] conceiving the autoencoder, and by the Generalized Recirculation which is an adaptation of Contrastive Hebbian Learning (CHL) [7].

UBAL shares the features with other models of this new wave of biologically inspired neural models. Random synaptic weights are proposed to avoid backpropagating via the same neural pathway in the Feedback Alignment model [5]. The Equilibrium propagation [9] as well as the whole family of Target propagation models [8] make use of the target clamping which is essential in contrastive learning.[1]

[1] The target (ground truth) is directly inserted (clamped) into the output layer of a model as a neural activation which can be propagated backwards in the network. Instead of using error derivatives, the weights are adapted based on local differences between forward (estimated) and backward (clamped) activation variables.

I. Farkaš et al. (Eds.): ICANN 2021, LNCS 12893, pp. 80–83, 2021.
https://doi.org/10.1007/978-3-030-86365-4_7

2 UBAL Model

UBAL is a heteroencoder model that maintains separate weight matrices for two different activation propagation directions between the visible layers. In the context of classification the propagation of input activation would be called the prediction in the forward direction and propagation of the clamped targets would be called the prediction in the backward direction. Inspired by [3] we also propagate the network's immediate prediction in the opposite direction which we call the *echo*.

Since the same learning rule applies for any two consecutive bidirectionally connected layers, hidden or visible, we define the model here for just two connected layers $\boldsymbol{p}$ and $\boldsymbol{q}$ and their synaptic weights $\boldsymbol{W}_{pq}$ for the forward direction and $\boldsymbol{M}_{qp}$ for the backward direction (Fig. 1). As listed in Table 1, the activation of propagation is expressed as the product of the presynaptic activation and the bias and the synaptic weight with the activation function f applied. The biases are added in both directions and labeled $\boldsymbol{b}$ or $\boldsymbol{d}$. Their synaptic connections are already assumed in the weight matrices.

The resulting activation variables are combined using the hyperparameters β and γ defined for each direction of activation propagation (F and B) into learning rule terms (Table 2). The learning rule in Eq. 1 and 2 is formed by these intermediate terms with the aim to emphasize its relationship to the contrastive Hebbian learning.

$$\Delta \boldsymbol{W}_{pq} = \lambda\, \boldsymbol{t}_p^{\mathrm{B}}(\boldsymbol{t}_q^{\mathrm{F}} - \boldsymbol{e}_q^{\mathrm{F}}) \tag{1}$$

$$\Delta \boldsymbol{M}_{qp} = \lambda\, \boldsymbol{t}_q^{\mathrm{F}}(\boldsymbol{t}_p^{\mathrm{B}} - \boldsymbol{e}_p^{\mathrm{B}}) \tag{2}$$

Table 1. Activation propagation.

Activation phase	Note	Computation
Forward prediction	$\boldsymbol{q}^{\mathrm{FP}}$	$f(\boldsymbol{W}_{pq}\boldsymbol{p}^{\mathrm{FP}} + \boldsymbol{b}_p)$
Forward echo	$\boldsymbol{p}^{\mathrm{FE}}$	$f(\boldsymbol{M}_{qp}\boldsymbol{q}^{\mathrm{FP}} + \boldsymbol{d}_q)$
Backward prediction	$\boldsymbol{p}^{\mathrm{BP}}$	$f(\boldsymbol{M}_{qp}\boldsymbol{q}^{\mathrm{BP}} + \boldsymbol{d}_q)$
Backward echo	$\boldsymbol{q}^{\mathrm{BE}}$	$f(\boldsymbol{W}_{pq}\boldsymbol{p}^{\mathrm{BP}} + \boldsymbol{b}_p)$

Table 2. Learning rule terms.

Learning rule term	Note	Computation
Forward target	$\boldsymbol{t}_q^{\mathrm{F}}$	$\beta_q^{\mathrm{F}}\boldsymbol{q}^{\mathrm{FP}} + (1-\beta_q^{\mathrm{F}})\boldsymbol{q}^{\mathrm{BP}}$
Forward estimate	$\boldsymbol{e}_q^{\mathrm{F}}$	$\gamma_q^{\mathrm{F}}\boldsymbol{q}^{\mathrm{FP}} + (1-\gamma_q^{\mathrm{F}})\boldsymbol{q}^{\mathrm{BE}}$
Backward target	$\boldsymbol{t}_p^{\mathrm{B}}$	$\beta_p^{\mathrm{B}}\boldsymbol{p}^{\mathrm{BP}} + (1-\beta_p^{\mathrm{B}})\boldsymbol{p}^{\mathrm{FP}}$
Backward estimate	$\boldsymbol{e}_p^{\mathrm{B}}$	$\gamma_p^{\mathrm{B}}\boldsymbol{p}^{\mathrm{BP}} + (1-\gamma_p^{\mathrm{B}})\boldsymbol{p}^{\mathrm{FE}}$

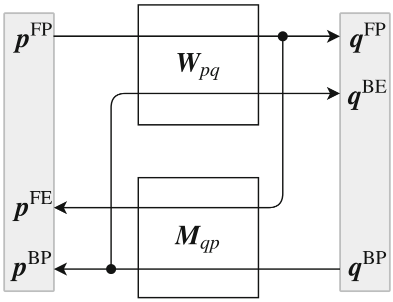

Fig. 1. UBAL connectivity of two connected layers $\boldsymbol{p}$ and $\boldsymbol{q}$.

The hyperparameters β mediate the clamping and trade-off the predicted and the clamped activation contributing to the weight change. The hyperparameters

γ trade-off the prediction and the echo activation variables. The values of βs and γs differ across the tasks that the network has to learn and are under continuous examination. Our general observation so far is that, setting all γs to 0.5 works well for associative tasks (encode-retrieve) and values around 0.0 and 1.0 enable UBAL to master classification.

3 Classification and Generative Properties

With the hyperparameter setups in Table 3 (input–hidden–output) and large-enough hidden layer (1500 neurons), the performance of UBAL in the MNIST [4] benchmark is comparable to the related models. UBAL can reaches up to 96% accuracy on the testing set, without any kind of image augmentation or any supplementary regularization techniques. We use a 3-layer network with standard sigmoidal units (softmax for output layer) and Gaussian weight initialization $\mathcal{N}(0.0, 0.5)$ and learning rate 0.05. MNIST digit targets are encoded as one-hot vectors and images are normalized to $(0, 1)$.

Table 3. Two setups of UBAL hyperparameters for MNIST.

	Setup A	Setup B
β^{F}	0.0–1.0–0.0	1.0–1.0–0.9
γ^{F}	1.0–1.0	1.0–1.0
γ^{B}	1.0–1.0	0.9–1.0
β^{B}	1.0–0.0–1.0	0.0–0.0–0.1

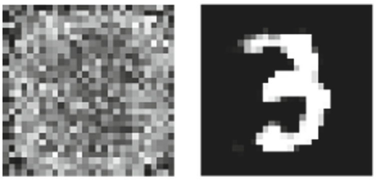

Fig. 2. Example of projected digit 3 with Setup A (left) and Setup B (right).

Hence UBAL is a heteroencoder, apart from classifying the digits, it also naturally makes projections of those digits in its input layer, which could be understood as the network's imagination as shown in Fig. 3. Our preliminary results suggest that these images differ among network initializations and they are different from the computed averages of all images in the dataset. Decreasing the hidden layer β^{F} from 1.0 to a smaller value (0.995–0.999999) yields slight decrease in accuracy, but more variable and graded images with soft edges.

A natural step in exploration of generative properties is to introduce noise into the network. Currently, we are adding small Gaussian noise to the labels. Our experiments show that the generalization ability of UBAL is not much impeded by a very low-variance noise, yet it yields more diverse backward projections in case the noise is added to the targets when gathering the backward projections. The ability of UBAL to generate patterns in the backward direction while trained for the classification task is mostly influenced by the β and γ hyperparameter setup. There are setups that do work well in terms of classification accuracy, but do not allow generation of legible numbers (Fig. 2). We will

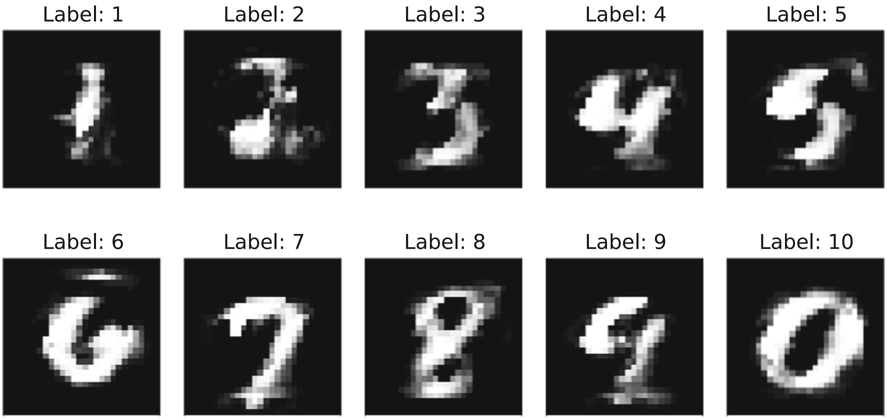

Fig. 3. MNIST digits generated by UBAL.

further explore the properties of the projected images and how they are classified by a UBAL and by other models. This relates to explainable AI, where there is a prospect of using UBAL for generating noise for adversarial examples [1]. In this line we plan to investigate the robustness of UBAL against adversarial attacks.

Acknowledgments. This research was supported by KEGA grant no. 042UK-4/2019.

References

1. Goodfellow, I.J., Shlens, J., Szegedy, C.: Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572 (2014)
2. Grossberg, S.: Competitive learning: from interactive activation to adaptive resonance. Cogn. Sci. **11**(1), 23–63 (1987)
3. Hinton, G.E., McClelland, J.L.: Learning representations by recirculation. In: Neural Information Processing Systems, pp. 358–366 (1988)
4. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998)
5. Lillicrap, T.P., Cownden, D., Tweed, D.B., Akerman, C.J.: Random synaptic feedback weights support error backpropagation for deep learning. Nat. Commun. **7**, 13276 (2016)
6. Malinovská, K., Malinovskỳ, L., Krsek, P., Kraus, S., Farkaš, I.: UBAL: a Universal Bidirectional Activation-based Learning Rule for Neural Networks. In: Proceedings of the 2019 2nd International Conference on Computational Intelligence and Intelligent Systems, pp. 57–62 (2019)
7. O'Reilly, R.C., Munakata, Y., Frank, M., Hazy, T., et al.: Computational Cognitive Neuroscience. PediaPress, Mainz (2012)
8. Ororbia, A.G., Mali, A.: Biologically motivated algorithms for propagating local target representations. arXiv preprint arXiv:1805.11703 (2018)
9. Scellier, B., Bengio, Y.: Equilibrium propagation: bridging the gap between energy-based models and backpropagation. Front. Comput. Neurosci. **11**, 24 (2017)