

自然语言处理技术——人工智能与数学模型

姓名：舒翔宇 班级：软件中外12班 学号：222021321062016

摘要

自然语言处理（NLP）是人工智能的一个重要分支，为了是利用计算机对人类自然语言的智能化处理。包括语音识别，机器翻译，自动回答，自动摘要等等。人类在人工智能的方面的研究从最开始的“让机器完全理解人类语言”到对语言的统计。从自然语言的研究，我们可以看到人工智能的研究经历了最开始的“走弯路”到现在的利用各种数学模型和统计方法的转变。

关键字

人工智能发展 自然语言处理 数学模型和统计方法

正文

1946年，以冯·诺依曼的ENIAC为标准的现代电子计算机出现后，人类就不断想象计算机能够做到人类无法做到的事情。但是事实却没有像想象中那样美好，计算机不比人类聪明多少，所以说人类又开始不断琢磨，怎么让计算机做到比人类更好。

一、最开始的“弯路”

图灵曾在上世纪50年代便思考机器是否智能，他提出一个判断机器是否智能的一个方法就是：让人与机器交流，如果人无法辨别对方是机器还是人，就说明这个机器拥有了智能。随后的达特茅斯会议，讨论了人工智能的概念、自然语言处理和神经网络等内容。但是人们却走了弯路，当时的科学家普遍认为，要实现人工智能，需要让机器拥有像人类一样的智慧。在自然语言处理上，最开始的研究便是对我们人类语言的语法分析和语义分析上。尽管科学家们在相关研究上花费很多经历和时间，但是所得到的成果十分有限。

二、数学模型的强大

对自然语言的语义分析和语法分析研究持续了几十年后，出现了一种新思路——统计语言模型。这基于的是一个很简单概率事实：一个句子是否合理，就看它的可能性大小如何。于是人们通过对语言学统计的研究，建立了合理的数字模型和统计模型“二元模型”，并且为了处理零概率等不平滑问题，又提出了“古德-图灵估计”。后来人们不断完善机器训练的方法，到后来的隐马尔可夫模型和贾里尼克的现代语言处理，RNN模型。人们通过几个形式简洁的统计语言模型实现了复杂的语言识别，机器翻译等等，在后来的实践中也证明了概率统计数学模型强大的能力。同时对于人工神经网络的研究，基于每一层级的不同弧的加权，训练出强大的人工神经网络，或找出其近似解，在多维空间进行模式分类等，在机器翻译，语音识别上展现出强大的能力。

三、数学与人工智能

人工智能发展到现在人们认可了数学模型在人工智能研究的重要作用。例如如今最火的机器学习，就广泛使用了梯度下降法。梯度下降法是在解决一些线性回归问题上的

一个最优解。我们提到的自然语言处理中的统计数学模型，也需要解决在统计学中的线性回归问题。当然这些形式简单的统计语言模型，里面还涉及到各种各样的学问，二项分布、多项分布、beta 分布、dirichlet 分布，gibbs 采样等，来处理在实际研究中的各种问题。

四、自己的心得体会

最开始作为计算机门外汉，我也一度认为要实现人工智能，就是要让机器拥有和人类一样的智慧，并且认为强大的计算机理所应当地应该可以代替我们去思考。但是开始逐渐了解这些方面后我才明白，真正让计算机拥有智慧的是数学。数学中的许多分支，例如统计学，运筹学，离散数学，线性代数等，构成了如今计算机强大的智慧。而且我们可以看到人们已经在人工智能的研究上取得了巨大的成果。自然语言处理上，我们身边充满了各种翻译软件，语音助手，各种 AI 机器逼真的声音，人们已经写出可以通过图灵测试的程序了。同时，像 alpha GO，谷歌无人驾驶等等，似乎一个“智能地球”正在到来。最本质的是，这是一个“数学地球”。这是一个由“数学规划，优化，最优化理论，决策，统计”构成的世界。这是一门大学问，也是很复杂的一门学科，是数学与工程的结合。虽然是很复杂，但是目前看来，我们也算是走到了一条正确的道路上。现在的比如基于深度学习的“人工神经网络”和“贝叶斯网络”。所以说我们不妨想象一个未来，未来不只是一只会下象棋的狗，而是一个充满了 AGI 的一个世界。

五、总结

在了解自然语言处理的时候，事实上还是很痛苦的。最大的问题在于各种论文中那些夸张的希腊字符和数学公式。不过痛苦归痛苦，这还是很令人兴奋的，至少我们还要那么多东西可以继续深入研究，还要那么多挑战可以面对。我们需继续学习下去，事在人为，休言万般皆是命；境由心造，退后一步自然宽。

参考资料：

- [1] 吴军. 数学之美[M]. 2021(6)
- [2] Ruobing Shen. 运筹学：一门建模、优化、决策的科学[J/OL]. 知乎日报 <https://daily.zhihu.com/story/9360390> . 2021-10-05
- [3] 张志峰 刘洁. 人工智能入门：如何最快下山？让梯度下降算法告诉你. 知乎 <https://zhuanlan.zhihu.com/p/67491592> . 2019-05-30