

well because they lie in the receptive field of fewer hidden units. However, if we do implicitly zero-pad the input, then the hidden units at the boundary are driven by fewer input pixels, and may fail to activate when needed.

20.7 Boltzmann Machines for Structured or Sequential Outputs

In the structured output scenario, we wish to train a model that can map from some input \mathbf{x} to some output \mathbf{y} , and the different entries of \mathbf{y} are related to each other and must obey some constraints. For example, in the speech synthesis task, \mathbf{y} is a waveform, and the entire waveform must sound like a coherent utterance.

A natural way to represent the relationships between the entries in \mathbf{y} is to use a probability distribution $p(\mathbf{y} \mid \mathbf{x})$. Boltzmann machines, extended to model conditional distributions, can supply this probabilistic model.

The same tool of conditional modeling with a Boltzmann machine can be used not just for structured output tasks, but also for sequence modeling. In the latter case, rather than mapping an input \mathbf{x} to an output \mathbf{y} , the model must estimate a probability distribution over a sequence of variables, $p(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(\tau)})$. Conditional Boltzmann machines can represent factors of the form $p(\mathbf{x}^{(t)} \mid \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(t-1)})$ in order to accomplish this task.

An important sequence modeling task for the video game and film industry is modeling sequences of joint angles of skeletons used to render 3-D characters. These sequences are often collected using motion capture systems to record the movements of actors. A probabilistic model of a character's movement allows the generation of new, previously unseen, but realistic animations. To solve this sequence modeling task, [Taylor et al. \(2007\)](#) introduced a conditional RBM modeling $p(\mathbf{x}^{(t)} \mid \mathbf{x}^{(t-1)}, \dots, \mathbf{x}^{(t-m)})$ for small m . The model is an RBM over $p(\mathbf{x}^{(t)})$ whose bias parameters are a linear function of the preceding m values of \mathbf{x} . When we condition on different values of $\mathbf{x}^{(t-1)}$ and earlier variables, we get a new RBM over \mathbf{x} . The weights in the RBM over \mathbf{x} never change, but by conditioning on different past values, we can change the probability of different hidden units in the RBM being active. By activating and deactivating different subsets of hidden units, we can make large changes to the probability distribution induced on \mathbf{x} . Other variants of conditional RBM ([Mnih et al., 2011](#)) and other variants of sequence modeling using conditional RBMs are possible ([Taylor and Hinton, 2009](#); [Sutskever et al., 2009](#); [Boulanger-Lewandowski et al., 2012](#)).

Another sequence modeling task is to model the distribution over sequences