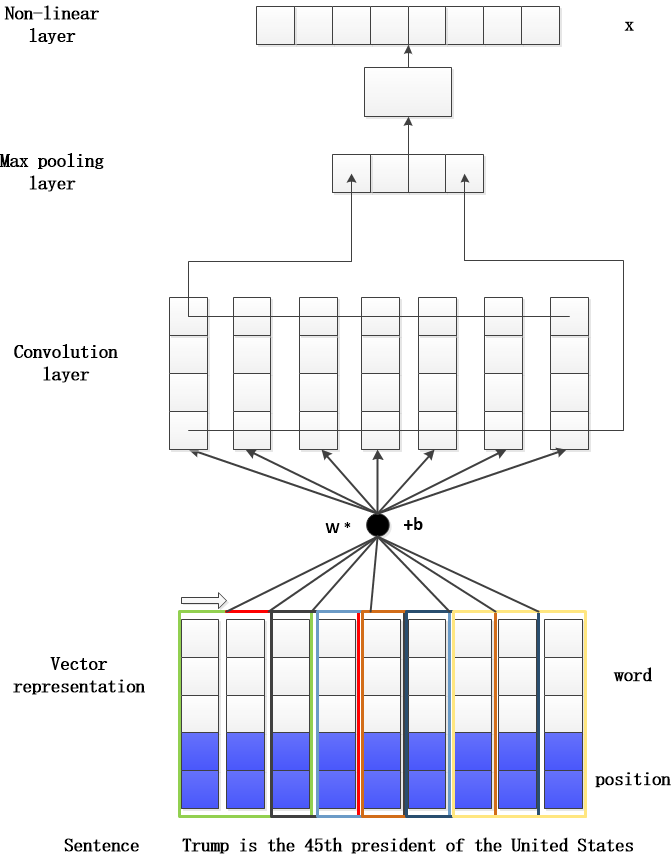
3 方法论述

本模型的基本任务是根据实体对（e1, e2）以及包含该实体对的句子集合{s1,s2,…,sn}预测出该实体对之间关于每个关系r的分布，下面介绍该模型由两个主要部分：

* **句子编码器** 通过卷积神经网络（CNN）将目标实体对相关的句子s编码成向量x的分布式表示。
* **基于实例的句子选择机制** 基于句子编码器得到的句子向量表达形式，使用句子级别的注意力选择机制去选择最能表达相关关系的句子。

3.1 句子编码器



如上图所示，通过卷积神经网络将句子转化为x的向量表示形式，首先我们将句子中的每个单词转化为特征向量，然后经过卷积层、池化层以及非线性转化层提取句子的关键特征最后形成关于该句子的低维向量表式。

3.1.1 输入形式

模型的输入是字符串形式的句子s,所以我们要根据预先学习好的单词编码矩阵，将字符串转化为对应的数字矩阵形式，除此之外，我们还要输入每个单子在句子中的位置信息，同样需要位置编码矩阵。

句子编码矩阵：该矩阵V，其中V代表所需单词的总个数，我们预先通过神经网络在训练语料库上进行学习，最后为语料库中的每个单词学习到能够体现对应单词语义特征的向量形式，例如句子s包括m个单词s = {w1,w2,…,wm} ,每个单词w最终会学习到一个向量的表达形式。

位置编码器矩阵：在关系抽取中，每个词在句子中的位置信息是非常重要的，而卷积神经网络不会关注文本中单词的位置信息，所以我们像（Zeng et al., 2014）一样，通过使用每个单词距离头尾实体的距离作为单词位置信息，帮助卷积神经网络捕捉到每个单词距离两个实体的相对距离，让它更能集中于位置信息关键的单词。例如 Trump is the 45th president of the United States，其中president距离头尾实体（Trump，United States）的距离分别为4和3，在图 中，假设dw是3并且dp是1，然后将句子中每个单词的编码与其对应的位置信息编码拼接在一起，最后每个单子的向量表示w={w1,w2,…wm},wi其中d= dw+2\*dp

3.1.2 卷积，池化以及非线性层

关系抽取的主要面临的问题是每个句子的长度不一样，这就导致句子的主要特征往往会出现在句子的不同位置，然而，通过卷积神经网络，我们可以有效的将文本的不同位置的特征结合在一起，首先卷积神经网络通过滑动窗口k，每次滑动一个单位距离，依次抽取句子的不同位置特征，然后融合在一起，在图中我们假设k=3,最后通过池化层从上述所有的句子的特征中形成句子s的向量形式x.

卷积操作其实就是向量序列w与卷积矩阵W ，其中df是最终句子s的向量形式x的大小同时也是过滤器的大小，为了更好了解卷积的过程，定义ti，代表第i个窗口中，k个单词的向量形式的拼接：

（ ）

在实际实验中，滑动窗口的边界可能会超越句子长度，我们的一般解决办法是在句子的两端用零填充，这样不仅可以保证不会越界，也会抽取出的特征融合长度与句子长度一致，因此第i个过滤器抽取的句子特征可以计算为：



其中b代表偏执向量，最终句子的向量表达x中第i个元素的计算方式如下：



3.2 基于实例的选择机制

先令集合S代表包含实体对（e1，e2）的句子集合即S={x1,x2,…,xm},其中xi为句子的最终向量形式。

为了利用到包中所有句子的特征，并且尽可能减少噪声标签的影响，我们首先利用每个句子的最终向量形式xi来计算关于该句子表达关系r的程度大小，具体计算方法如下：



其中ei代表句子xi关于关系r的得分函数，具体计算方法如下：

ei = xi**Ar**

其中A作为被训练的评分矩阵，训练标签**r**从A中查询该关系对应的向量然后与xi做乘积得到xi关于**r**评分ei，通过计算S中每个句子xi的评分，将包中每个句子通过评分权重相加得到S的向量形式**s,**计算如下：



通过选择机制我们可以尽量避免S中噪声对模型的影响，通过不断地学习，使得S中表达关系**r**的句子的评分相对高一些，最后通过softmax函数计算条件概率，



其中nr是总的关系数量，**o**是神经网络的输出，计算如下：

**o**= **M**s + d

其中，M是关系矩阵。

（Zeng et al., 2015）基于假设：每对实体对的句子集合中至少有一个句子表达实体之间的关系，就是只用S中的最有可能表达目标关系句子，从另一个角度来看，他们的方法在多实例学习中是我们选择机制的一种特殊情况，即评分高的句子分数权重等于1，其它为0的情况。

3.3 优化与实现细节

我们首先定义目标函数：



其中s是实体对的总个数，是我们要优化的参数，针对优化问题，我们采用随机梯度优化算法去减少损失函数，在训练过程中，我们随机从训练集中选择batch来训练模型，同时我们在输出层使用dropout函数来防止过拟合，具体实现如下：



其中f为满足伯努利分布的向量， 代表向量之间的点乘，s向量中元素以概率p变为0，这样可以有效地防止过拟合。

4. 实验

我们的实验主要是为了证明我们的神经网络模型可以充分的利用语句特征并且能够减少噪声标签带来的影响，下面我们首先介绍我们模型的数据集和评估项，然后我们利用十字交叉验证来跟新模型参数，最后我们与其他模型准确率与召回率上进行了比骄傲。

4.1数据集与评估项

我们评估模型用到的数据集是由(Riedel et al.,2010)提供的，该数据集在远监督关系抽取任务中被广泛的使用，该数据集的形成过程主要是通过对齐Freebase数据库与纽约时报，利用斯坦福大学提过的命名实体识别工具识别出句子中的实体，然后进一步与Freebase进行匹配。Freebase中的关系被分成两个部分，一部分用来训练，一部分用来测试。训练集主要是用到2005-2006年的纽约时报作为语料库，测试集主要使用2007年的纽约时报，一共存在53个关系，其中存在一个NA，代表实体对之间不存在关系或者存在关系却不在剩余的的52个关系中，最后形成的的训练集一共包括522611个句子，281270个实体对，以及18252个关系实例，测试集包括172448个句子，96678个实体对，以及1950个关系实例。

与(Mintz et al., 2009)工作相似，我们评估我们的模型在held-out评估集上，主要是将我们从测试集中发现实体关系下与Freebase中的关系比较，所以held-out评估集在不需要人工标注的情况下是近似准确，我们在我们的实验中提供了准确率和召回率曲线以及前N准确率。

4.2 实验设置

4.2.1单词编码

在我们的文章中。我们用word2vec工具在纽约时报上训练单词编码，我们只训练出现次数超过100次的单词，除此之外。，我们会将文本中出现的多个单词表示的实体看作一个词来训练。

4.4.2 参数设置

Window size k = 3

Sentence embedding size df = 230

Word dimension dw = 50

Position dimension dp=5

Batch size B = 160

Learning rate  = 0.5

Dropout probability p = 0.5

4.3基于句子级别的选择机制

为了证明我们模型的的影响，我们通过held-out评估比较我们的模型优于其它方法，我们选择(Zeng et al.,2014)提出的CNN模型作为我们的句子级别的编码器，分别比较CNN+ONE、CNN+AVE、CNN+ATT，第一种是基于至少有一个句子表达实体间关系假设前提，在(Zeng et al., 2015)中使用过，第二种是将所有的句子平等看待，取均值作为S的向量表示形式。

从图中我们可以观察到：（1）对于CNN模型来说，通过比较CNN+one我们知道直接预测句子的标签这样会使得噪声标签对模型影响较大，而只取一个的思想可以有效的减少噪声标签的影响（2）AVE方法比CNN有很大的提高，而且还要略高于ONE方法，这也说明了Ave在降噪上的作用要强于ONE方法，所以结合跟多句子的方法是对模型有用的；（3）AVE方法比ONE方法提高不是很大，这也说明了虽然AVE获取了更多句子的信息，但也引进了大量的噪声数据；（4）ATT方法相比较其他几种方法实现了较高的准确率，并且可以看出ATT的表现要远远高于ATT，说明ATT在引进更多句子信息的同时也减少了噪声数据的影响。我们分别例举了第100、200、300时数据的准确率：