**基于双注意力机制神经网络的**

**句子相似度识别的生成代码框架**

1.case需求的关键词描述，相当于：

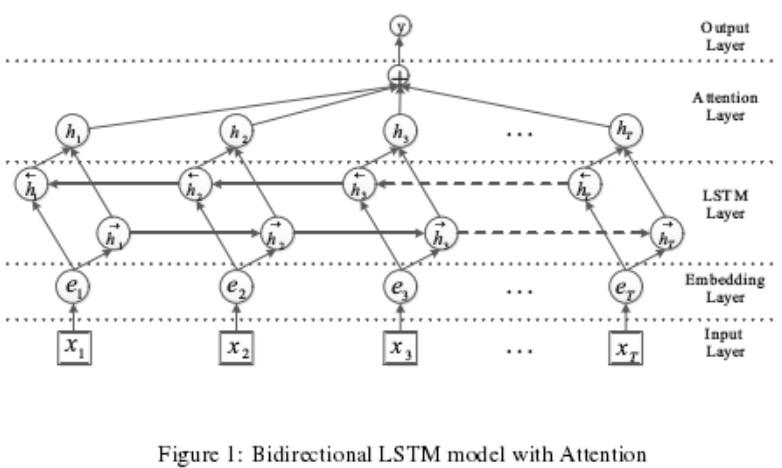
多事例描述（比如：一个事件有5个代码和5个描述）对应于相应的case的类别（比如信号，文件系统.....）

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **TAGS** | **事例描述** | **代码实现** | **类别** | **代码** |
|  | 在发送异步请求时，当希望能够停止异步I/O，则调用aio\_cancel | int fd, ret;  char buf[64] = { 0 };  struct aiocb my\_aiocb; | 信号 |  |
|  | 停止异步I/O（2） | ....... | 信号 |  |
|  | 停止异步I/O（3） | ...... | 信号 |  |
|  | 停止异步I/O（4） | ...... | 信号 |  |
|  | 停止异步I/O（5） | ...... | 信号 |  |

2.利用神经网络进行训练（RNN，基于word-level的实现）

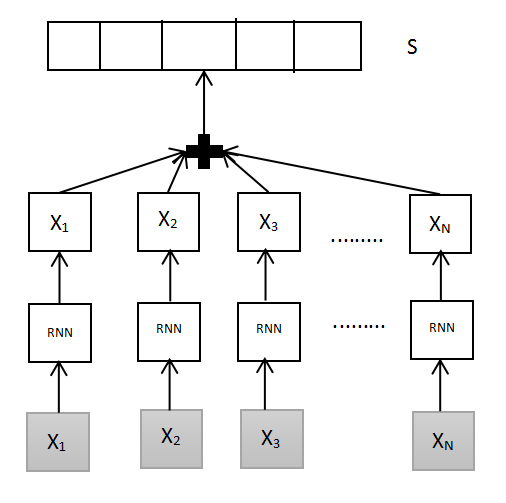
1. 把多个事例转化为自然语言，词向量作为输入
2. Case的类别作为标签
3. Attention-layer，把word合并和句子级别

其神经网络图如下：



3.基于句子级别（sentence-level）再用一次神经网络

1. 多事例（也即多个句子）合并
2. RNN神经网络
3. Attention-layer



4.当用户输入需求然后得到相应类别

5.在该类别中，首先利用用户输入与该类别下的多事例做句子相似度匹配

6.得到多事例排名，最后得到代码

参考的文献：《Neural Relation Extraction with Selective Attention over Instances 》(2016)

分工：

**1.收集case的描述和其类别。（王翔，李继昌）**

举例：相同事例的共5个，其类别一样。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **TAGS** | **Case描述** | **Api** | **类别** | **代码** |
|  | 读文件操作1 | read | 文件与系统 |  |
|  | 读文件操作2 | read | 文件与系统 |  |
|  | 读文件操作3 | read | 文件与系统 |  |
|  | 读文件操作4 | read | 文件与系统 |  |
|  | 读文件操作5 | read | 文件与系统 |  |

2.实现预处理（参考别人的代码：TensorFlow-NRE）（李洛勤）

3.神经网络实现算法（算法模型时：TensorFlow-NRE-master）（李洛勤）

4.关键词提取（词语相似度，跟TAG匹配，设阈值）（谢锦松）

5.实现句子相似度识别算法（找出一个句子和另外不同句子的相似度）（叶龙）

6.前台页面的构建（有输入和输出）（李林杰）、UI设计（郭美圆）

1. 说明文档（项目文档）（李继昌、王翔）

8.时间安排（郭美圆）