

0.1 Mathematical Equations

$$x^{(j)} = \sum_{i < j} o^{(i,j)} \left(x^{(i)} \right) \quad (1)$$

$$\bar{o}^{(i,j)}(x) = \sum_{o \in \mathcal{O}} \frac{\exp \left(\alpha_o^{(i,j)} \right)}{\sum_{o' \in \mathcal{O}} \exp \left(\alpha_{o'}^{(i,j)} \right)} o(x) \quad (2)$$

where $\beta^{(i,j)} = \left(\beta_{o_1^{(i,j)}}, \beta_{o_2^{(i,j)}}, \dots, \beta_{o_M^{(i,j)}} \right)$ is one-hot encoded alpha by Softmax, and $\alpha_o^{(i,j)} = \left(\alpha_{o_1^{(i,j)}}, \alpha_{o_2^{(i,j)}}, \dots, \alpha_{o_M^{(i,j)}} \right)$ is an architectural weight vector.

$$\text{Softmax} \left(\alpha_o^{(i,j)} \right) \approx \beta^{(i,j)} \quad (3)$$

$$\bar{o}^{(i,j)}(x) = \sum_{o \in \mathcal{O}} \frac{\exp \left(\alpha_o^{(i,j)} / T \right)}{\sum_{o' \in \mathcal{O}} \exp \left(\alpha_{o'}^{(i,j)} / T \right)} o(x) \quad (4)$$

$$\bar{o}^{(i,j)}(x_i) = \sum_{o \in \mathcal{O}} \frac{\exp \left(\alpha_o^{(i,j)} / T \right)}{\sum_{o' \in \mathcal{O}} \exp \left(\alpha_{o'}^{(i,j)} / T \right)} o(x_i) \quad (5)$$