Chapter 6

Summary & Outlook

The presented work was based on the observation that predictability must be assumed to play a central role in the data processing performed by every intelligent agent that is interacting with an environment. Accordingly, the central hypothesis has been twofold: Firstly, that the predictable aspects of time-structured data (interactive or non-interactive) can serve as a valuable (and largely overlooked) source of information for unsupervised machine learning. And secondly, that predictability may function as a first principle that—like slowness, sparseness, and others before—can be helpful in understanding neural self-organization.

With respect to unsupervised machine learning, Chapter 3 describes the development of GPFA, a new algorithm for unsupervised dimensionality reduction and feature learning that focuses on outputs that are predictive with respect to their own future. Preliminary experiments have shown its effectiveness on different datasets and compared it to similar learning algorithms that have been developed by others in parallel. Similarly, Chapter 5 describes a newly developed unsupervised learning algorithm that can learn a discrete set of predictive states or partitions for a real-valued time series or a continuous, interactive feature space.

Since the initial empirical results for the GPFA algorithm have been inconclusive in parts, Chapter 4 describes a more extensive empirical comparison of the algorithms that are available for predictable feature learning (including GPFA and SFA) and comes to the conclusion that the predictable aspects found in “real-world” datasets are mostly slow ones\(^1\). In other words, the empirical results suggest that—in practice—SFA actually

\(^1\)While high frequencies have been rare among the static datasets collected for the experiments in Chapter 4, the splitting algorithm from Chapter 5 quickly leads to high frequencies in the data (as discussed in Section 5.4.3). In that case, a GPFA-based splitting approach can be assumed to be useful.
functions as an efficient and robust approach to predictable feature learning (according to three different measures)—even though it was not originally designed for that purpose. This observation is important with respect to the understanding of predictability as a computational principle to explain neural self-organization. It implies that previous results of SFA on hippocampus and visual cortex can already be understood as cases of predictable feature learning and thus being in support of the initial hypothesis.

At the same time, there remains a difference between slow and predictable features since different solvers may learn somewhat different features (although they may not differ much in terms of a given predictability measure). It remains an open question if these differences may become relevant in specific tasks. In any case, one can assume that predictability serves at least as a necessary condition of a useful feature and a natural next step would be to test the learned features in actual reinforcement settings because that is closest to the interactive scenario that has been motivated the whole work from the start. To that end, the proposed algorithms are flexible enough (in particular through their formulation in terms of graph embedding) to lend themselves to many possible extensions and modifications in the future and, in a similar fashion, there is room to further analyze the relationship between the proposed measures of predictability and other measures and frameworks like information bottlenecks.

From a biological perspective, neural information processing is now often thought as a continuous interplay between bottom-up and top-down processes. It seems promising to take this into account as well for future explorations of predictability as a first principle for neural self-organization.