



Figure 10.11: Computation of a typical bidirectional recurrent neural network, meant to learn to map input sequences \mathbf{x} to target sequences \mathbf{y} , with loss $L^{(t)}$ at each step t . The \mathbf{h} recurrence propagates information forward in time (towards the right) while the \mathbf{g} recurrence propagates information backward in time (towards the left). Thus at each point t , the output units $\mathbf{o}^{(t)}$ can benefit from a relevant summary of the past in its $\mathbf{h}^{(t)}$ input and from a relevant summary of the future in its $\mathbf{g}^{(t)}$ input.