**阅读*Thirst…*的总结及大数据分析处理的相关想法**

左之睿

（南京大学 人工智能学院 南京 210000）

摘要：众所周知，人工智能的发展离不开对脑内神经的研究、分析，而在人工智能领域，对外界输入产生相应反应与大脑区域和身体器官之间的映射不谋而合。因此，通过对脑内神经元活动及其对活动的影响的研究，不仅在神经学领域是一个重大课题，在AI领域同样意义重要。

关键词：神经科学，动机驱动，人工智能，大数据处理分析

1 对*Thirst*…一文的总结

这篇论文的主要目的是通过观察处于不同动机状态下的动物（小鼠），确定这些状态在大脑中的区域如何表现以及这些状态如何影响大脑中的神经元的活动来改变行为。在作者的实验中，展示了口渴这一动机状态在大脑中的整体表现以及对行动的影响。

实验中，向小鼠植入headbar后让其恢复，之后对其限制供水并让其接受嗅觉“Go/No-Go”任务的训练，用Neuropixels微电极阵列进行记录每个阶段大脑区域神经元的活动。实验中，老鼠会执行几百次实验直至满意（不断喝水直到饱足），并在满意时继续执行几十次实验。在实验过程中，在与水获取有关的持续活动之前，对水预测的嗅觉提示具有快速而广泛的响应。而在具有饱腹感的小鼠中，相同的感觉提示仅产生瞬时的活动变化，而没有行为反应。此外，实验者还以光遗传学方式刺激小鼠下丘脑口渴神经元，并使小鼠达到饱腹感后进行更多试验，发现这种局灶性刺激使得行为以及整个脑内与饱腹状态相关联的单神经元均返回到了对应于口渴的状态。此外，刺激还按照部分大脑区域的有关饱腹感的模式调节了活动。

最终得出了如下结论：整个大脑的干渴动机状态似乎可以控制感觉信息在大脑中的传播及其随后转化为行为输出的过程。动机状态制定了初始条件，而这些条件又决定了全脑动力系统如何将感觉输入转化为行为输出。

2 大数据分析处理的相关想法

2.1数据预处理

得到实验结果时，并不能确保所有的数据都是准确无误的，这时如果盲目使用所有数据进行分析，可能得出的结果并不准确，因此，进行预处理以筛选出较为可靠数据是必要的。预处理又分为四个模块：1、数据清洗；2、数据集成；3、数据变换；4、数据归约（数值规约）。接下来依次讨论。

首先是数据清洗，这一步包括两种情况：缺失值和异常值。缺失值的处理较为简单，直接将其删除，或者填充固定值/均值/中位数/众数等等即可，而异常值的处理则相对麻烦，首先需要对其进行观测，除了简单统计和绘图观测，这里要介绍一种基于聚类的方法。所谓聚类，就是以相似性为基础把相似的对象通过静态分类的方法分成不同组别，其中K-MEANS算法（以空间中k个点为中心进行聚类）较为常用，在实验数据清洗中，只需挑出一些合理数据并规定距离上限，就可以轻松挑选出超出规定范围的异常值。

数据集成则是为了简化一些不必要的属性，缩小数据的维度，比如记录学生身高，往往只需要姓名或者学号及对应身高即可，因此同时出现姓名学号属性时就在数据集成这一步骤中进行简化。

接下来是数据规范化，即把数据变换成合适的形式，常见的方式有规范化、连续属性离散化、特征二值化。其中，规范化可以解决量纲不同或者数值范围差异较大的影响，便于后续绘制、观察图表；连续属性离散化则是将连续的数据近似为离散的点，采用之前提到的聚类就可以做到；特征二值化则更为简单，直接将数据分为0/1，这一方法在机器学习领域很常用。

最后是数据归约：对属性和数值进行规约以获得一个原数据集的小的多的规约表示，但仍接近原数据的完整性，在规约后数据集上进行分析仍然能得到近似的结果。其中属性规约包括：向前选择、向后删除、决策树、PCA等方法；数值规约则可以通过有参方法（回归法，对数线性模型），无参法（直方图，聚类，抽样）完成。

2.2数据可视化

得到了较为可靠的数据，为了对其进行分析，除了进行回归计算等数学方法，对其进行可视化操作也是一个不错的选择。其中较为简单的，则是直接使用数据集绘制直方图、饼图、折线图等，但这样虽然能给人以直观印象，但同时可能无法反映数据深层次的关联。想要在利用图片表示更深层次的信息，可以利用seaborn模块（python语言）对二维数据绘制heatmap（热图），pairplot（多变量图）等等，便于发现数据之间的联系，研究人员也可以据此提出更多猜想或是结论。

1. 结束语

本文大致介绍了对Thirst…一文阅读后的总结，以及有关大数据处理的一些想法。希望神经科学在未来能够为人工智能的发展提供更多的参考以及启发，也希望在未来先进的数据处理方法可以为各学科的科研提供更加准确、高效的数据分析，得出更加精确的结论。