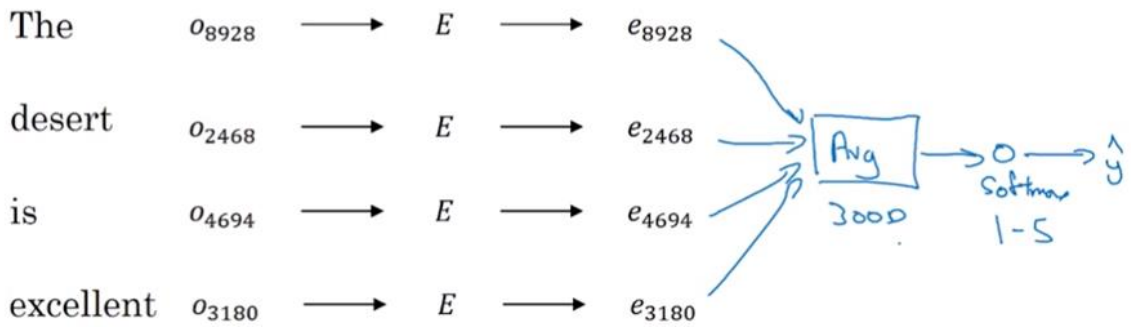
1、情感分类

之前学习了词嵌入，其目的是更好的提高模型泛化性（因为不同词汇之间有了特征上的关联），同时学习了 word2vec、skip-gram 以及 glove 算法，更好的解决了常见与非常见词频问题和预测效率低下问题。而本次的实战环节就是通过之前学的内容做一个情感预测问题，具体而言，输入一段文字，然后给出该语句的评分等级。

2、神经网络架构

架构 1：

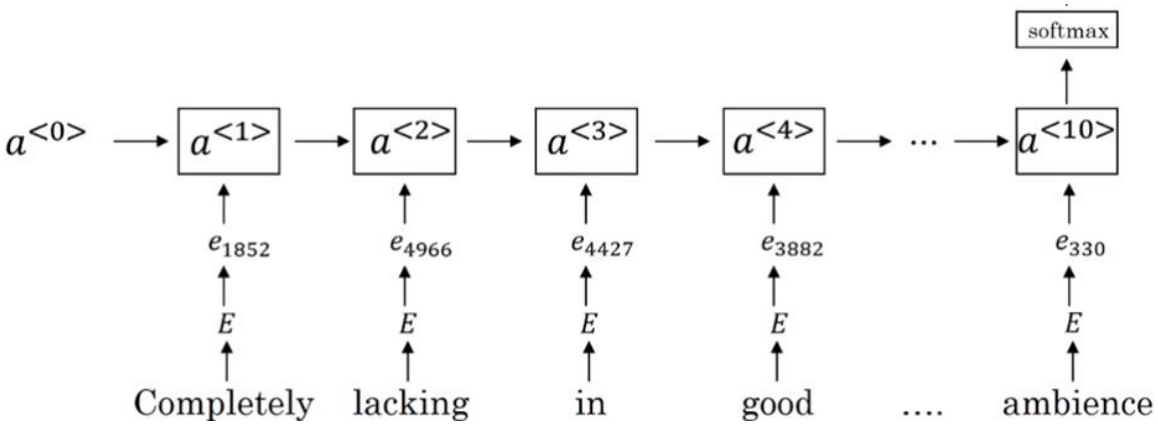
一对一神经网络，如图所示：



该方式是将单词的 embed 编码求平均，然后输入到一个 softmax 预测分类网络里面，优点：算法简单，适用于任意长度文字；缺点：缺失了语句顺序特性，比如有很多 not good，但是该方法输入时会保留很多 good 特征，分类器极大可能误认为就是 good。

架构 2：

多对一 RNN 神经网络，如图所示：



该方式是将语句中所有的单词输入，通过一个 RNN 网络，将信息不断流向下游，最终通过 softmax 预测情感分类。优点：泛化性增强，适用于单词、顺序改变的情况，例如将：lacking of 变更为 obsent of，网络也能正确的预测出含义。缺点：结构复杂，计算次数和输入语句长度有关。

3、机器决策与社会问题

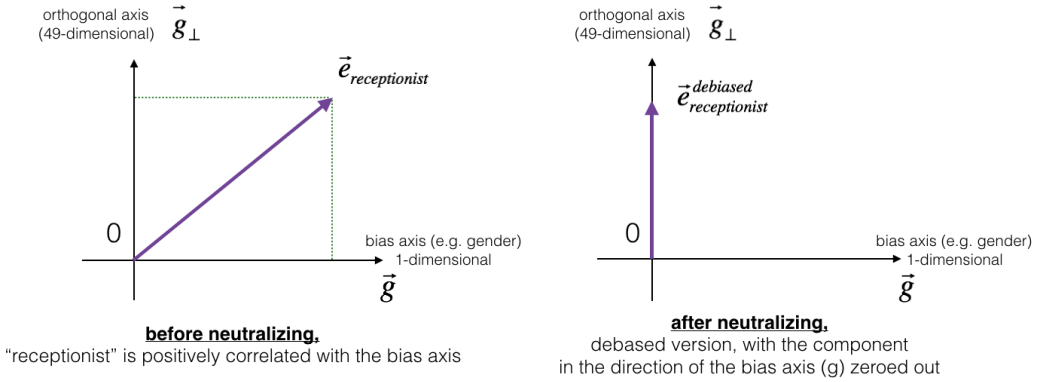
在 NLP 领域中，一个社会伦理问题就是算法学习过程中极易受训练样本的影响，一个简答的例子，在性别方面，通常男性和 doctor、computer_Programmer 出现，而女性经常和 nurse、homemaker 出现，这就导致了在利用 NLP 进行决策的时候，会出现性别歧视方面的问题，男性更趋向于技术性劳动，女性更趋向于重复性劳动，这是不应当出现的。通过 g=Ewoman−Eman，可以得到指向女性词汇的嵌入特征编码，而根据余弦相似度计算模型我们可以看到 guns、science、technology 更接近于男性，arts，teacher 更接近于女性，这是相当可怕的。

解决该问题的方案分为两个步骤：

A、消除与性别无关的词汇的偏差（中和）

B、性别词汇的均衡算法（均衡）

中和：通过将非性别指代词进行性别上的中和，我们可以使该类词汇不再具有很强

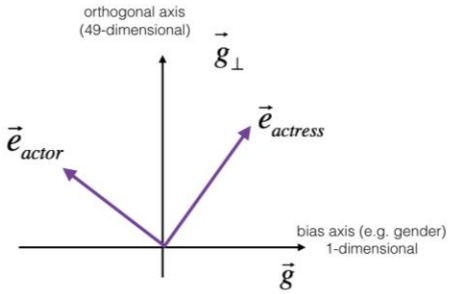


$$e^{bias\_component} = \frac{e \cdot g}{||g||_2^2} * g$$

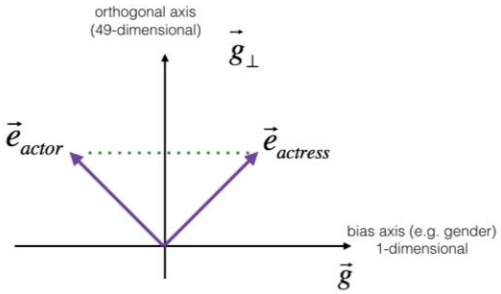
$$e^{debiased} = e - e^{bias\_component}$$

的性别内涵指代，手段是通过词向量 - 词向量在性别方向上的投影，

均衡：通过中和算法，已经将非性别指代词汇进行了性别消偏，比如男演员和女演员，与保姆一词更接近的是女演员，之前虽然进行了保姆一词的性别消偏，但还是无法保证男演员与女演员与保姆一词的距离，具体的原理是消除男演员和女演员在其他维度上的偏差，原理如图所示：



before equalizing,
"actress" and "actor" differ
in many ways beyond the
direction of $\vec{g}$



after equalizing,
"actress" and "actor" differ
only in the direction of $\vec{g}$, and further
are equal in distance from $\vec{g}_\perp$

这样一来男演员和女演员就离保姆一词一样近了。

经过以上中和和均衡以后，男演员和女演员与保姆的余弦相似度就一样了。