

机器学习如何发现生物拟态证据？

生物进化理论已经得到广泛证明，然而当前研究中的很多方式还是主观性的。如对种属的分类和描述，对进化过程的分析，多依靠研究者自身的知识储备和主观分析，这使得一些理论存在长期争议。随着数据量的积累，生物学家开始使用机器学习工具来寻找更具量化性的、客观的证据来支持自己的理论，近年来取得了很大成功。

一个例子是 2019 年 8 月份发表在 *Science* 杂志上关于生物拟态的论文[2]。早在 1878 年，德国自然学家 **Fritz Muller** 就提出了一个理论，认为生活在同一个地区的某些个体会互相学习，从而实现更好的协同进化[4]。例如，生活在同一个环境下的蝴蝶会互相模仿，使得捕食者不易发现。同样的模仿也发生在毒蛙和珊瑚蛇身上，称为 **Muller** 拟态。**Muller** 拟态的假设是如果捕食者因某种原因（如有毒）放弃对某一类对象的捕猎，则也会因同样的原因放弃对具有拟态特征的其他对象的捕猎。



图 1: [Fritz Müller](#) 的拟态理论，副王峡蝶（上）和黑脉金斑蝶（下）协同进化出同样的特征，包括翅膀花纹和不好的味道[1]。

Muller 的拟态理论是生物进化的第一个数学理论，但一直没有量化证明。*Science* 这篇文章利用一个深度卷积神经网络（称为 **ButterflyNet**），通过蝴蝶的原始图片来预测蝴蝶的一个亚类。它们选用两类蝴蝶，分别称为 **H.Erato** 和 **H.Melpomene**，共 38 个亚类，1234 个标本，2468 张照片。他们用 1500 张照片训练神经网络，训练的目标是识别蝴蝶的亚类。

训练完成后，这一网络可以将任意一个蝴蝶的照片映射到一个向量空间，在该空间中可以直观看到各种蝴蝶之间的相似性，如图 2 所示。

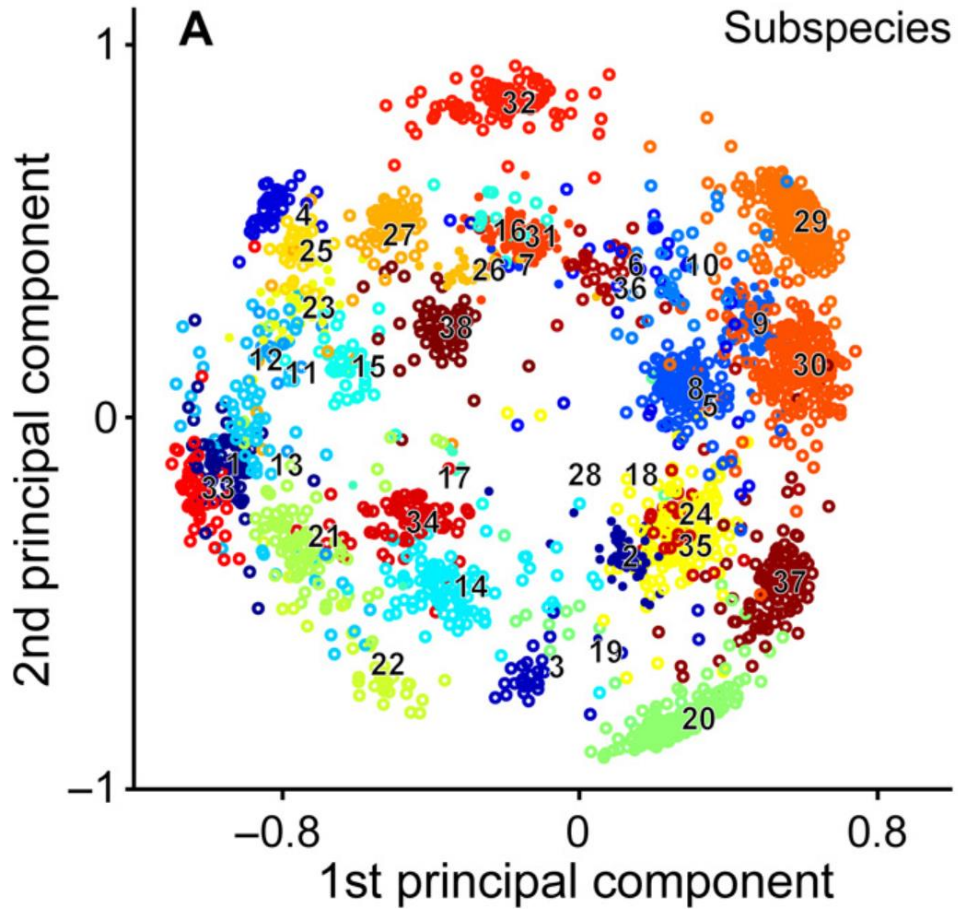


图 2：利用 ButterflyNet 将不同种类蝴蝶的 2468 张照片映射到一个向量空间。每种颜色代表一个蝴蝶亚类[2]。

进一步，论文作者还利用邻接法[3]将上述向量空间表示成一个树状结构，如图 3 所示，其中每个圆圈代表一个亚类，圆圈边上的文字为黑色的代表 *H. erato* 类，文字标记为灰色的代表 *H. melpomene* 类。分析表明，在同一种类的蝴蝶内部，具有互相拟态的两个亚类比不具有互相拟态的两个亚类相似性更为明显；更为重要的是，同属不同种类，但互相拟态的两个亚类的相似性甚至比同一种类的两个亚类相似性更高。这一结果清晰证明了拟态使物种互相接近，从而实现协同进化。

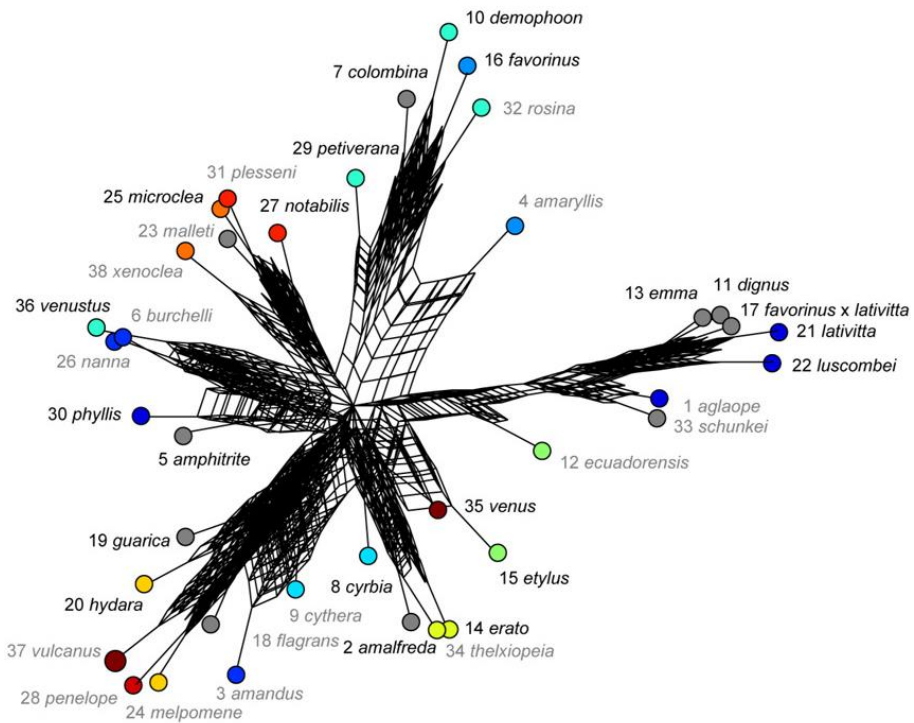


图 3: ButterflyNet 生成的邻接树，其中每个圆圈代表一个亚类，黑色和灰色文字标记分别代表 *H. erato* 和 *H. melpomene* 两个大类[2]。

1. Müllerian mimicry, https://en.wikipedia.org/wiki/M%C3%BCllerian_mimicry
2. Cuthill J F H, Guttenberg N, Ledger S, et al. Deep learning on butterfly phenotypes tests evolution's oldest mathematical model[J]. *Science advances*, 2019, 5(8):
3. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees[J]. *Molecular biology and evolution*, 1987, 4(4): 406-425.
4. F. Müller, *Ituna* and *Thyridia*; a remarkable case of mimicry in butterflies. *Trans. Entomol. Soc. Lond.* 12–29 (1879)