

谈谈软件与深度学习

阮海航 16231015 计科 1606

关于软件与深度学习，我个人觉得软件是可以学习的，也是具有深度学习功能的。

关于深度学习，深度学习是机器学习的分支，它试图使用包含复杂结构或者由多重非线性变换构成的多个处理层对数据进行高层抽象的算法，简单的说，深度学习就是一个函数集，如此而已。深度学习将特征提取和分类结合到一个框架中，用数据学习特征，是一种可以自动学习特征的方法。深度学习起源于神经网络，但现在已超越了这个框架。至今已有数种深度学习框架，如深度神经网络、卷积神经网络和深度置信网络和递归神经网络等，已被应用计算机视觉、语音识别、自然语言处理、音频识别与生物信息学等领域并获取了极好的效果。深度学习的动机在于建立可以模拟人脑进行分析学习的神经网络，它模仿人脑的机制来解释数据，例如图像、文本和声音等。深度学习通过学习一种深层非线性网络结构，只需简单的网络结构即可实现复杂函数的逼近，并展现了强大的从大量无标注样本集中学习数据集本质特征的能力。深度学习能够获得可更好地表示数据的特征，同时由于模型的层次深（通常有 5 层、6 层，甚至 10 多层的隐层节点，“深”的好处是可以控制隐层节点的数目为输入节点数目的多项式倍而非多达指数倍）、表达能力强，因此有能力表示大规模数据。

关于软件，是基于操作系统运行的程序，软件是可以实现深度学习的，毕竟深度学习是一套算法，软件基于这个算法运行是完全可以的。构建深度学习研究环境首先要考虑硬件环境的支持，这也是深度学习算法区别于其他机器学习算法的重要一点，同时还要依据实际需求对深度学习软件框架进行规划和选择。

比如网络爬虫，网络爬虫是获取各类网站公开数据的有效工具，使用网络爬虫是研究深度学习技术的一项必备技能。网络采集器属于网络爬虫的一种，通过软件的形式实现简单快捷地采集网络上分散的内容，具有很好的内容收集作用，而且不需要技术成本，可被用作为初级的数据采集工具，例如火车采集器、八爪鱼、集搜客等。此外，针对特定网站的特定数据采集需求，也可结合开源系统自行开发爬虫工具，这样具有更高的自由性和自主性。利用爬虫可以获取一些从其它渠道获取不到的数据资源，更重要的是帮助研究者打开寻找和搜集数据的思路。