column vector

$$x_t = \begin{bmatrix} \\ \\ \end{bmatrix}_{d_x \times 1} \qquad h_t = \begin{bmatrix} \\ \\ \end{bmatrix}_{d_h \times 1}$$

$$\sigma\left(\begin{bmatrix} 0.1 \\ -0.5 \\ 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} \sigma(0.1) \\ \sigma(-0.5) \\ \sigma(0) \\ \sigma(1) \end{bmatrix} = \begin{bmatrix} 0.525 \\ 0.378 \\ 0.5 \\ 0.73 \end{bmatrix}$$

$$z_t : d_h \times 1$$

$$r_t : d_h \times 1$$

$$W^{(z)}, W^{(r)}, W^{(h)} : d_h \times d_x$$

$$U^{(z)}, U^{(r)}, W^{(h)} : d_h \times d_h$$

Conv 1   1-D (unigram)
filter #1

$$\begin{bmatrix} 0.03 \\ 0.02 \\ 0.1 \\ 0 \\ 1.1 \\ 0 \end{bmatrix}$$

$$\begin{array}{l} 0.03 = \\ 0.02 = \\ 0.1 = \\ 0 \\ 1.1 = \\ 0 = \end{array}$$

$\boxed{1 \mid 1 \mid 0}$ ×the $\boxed{0 \mid 0.03 \mid 0.01}$

movie $\boxed{0.01 \mid 0.01 \mid 0.04}$

was $\boxed{0 \mid 0.1 \mid 0}$

not $\boxed{0.1 \mid 0 \mid 0.4}$

great $\boxed{0.0 \mid 0.81 \mid 0}$

good $\boxed{0.6 \mid 0.5 \mid 0}$

. $\boxed{0 \mid 0 \mid 0}$

Conv 1   1-D
filter 2 $\boxed{1 \mid -1 \mid 0}$

Conv 2
$$\begin{array}{c} 0.03 \\ 0.02 \\ 0.1 \\ 0 \\ 1.1 \\ 0 \end{array} \quad \begin{array}{c} -0.03 \\ 0 \\ -0.1 \\ 0 \\ 0.1 \\ 0 \end{array}$$

$$\begin{bmatrix} 1.1 \\ 0.1 \end{bmatrix} \begin{bmatrix} \\ \\ 1.5 \end{bmatrix}$$

k-dim vectors

(using k-filters)

2-D filter #5
(bigram)

$$X \begin{array}{|c|c|c|} \hline 0 & 0 & 1 \\ \hline 1 & 1 & 0 \\ \hline \end{array}$$

Conv 5
$$\begin{bmatrix} 0.03 \\ 0.14 \\ 0.1 \\ 1.5 \\ 0.0 \end{bmatrix}$$

max pooling
perform
a sub-sampling
(same as avg. pooling)

max — only features
that most strongly
triggering