# Attention classifier

## 公式

分类器的输入为encoder层的输出，为句子的长度。分类器先将输入到LSTM模型中得到，即。

然后开始利用attention模型，attention模型分为两步：一是在所有输入信息上计算注意力分布，二是根据注意力分布来计算输入信息的加权平均。

**计算注意力分布**：给定一个和任务相关的查询向量**q**，我们用注意力，来表示被选择信息得索引位置，即表示选择了第个输入向量。那么给定和下，选择第个输入向量的概率

其中为注意力分布，为注意力打分函数。

**计算加权平均：**注意力分布 可以解释为在给定任务相关的查询时，第个输入向量受关注的程度。然后我们使用加权平均对信息进行汇总：

然后对于我们的多分类问题，需要让输出可以分类出零售商、生产商、其他结果，我们将输入到一个输出为3维的全连接层，并用softmax激活函数来得到每一类的概率。

那么最终得到每一类的概率（用表示第k类，MLP表示全连接层）

最后用分类交叉熵的损失函数来训练网络。