

Elman's Recurrent Neural Networks

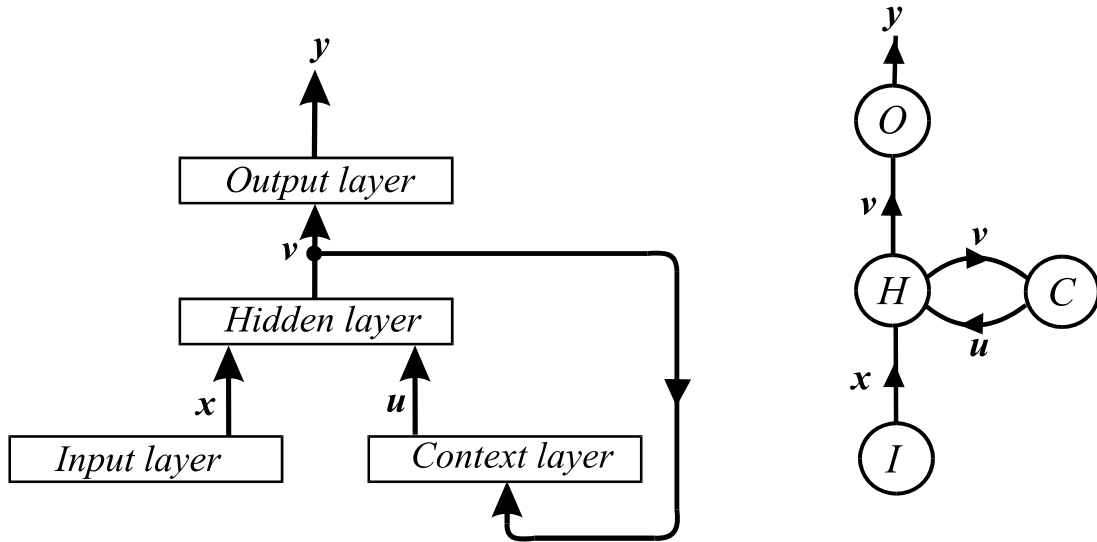
Vladimír Kvasnička

Department of Mathematics
Slovak Technical University
812 37 Bratislava, Slovakia



March 1999

Architecture of Elman's recurrent neural network



$$\left. \begin{aligned}
 \mathbf{x} &= \text{input} \in \{a, b, c, d\}^* \\
 \mathbf{u} &= C(\mathbf{v}) \\
 \mathbf{v} &= H(\mathbf{x}, \mathbf{u}) \\
 \mathbf{y} &= O(\mathbf{v}) \in R^n
 \end{aligned} \right\} \Rightarrow$$

$$\Rightarrow \begin{cases}
 \mathbf{x} = \text{input} \in \{a, b, c, d\}^* \\
 \mathbf{v} = H(\mathbf{x}, C(\mathbf{v})) \quad (\text{a nonlinearity}) \\
 \mathbf{y} = O(\mathbf{v}) \in R^n
 \end{cases}$$

An iterative solution of

$$\mathbf{x} = \text{input} \in \{a, b, c, d\}^*$$

$$\mathbf{u} = C(\mathbf{v})$$

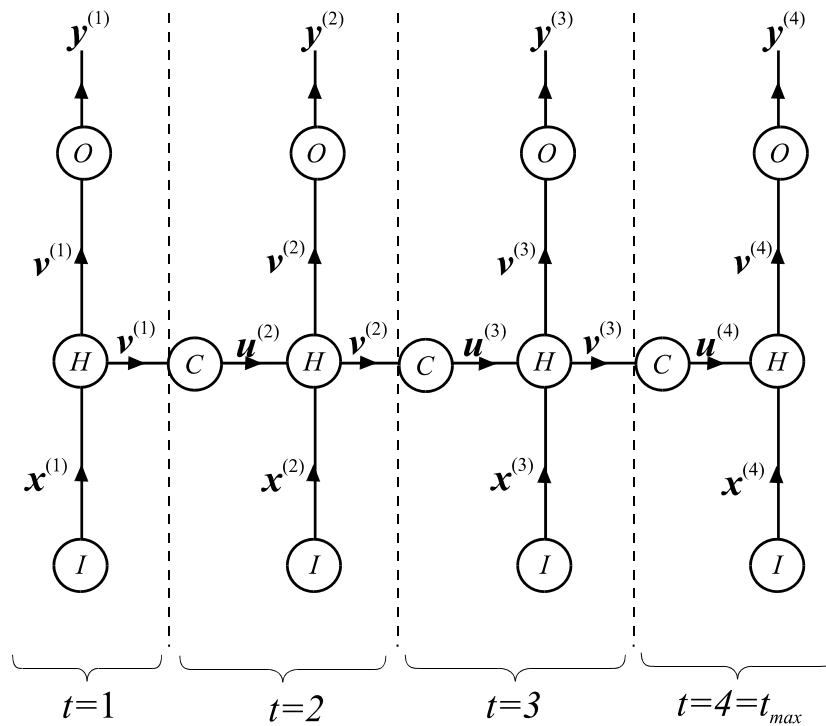
$$\mathbf{v} = H(\mathbf{x}, \mathbf{u})$$

$$\mathbf{y} = O(\mathbf{v}) \in R^n$$

specifies activities of neurons recurrently as follows

$$\left. \begin{array}{l} \mathbf{x}^{(t)} = \text{input} \\ \mathbf{u}^{(t)} = \begin{cases} 0 & (\text{for } t = 1) \\ C(\mathbf{v}^{(t-1)}) & (\text{for } t \geq 2) \end{cases} \\ \mathbf{v}^{(t)} = H(\mathbf{x}^{(t)}, \mathbf{u}^{(t)}) \\ \mathbf{y}^{(t)} = O(\mathbf{v}^{(t)}) \end{array} \right\} \text{for } t = 1, 2, \dots, t_{max}$$

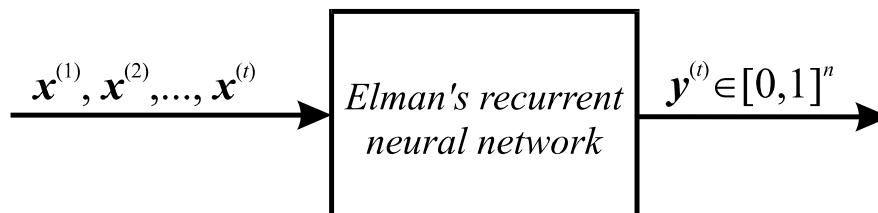
Unfolded recurrent neural network



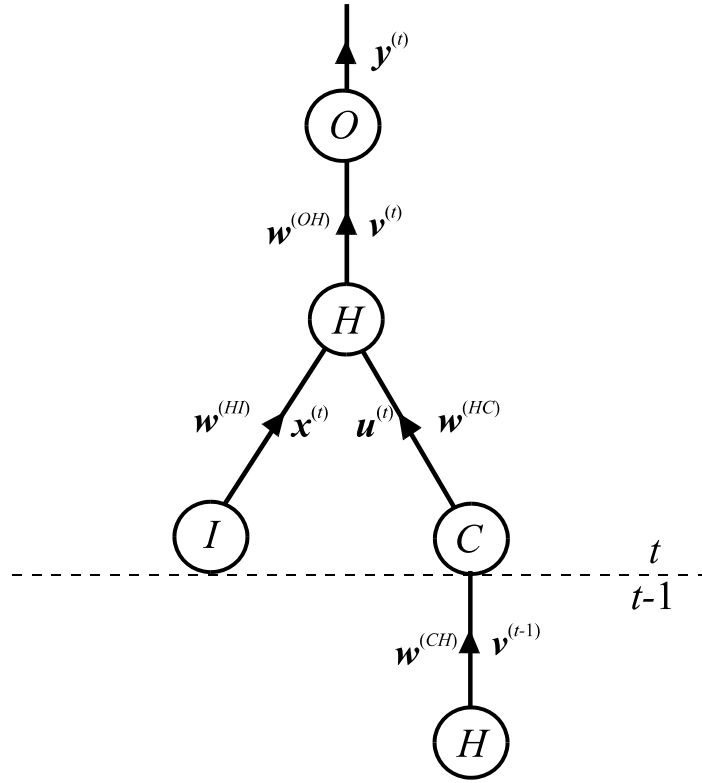
Unfolded Elman's recurrent neural network may be considered as a parametric mapping that maps a sequence of input vectors onto an output vector

$$\mathbf{y}^{(t)} = G(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(t)}; \mathbf{w}, \mathfrak{D})$$

for $t=1, 2, \dots, t_{max}$.



Activities of single neurons



$$x_i^{(t)} = \text{input}$$

$$u_i^{(t)} = \begin{cases} 0 & (t = 1) \\ t \left(\mathfrak{G}_i^{(C)} + \sum_{j=1}^{N_H} w_{ij}^{(CH)} v_j^{(t-1)} \right) & (t \geq 2) \end{cases}$$

$$v_i^{(t)} = t \left(\mathfrak{G}_i^{(H)} + \sum_{j=1}^{N_I} w_{ij}^{(HI)} x_j^{(t)} + \sum_{j=1}^{N_C} w_{ij}^{(HC)} u_j^{(t)} \right)$$

$$y_i^{(t)} = t \left(\mathfrak{G}_i^{(O)} + \sum_{j=1}^{N_H} w_{ij}^{(OH)} v_j^{(t)} \right)$$

Adaptation (learning) of the recurrent neural network

$$A_{train} = \left\{ \left(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(t_{max})} \right) / \left(\mathbf{y}_{req}^{(1)}, \mathbf{y}_{req}^{(2)}, \dots, \mathbf{y}_{req}^{(t_{max})} \right) \right\}$$

$$E = \frac{1}{2} \sum_{t=1}^{t_{max}} \omega_t \left(\mathbf{y}^{(t)} - \mathbf{y}_{req}^{(t)} \right)^2$$

An adaptation is equivalent to a minimization of the objective function E with respect to weight and threshold coefficients

$$\left(w_{opt}, \vartheta_{opt} \right) = \arg \min_{(w, \vartheta)} E(w, \vartheta)$$

This optimization problem is most frequently solved by the so-called **steepest descent** gradient method of

$$w_{k+1} = w_k - \lambda \text{grad } E(w_k)$$

where $\lambda > 0$.

**How to calculate partial derivatives
of the objective function with respect to
threshold and weight coefficients?**

$$\left(\frac{\partial E}{\partial \vartheta_i} \right) = t'(\xi_i) \left(g_i + \sum_k \frac{\partial E}{\partial \vartheta_k} w_{ki} \right)$$

$$t'(\xi_i) = t(\xi_i)[1 - t(\xi_i)]$$

$$g_i = \begin{cases} x_i - x_{i,req} & (i \in O) \\ 0 & (i \notin O) \end{cases}$$

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial \vartheta_i} x_j$$

These general formulae (they form a background of the so-called **back-propagation approach**) are immediately applicable to unfolded recurrent neural networks for calculation of partial derivatives of the objective function E .

Partial derivatives with respect to threshold coefficients

(1) Initialization, $t=t_{max}$

$$\left(\frac{\partial E}{\partial \vartheta_i^{(O)}} \right)^{(t_{max})} = y_i^{(t_{max})} (1 - y_i^{(t_{max})}) \omega_{t_{max}} (y_i^{(t_{max})} - y_{i,req}^{(t_{max})})$$

$$\left(\frac{\partial E}{\partial \vartheta_i^{(H)}} \right)^{(t_{max})} = v_i^{(t_{max})} (1 - v_i^{(t_{max})}) \sum_{j=1}^{N_O} \left(\frac{\partial E}{\partial \vartheta_j^{(O)}} \right)^{(t_{max})} w_{ji}^{(OH)}$$

$$\left(\frac{\partial E}{\partial \vartheta_i^{(C)}} \right)^{(t_{max})} = u_i^{(t_{max})} (1 - u_i^{(t_{max})}) \sum_{j=1}^{N_H} \left(\frac{\partial E}{\partial \vartheta_j^{(H)}} \right)^{(t_{max})} w_{ji}^{(HC)}$$

(2) Iteration, $1 \leq t < t_{\max}$

$$\left(\frac{\partial E}{\partial \vartheta_i^{(O)}} \right)^{(t)} = y_i^{(t)} (1 - y_i^{(t)}) \omega_t (y_i^{(t)} - y_{i,req}^{(t)})$$

$$\left(\frac{\partial E}{\partial \vartheta_i^{(H)}} \right)^{(t)} = v_i^{(t)} (1 - v_i^{(t)}) \left(\begin{array}{c} \sum_{j=1}^{N_O} \left(\frac{\partial E}{\partial \vartheta_j^{(O)}} \right)^{(t)} w_{ji}^{(OH)} + \\ \sum_{j=1}^{N_C} \left(\frac{\partial E}{\partial \vartheta_j^{(C)}} \right)^{(t+1)} w_{ji}^{(CH)} \end{array} \right)$$

$$\left(\frac{\partial E}{\partial \vartheta_i^{(C)}} \right)^{(t)} = u_i^{(t)} (1 - u_i^{(t)}) \sum_{j=1}^{N_H} \left(\frac{\partial E}{\partial \vartheta_j^{(H)}} \right)^{(t)} w_{ji}^{(HC)}$$

Partial derivatives with respect to weight coefficients

(1) Initialization, $t=t_{\max}$

$$\left(\frac{\partial E}{\partial w_{ij}^{(OH)}} \right)^{(t_{\max})} = \left(\frac{\partial E}{\partial \vartheta_i^{(O)}} \right)^{(t_{\max})} v_j^{(t_{\max})}$$

$$\left(\frac{\partial E}{\partial w_{ij}^{(HC)}} \right)^{(t_{\max})} = \left(\frac{\partial E}{\partial \vartheta_i^{(H)}} \right)^{(t_{\max})} u_j^{(t_{\max})}$$

$$\left(\frac{\partial E}{\partial w_{ij}^{(HI)}} \right)^{(t_{\max})} = \left(\frac{\partial E}{\partial \vartheta_i^{(H)}} \right)^{(t_{\max})} x_j^{(t_{\max})}$$

(2) Iteration, $1 \leq t < t_{\max}$

$$\left(\frac{\partial E}{\partial w_{ij}^{(OH)}} \right)^{(t)} = \left(\frac{\partial E}{\partial \vartheta_i^{(O)}} \right)^{(t)} v_j^{(t)}$$

$$\left(\frac{\partial E}{\partial w_{ij}^{(CH)}} \right)^{(t)} = \left(\frac{\partial E}{\partial \vartheta_i^{(C)}} \right)^{(t)} v_j^{(t-1)}$$

$$\left(\frac{\partial E}{\partial w_{ij}^{(HC)}} \right)^{(t)} = \left(\frac{\partial E}{\partial \vartheta_i^{(H)}} \right)^{(t)} u_j^{(t)}$$

$$\left(\frac{\partial E}{\partial w_{ij}^{(HI)}} \right)^{(t)} = \left(\frac{\partial E}{\partial \vartheta_i^{(H)}} \right)^{(t)} x_j^{(t)}$$

Total partial derivatives

$$\frac{\partial E}{\partial \vartheta_i^{(O)}} = \sum_{t=1}^{t_{\max}} \left(\frac{\partial E}{\partial \vartheta_i^{(O)}} \right)^{(t)}, \quad \frac{\partial E}{\partial \vartheta_i^{(H)}} = \sum_{t=1}^{t_{\max}} \left(\frac{\partial E}{\partial \vartheta_i^{(H)}} \right)^{(t)}$$

$$\frac{\partial E}{\partial \vartheta_i^{(C)}} = \sum_{t=2}^{t_{\max}} \left(\frac{\partial E}{\partial \vartheta_i^{(C)}} \right)^{(t)}$$

$$\frac{\partial E}{\partial w_{ij}^{(OH)}} = \sum_{t=1}^{t_{\max}} \left(\frac{\partial E}{\partial w_{ij}^{(OH)}} \right)^{(t)} = \sum_{t=1}^{t_{\max}} \left(\frac{\partial E}{\partial \vartheta_i^{(O)}} \right)^{(t)} v_j^{(t)}$$

$$\frac{\partial E}{\partial w_{ij}^{(HC)}} = \sum_{t=2}^{t_{\max}} \left(\frac{\partial E}{\partial w_{ij}^{(HC)}} \right)^{(t)} = \sum_{t=2}^{t_{\max}} \left(\frac{\partial E}{\partial \vartheta_i^{(H)}} \right)^{(t)} u_j^{(t)}$$

$$\frac{\partial E}{\partial w_{ij}^{(CH)}} = \sum_{t=1}^{t_{\max}-1} \left(\frac{\partial E}{\partial w_{ij}^{(CH)}} \right)^{(t)} = \sum_{t=1}^{t_{\max}-1} \left(\frac{\partial E}{\partial \vartheta_i^{(C)}} \right)^{(t+1)} v_j^{(t)}$$

$$\frac{\partial E}{\partial w_{ij}^{(HI)}} = \sum_{t=1}^{t_{\max}} \left(\frac{\partial E}{\partial w_{ij}^{(HI)}} \right)^{(t)} = \sum_{t=1}^{t_{\max}} \left(\frac{\partial E}{\partial \vartheta_i^{(HI)}} \right)^{(t)} x_j^{(t)}$$