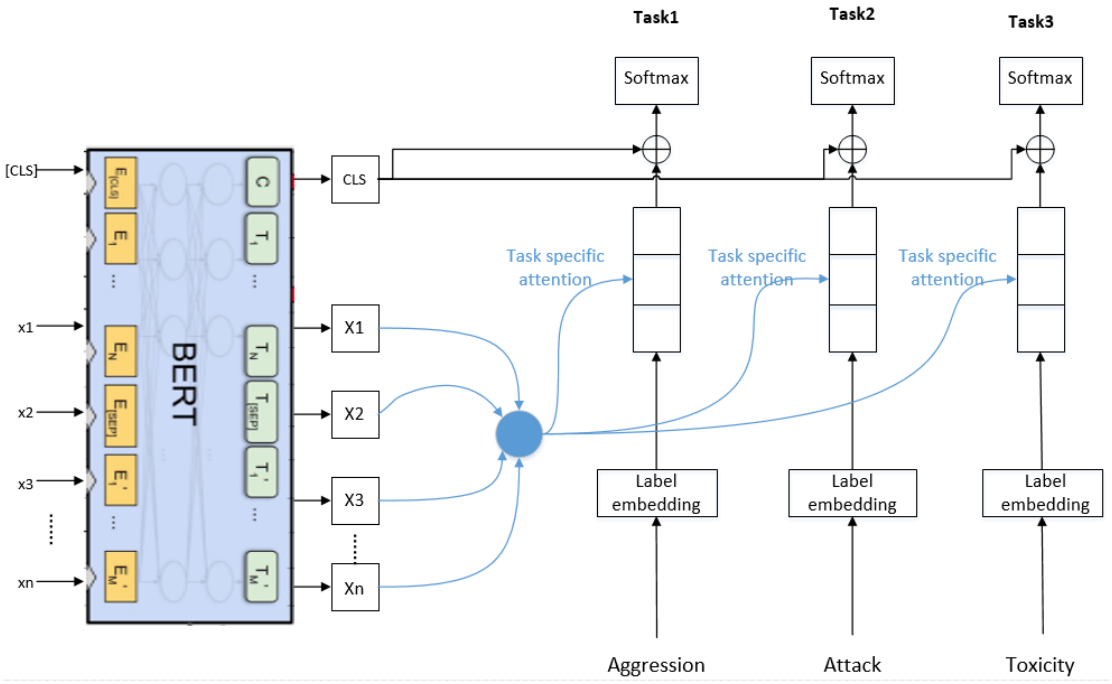

模型说明：

1. BERT 作为 encoder，得到输入单词的特征表示和句子级[cls]表示
2. (1) 将 Aggression,Attack,Toxicity 三个类别分别做 label embedding
   (2) 对于每个 task,将 label embedding 和单词特征表示做 attention,得到c_i

$$c_i = \sum_{j=1}^{n} \alpha_{ij} h_j$$

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{n} \exp(e_{ik})}$$

$$e_{ij} = v^T \tanh(W L_s + U h_j)$$

   其中，h_j为 encoder 输出，L_s为 label embedding
3. 对于每个任务，将不同任务得到的c_i分别与[cls]表示拼接，之后 softmax 进行分类