Word2Vec

□□CBOW□□skip-gram□□□□□□□□□□□□□□□□

□□□□□□□□□□□□

□□□□$v=W \times one_hot(v)$

□□W□$Word2Vec$□□□□□□

CBOW

□□□□□□□□□□□

□□□

□□□□□□□□□ `one-hot` □□□

image-20250705224627824

□□□

□□□□□□□□□□□□□□□□□C□□□□ $$ X =(x_{i-c},x_{i-c+1},\dots,x_{i-1},x_{i+1}\dots,x_{i+c}) \in \mathbb{R}^{V\times2C} $$ □□□x_i□□□□□$x_i \in \mathbb{R}^{V \times 1}$

□□□□

□□□□x_i□□□□□□□□□□□□W□□□□□□$b\in \mathbb{R}^{D \times 1}$□□$x_j'$□□$X_j'= WX_j + b$□□□□□□□ $$ X' = WX + b $$ □□ $$ X =[x_{i-c},x_{i-c+1},\dots,x_{i-1},x_{i+1}\dots,x_{i+c}] \ X' = [x_{i-c}',x_{i-c+1}',\dots,x_{i-1}',x_{i+1}'\dots,x_{i+c}'] $$

□□□□

□□□□□□□X_j'□□□□□□□h $$ h=\sum_{j=i-C,j\neq i}^{i+C}{\frac{1}{2C}(x_{i-c}'+x_{i-c+1}'+x_{i-1}'+x_{i+1}'+x_{i+c}')} $$

□□□□

□□□h□□□□□□□□□□□□□□P□$P = (P_1,P_2,\dots,P_V)^T$□$P_i \in R$□□□□□□□$i$□□□□□ $$ P = W'h + b' $$

Softmax□

□□□□□□□□Softmax□□□□□□P'□$P' = (p_1',p_2',\dots,p_V')$□$p_i'$□□□□□□i□□□□□ $$ p_i' = softmax(p_i) = \frac{exp(p_i)}{\sum_{k=1}^Vexp(p_k)} $$ □□□□P'□□□□□□□□□□□□□□1□□□□□0□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□

□□□□□□□**Cross Entropy**□□□□□□□□□Softmax□□□□□□□□P'□□□□□□□□□□T□□□$P'= (P_1',P_2',\dots,P_V')^T$□□□□□□□$T=(t_1,t_2,\dots,t_V)$□□□□□□□□□□□□ $$ Loss = -\sum_{i=1}^Vt_i\log (P_i') $$

skip-gram

□□□□□□□□□□□

□□□

□□

image-20250705224657637

[illegible]

[illegible]$x_i\in \mathbb{R}^{V \times 1} $[illegible]i[illegible]

[illegible]

[illegible]$x_i\in \mathbb{R}^{V \times 1} $[illegible]$h$ $$ h = Wx_i + b $$

[illegible]

[illegible]h[illegible]W_j'[illegible]b_j'[illegible]$S_j\in \mathbb{R}^{V\times 1}$[illegible]$S = (S_1,S_2,\dots,S_{2C})^T$[illegible] $$ S_j=W_j'h+b_j' $$

Softmax[illegible]

[illegible]S_j[illegible]softmax[illegible]P_j[illegible]$P_j = (P_j(0),P_j(1),\dots,P_j(V-1))^T$ $$ P_j(k) = Softmax(S_j) = \frac{exp(S_j(k))}{\sum_{l=0}^{V-1}{exp(S_j(l))}} $$

[illegible]

[illegible]P[illegible]1[illegible]0[illegible]

[illegible]

[illegible]Cross Entropy[illegible]$P_j(k)$[illegible]T[illegible]$P_j = (P_1,P_2,\dots,P_{2C})$[illegible]$T_j=[t_j(1),t_j(2),\dots,t_j(V)]$[illegible] $$ L_j = -\sum_{k=1}^{V}t_k\log(P_j(k)) \ Loss = \sum_{j=1}^{2C}L_j $$

[illegible]

[illegible]word2vec[illegible]V[illegible]**Huffman**[illegible]$\log V$[illegible]

[illegible]Softmax

[illegible]Huffman[illegible]

- [illegible]
- [illegible]

[illegible]

image-20250705162047281

[illegible]

1. [illegible]**softmax**[illegible]P[illegible]
2. [illegible]P[illegible]$1-P$
3. [illegible]
4. [illegible]

[illegible]

[illegible]

[illegible]

□□□□□□□□□□□□□□□□“□□□”□□□**1**□□□□□□□□□□□□**0**□

□□□□

□□□□□□□□□□□□□□□□□□□□□□

□□□□□

□□□□□□T_1□□□□□□□$T_1 = (t_1,t_2,\dots,t_V)^T$□$P$□□□□□□□□□□□□$P=(P_1,P_2,\dots,P_V)$□□□□□□□□□$T$□□□□□□□□□□**1**□□$S_1 = (\theta_1h)^TT_1$

□□□□□□□□□□□□□□□□□□□□□sigmoid□□□□□□□Cross Entropy $$ P_1 = \sigma(S_1) = \sigma((\theta_1h)^TT_1) = \frac{1}{1+e^{-((\theta_1h+b_1')^TT_1)}}\ Loss_+ = -log(P_1) = -log(\frac{1}{1+e^{-((\theta_1h+b_1')^TT_1)}}) $$ □□□□

□□□□□□□□□□□□□□□□□□□□□ $$ f(w) = \frac{[count(w)]^{\frac{3}{4}}}{\sum_{i=1}^V[count(i)]^{\frac{3}{4}}} $$ □□□$count(index)$□□□□□□□index□□□□□□□□□□w□□□□□□□□V□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□k□□□□5□□□□□□□□□□□□5~20□□

□□□□□□□□□ □□□□□□□□□□□□□□□□□$Sigmoid$□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□θ_0□□□□□□□h□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□ $$ S_{0,i} = -(\theta_0h+b_0')^TT_{0,i} $$ □□□$T_{0,i}$□□□□□□□□□□□□Sigmoid□□□□□□□□ $$ P_1 = \sigma(S_{0,i}) $$ □□□□□□□□ $$ Loss_- = \sum_{i=1}^k\log(P_{0,i}) $$ □□□$Loss_+$□$Loss_-$□□□□□□□$Loss$ $$ Loss = Loss_-+Loss_+ $$