

人工智能如何发现新药

人类使用化学药物的历史可以追溯到 19 世纪，其中阿司匹林是杰出的代表。目前，人们使用的化学药物大约有 4000 种，每年新增 20-30 种。

一种新药的开发极其复杂，其中第一步是寻找具有某种属性的先导化合物。在很长一段时间里，人们基于经验开发先导化合物，成本高，周期长。自上个世纪 90 年代以来，人们就尝试利用人工智能的方法来辅助新药开发，如预测药物的属性和化合过程等。近年来，人们利用各种机器学习方法直接合成新的药物分子，取得了长足进展。

例如，2020 年 5 月 Nature 杂志发表了一篇论文[1]，利用循环神经网络（RNN）模型合成具有某种目标属性的分子。研究者首先将分子式重构成一个从左到右的数字/字母序列，再将该分子的各种属性提取出来，之后建立 RNN 模型，基于这些属性生成对应的数字/字母序列，最后将生成的序列还原成分子结构。

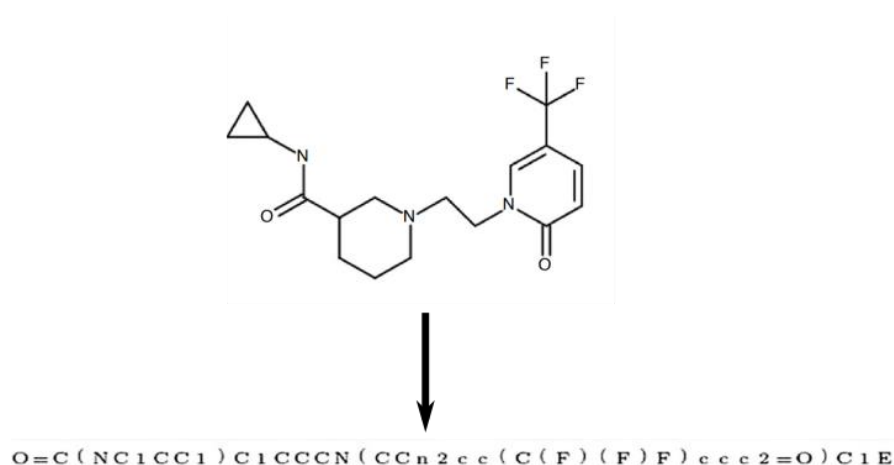


图 1：将药物分子表示成数字/字母序列 [2]

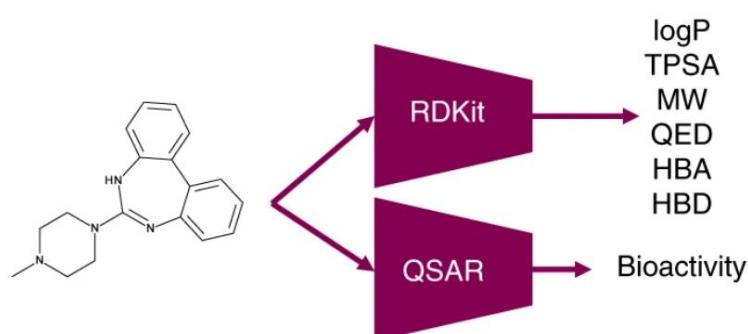


图 2：提取分子属性 [1]

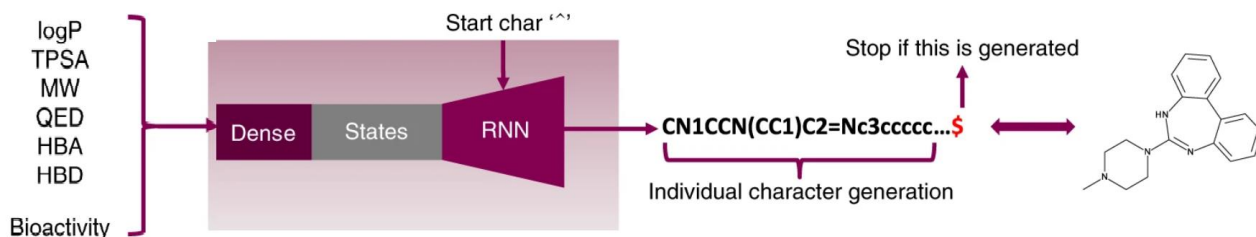


图 3: 建立 RNN 模型, 生成符合属性的分子结构 [1]

值得说明的是, 上述生成过程是随机的, 即由一种属性设定可以生成多种分子结构, 越是符合目标属性的分子, 生成的可能性越大。如图 4 所示, 左右两图分别是基于两种属性设定生成的分子结构。可以看到每种设定都可以生成多种分子结构, 这些结构具有相似属性, 因而都是可能的候选。制药师可以对这些候选进行筛选, 选择可能性较大且成本较低的结构进行试制。这一方法可充分启发研究人员的思路, 从而有效缩短发现新药的时间。

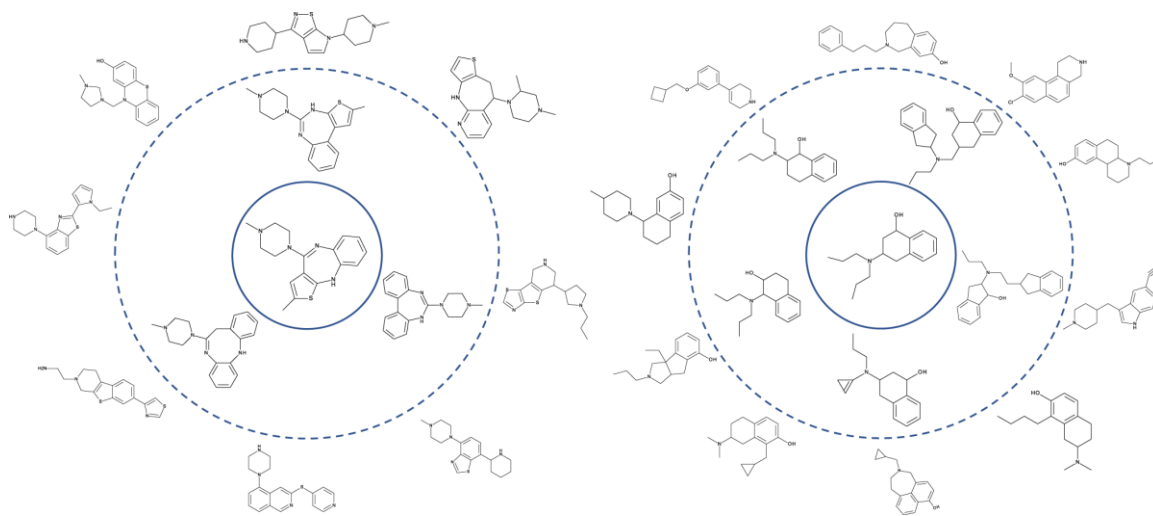


图 5: 由两种不同的属性设定生成的新药候选 [1]

1. Kotsias et al., Direct steering of de novo molecular generation with descriptor conditional recurrent neural networks, Nature machine Intelligence, 2020.
2. Chen et al, The rise of deep learning in drug discovery, Drug Discovery Today, vol23, no. 6, 2018.