

$$\begin{aligned}
M &:= 100 \\
C &:= 20000 \\
X_{i,t} &\in N, \ i = 1...n, \ t = 1...L_i \\
Y_i &\in N, \ i = 1...n \\
E &\in R^{C,M} \\
Wz, Rz, Wi, Ri, Wf, Rf, Wo, Ro &\in R^{M,M} \\
bz, pi, bi, pf, bf, po, bo &\in R^{1,M} \\
x_{i,t} &= E_{X_{i,t}}, \ i = 1...n, \ t = 1...L_i \\
z_{i,1} &= \text{Tanh}(x_{i,1} \cdot Wz + bz), \ i = 1...n \\
ig_{i,1} &= \text{Sigmoid}(x_{i,1} \cdot Wi + bi), \ i = 1...n \\
fg_{i,1} &= \text{Sigmoid}(x_{i,1} \cdot Wf + bf), \ i = 1...n \\
c_{i,1} &= ig_{i,1} * z_{i,1}, \ i = 1...n \\
o_{i,1} &= \text{Sigmoid}(x_{i,1} \cdot Wo + c_{i,1} * po + bo), \ i = 1...n \\
y_{i,t} &= o_{i,t} * \text{Tanh}(c_{i,t}), \ i = 1...n, \ t = 1...L_i \\
z_{i,t} &= \text{Tanh}(x_{i,t} \cdot Wz + y_{i,t-1} \cdot Rz + bz), \ i = 1...n, \ t = 2...L_i \\
ig_{i,t} &= \text{Sigmoid}(x_{i,t} \cdot Wi + y_{i,t-1} \cdot Ri), \ i = 1...n, \ t = 2...L_i \\
fg_{i,t} &= \text{Sigmoid}(x_{i,t} \cdot Wf + y_{i,t-1} \cdot Rf), \ i = 1...n, \ t = 2...L_i \\
c_{i,t} &= ig_{i,t} * z_{i,t} + fg_{i,t} * c_{i,t-1}, \ i = 1...n, \ t = 2...L_i \\
o_{i,t} &= \text{Sigmoid}(x_{i,t} \cdot Wo + y_{i,t-1} \cdot Ro + c_{i,t} * po + bo), \ i = 1...n, \ t = 2...L_i \\
r_i &= \sum_{t=1}^{L_i} y_{i,t}, \ i = 1...n \\
U &\in R^{M,2} \\
b &\in R^{1,2} \\
P_i &= r_i \cdot U + b, \ i = 1...n \\
loss &= \sum_{i=1}^n \text{CrossEntropy}(P_i, \text{OneHot}(Y_i, 2)) \\
X, Y &\sim \text{"imdb.data"} \\
E, Wz, Rz, Wi, Ri, Wf, Rf, Wo, Ro, pi, pf, po, U &\sim \text{Gaussian}(0, 0.1) \\
bz, bi, bf, bo, b &\sim \text{Zero}() \\
\min_{E,Wz,Rz,Wi,Ri,Wf,Rf,Wo,Ro,bz,pi,bi,pf,bf,po,bo,U,b} & \quad loss
\end{aligned}$$