基于双向attention flow的机器阅读理解

摘要

机器阅读理解（Machine Comprehension）的实质是要求计算机能够深入理解所给文章内容，并能够正确回答一系列基于所给文章的问题。在近期的相关研究中，attnetion机制已经被成功运用到了机器阅读理解领域中，并且取得了很好的效果。具体而言，这种方法运用attention的方法将文章（或者问题）转换成一个向量。目前绝大多数的attention机制都是单向的attention。这篇文章中，我们引入了带有双向attention flow(BIDAF)的人工神经网络。这种网络结构能够实现一段文本在不同层面粒度上的表示并且能够运用双向attention机制实现一个针对特定问题的文本表示。我们的实验结果表明我们的模型在斯坦福问答数据集（SQuAD）和CNN/Daily Mail数据集上实现了最好效果。

1. 机器阅读理解介绍

机器阅读理解与问答系统近些年来在自然语言处理和计算机视觉领域获得了很高的关注度。基于端到端训练的系统在许多领域都取得了十分不错的效果，包括文本与图像领域。之所能取得这么大的成果，很大一部原因在于这些模型使用了神经网络attention机制。在机器阅读理解领域，这种机制能够使系统着重关注自然语句的某些部分，在图像处理领域，则能够关注图片的某些部分(Weston et al., 2015; Antol et al., 2015; Xiong et al., 2016a)。attention机制主要有几个特点：一是经过计算得到的attention权重能够用来对每一个句子进行编码，将其变成一个向量。比如从一个语段中抽取出最关键的信息。二是attention的权重是动态调整的，当前时间步的attention权重是基于之前时间步的attention向量的函数。第三，attention机制一般是单项传递的，即基于问题在原文文章中施加attention。

在这篇论文中，我们引入基于双向attnetion Flow的神经网络结构（BIDAF）。BIDAF包括了字符、单词和段落三个层面的编码，并且通过一个双向的attnetion flow对所给文章进行基于问题的编码表示。我们对之前十分流行的attention机制的实现方式提出了一些改进。首先，我们的attention层并不是简单地将一段文本概括成一个向量，而是每一个时间步都计算attention，而且添加了attention的向量以及从前面的神经网络层生成的对文本的表示能够在每一个时间步在RNN中流动，这样就减少了在早期概括一段文本时的信息损失。第二，我们使用了一种不依赖记忆的attention机制，也就是说每一个时间步的attnetion计算只依赖于我们的问题文本与原文文本，而与之前时间步计算出的attention无关。我们认为这样的简化能够将attention层和modeling层的工作分开，这样就使得attention层专注于学习问题与原文之间的attention表示。同样能够使得最新计算出的attention能够避免受到之前时间步可能存在错误的attention的影响。我们的实验结果表明这种不依赖记忆的attention机制明显优于动态attention机制。第三，我们计算两个方向的attention，既包括问题到原文的，也包括原文到问题的，这样就使得双方能为彼此提供更多的信息。

我们的模型在斯坦福公开问答数据集上取得了我们提交时刻的最好结果，超过了之前的所有方法该模型同样在CNN/DailyMail数据集上取得了最好的结果。另外，我们还在斯坦福公开问答数据集上深入研究了我们的模型，分析了我们的模型相对于传统机器阅读模型表现上的不同。

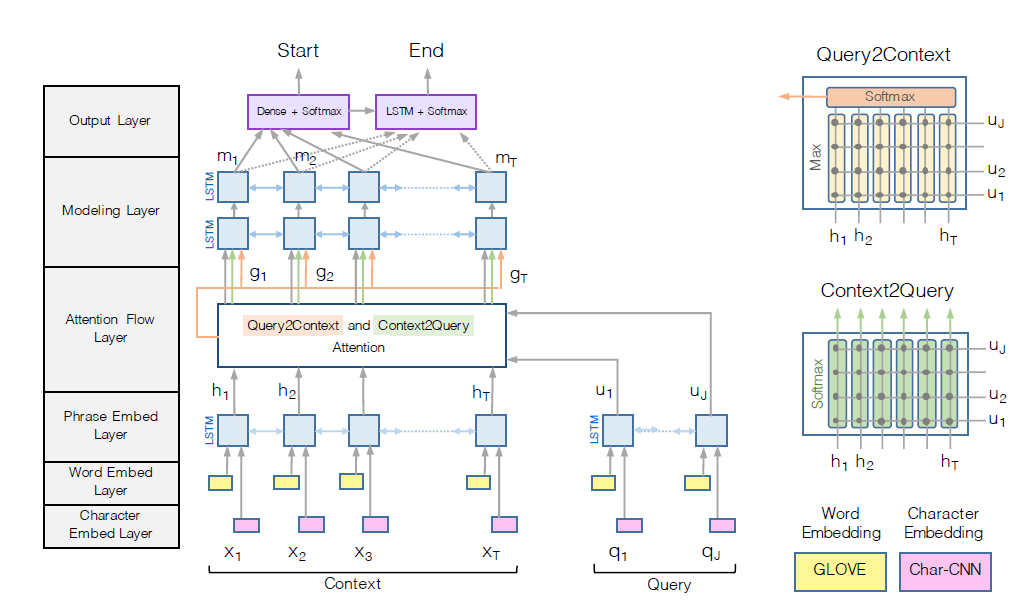


图1：双向attention Flow 模型

2. 模型介绍

我们的机器阅读理解模型是一个多阶段、多层次的计算模型，共分为六个层次（如图1所示）：

1. 字符编码层使用字符层次的卷积神经网络（CNN）将每一个单词映射到一个向量空间。
2. 词语编码层使用预先训练过的单词编码模型将每个单词映射到一个向量空间。
3. 段落编码层使用上下文信息对词语编码做出调整和提升。前三层同时运用到问题文本与原文文本中。
4. Attention Flow层将问题编码向量与原文编码向量组合起来并产生一系列对问题敏感的特征向量
5. 建模层使用一个循环神经网络（RNN）来扫描整个原文文本
6. 输出层将针对问题的答案输出

2.1 字符编码层。字符编码层的作用是将每一个单词映射到一个高维的向量空间。我们用 和 分别表示输入的原文文本和问题文本。我们对每一个单词使用均积神经网络（CNN）进行字符级别编码。字符可以看作CNN的一维输入

2.2 词语编码层