Using Deep Learning Techniques to Analyze and Classify EEG Recordings

Sebastian Stober

Brain and Mind Institute, Department of Psychology
Western University
London, Ontario, Canada, N6A 5B7
sstober@uwo.ca

Abstract

Deep learning is a recently emerged field within machine learning which is gaining more and more attention as it significantly outperforms other approaches in an increasing number of applications – from speech detection to recognizing objects in images.[1] The term "deep learning" originally referred to artificial neural networks (ANNs) with more than one hidden layer but can be generalized to architectures composed of multiple non-linear transformations. One major advantage of deep learning compared to traditional approaches is that they can work directly on raw data and do not require any tuning or hand-crafting of features. Instead, they are able to learn their own feature representations (see, e.g., [1]).

Up to now, applications of deep learning techniques within neuroscience have been very limited. They have been successfully applied to detect anomalies related to epilepsy in electroencephalography (EEG) recordings [6] and to classify sleep stages from EEG as well as recordings of eye movements and skeletal muscle activity [3]. Very recently, the potential of deep learning techniques for neuroimaging has also been demonstrated for functional and structural magnetic resonance imaging (MRI) data [4].

In our work, we have applied two popular deep learning techniques – stacked denoising autoencoders (SDAs) and convolutional neural networks (CNNs) – to classify and analyze EEG recordings of auditory rhythm perception. As proposed in [5], SDAs are a nested stack of autoencoders – ANNs with an encoder layer that transforms the input into an internal representation and a corresponding decoder layer for the inverse transformation. The dimensionality of the internal representation usually decreases with each layer to create an artificial encoding "bottleneck" or is otherwise regularized to encourage sparsity. Additionally, the inputs are corrupted during training for higher robustness to noise. An example for a compression of 1:10 through 3 layers is given in Figure 1. The compressed internal representation of a SDA can be used as features for further processing steps such as classification. Moreover, a SDA trained on EEG data can be applied for automatic artifact rejection as such inputs will result in a high reconstruction error (Figure 2).

[1] An up-to-date reading list for an overview of the field of deep learning is provided at http://deeplearning.net/reading-list/.

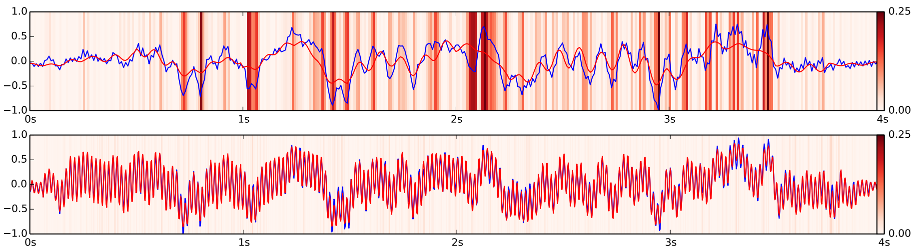

Fig. 1. Input (blue) and reconstruction (red) for 4s of EEG data sampled at 100Hz (top) / 400Hz (bottom). Background color indicates the squared sample error (right scale).

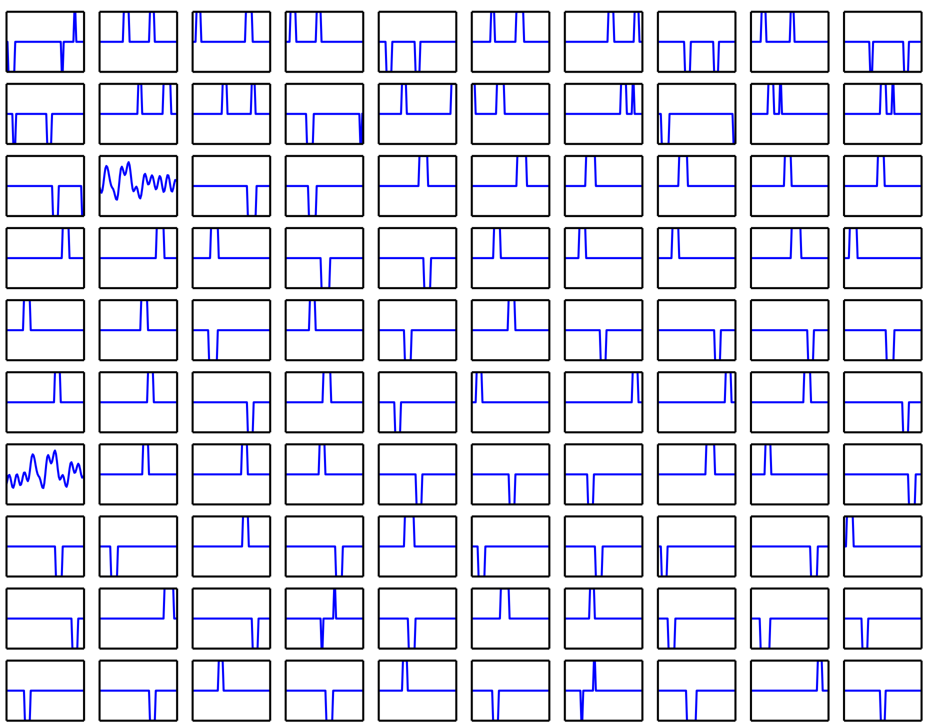

Fig. 2. Top 100 input segments (200ms at 400Hz) with highest reconstruction error and clearly visible artifacts.

CNNs (cf. e.g. [2]), are ANNs with one or more convolutional layers. In such a layer, a linear convolution operation is applied for local segments of the input followed by a non-linear transformation and a pooling operation over neighboring segments. Thus, such a layer can be considered as trainable local feature detector. Forming hierarchies of such detectors allows to recognize more complex, higher-level patterns in the inputs. We have successfully applied such networks to classify perceived rhythms into types (African vs. Western) as well as to identify individual rhythms. By studying the learned feature detectors, we also hope to gain more knowledge about the underlying EEG signals as well as the corresponding cognitive processes.
Acknowledgments: This work was supported by a fellowship within the Postdoc-Program of the German Academic Exchange Service (DAAD).

References

1. Y. Bengio, A. Courville, and P. Vincent. Representation learning: A review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(8):1798–1828, 2013.
2. A. Krizhevsky, I. Sutskever, and G. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, pages 1097–1105, 2012.
3. M. Längkvist, L. Karlsson, and M. Loutfi. Sleep stage classification using unsupervised feature learning. *Adv. Artif. Neu. Sys.*, 2012:5:5–5:5, Jan 2012.
4. S. Plis, D. Hjelm, R.Salakhutdinov, and V. Calhoun. Deep learning for neuroimaging: a validation study. *arXiv preprint arXiv:1312.5847*, 2013.
5. P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res.*, 11:3371–3408, Dec. 2010.
6. D. Wulsin, J. Gupta, R. Mani, J. Blanco, and B. Litt. Modeling electroencephalography waveforms with semi-supervised deep belief nets: fast classification and anomaly measurement. *J Neural Eng*, 8(3):036015, Jun 2011.