

复杂系统和复杂网络中两个相变例子

蔡怀广 202228014628021 模式识别与智能系统

摘要与引言：复杂系统和复杂网络几乎在所有领域广泛存在，在这篇综述中我们通过 SAT 和 Percolation 两个相变问题，展现复杂系统和复杂网络中的典型研究思路。

1. 复杂系统综述

复杂系统的定义： 微观个体的行为是简单的，但是它们非线性交互、数量多，宏观上涌现出规律性的行为。一个典型的复杂系统是原胞自动机，微观上个体仅仅有着有限的行为，但宏观上就会涌现出一些规律的现象。一个复杂系统中经常出现的现象是相变，比如生态学中水华的暴发，物理学中的三态变化。

2. 复杂网络综述

复杂网络在各个研究领域（比如数学、物理、计算机、统计、社会科学等）广泛存在，比如因特网、生化网络、社交网络。复杂网络用到的研究工具也来自很多领域，比如线性代数、谱理论、概率论、组合优化、随机模型、统计和机器学习，还有很多数值方法^[1]。

在推荐系统或者社交网络中，用到的研究工具中理论性强一点的是线性代数、优化理论、图谱理论，理论性弱一点的则是神经网络。前者一般引入“交互是线性的”的假设来降低研究难度，后者交互可以是非线性的（比如图注意力网络）。

在深度学习中，关于对复杂网络建模一个有趣的经验是，假设只有网络中节点的邻接关系，那么用简单的线性交互的模型（图卷积神经网络或者 PageRank）就能达到很好的效果。但如果除了邻接关系，还有节点丰富的特征信息，我们使用非线性的交互、更复杂的模型效果大概率比线性模型好。

3. 两个研究案例

复杂系统和复杂网络的联系是很紧密的。网络为复杂系统的连接或者交互提供了便利的表示工具^[1]。接下来介绍两个相变的例子——SAT 和 Percolation，SAT 可以归属为复杂系统中，Percolation 可以归属为复杂网络中。

3.1 SAT

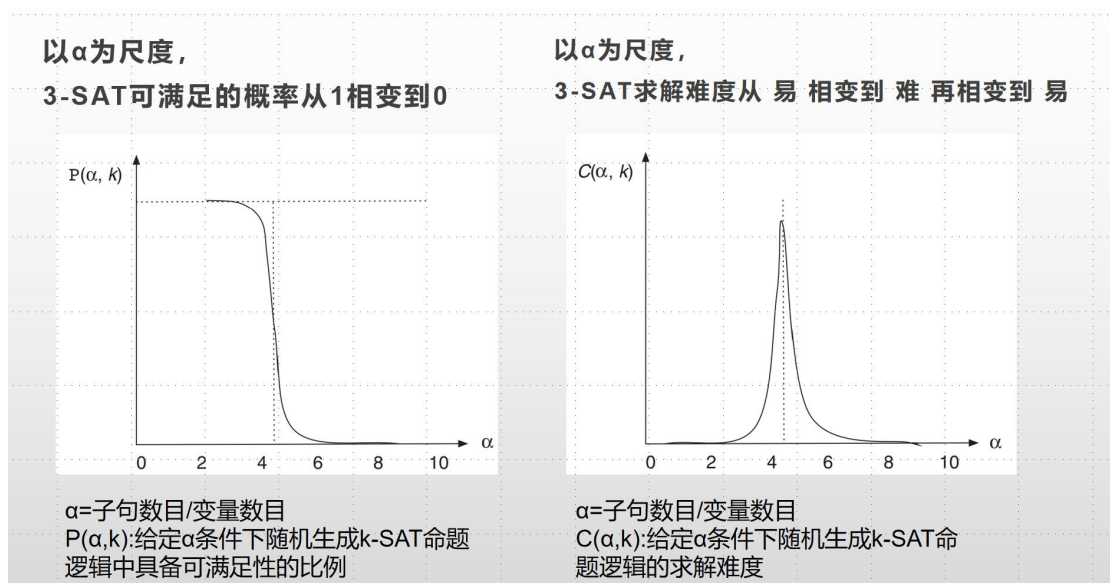
SAT 问题（propositional satisfiability problem, 也称 Boolean satisfiability problem），中文名叫命题逻辑可满足性或者布尔逻辑电路问题，是人们发现的第一个 NP 完全问题，是计算机复杂性理论的最核心问题之一，与世纪 7 大难题中的“P=NP?”有着深刻的联系。

SAT 问题要解决的是判断一个命题逻辑是可满足的还是不可满足的。比如在 $(A \vee B) \wedge C$ 的命题逻辑中，我们可以让 C 取 True，A 和 B 中一个取 True 即可使得 $(A \vee B) \wedge C$ 输出为 True。因此 $(A \vee B) \wedge C$ 是可满足的。而能否通过对 A 赋值使得 $A \wedge \neg A$ 输出为 True？这是不行的，因此 $A \wedge \neg A$ 是不可满足的。值得注意的是，我们关注的不是某个命题逻辑能否被满足，而是给出这个命题逻辑可满足性的判断结果的计算复杂性。

更进一步地说，SAT 中一类问题叫 k -SAT^[2]，指每个子句（用括号括起来表示）中只有 k 个变量。比如 $(A \vee B) \wedge (C \vee \neg A) \wedge \dots$ 就是 2-SAT 问题。以 k 为尺度，问题难度从 P 相变到了 NP。具体来说，1-SAT、2-SAT 都是 P 问题，但是 3-SAT 及以上都是 NP 问题。我们可以将这种相变看做是问题难度（P 还是 NP）随着问题复杂性（用 k 衡量）发生的变化。

我们可以命题逻辑看做是复杂系统，而将每个子句看做是复杂系统中的单个个体。 k -SAT 中（无论 k 取什么值）有着无限的子句，也就是说命题逻辑复杂系统中可以有无限的个体。但是 k -SAT 是否为 NP 仅仅和 k 相关而与子句的数量无关，也就是说命题逻辑复杂系统的整体的某个性质和个体的数量无关、而仅仅和单个个体的性质（也许包含它们之间的交互）相关。

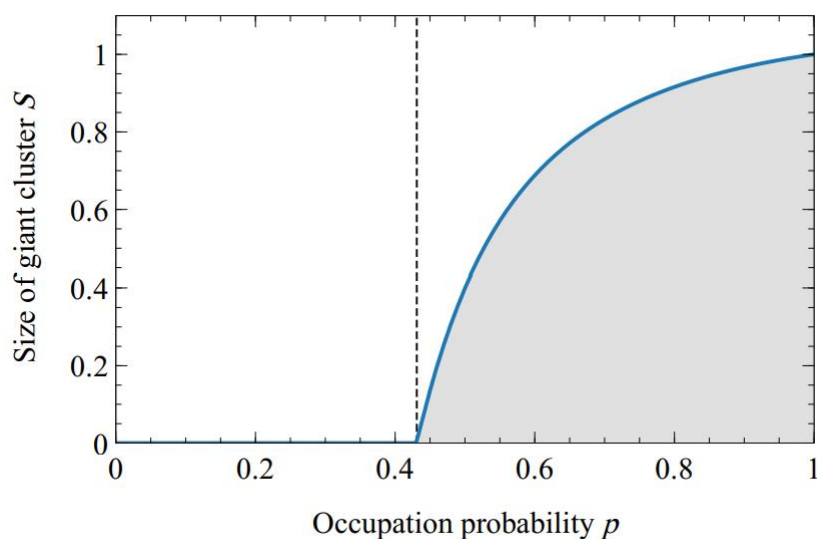
K -SAT 中还有一些相变现象，这里就不详述了，示意图如下。值得注意的是，在坐标 4.3 左右，相变的剧烈程度最大。



3.2 Percolation

另外一个例子是 Percolation^[1]，用于建模疾病在网络中的传播，也用于观察网络对链接失效的情况下的稳定性。具体来说，对于一个网络，我们以 p 的概率随机占据每条边，接着去研究这些被占据的边组成大的连通分支的大小。我们使用 Message passing 的方法来建模这种现象。Message passing 方法认为一个节点的性质是由其邻居传递过来的 Message 来决定，这与深度学习目前流行的图神经网络有异曲同工之妙——几乎所有的图神经网络都可归类于是 Message passing，一方面方便理论建模，另一方面方便计算机的实现（在图上的运算可以不将整个图加载到内存中）。

Percolation 中一个令人惊讶的理论结果是，通过简单的运算，我们可以从理论推导出相变的发生。下面的图描述的是 p 和形成的连通子图大小的关系，一个巧合的结果是相变发生在 0.43。



4. 总结

SAT 和 Percolation 是复杂系统和复杂网络两个相变问题,也代表着两种不同的研究思路。在 SAT 中,我们定义了微观个体(子句)的行为,通过大量实验模拟总结出曲线,发现了相变现象。而在 Percolation 中,我们从理论推导出连通子图的大小,并且可以从理论上预言相变出现的位置。前者是统计的思路,后者是建模的思路。统计的思路的好处是贴合实际,但是对观察到的现象通常不能给出令人信服的解释;建模的思路的好处是通常可以写出清晰的数学表达式,但也引入了人为的假设,此种假设在不同情况下可能不会满足。

任何科学问题后面都存在着一些简明的法则。目前复杂系统和复杂网络的研究还大多处于观察和实验阶段,研究工具也是从其他各个领域汲取,这些简明的法则还较少被人以理论的方式总结出来。

5. 引用

1. M. E. J. Newman(2022).Message passing methods on complex networks. arXiv:2211.05054
2. k-SAT Phase Transitions. <https://www.youtube.com/watch?v=8nh9OTfAHMY>
3. Saitta, L., Sebag, M. (2011). Phase Transitions in Machine Learning. In: Sammut, C., Webb, G.I. (eds) Encyclopedia of Machine Learning. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-30164-8_635