

# 机器如何理解人的语言

机器能理解人类的语言吗？很难。

人工智能的创始人图灵在设计图灵测试时给出的智能标准即是对语言的理解和运用。他认为，如果一个人和计算机进行语言交流，如果无法判断出对方是人还是机器，那就说明机器拥有了足够的智能。可见语言在代表人类智能方面处于非常重要的地位。

## 1. 语言理解为什么如此困难

为什么理解语言这么困难呢？这是因为语言具有极高的信息量，要理解一句话需要大量基础知识和常识经验。例如下面一句话：

**“马拉多纳利用空气动力学原理踢出了香蕉球”**

要理解这句话，首先我们得知道“马拉多纳”是个球星的名字，“香蕉球”是一种弧形的球，“空气动力学”是一门关于空气运动的学问，然后会再把这些概念串起来，才理解了，哦，原来是一个人踢了一脚厉害的球。如果没有必要的知识储备和语法经验，我们是无法从这串符号中把这些事情解析出来的。同时，语言又非常复杂，一件事有多种说法，同一串符号又可能表示不同意思，新的词汇不断产生，不同语言之间互相渗透...对于这样一个充满不确定性、动态的符号系统，理解起来是非常困难的。

## 2. 基于分析的理解

如何让计算机理解人的语言呢？人们首先想到的是模仿语言学家对句子进行拆分，分析出主要成分以及成分之间的关系，这种方法可以称为“结构分析法”。一种分析方法是找到句子的中心动作以及动作的主体和目标，形成一个“动作（主体，目标）”这样的三元组。例如，可以将“我有一辆自行车”表示成如下三元组：

**拥有（我，自行车）**

这个三元组代表了原句子的主要语义信息。基于三元组来表达语义简洁清晰，在知识图谱中有广泛应用。更复杂的语义分析方法包括构造语义依存树等。

## 3. 基于语义嵌入的理解

对于简单的句子，结构分析法可以实现有效的语义理解，但对于复杂的句子，这种方法还是非常吃力的。首先，对句子结构拆分本身就比较困难，对每个成份的语义角色进行判定也不容易，而确定各个成份的之间关系也有不少问题。近年来，随着深度学习的兴起，基于神经模型的语义理解取得长足进步。这一方法通过大量数据来学习一个词或一句话的语义向量，通常称为“语义嵌入”。

我们以词为例来说明。什么叫一个词的词义呢？传统方法会把一个词表示成一个固定的符号，然后对这个符号加以解释。比如“苹果”，首先是一种水果，然后是圆形，味酸甜，口感脆。有了这些描述，就可以把“苹果”和其他东西区分开。深度学习不理睬这些具体化的描述，而是通过一个神经网络，将每个单词表示成一个向量，称为“词向量”。学习的目的是使得语义相似的词对应的词向量更加接近，而语义无关的词对应的词向量离的更远。这样就相当于把词的语义注入到了对应的词向量中，这一过程称为“语义嵌入”。一个词向量的例子如图 1 所示。

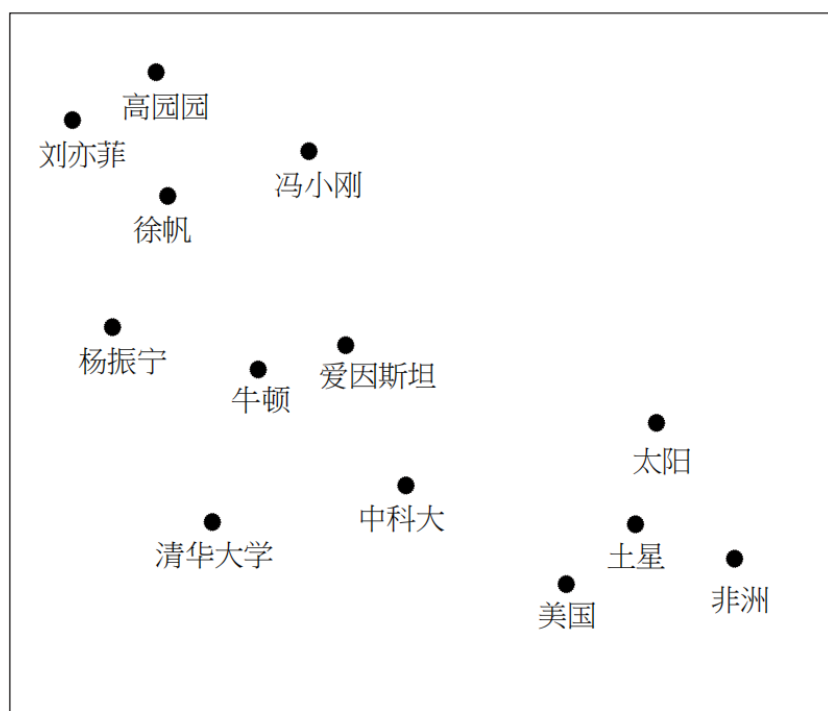


图 1: 词向量的例子 [1]

和词的语义嵌入类似，我们也可以定义句子之间的语义相似性，实现句子级的语义嵌入。图 2 是一个对句子进行嵌入的例子。在这一模型中，我们定义前后相连的句子在语义上具有相似性，因此可以将前一句首先映射成为一个句子向量  $s$ ，再基于  $s$ ，使得生成下一句的概率最大化（这意味着两句话在语义上是相近的）。这一模型可以用于网站上的自动回贴机器人，对用户的发贴给出一些有趣的回复，如图 3 所示。

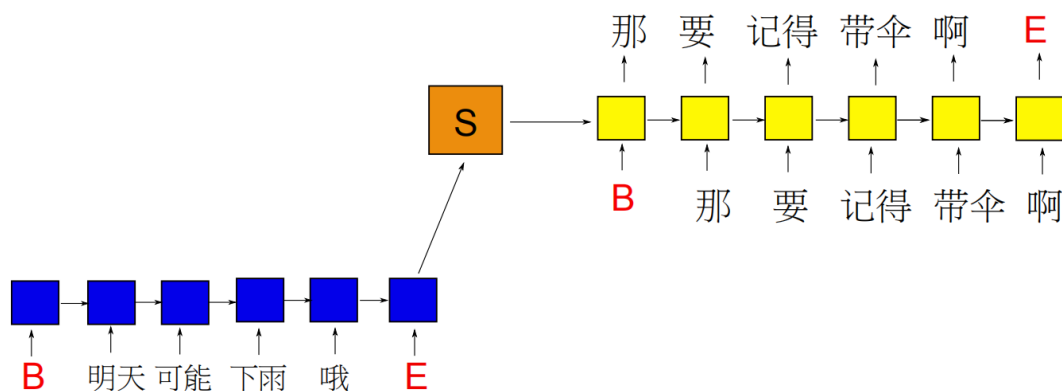


图 2: 句子的语义嵌入 [2]

<b>Post</b>	意大利禁区里老是八个人... 太夸张了 There are always 8 players at the Italian restricted area. Unbelievable!
<b>Response1</b>	我是意大利队的球迷, 等待比赛开始。 I am a big fan of the Italy team, waiting for the football game to start
<b>Response2</b>	意大利的食物太美味了 Italian food is absolutely delicious.
<b>Response3</b>	太夸张了吧! Unbelievable!
<b>Response4</b>	哈哈哈仍然是0:0。还没看到进球。 Ha!Ha!Ha! it is still 0:0, no goal so far.
<b>Response5</b>	这正是意大利式防守足球。 This is exactly the Italian defending style football game.

图 3: 将图 2 的模型用于“自动回帖”, 第 1 行是原帖, 下面 5 行是自动回帖 [2]。

#### 4. 深度学习真的理解了语言吗?

和传统基于结构分析的语义理解方法相比, 基于语义嵌入的理解是一种较为“模糊”的理解。它首先确定一个语义远近的度量标准, 如两个词是否同现在某一环境下, 两句话是否前后相邻等等; 模型训练的任务是使词或句子的语义向量在这一度量标准下与实际观测数据更加一致。计算出这些语义向量显然不能说计算机真的理解了一个词或一句话的意义, 但因为这些向量是基于约定好的语义度量标准训练的, 如果目标任务和该语义度量标准一致, 则这种模糊的“理解”往往可取得非常好的效果。

1. 王东, 利节, 许莎, 《人工智能》第四章, 理解你的语言, 清华大学出版社, 2019.10.
2. Vinyals O, Le Q (2015) A neural conversational model. arXiv preprint arXiv:150605869