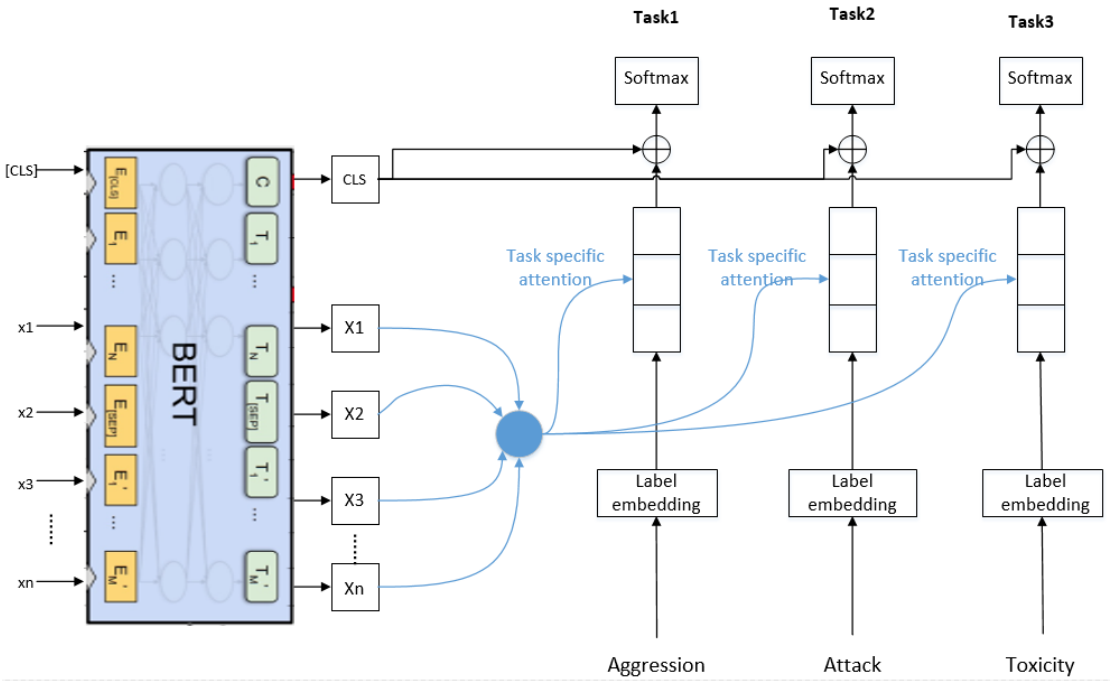

模型说明：

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

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

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

$$e_{ij} = v^T \tanh(WL_s + Uh_j)$$

   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
4. Loss=0.3*task1_loss+0.3*task2_loss+0.3*task3_loss

之前论文中的结果

	Acc	P	R	F1	AUC
ST ATT	.888	.544	.725	.615	.754
ST AGG	.866	.628	.638	.627	.752
ST TOX	.884	.611	.710	.652	.769
LSTM ATT	.897	.626	.754	.672	**.804**
LSTM AGG	.891	.611	**.735**	.655	.780
LSTM TOX	**.909**	**.669**	.694	.666	.794
MT ATT	**.903**	**.610**	**.800**	**.692**	.787
MT AGG	**.894**	**.645**	.714	**.678**	**.783**
MT TOX	.905	.626	**.753**	**.683**	**.798**

Bert Single task score

	Accuracy	Precision	Recall	F1
aggression	0.924	0.786	0.786	0.787
attack	0.945	0.808	0.793	0.800
toxicity	0.940	0.823	0.837	0.830

Bert Multi task score

	Accuracy	Precision	Recall	F1
aggression	0.930	0.830	0.763	0.795
attack	0.938	0.814	0.808	0.811
toxicity	0.943	0.861	0.802	0.831