

机器学习有哪些基本方法？

机器学习是实现人工智能的基本方法之一，其基本思路是让机器自动从数据中学习规律和知识。

机器学习的方法有很多，如果考虑学习的基础原则，可以大致将这些方法分为四种：规则方法、概率方法、神经网络方法、进化仿生方法。

规则方法：这种方法来源于传统人工智能系统，这些系统将人类知识表达成可计算的符号系统，如果存在经验数据，则可依这些数据对该符号系统进行学习，包括对现有知识进行排序，从数据中统计抽取出新规则，等等。这种学习通常非常有限，因为过度学习可能会破坏原有系统的精确性。

概率方法：这一方法来源于统计学，将知识形式化为事件之间的概率关系，以这一关系作为基础框架，学习时依实际数据对概率函数的具体参数进行更新。这种方法结合了人类知识和数据的双重信息，理论严谨，高效可靠。

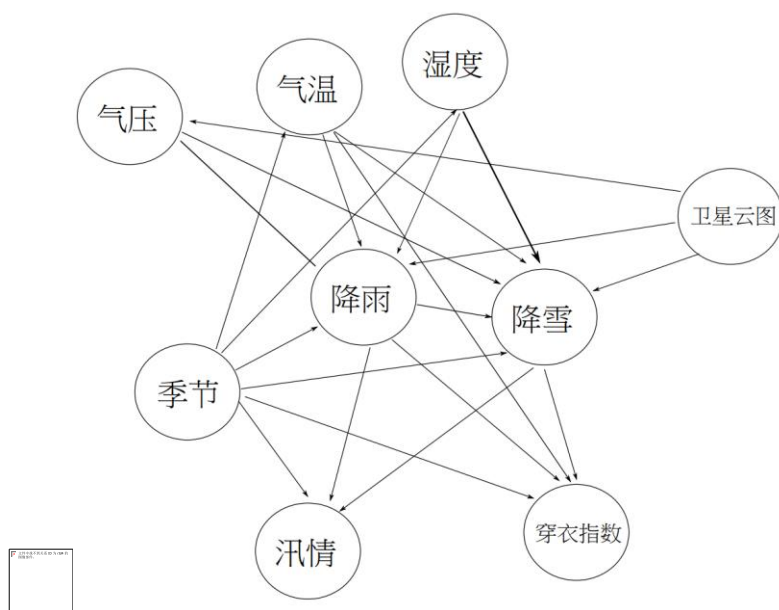


图 1：概率方法将知识形式化为事间之间的概率关系

神经网络方法：这一方法来源于对人类神经系统的模拟，通过互相联接的神经元来实现所需要的功能，并通过调整神经元之间的连接强弱实现学习。近年来大放异彩的深度学习方法就是神经网络方法的特例。和概率方法相比，神经网络方法更灵活，学习能力更强，对数据的依赖也更大。当有大量数据可用时，这一方法有明显优势。例如，深度学习的其成功很大程度上应归功于近年来数据规模的大幅增长，让这种纯数据驱动的方法更有用武之地。

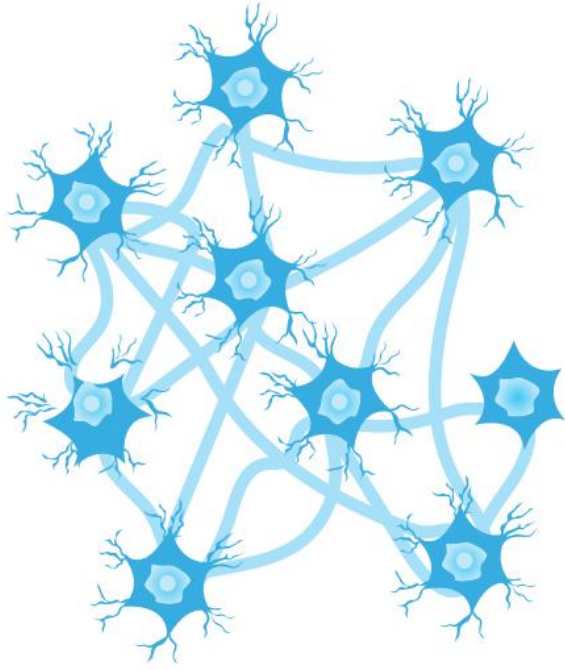


图 2：神经网络利用大量神经元互相连接来实现功能

进化仿生方法：这一方法来源于生物进化理论。生物进化依靠的是自然选择，机器也可以通过这种方法进化。例如，可以尝试让机器设计各种方案来完成目标任务，出色的方案留下，不够好的淘汰。经过长期的选择后，就可以得到优化的方案。

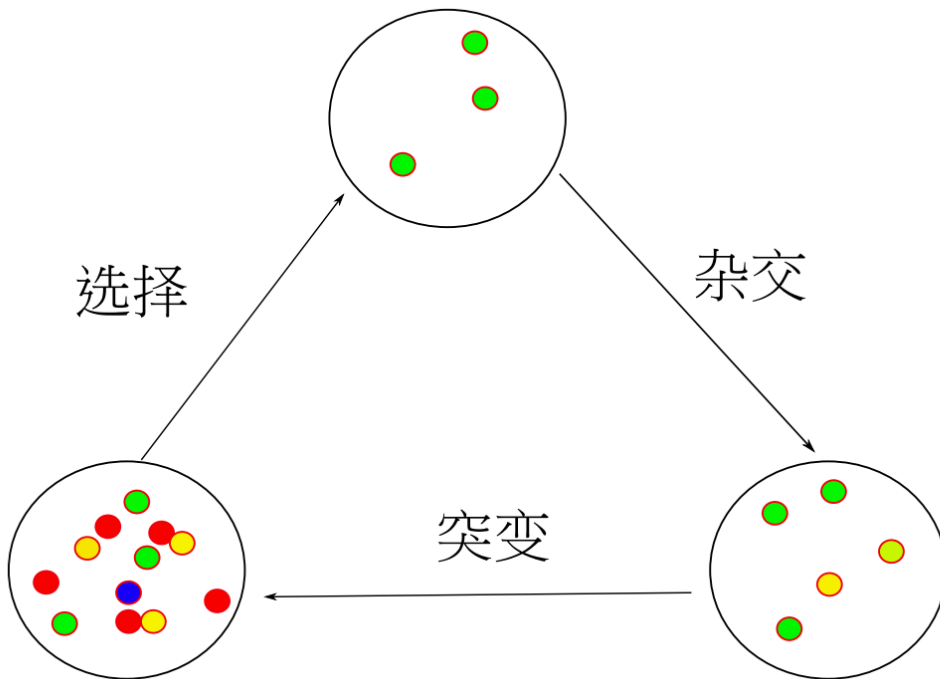


图 3：进化仿生方法模拟生物进化的自然选择原理进行学习

上述这四类方法代表了“让机器自己学习”这件事的四种基本思路。当然，这一分类并不能囊括机器学习里的所有算法；同时，很多方法兼具多重思路，例如当前绝大多数神经网络方法都具有概率解释。

[1] Domingos P (2015) The Master Algorithm: How The Quest for The Ultimate Learning Machine Will Remake Our World. Baisc Books, Inc.

[2] 王东，机器学习导论，清华大学出版社，2021.2