

麻省理工团队开发检测自主系统故障的方法

| 逸文

从汽车防撞系统到航空调度系统，再到电网，我们所依赖的许多服务都是由计算机管理的。随着这些自主系统的复杂性和普遍性不断提高，它们出现故障的方式也会随之增多。

现在，麻省理工学院的工程师们开发出了一种方法，可以与任何自主系统配对使用，在它们被部署到现实世界之前，就能快速识别出系统中的一系列潜在故障。更重要的是，这种系统可以找到故障的修复方法，并提出维修建议，以避免系统崩溃。

研究小组的研究表明，这种方法可以根除各种模拟自主系统的故障，包括小型和大型电网网络、飞机防撞系统、救援无人机团队和机器人机械手。在每个系统中，新方法都能以自动采样算法的形式，快速识别出一系列可能出现的故障，以及避免这些故障的修复方法。

新算法采取了与其他自动搜索不同的方法，其他自动搜索的目的是发现系统中最严重的故障。研究小组说，其他方法可能会漏掉一些更微小但却很重要的漏洞，而新算法可以捕捉到这些漏洞。

麻省理工学院航空航天系研究生查尔斯·道森（Charles Dawson）说：“在现实中，这些更复杂的系统可能会出现一系列混乱。我们希望能够信任这些系统，让它们为我们开车、驾驶飞机或管理电网。了解它们的极限以及在什么情况下可能会出现故障，这一点非常重要。”

道森和麻省理工学院航空航天学助理教授范楚楚（Chuchu Fan）将在机器学习大会上介绍他们的研究成果。

超越现实故障的敏感性

2021年，得克萨斯州的一次重大系统崩溃引发了范楚楚和道森的思考。当年2月，冬季风暴席卷该州，带来了意想不到的低温，引发了整个电网的故障。这场危机导致450多万

户家庭和企业停电多日。这次系统性故障是德克萨斯州历史上最严重的能源危机。

道森说：“那是一次相当严重的故障，让我不禁怀疑我们是否能预测到它。我们能否利用电网的物理知识，了解电网的薄弱环节，然后有针对性地进行升级和软件修复，在灾难性事件发生之前加强这些薄弱环节？”

道森和范楚楚的工作重点是研究



机器人系统，并寻找使其在环境中更具弹性的方法。在得克萨斯州电力危机的部分推动下，他们开始扩大研究范围，以发现和修复其他更复杂、更大规模的自主系统的故障。为此，他们意识到必须改变发现故障的传统方法。

设计人员通常通过确定最可能、最严重的故障来测试自主系统的安全性。他们首先对系统进行计算机模拟，模拟系统的基本物理特性以及可能影响系统行为的所有变量。然后，他们使用一种进行“对抗优化”的算法运行模拟，这种算法通过对系统进行一次又一次的微小改动，自动优化最坏的情况，直到能够缩小与最严重故障相关的改动范围。

道森指出：“把所有这些变化都

缩小到最严重或最可能发生的故障中，你会失去很多可以看到的复杂行为。相反，我们希望优先识别故障的多样性。”

为此，团队采取了一种更加“敏感”的方法。他们开发了一种算法，可以自动生成系统内的随机变化，并评估系统对这些变化的敏感度或潜在故障。系统对某种变化越敏感，这种变化就越有可能导致故障。



这种方法能让团队找出更多可能出现的故障。通过这种方法，该算法还能让研究人员通过回溯导致特定故障的变化链来确定修复方法。

“我们认识到这个问题实际上具有两面性，”范楚楚说，“硬币有两面。如果你能预测故障，就应该能预测如何避免故障。现在，我们的方法正在闭合这一循环。”

隐藏的故障

研究小组在各种模拟自主系统上测试了这种新方法，包括小型和大型电网。在这些情况下，研究人员将他们的算法与区域规模的通用电网进行了拟配对。结果表明，传统的方法只针对最容易发生故障的单个电线上，而研究小组的算法发现，如果再加上

第二条电力线发生故障，就可能发生全面停电。

“我们的方法可以发现系统中隐藏的相关性，”道森说，“因为我们在探索故障空间方面做得更好，所以我们可以发现各种故障，有时甚至包括比现有情况更严重的故障。”

研究人员在其他自主系统上也展示了类似的不同结果，包括避免飞机相撞和协调无人机的模拟。为了了解他们在模拟中预测的失败是否会在现实中得到证实，他们还在一个机器人操纵器上演示了这一方法，这是一种专门用来推拿物体的机械臂。

研究小组首先在一个机器人的模拟环境中运行了他们的算法，该机器人的任务是将一个瓶子推开而不是推倒它。当他们在实验室中用实际机器人运行同样的场景时，发现机器人以算法预测的方式失败了——例如，推倒了瓶子或没有够到瓶子。当他们应用算法建议的修复方法时，机器人成功地推开了瓶子。

道森说：“这表明，在现实中，当我们预测这个系统会失败时，它就会失败，而当我们期望它成功时，它就会成功。”

原则上讲，该团队的方法可以发现并修复任何自主系统的故障，只要能准确模拟其行为即可。道森设想，有朝一日，这种方法会被制作成一款应用程序，供设计师和工程师下载使用，以便在现实世界中进行测试之前，调整和强化他们自己的系统。

道森说：“随着我们对这些自主决策系统的依赖程度越来越高，我认为故障的类型将会发生变化。与其说是系统内部的机械故障，但实际上看到更多由自主决策和物理世界的相互作用导致的故障。我们正试图通过识别不同类型的故障来应对这种转变，现在就着手解决它们。”

这项研究部分得到了美国国家航空航天局、美国国家科学基金会和美国空军科学研究办公室的支持。



DARPA将与商业公司合作开展新一轮月球经济研究

要在月球上建立商业市场，就必须换一种思维方式，这就是美国国防部预先研究计划局（DARPA）的看法，该局是美国国防部的研发机构，他们刚刚挑选了14家公司参与一项新的研究，为月球的未来开发技术框架。

该机构在一份声明中说：“未来十年，月球经济可能会蓬勃发展。要实现这一目标，就必须超越目前孤立、自给自足系统的技术范式，这些系统必须有机地支持所有必要的资源——如电力和通信——并面向一个强调商业活动综合模式的未来框架。”

入选“月球十年架构”（LunA-10）能力研究的公司将在为期7个月的时间里共同努力，为物流、建筑和通信等核心月球服务设计可互操作的综合解决方案。最重要的是，这项研究不会为月面技术演示或技术建设提供资金，相反，可以将这项研究视为未来所有活动的分析框架。

这14家公司包括大型航空航天公司和较小的太空初创公司。它们是蓝色起源、Cislunar Industries LLC、Crescent Space Services LLC、Fibertek、Firefly Aerospace、Gitai、Helios、Honeybee Robotics、ICON、诺基亚美国公司、诺斯罗普·格鲁曼、Redwire、

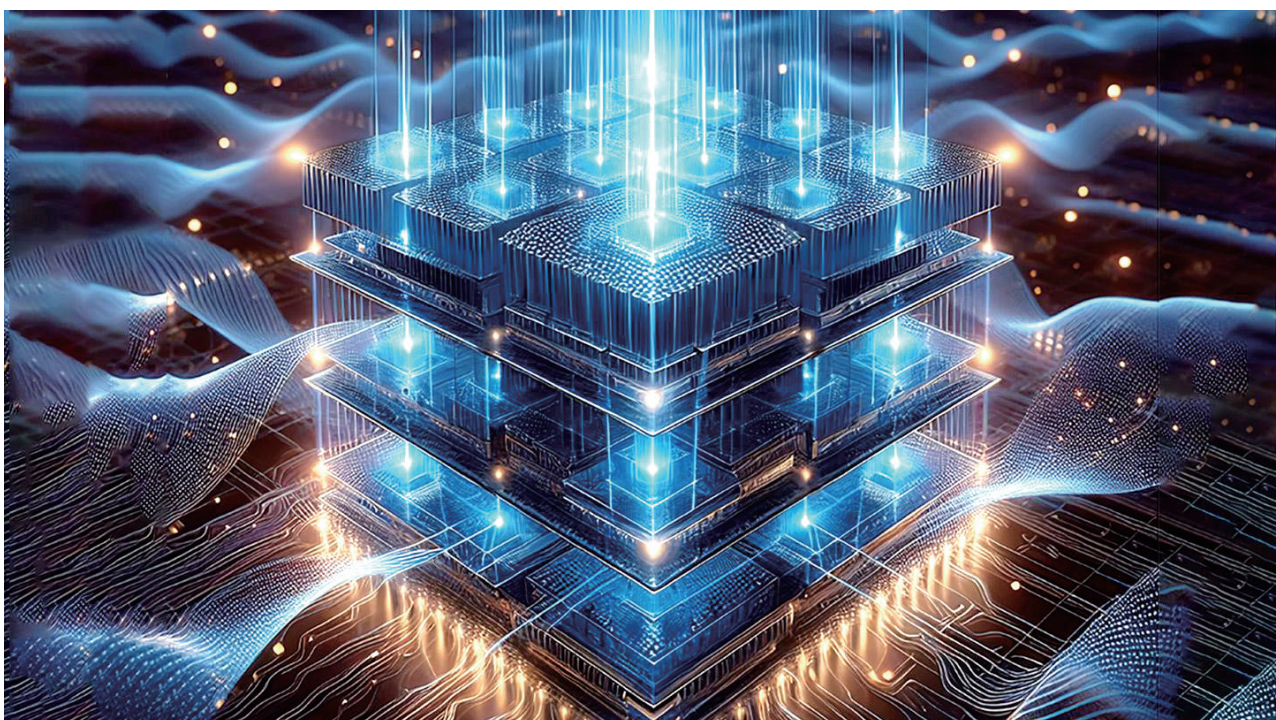
Sierra Space和SpaceX公司。

DARPA表示，这些公司将在“高度协作的环境”中开展工作，为月球服务设计系统级解决方案。可以探索的其他此类服务包括商业原地资源利用、定位、导航和授时（PNT）、移动和运输以及机器人技术。

DARPA没有透露每家公司将专注于哪些领域，但一些参与者已分别公布了其贡献的详细信息。例如，Firefly公司表示，它将在其Elytra航天器系列的基础上开发一个“在轨航天器枢纽框架”。

该公司在一份新闻稿中说：“基于Firefly公司Elytra飞行器的能力，该框架将定义航天器如何在轨对接并提供按需服务，如加油、交付、运输和离轨。该框架的目标是通过可扩展的航天器枢纽，帮助在轨任务响应时间从数年缩短到数天，这些航天器枢纽可以为整个日月空间的航天器提供托管和整合。”

Gitai的提案涉及其Inchworm机器人，该机器人配备了可更换工具的末端效应器，旨在处理月球和空间站上的劳动。Redwire公司的贡献将集中在高速通信和PNT服务上，该公司指出：“提供强大服务和广泛月球覆盖范围的半月轨道平台星座将是繁荣的商业月球生态系统的重要组成部分。”（航柯）



研究人员开发可打破深度物理神经网络障碍的算法

瑞士洛桑联邦理工学院（EPFL）的研究人员开发出一种算法，可以像训练神经网络一样精确地训练模拟神经网络，从而开发出更高效的深度学习硬件替代品。

通过算法“学习”而非传统编程来处理海量数据，ChatGPT等深度神经网络的潜力似乎是无限的。但是，随着这些系统的范围和影响不断扩大，其规模、复杂性和能耗也随之增加——后者的重要性足以引起人们对全球碳排放的担忧。虽然我们通常认为技术进步是从模拟到数字的转变，但研究人员现在却在数字深度神经网络的物理替代品中寻找这一问题的答案。

EPFL工程学院波工程实验室的罗曼·弗勒里（Romain Fleury）就是这样一位研究人员。在发表于《科学》（Science）的一篇论文中，他和他的同事描述了一种训练物理系统的算法，与其他方法相比，该算法提高了速度、增强了鲁棒性并降低了功耗。“我们成功地在三个基于波的物理系统上测试了我们的训练算法，这些系统使用声波、光波和微波而不是电子来携带信息。但我们的多功能方法可用于训练任何物理系统。”第一作者、波工程实验室（LWE）研究员阿里·莫梅尼（Ali Momeni）说。

更符合生物学原理的方法

神经网络训练是指帮助系统学习如何为图像或语音识别等任务生成最佳参数值。它传统上包括两个步骤：前向传递，即通过网络发送数据，并根据输出计算误差函数；后向传递（也称为反向传播，或BP），即计算误差函数相对于所有网络参数的梯度。

在反复迭代的过程中，系统会根据这两个计算结果进行自我更新，从而返回越来越精确的数值。问题出在哪里？除了非常耗能之外，后向传递还不适合物理系统。事实上，训练物理系统通常需要一个数字孪生来完成向后传递步骤，这不仅效率低下，而且还存在现实与模拟不匹配的风险。

科学家们的想法是用物理系统的第二次前向传递取代后向传递步骤，在本地更新每个网络层。除了降低功耗和无需数字孪生外，这种方法还能更好地反映人类的学习情况。“神经网络的结构受到大脑的启发，但大脑不太可能通过后向传递学习。”莫梅尼解释道，“新的想法是，如果我们对每个物理层进行本地训练，我们就可以使用实际的物理系统，而不是首先建立一个数字模型。因此，我们开发了一种更符合生物学原理的方法。”

EPFL的研究人员与法国国家科学研究中心电子与数字技术研究所（CNRS IETR）的菲利普·德尔·乌尔涅（Philipp del Hougne）和微软研究院的巴巴克·拉赫马尼（Babak Rahmani）一起，利用他们的物理局部学习算法（PhyLL）训练实验声学 and 微波系统以及一个模型光学系统，对元音和图像等数据进行分类。

除了显示出与基于后向传递的训练相当的准确性外，该方法还具有鲁棒性和适应性——即使是在受到不可预测的外部扰动的系统中——与目前的技术水平相比也是如此。

未来的模拟

虽然EPFL波工程实验室的方法首次实现了深度物理神经网络的无后向传递训练，但仍然需要对参数进行一些数字更新。

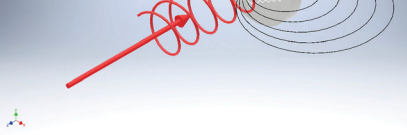
“这是一种混合训练方法，但我们的目标是尽可能减少数字计算。”莫梅尼说，“在我们的实验中，我们使用了最多10层的神经网络，但如果我们有100层、数十亿个参数，它还能工作吗？这是下一步需要克服物理系统的技术限制。”

研究人员现在希望在小型光学系统上实现他们的算法，最终目标是提高网络的可扩展性。（逸文）

研究发现可用超薄碳制成微型电磁铁

辐射脉冲搅动电子

实验成功的先决条件是研究人员必须以特定的方式极化太赫兹闪光。专门的光学仪器改变了辐射的振荡方向，使其在空间中螺旋式移动。当这



些圆偏振闪光击中微米大小的石墨烯圆盘时，就产生了决定性的效果：在辐射的刺激下，石墨烯圆盘中的自由电子开始旋转，就像用木勺搅拌桶中的水一样。根据基本的物理定律，循环电流总是会产生磁场，因此石墨烯圆盘变成了微小的电磁铁。

杜伊斯堡-埃森大学教授马丁·米滕多夫（Martin Mittendorff）说：“这个想法其实很简单。事后看来，我们对以前没有人这样做感到惊讶。”同样令人惊讶的是这一过程的效率：与用光照射纳米金粒子的实验相比，HZDR的实验效率要高出一百万倍，这是一个令人印象深刻的提升。这种新现象最初可用于科学实验，将材料样品暴露在短而强的磁脉冲中，以更详细地研究某些材料特性。

“我们能够产生0.5特斯拉的磁场，大约是地球磁场的一万倍。这些磁脉冲很短，只有大约10皮秒或十亿分之一秒长。”温纳尔说。

“采用我们的方法，磁场不会像许多其他方法那样发生极性反转，”温纳尔解释说，“因此，它仍然是单极的。换句话说，在来自石墨烯磁盘的磁脉冲持续的10皮秒内，北极仍然是北极，南极仍然是南极——这对某些系列实验来说是一个潜在的优势。”

磁性电子产品的梦想

从长远来看，这些微小的磁体甚至可能对未来的某些技术有用：当超短辐射闪光产生这些磁体时，石墨烯圆盘就能进行极其快速和精确的磁切换操作。例如，这不仅对磁性存储技术很有意义，而且对所谓的自旋电子学——一种磁性电子学——也很有意义。

在这里，处理器中流动的不是电荷，而是以电子自旋形式存在的微弱磁场。人们希望这能再次大大加快开关过程。可以想象，石墨烯圆盘可以用于可切换电磁铁，控制未来的自旋电子芯片。

然而，专家们必须为此发明非常小的、高度微型化的太赫兹源——当然，这还有很长的路要走。温纳尔评论说：“你无法使用我们在实验中的那种完全自由电子激光器。”不过，实验室桌子上的辐射源应该足以满足未来科学实验的需要。在一些研究设施中已经可以找到这种更为紧凑的太赫兹辐射源。（航柯）

