

5. Sequence modeling with recurrent neural networks

5.5. Gated Recurrent Units

Manel Martínez-Ramón

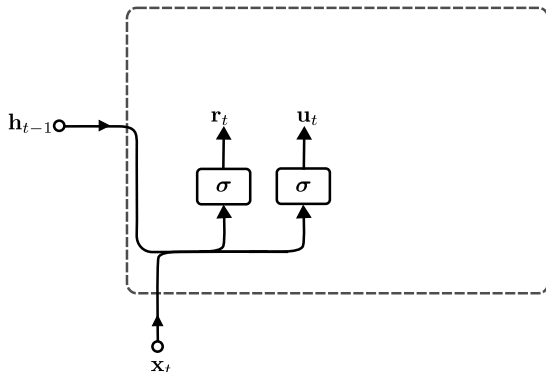
Meenu Ajith

Aswathy Rajendra Kurup

- The LSTM architecture was successful by avoiding the unwanted Elman's RNN phenomena.
- After its inception, a number of authors published attempts to simplify this structure.
- The Gated Recurrent Unit (GRU) was introduced in 2014 by K. Cho, K. et al (including Yoshua Bengio) whose performance is similar to the one of the LSTM, but with less computational time.

Structure

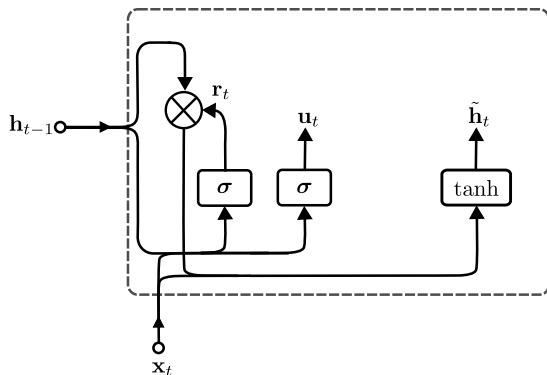
Reset and update gates \mathbf{r}_t and \mathbf{u}_t



- The reset gate \mathbf{r}_t has the role of the forgetting gate in an LSTM. It modulates how much of the previous state is passed.
- The update gate \mathbf{u}_t is used to update the previous state modulated by \mathbf{r}_t .

Structure

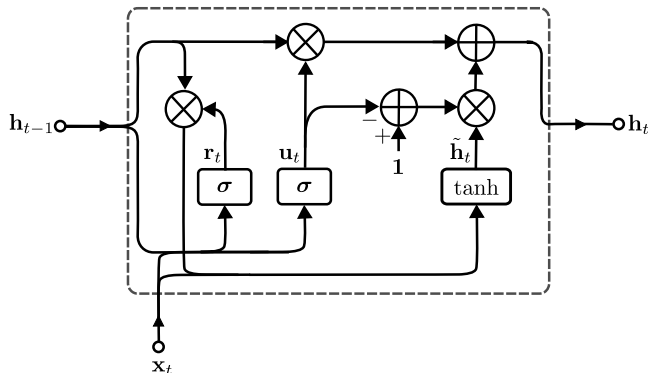
Candidate hidden state $\tilde{\mathbf{h}}_t$



- The combination of the current input \mathbf{x}_t and the previous modulated state $\mathbf{r}_t \odot \mathbf{h}_{t-1}$ is used to estimate a new candidate $\tilde{\mathbf{h}}_t$ for the new hidden state.

Structure

New hidden state \mathbf{h}_t

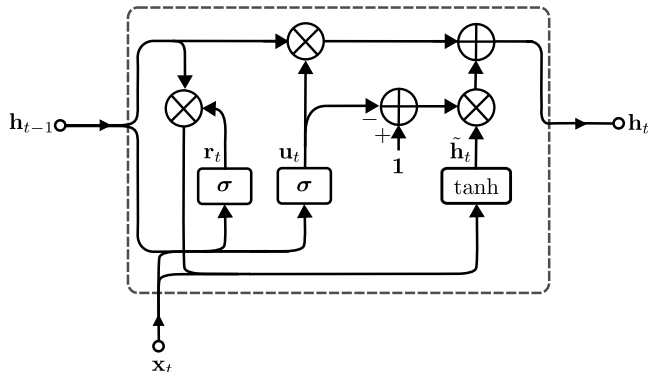


- The new state \mathbf{h}_t is a combination of the previous state and the new candidate $\tilde{\mathbf{h}}_t$. Note the unit gain feedback.

$$\mathbf{h}_t = \mathbf{u}_t \odot \mathbf{h}_{t-1} + (\mathbf{1} - \mathbf{u}_t) \odot \tilde{\mathbf{h}}_t \quad (1)$$

Structure

GRU equations \mathbf{h}_t



$$\begin{aligned} \mathbf{r}_t &= \sigma(\mathbf{W}_{xr}^\top \mathbf{x}_t + \mathbf{W}_{hr}^\top \mathbf{h}_{t-1} + \mathbf{b}_r) & \tilde{\mathbf{h}}_t &= \sigma(\mathbf{W}_{xh}^\top \mathbf{x}_t + \mathbf{W}_{hh}^\top (\mathbf{r}_t \odot \mathbf{h}_{t-1}) \\ & & & + \mathbf{b}_h) \\ \mathbf{u}_t &= \sigma(\mathbf{W}_{xu}^\top \mathbf{x}_t + \mathbf{W}_{hu}^\top \mathbf{h}_{t-1} + \mathbf{b}_u) & \mathbf{h}_t &= \mathbf{u}_t \odot \mathbf{h}_{t-1} + (1 - \mathbf{u}_t) \odot \tilde{\mathbf{h}}_t \end{aligned} \quad (2)$$