Learning a peripersonal space representation using Conditional Restricted Boltzmann Machine

Zdenek Straka and Matej Hoffmann

Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University in Prague,
Karlovo namesti 13, 121 35 Prague 2, Czech Republic
Email: {zdenek.straka, matej.hoffmann}@fel.cvut.cz

Abstract

We present a neural network learning architecture composed of a Restricted Boltzmann Machine (RBM) and a Conditional RBM (CRBM) that performs multisensory integration and prediction, motivated by the problem of learning a representation of defensive peripersonal space. This work follows up on our previous work (Straka and Hoffmann 2017) where we proposed a network composed of a RBM and a feedforward neural network (FFNN). In this work, with a similar 2D simulated scenario, we sought to replace the FFNN with an RBM-like module and opted for the CRBM which is responsible for making a temporal prediction. We demonstrate that the new architecture is capable of learning to map from visual and tactile inputs at a previous time step (without tactile activation) to future activations with the visual stimulus at the "skin" and corresponding tactile activation, including the confidence of the predictions.

1 Introduction

Defensive peripersonal space (PPS) (e.g., Cléry et al. (2015) is a kind of safety margin surrounding our bodies that draws on visuo-tactile interactions: approaching stimuli are registered by vision and processed, producing anticipation or prediction of contact in the tactile modality. The mechanisms of this representation and its development are not understood. This work follows up on our previous work Straka and Hoffmann (2017) where we proposed a neural network composed of a RBM, which learns in an unsupervised manner to represent position and velocity features of a stimulus, and a feedforward neural network (FFNN) trained in a supervised way to predict the position of touch (contact). In this work, with a similar 2D simulated scenario, we sought to replace the FFNN with an RBM-like module and opted for the CRBM which is responsible for making the temporal prediction. We demonstrate that the new architecture is capable of learning to map from visual and tactile inputs at a previous time step (without tactile activation) to future activations with the visual stimulus at the "skin" and corresponding tactile activation, including the confidence of the predictions.

2 Methods

Fig. 1 provides an overview. There is a rectangular 2D input space, completely covered by the receptive fields (RF) of visual unimodal neurons (see left part of panels B/C for illustration of RF center coordinates). The central part of the space is also covered by RFs of tactile unimodal neurons (right part of panels B/C) – the "skin". The architecture consists of two identical copies of an RBM and a CRBM on top. The left RBM serves to represent the visual and tactile inputs pertaining to objects approaching the "skin" up to the moment of contact; the right RBM represents the same inputs at the moment of contact. The CRBM on top eventually serves to predict the contact with skin and its location. The input layers (panels B or C) encode positions of an object perceived by visual (orange neurons) and tactile (green neurons) modalities using probabilistic population coding which encodes also confidence of both percepts (see Makin et al. (2013); Straka and Hoffmann (2017)). The hidden layers of the RBMs, which integrate both visual and tactile inputs, are used as inputs of the CRBM Taylor et al. (2007). Training consists of simulated objects crossing the visual field and eventually contacting the "skin". Their trajectory is sampled and stored in a buffer and then fed in the RBMs—the left RBM with stimuli crossing the visual field before and up to contact (hence $r_{outskin}$) and the right RBM with the final time step when the stimulus reached the tactile field (hence r_{inskin}). Firstly, it is necessary to train the RBM to integrate the visual and tactile inputs. Then the CRBM is trained.

During testing, only $r_{outskin}$ is given and r_{inskin} is inferred using k steps of Gibbs sampling. Then, $v_{outskin}$ and r_{inskin} were obtained averaging over 15 samples. For getting 2D position from the activated neural population r_{inskin}, both visual and tactile subpopulations were combined taking confidence of each subpopulation into account (see Makin et al. (2013)).

3 Results

After training, the network was successfully able to predict stimuli corresponding to the tactile stimulation. As

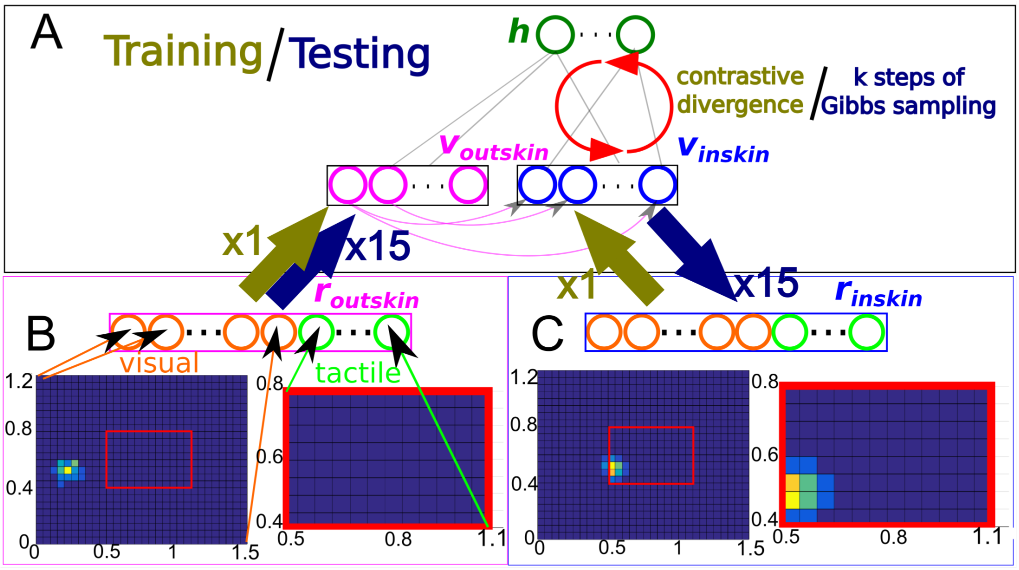

Fig. 1: Scenario, neural network architecture, and schematic illustration of training and testing.

would be expected, mean prediction error (distance between predicted and actual stimulus) was decreasing with decreasing distance of the object from the "skin". From a certain distance from the border of the tactile modality, the prediction error was nearly constant and small. For example, for 200 hidden neurons of the CRBM, the distance was approximately 0.2 and the mean error was about 0.02. The confidence (see Straka and Hoffmann (2017)) of the predictions was increasing with the decreasing distance—negatively correlated with the error.

However, the scenario is still highly simplified and the stimulus velocity not explicitly accounted for—this will be one of our directions for future work.

Acknowledgments

Code related to CRBM was taken from supporting website to Taylor et al. (2007). We also thank Joseph Makin for sharing code based on Makin et al. (2013). This work was supported by the Czech Science Foundation under Project GA17-15697Y. Z.S. was also partly supported by the Czech Technical University in Prague, grant No. SGS18/138/OHK3/2T/13.

References

Cléry, J., Guipponi, O., Wardak, C. a Hamed, S. B. (2015). Neuronal bases of peripersonal and extrapersonal spaces, their plasticity and their dynamics: knowns and unknowns. *Neuropsychologia*, 70:313–326.

Hinton, G. E. (2002). Training products of experts by minimizing contrastive divergence. *Neural Computation*, 14(8):1771–1800.

Ma, W. J., Beck, J. M., Latham, P. E. a Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, 9(11):1432–1438.

Magosso, E., Zavaglia, M., Serino, A., Di Pellegrino, G. a Ursino, M. (2010). Visuotactile representation of peripersonal space: a neural network study. *Neural Computation*, 22(1):190–243.

Makin, J. G., Fellows, M. R. a Sabes, P. N. (2013). Learning multisensory integration and coordinate transformation via density estimation. *PLoS Comput Biol*, 9(4):e1003035.

Straka, Z. a Hoffmann, M. (2017). Learning a peripersonal space representation as a visuo-tactile prediction task. Lintas, A., Rovetta, S., Verschure, P. F. a Villa, A. E. (zost.), V *Artificial Neural Networks and Machine Learning – ICANN 2017: 26th International Conference on Artificial Neural Networks, Alghero, Italy, September 11-14, 2017, Proceedings, Part I*, str. 101–109, Cham. Springer International Publishing.

Taylor, G. W., Hinton, G. E. a Roweis, S. T. (2007). Modeling human motion using binary latent variables. V *Advances in neural information processing systems*, str. 1345–1352.

Welling, M., Rosen-Zvi, M. a Hinton, G. E. (2004). Exponential family harmoniums with an application to information retrieval. V *NIPS*, vol. 4, str. 1481–1488.