Signed Network Embedding in Social Media Supplementary Material

Suhang Wang∗ Jiliang Tang† Charu Aggarwal† Yi Chang‡ Huan Liu∗

S1 Optimization of SINE

Following the common way, we employ the backpropagation to optimize the deep network for SINE [1]. The key idea of backpropagation is to update the parameters in a backward direction by propagating "errors" backward to efficiently calculate the gradients. Basically, we want to optimize Eq. (3.3) w.r.t to $\mathbf{X}, \mathbf{x}_0$ and $\theta$. The key step of optimizing Eq. (3.3) is to get the gradient of $\max(0,f(\mathbf{x}_i,\mathbf{x}_k)+\delta-f(\mathbf{x}_i,\mathbf{x}_j))$ respect to the parameters, $\mathbf{X}, \mathbf{x}_0$ and $\theta$. With the gradient, we can then update the parameters using gradient descent method. Let’s first analyze $\max(0,f(\mathbf{x}_i,\mathbf{x}_k)+\delta-f(\mathbf{x}_i,\mathbf{x}_j))$.

- If $\max(0,f(\mathbf{x}_i,\mathbf{x}_k)+\delta-f(\mathbf{x}_i,\mathbf{x}_j)) = 0$, or equivalent, $f(\mathbf{x}_i,\mathbf{x}_k)+\delta-f(\mathbf{x}_i,\mathbf{x}_j) \leq 0$, the parameters have already been optimized for the inputs $\mathbf{x}_i$ and $\mathbf{x}_j$. In other words, the gradient of $\max(0,f(\mathbf{x}_i,\mathbf{x}_k)+\delta-f(\mathbf{x}_i,\mathbf{x}_j))$ is 0 when $f(\mathbf{x}_i,\mathbf{x}_k)+\delta-f(\mathbf{x}_i,\mathbf{x}_j) \leq 0$.

- If $\max(0,f(\mathbf{x}_i,\mathbf{x}_k)+\delta-f(\mathbf{x}_i,\mathbf{x}_j)) > 0$, $\max(0,f(\mathbf{x}_i,\mathbf{x}_k)+\delta-f(\mathbf{x}_i,\mathbf{x}_j))$ is equal to $f(\mathbf{x}_i,\mathbf{x}_k)+\delta-f(\mathbf{x}_i,\mathbf{x}_j)$.

The same idea can be applied to $\max(0,f(\mathbf{x}_i,\mathbf{x}_0)+\delta_0-f(\mathbf{x}_i,\mathbf{x}_j))$. Based on the aforementioned analysis, we only need to take gradient of $f(\mathbf{x}_i,\mathbf{x}_j)$ w.r.t the parameters. Then we are able to get the gradient of Eq. (3.3) with some calculations. We will start from the parameters of the N-th layer and go backward to get derivatives for other layers. First, using Eq (3.8), the derivative of $f(\mathbf{x}_i,\mathbf{x}_j)$ w.r.t $\mathbf{w}$ is given as

\[
\frac{\partial f(\mathbf{x}_i, \mathbf{x}_j)}{\partial \mathbf{w}} = (1 - f^2(\mathbf{x}_i, \mathbf{x}_j)) \frac{\partial}{\partial \mathbf{w}} [\mathbf{w}^T \mathbf{z}^{N1} + b] \\
= (1 - f^2(\mathbf{x}_i, \mathbf{x}_j)) \mathbf{z}^{N1}
\]

and similarly, the derivative of $f(\mathbf{x}_i, \mathbf{x}_j)$ w.r.t $b$ is

\[
\frac{\partial f(\mathbf{x}_i, \mathbf{x}_j)}{\partial b} = 1 - f^2(\mathbf{x}_i, \mathbf{x}_j)
\]

Next, the gradient of $f(\mathbf{x}_i, \mathbf{x}_j)$ w.r.t $\mathbf{z}^{N1}$ is given as

\[
\frac{\partial f(\mathbf{x}_i, \mathbf{x}_j)}{\partial \mathbf{z}^{N1}} = [1 - f^2(\mathbf{x}_i, \mathbf{x}_j)] \frac{\partial}{\partial \mathbf{z}^{N1}} [\mathbf{w}^T \mathbf{z}^{N1} + b] \\
= [1 - f^2(\mathbf{x}_i, \mathbf{x}_j)] \mathbf{w}
\]

Let $\delta^{N1}$ be a vector with its $s$-th element $\delta_s^{N1}$ defined as

\[
\delta_s^{N1} = \frac{\partial f(\mathbf{x}_i, \mathbf{x}_j)}{\partial \mathbf{z}^{N1}} [1 - (z_s^{N1})^2] \\
= [1 - (z_s^{N1})^2](1 - f^2(\mathbf{x}_i, \mathbf{x}_j))w_s
\]

where $z_s^{N1}$ is the $s$-th element of $\mathbf{z}^{N1}$. $\delta^{N1}$ is the “error” generated by the output layer and will propagate back to the N-th layer as shown later. Using the chain rule and Eq. (3.7), the derivative of $f(\mathbf{x}_i, \mathbf{x}_j)$ w.r.t. $\mathbf{W}^{N}$ is given as:

\[
\frac{\partial f(\mathbf{x}_i, \mathbf{x}_j)}{\partial \mathbf{W}^{N}_{st}} = \sum_k \frac{\partial f(\mathbf{x}_i, \mathbf{x}_j)}{\partial \mathbf{z}_k^{N1}} \frac{\partial \mathbf{z}_k^{N1}}{\partial \mathbf{W}^{N}_{st}} \\
= \frac{\partial f(\mathbf{x}_i, \mathbf{x}_j)}{\partial \mathbf{z}_s^{N1}} [1 - (z_s^{N1})^2] z_s^{(N-1)1}
\]

With Eq. (S4), the above equation is simplified as

\[
\frac{\partial f(\mathbf{x}_i, \mathbf{x}_j)}{\partial \mathbf{W}^{N}_{st}} = \delta_s^{N1} z_s^{(N-1)1}
\]

Similarly, the derivative of $f(\mathbf{x}_i, \mathbf{x}_j)$ w.r.t. $\mathbf{b}^{N}$ is given as:

\[
\frac{\partial f(\mathbf{x}_i, \mathbf{x}_j)}{\partial \mathbf{b}^{N}_{s}} = \delta_s^{N1}
\]

From Eqs (S6) and (S7), we can see that the “error” $\delta^{N1}$ is propagated backwards, i.e., it is used for the calculation of the gradients of the parameters for the N-th layer.

Generally, the “error” for the $n$-th layer is denoted as $\delta^{n1}, 1 \leq n < N$, with it’s $s$-th element defined as

\[
\delta^{n1}_s = \frac{\partial f(\mathbf{x}_i, \mathbf{x}_j)}{\partial \mathbf{z}_s^{n1}} [1 - (z_s^{n1})^2]
\]
where the derivative of $f(x_i, x_j)$ w.r.t $z^{n1}$ is given as

$$\frac{\partial f(x_i, x_j)}{\partial z^{n1}} = \sum_k \frac{\partial f(x_i, x_j)}{\partial z^k} \frac{\partial z^k}{\partial z^{n1}} = \sum_k \frac{\partial f(x_i, x_j)}{\partial z^k} [1 - (z^{n1})^2] W_{ks}^{n1} = \sum_k \delta_k^{(n1)} W_{ks}^{n1}$$

(S9)

Thus, we have

$$(S10) \quad \delta_{s}^{n1} = [1 - (z_{s}^{n1})^2] \sum_k \delta_k^{(n1)} W_{ks}^{n1}, 1 \leq n < N$$

It is clear from the above equation that the “error” $\delta_k^{(n1)}$ from the $(n + 1)$-th layer is back propagated to the $n$-th layer for the calculation of $\delta_{s}^{n1}$. With $\delta_{s}^{n1}$, the derivative of $f(x_i, x_j)$ w.r.t $W^n$ and $b^n$ is given as

$$\frac{\partial f(x_i, x_j)}{\partial W_{st}^{n}} = \sum_k \frac{\partial f(x_i, x_j)}{\partial z^k} \frac{\partial z^k}{\partial W_{st}^{n}} = \frac{\partial f(x_i, x_j)}{\partial z^{n1}} [1 - (z^{n1})^2] W_{st}^{n1} = \delta_{s}^{n1} W_{st}^{n1}, 1 < n < N$$

(S11)

and

$$(S12) \quad \frac{\partial f(x_i, x_j)}{\partial b_{s}^{n}} = \delta_{s}^{n1}, 1 \leq n < N$$

The derivative of $f(x_i, x_j)$ w.r.t $W^{11}$ and $W^{12}$ are

$$\frac{\partial f(x_i, x_j)}{\partial W^{11}} = \sum_k \frac{\partial f(x_i, x_j)}{\partial z^k} \frac{\partial z^k}{\partial W^{11}} = \delta^{11}(x_i)^T$$

(S13)

and

$$\frac{\partial f(x_i, x_j)}{\partial W^{12}} = \delta^{11}(x_j)^T$$

(S14)

Finally, the derivative of $f(x_i, x_j)$ w.r.t $x_i$ is

$$\frac{\partial f(x_i, x_j)}{\partial x_i} = \sum_k \frac{\partial f(x_i, x_j)}{\partial z^k} \frac{\partial z^k}{\partial x_i} = (W^{11})^T \delta^{11}$$

(S15)

and the derivative of $f(x_i, x_j)$ w.r.t $x_j$ is

$$\frac{\partial f(x_i, x_j)}{\partial x_j} = (W^{12})^T \delta^{11}$$

(S16)

Similarly, for $f(x_i, x_k) = \tanh((W^{N1})^T z^{N2} + b^{N1})$, we define $\delta_{s}^{n2}$ as

$$\delta_{s}^{n2} = \frac{\partial f(x_i, x_k)}{\partial z^{n2}} [1 - (z^{n2})^2]$$

(S17)

With the same procedure as $f(x_i, x_j)$, we can get the derivatives of $f(x_i, x_k)$ w.r.t the parameters. We omit the details here and just

With these derivatives, it’s easy to get the derivatives of the objective in Eq (3.3) w.r.t to the parameters. We denote the objective as $\mathcal{L}(X, x_0, \theta)$. In each iteration, the parameters are updated using gradient descent. Taking $x_0$ as an example, the update rule is given as

$$x_0 \leftarrow x_0 - \gamma \frac{\partial \mathcal{L}(X, x_0, \theta)}{\partial x_0}$$

(S18)

where $\gamma$ is the learning rate.

**S2 Summary of Derivatives**

In this section, we summarize the derivatives. For $f(x_i, x_j) = \tanh(W^T z^{N1} + b)$, we have

$$\delta_{s}^{n1} = \begin{cases} \frac{[1 - (z_{s}^{n1})^2] \sum_k \delta_k^{(n1)} W_{ks}^{(n1)}}{w_s[1 - (z_{s}^{N2})^2][1 - f^2(x_i, x_j)]}, & n = N \end{cases}$$

and the derivatives of $f(x_i, x_j)$ w.r.t $\theta$ are given as

$$\frac{\partial f(x_i, x_j)}{\partial W} = \frac{[1 - f^2(x_i, x_j)] z^{N1}}{w_s[1 - (z_{s}^{N2})^2][1 - f^2(x_i, x_j)]}, \quad n = N$$

(S20)

the derivatives of $f(x_i, x_j)$ w.r.t $x_i, x_j$ are given as

$$\frac{\partial f(x_i, x_j)}{\partial x_i} = (W^{11})^T \delta^{11}$$

(S21)

and

$$\frac{\partial f(x_i, x_j)}{\partial x_j} = (W^{12})^T \delta^{11}$$

(S22)

For $f(x_i, x_k) = \tanh((W^{N1})^T z^{N2} + b)$, we have

$$\delta_{s}^{n2} = \begin{cases} \frac{[1 - (z_{s}^{n2})^2] \sum_k \delta_k^{(n2)} W_{ks}^{(n2)}}{w_s[1 - (z_{s}^{N2})^2][1 - f^2(x_i, x_k)]}, & 1 \leq n < N \end{cases}$$

(S23)
and the derivatives of $f(x_i, x_k)$ w.r.t $\theta$ are given as

$$\frac{\partial f(x_i, x_k)}{\partial w} = [1 - f^2(x_i, x_k)]z^2$$

$$\frac{\partial f(x_i, x_k)}{\partial W_{st}} = \delta^2_{s} z_{t}^{(n-1)2}, 1 < n < N$$

(S23)

$$\frac{\partial f(x_i, x_k)}{\partial W_{11}} = \delta^{12}(x_i)^T$$

$$\frac{\partial f(x_i, x_k)}{\partial W_{12}} = \delta^{12}(x_j)^T$$

$$\frac{\partial f(x_i, x_k)}{\partial b} = 1 - f^2(x_i, x_k)$$

$$\frac{\partial f(x_i, x_k)}{\partial b_n} = \delta^2_{s}, 1 \leq n < N$$

the derivatives of $f(x_i, x_k)$ w.r.t $x_i, x_k$ are given as

(S24)

$$\frac{\partial f(x_i, x_k)}{\partial x_i} = (W_{11})^T \delta^{12} \text{ and } \frac{\partial f(x_i, x_k)}{\partial x_k} = (W_{12})^T \delta^{12}$$

References