Assignment 3 for CS224d

Lifu Huang

April 2016

1 RNN's(Recursive Neural Network)

(a)

$$\delta^{(s)} = \hat{y} - y$$
$$\delta^{(1)} = f'(h^{(1)}) \circ (U^T \delta^{(s)} + \delta_{above})$$
$$\delta_{below} = (W^{(1)})^T \delta^{(1)}$$
$$\nabla_U J = \delta^{(s)}(h^{(1)})^T$$
$$\nabla_{b^{(s)}} J = \delta^{(s)}$$
$$\nabla_{W^{(1)}} J = \delta^{(1)} \left[h^{(1)}_{left}; h^{(1)}_{right} \right]^T$$
$$\nabla_{b^{(1)}} J = \delta^{(1)}$$
$$\nabla_{[L_{left}; L_{right}]} J = \delta_{below}$$

(b)
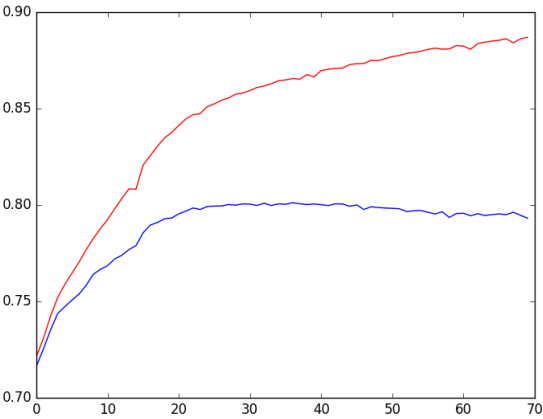
Please see code files for details.

(c)

(a)

Figure 1: Accuracy on Training and Dev Set over Epochs

(b)

Beacause training for too many epochs may lead to the problem of over fitting.
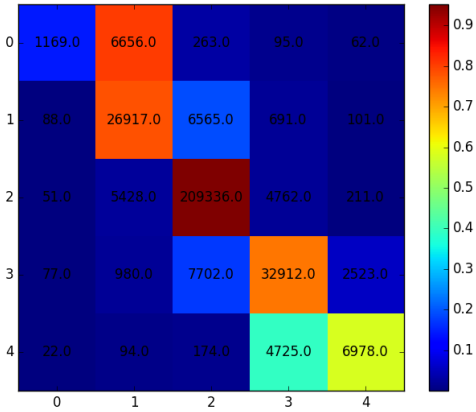
(c)
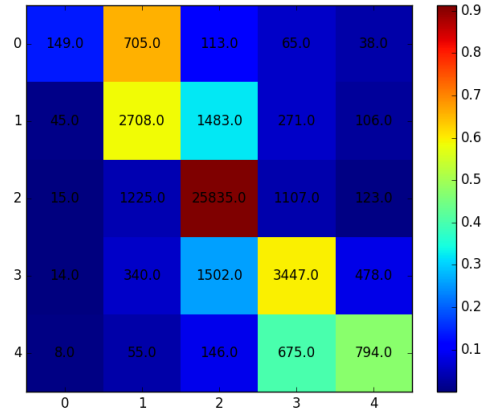

Figure 2: Confusion Matrix on Training Set
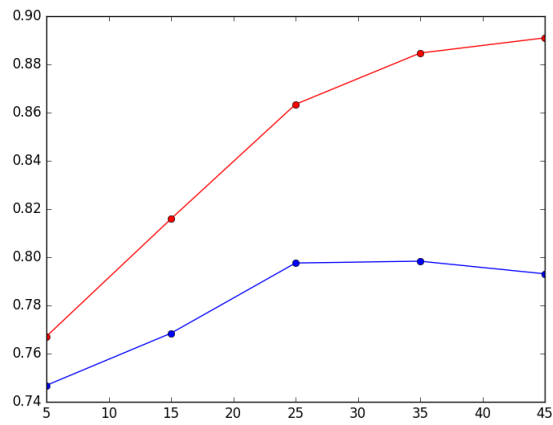
Figure 3: Confusion Matrix on Dev Set

(d)

Figure 4: Accuracy on Training and Dev Set over wvecDims

3

2   2-Layer Deep RNN's

(a)

$$\delta^{(s)} = \hat{y} - y$$
$$\nabla_U J = \delta^{(s)}(h^{(2)})^T$$
$$\nabla_{b^{(s)}} J = \delta^{(s)}$$

$$\delta^{(2)} = f'(h^{(2)}) \circ (U^T \delta^{(s)})$$
$$\nabla_{W^{(2)}} J = \delta^{(2)}(h^{(1)})^T$$
$$\nabla_{b^{(2)}} J = \delta^{(2)}$$

$$\delta^{(1)} = f'(h^{(1)}) \circ ((W^{(2)})^T \delta^{(2)} + \delta_{above})$$
$$\nabla_{W^{(1)}} J = \delta^{(1)} \left[h^{(1)}_{left}; h^{(1)}_{right} \right]^T$$
$$\nabla_{b^{(1)}} J = \delta^{(1)}$$

$$\delta_{below} = (W^{(1)})^T \delta^{(1)}$$
$$\nabla_{[L_{left};L_{right}]} J = \delta_{below}$$

(b)

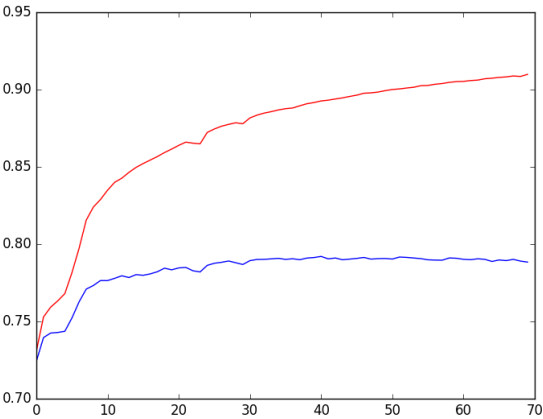Please see code files for details.

(c)

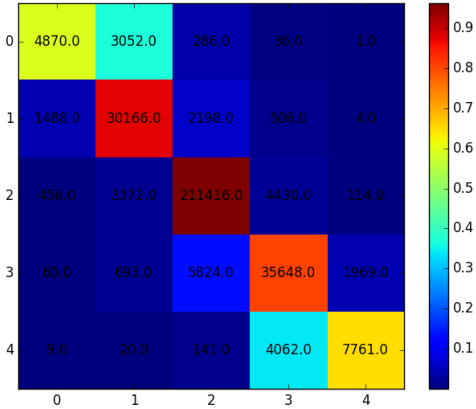**(a)**

Figure 5:

(b)

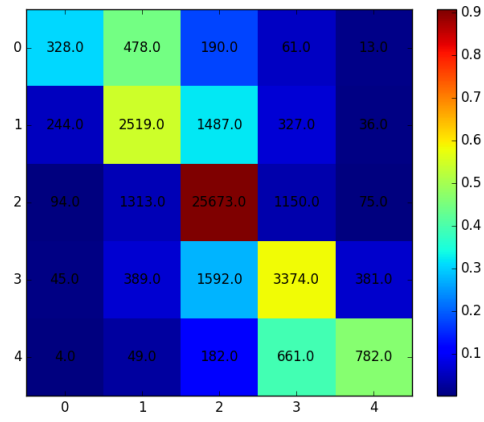Figure 6: Confusion Matrix on Training Set

Figure 7: Confusion Matrix on Dev Set

(c)

Model RNN2 performs better on training set while worse on dev set, which indicates a typical problem of over fitting.
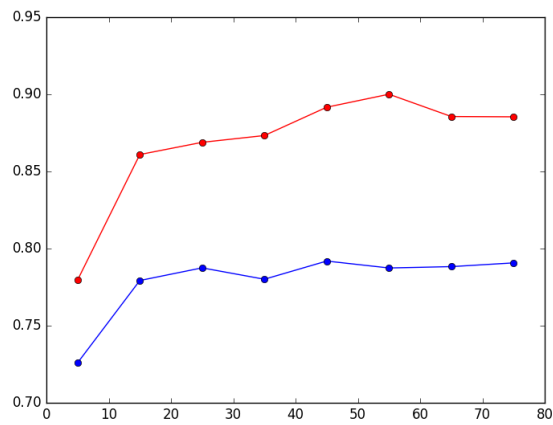
(d)

Figure 8: Accuracy on Training and Dev Set over middleDims

(d)

To be updated.

(e)

To be updated.

(f)

To be updated.

3 Extra Credit: Recursive Neural Tensor Networks

To be updated.