

局部随机行为机器人可促进人类整体协作

遵循原则办事一直以来被认为是人类高效合作的基础,然而最近一项研究表明,一定程度上不按常理出牌反而可以促进人类合作,实现大同。

2017年5月17日,来自耶鲁大学的学者 Hirokazu Shirado 和 Nicholas A. Christakis,在《Nature》上发表了一篇题为《Locally noisy autonomous agents improve global human coordination in network experiments》^[1]的文章(图1),引起了广泛关注。



图1 《Nature》2017年第7654期封面

1 机器人Bot与人协作实验

在论文中设计了一个实验,要求参与的志愿者玩一个在线游戏,游戏中搭建了一个由20个节点构成的网络,每位志愿者控制其中一个网络节点,每个节点可以在橙色、绿色和紫色3种颜色之间转换。当相邻两个节点颜色相同时,认为两个节点之间存在冲突,如图2(a)中的红色边所示。游戏中,每位志愿者只能看到自己以及和自身相邻节点的颜色,从而调整自己的颜色。游戏最终目的是要使网络变为不存在冲突的状态,如图2(b)所示。如果整个网络在5 min之内实现了目标(相邻节点颜色不同),所有玩家都会受到额外的奖励。

研究人员一共招募了4000名人类玩家,并将230个随机生成的机器人Bot置于该网络中,实验将由这些人类玩家以及机器人共同协作完成。实验被分为多组,以便进行对比。第1组,20个节点全部由人控制;第2组,17个节点由人控制,3个节点颜色固定;第3组,17个节点由人控制,3个节点由机器人Bot控制。这第3组实验可以根据机器人Bot位置和动作策略的不同而细分的。根据位置划分,分为3种类型:1) 机器人Bot部署在网络中心;2) 机器人Bot部署在网络边缘;3) 机器人Bot部署位置随机。根据机器人Bot采用的动作策略又可分为3种类型:1) 机器人Bot每隔1.5 s选择一种颜色,使得自己和周围相邻节点冲突最少;2) 机器人Bot在70%的时间里采用第一种策略,剩下的30%时间里随机选择颜色;3) 机器人Bot在90%的时间里采用第1种策略,剩下的10%时间里随机选择颜色。以上根据机器人Bot的位置和策略的不同,可以组合构成第3组实验方案的9类实验。

实验结果如图3所示。图3中的纵轴表示尚未被解决冲突的实验数量占同类实验总数的比率,横轴表示游戏进行的时长。深蓝色线条表示由人和机器人Bot协作的结果,浅蓝色线条表示全部由人控制的结果。图3中的 P 值是对数秩检验结果,因此该组实验存在显著差异

的置信度可表示为 $1-P$ 。由于上述未加入机器人Bot的第1组和第2组实验结果非常接近,在图2的实验中只给出了第1组实验和第3组实验的对比结果。图2中的第1行,第2行和第3行分别给出了机器人Bot位置对结果的影响,分别对应为机器人Bot随机摆放,放在网络中间以及放在网络边缘3种位置情况;图2中的第1列,第2列和第3列分别给出机器人Bot采用不同智能策略(按照0, 10%, 30%的比例随机选择颜色)的实验结果;以图2左上方第1张子图为例,深蓝色线条表示3个机器人Bot参与游戏的实验结果,其中参与实验的3个机器人Bot位置随机摆放,机器人Bot每隔1.5 s选择1种颜色,使得自己和周围相邻节点冲突最少。浅蓝色曲线表示20个节点全部由人类控制的结果。由该子图中的 P 值可知,有74.3%($1-0.257$)的概率可以认为该子图对应的实验结果是有显著性差异的。

2 实验结果分析与评价

实验结果表明,混有机器人Bot的网络测试结果和全部由人类控制的网络测试结果相比互有优劣,存在一定随机性。但是注意看最中间的实验结果,具有10%行为噪声且位于网络中心的机器人Bot反而可以显著提升游戏的可解性,同时具有98.5%的概率说明两组实验存

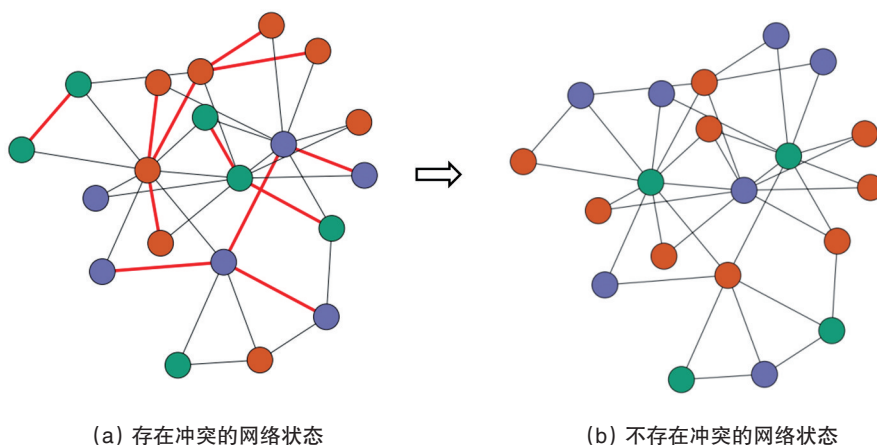


图2 在线游戏实验中网络节点变化示意

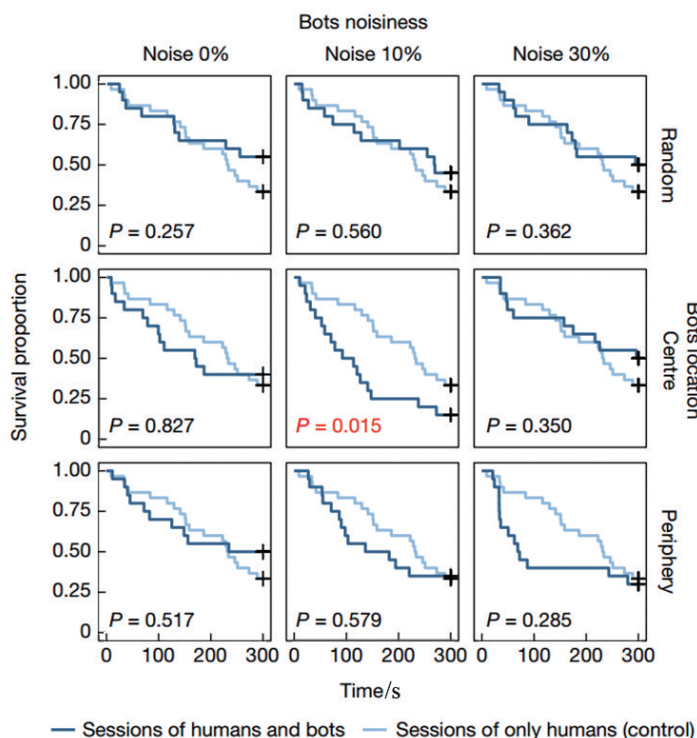


图3 机器人Bot与人协作实验结果

在显著性差异。在全部由人类控制的总计30次的实验中,共有20次实验在5 min内解决了所有冲突,这20次实验的平均时间为232.4 s,而有10%行为噪声且机器人Bot位于网络中心的这组实验一共进行了20次,其中17次实验在5 min内解决了所有冲突,这17次实验的平均时间为103.1 s,相比全由人控制的实验,其解决问题的比率提升了26.9%,其平均速度提升了55.6%。

乍一看来,上述实验结果和人类直觉似乎相悖,直觉通常会认为噪声、随机动作是无意义的,只是浪费资源的一种行为。然而,这一直觉却被日益深入发展的人工智能所颠覆。以最近全球热切关注的围棋人机大战为例,AlphaGo^[2]的一些落子与人类棋手的直觉和经验相差甚远,被认为是败笔,却在最后成为了整盘棋的转折点。事实上,人类棋手往往会以自己主观的最优解法去理解整盘棋,和对手对抗。在训练过程中潜移默化的认为某些落子会使整盘棋更有胜算,最后就固化这样的思维,形成了自身的直觉。这样做的好处是在围棋这一超高维度运算中可以快速地降低问题的复杂性和维度,进而做出判断。但是,缺陷也是显而易见的,上述直觉引导下的策略会造成解空间覆盖不全,尽管直觉可

以比较快地收敛到一个相对不错的解,但是这样的解往往还存在比较大的提升空间。这也就不难理解,AlphaGo当时看似失误的走棋会成为最后的关键。其原因就在于这是对更大的解空间进行探索而得出的,而这样的探索过程往往伴有一定程度的随机性,解空间大了,可以选择的策略自然就会更多,获得更优解的概率也就相应上升了。

其实在人工智能领域,利用噪声来探索未知空间早已有所应用。例如强化学习领域中对未知领域的探索常见的做法就是采用e-greedy方法^[3],即在智能体学习的前期采用完全随机的探索,去充分的探索未知领域,到了学习后期,则以90%概率选择当前最优,10%的概率仍然进行随机探索,这种做法不仅利用了自己本身的知识,也不断继续学习。随着学习时间的推移,智能体会学得越来越好,设想一下,如果不对未知进行探索,仅仅只满足现在所得,故步自封,自然也就不会进步了。

但是之前的探索仅仅对于自身进行,也就是优化自我,而上述文章中的实验是基于随机噪声优化协同合作,这是第一次被提出的,是这篇文章的一大亮点。人与人之间合作一般有两种形式,第一是自上而下,由leader指定任务,分

配工作,这样并不一定高效;第二每个人发挥各自特长,但这又会引入个人目标与最终目标的矛盾。本文对两种合作方式都做了实验,在实验初期指定3个点的颜色就是对第1种方法的模拟,结果和全由人类控制的效果类似。但是对第2种合作方式提出了一种非常大胆的想法——一定程度的随机。在实验中玩家往往会选择使得自己和周围节点冲突最少的颜色(个人目标),但是这种方法和全局无冲突(全局目标)存在一定矛盾。通过在网络中引入能够产生随机行为的机器人Bot,很好的协调了相互之间的合作,也就是机器人帮助人类实现了协作提升。

3 未来展望

《Locally noisy autonomous agents improve global human coordination in network experiments》一文对合作共赢提出了新的想法,期待这一想法可以在后续可以进一步展开。下一步的研究重点可能在于以下几个方面:

正如作者所说,本实验和真正人类协作还存在一定的差距,后续工作将会在网络合作更大,任务更加复杂的环境中进行,来拓展该研究的实用性。

理论上进行推导,为什么是10%而不是其他的噪声程度,可以达到促进作用,是否存在一定的极限,极限是多少,这样将会更有说服力,更具泛化能力。

参考文献

- [1] Shirado H, Christakis N A. Locally noisy autonomous agents improve global human coordination in network experiments[J]. Nature, 2017, 545(7654): 370-374.
- [2] Silver D, Huang A, Maddison C J, et al. Mastering the game of Go with deep neural networks and tree search[J]. Nature, 2016, 529(7587): 484-489.
- [3] Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529-533.

文/刘勇,卢宇鹏

作者简介:刘勇,浙江大学智能系统与控制研究所,教授;卢宇鹏,浙江大学智能系统与控制研究所,硕士研究生。

(责任编辑 刘志远)