Training RNNs as Fast as CNNs

Model

Main motivations

1. ***process the input at each step independently of the other inputs.***
2. ***do the recurrent combination with relatively lightweight computaion (element-wise operations that can be fused into a single kernel function call).***

Euqatioins of Simple Recurrent Units (SRU)

- ***linear transformation of the input***

$$\tilde{\mathbf{x}}_t = \mathbf{W}\mathbf{x}_t$$

- ***forget gate***

$$\mathbf{f}_t = \sigma(\mathbf{W}_f\mathbf{x}_t + \mathbf{b}_f)$$

- ***reset gate***

$$\mathbf{r}_t = \sigma(\mathbf{W}_r\mathbf{x}_t + \mathbf{b}_r)$$

- ***internal state***

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + (\mathbf{1} - \mathbf{f}_t) \odot \tilde{\mathbf{x}}_t$$

- ***output state***

$$\mathbf{h}_t = \mathbf{r}_t \odot g(\mathbf{c_t}) + (\mathbf{1} - \mathbf{r}_t \odot \mathbf{x}_t)$$