

浅谈控制中的共享信息和共享自主

赵云波 康宇 朱进 中国科学技术大学自动化系

控制系统是由受控对象、传感器、控制器、执行器等不同组件有机构成的随时间演化的一个整体,控制理论研究控制系统的动态行为,和前馈、反馈等控制机制对系统动态行为的改造,而控制工程则依托控制理论解决各种实际应用领域的控制问题。与绝大多数信息和工程技术不同的是,控制是一种使能技术,并不局限于某一特定应用领域,而是在现代科技的几乎所有方面都起着某种关键基础性作用^{[1][2]}。

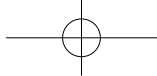
从技术发展的宏大角度来看,控制的发展与人类历史的技术变革相伴相生,而技术变革也为控制带来了不同的机遇和相伴而来的要求和挑战。首先,起于19世纪末的工业革命掀起了轰轰烈烈的以机器替代人的劳动的革命,在其中控制赋能的自动化起到了关键基础性的作用;其次,伴随着始于20世纪末、至今仍方兴未艾的信息革命的发展,远程、大规模、网络化的控制系统和自动化应用极大地拓展了控制的版图^[3];最后,近些年才重新兴盛起来但潜力无限的人工智

能(Artificial Intelligence, AI)技术,则带来了变革性的机器自主性革命,将控制系统的研究带入了一个崭新的领域^[6,5,4]。

本文试图简单讨论信息革命和AI自主革命带给控制的若干机遇和挑战,尤其分别关注其中的“共享信息”和“共享自主”控制范式。本文并不打算对这两个相关领域给出详尽的理论和方法上的介绍,而是较为科普性质,目的是通过对问题的来源和其重要性的讨论,希望能够吸引更多研究者投入到这两个相关联领域的研究当中来。

1. 从传统控制到共享信息的控制

任何控制系统都需要在传感器和控制器之间、控制器和执行器之间进行信息传输,前者负责将传感器采集的受控对象的输出信息传递给控制器,后者则负责将控制器产生的控制信号传递给执行器。本节将首先探讨控制系统中信息传输方式的演变过程,进而讨论在信息化时代产生的共享信息的特点及其相关控制问题。



1.1 控制系统中信息传输方式的演变

从控制系统中信息传输方式演变的角度来看,控制系统的形态经历了如下三个不同的发展阶段。

- 在最早期,控制系统是小型化的、局部的,各个组件之间由专用和私有的信道直接相连。因为这种“私有专线”的特点,控制系统中的信息传输是完整的、几乎无延时的,因而也就不需要考虑信道本身对控制系统的影响。
- 在第二阶段,控制系统规模开始变大、结构变复杂,一个系统中可以包含大量的传感、控制、执行组件,甚至可能由若干子系统构成。对这类系统,要在任意两个组件之间建立私有专线连接一方面成本太高,另一方面也使系统变得太过复杂,因此一种“总线”结构的通信信道在这类控制系统中开始大规模采用。其基本思想是建立一条骨干通信线路,然后所有的控制系统相关组件都连接到这条骨干线路上去,从而任意两个组件都可以通过这个骨干线路互通。可以看出,总线结构的通信信道仍然是控制系统私有的,但对每两个组件之间的通信来说,它并不是一条专线,而是需要跟其他的通信需求共享同一条信道,因此,其基本特点是“私有共享”。
- 随着通信网络技术的飞速发展,公共数据通信网络(典型的如互联网)变得非常普遍,如果能够利用公共



图1 网络化控制系统示意图

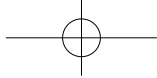
数据通信网络作为控制系统的信息传输信道,就可以在很低的成本下把最大范围的控制组件连接在一起,极大地拓展控制系统的适用范围和规模。但是,公共数据通信网络并非是为控制系统特别设计的,也从本质上难以满足控制系统所需要的数据传输的实时性、准确性等要求,这样,“公有共享”的数据传输特点就为控制系统设计带来了很大的挑战。

1.2 信息共享的控制:网络化控制系统

上述第三种控制系统形态,即利用公用数据通信网络作为信息传输通道的控制系统,一般称为“网络化控制系统”,在近些年得到了广泛的关注和研究,其系统架构见图1。

显然,网络化控制系统面临挑战的缘由,在于其“公有共享”数据传输的本质特点。这一本质特点可简单分析如下(详细讨论可参见文献[7])。

- 数字化。网络化控制系统中的所有数据传输经由数据通信网络进行,而数据通信网络中的数据是数字化的。这一特点使得网络化控制系统需要考虑如下问题:1)量化。控制系统中的信号,不管是系统输出还是控制信号,很多情况下都是连续模拟信号,数字化传输的要求使得必须对模拟信号进行量化,而量化中的信息丢失不可避免地会对系统性能产生影响。2)时间同步。网络化控制系统的各个组件必然处于数据通信网络的不同节点处,



各个通信节点的时间并不必然是同步的，而不同步的通信节点对网络化控制系统而言则意味着控制系统中传输的数据的发送时间、接收时间和经历延时等无法准确探知，但这些信息对控制系统设计和分析而言都是至关重要的。3) 驱动机制。控制系统的数据传输既可以是时间驱动的，即在某一事先确定的时刻进行传输（一般是给定某一固定周期），也可以是事件驱动的，即在某一事件发生时才进行传输。网络化控制系统的架构天然允许这两种驱动方式的使用，而如何有效利用合适的驱动机制提升系统性能就成为一个研究的热点。

- 包传输。通信网络中的数据通过数据包进行传输。这一特点造成如下问题：1) 网络诱导延时。数据包从等待传输到传输成功总是需要一定时间的，这就造成控制系统的数据传输带有不可避免的延时。2) 数据包丢失。实际通信网络中数据包可能会因为各种原因丢失，意味着打包在内的控制系统数据以某种不可预见的可能性丢失。3) 数据包错序。每个数据包的传输经历不可控且随机的延时，意味着先发的数据包可能后到，后发的数据包也可能先到。4) 单包 / 多包传输。数据通信网络的数据包大小并非为控制系统专门设计，这就意味着控制系统所要传输的数据或者少于或者多于单个数据包的大小（一般是前者）。
- 通信资源限制。数据通信网络的资源自然是有限的，对于网络化控制系统来讲又是公有不透明的，这就产生了如下问题：1) 采样周期和系统性能的平衡。传统采样控制系统中加大采样频率一般总是会导向更好的系统性能，因为有更多的系统信息可用；但在网络

化控制系统的情况下，加大采样频率，会增加信息传输需求，后者反过来可能会拥塞网络，反而会造成系统性能下降。2) 网络接入限制。因为通信网络是公有的并且有限时间内只能接受有限的传输要求，这就会使得控制系统的数据传输需要考虑通信网络的接入限制。3) 能量限制。在通信节点使用电池供电的情况下，也需要考虑减少信息传输以节省能量消耗，增加系统生存时间。

- 网络化。使用数据通信网络作为信息传输通道使得大规模控制系统的建设更为容易，但也带来了一些问题，比如在无线网络情况下存在相邻节点的信道干扰等问题，需要对局部信息传输进行有效调度；在多智能体研究中，考虑通信受限下的系统同步和协同等问题也是很大的挑战。

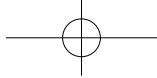
针对上述共享信息所带来的问题，近十几年来网络化控制系统领域已经有了大量的研究，在此不多赘述，感兴趣的读者可参考文献 [11,10,9,8]。

2. 从传统控制到共享自主的控制

2.1 AI 赋能的控制

基于深度学习的人工智能技术在近十年得到了井喷式的发展。这一发展得益于信息技术支撑下各应用领域中大数据可得性提升、面向人工智能算法的特定芯片算力的提升和深度学习算法的突破等方面因素^[12]。在各国政府、大型跨国公司和科技界的关注下，有理由相信人工智能相关理论在可见的未来还将迅猛发展，相关技术将会用到越来越多的领域中，产生越来越大的社会、经济、军事等方面的影响^[13]。

将 AI 技术应用于控制系统，就形成所谓 AI 赋能的控制。比如，可以利用 AI 技术对复杂系统中大量传感数据进行处理，或在网络化控制系



统中的对网络特性进行学习和刻画，或通过输入输出数据对系统进行无模型控制，等等。

上述角度的研究着眼于利用 AI 增强控制系统的某个组件，然而更为本质的挑战也许在于 AI 所带来的潜在的自主性能力，这意味着 AI 可在合适的场景下自行设定目标、决定控制策略、主导系统运行过程。将这一自主性能力用于控制系统，意味着控制将会变得越来越智能。

但基于深度学习的 AI 的局限性使得 AI 赋能的控制走向自主化的道路上面临着重大的挑战。比如，深度学习算法的训练需要大量的数据，很多控制系统难以在运行之前就采集足够的数据进行训练；深度学习算法的训练和执行对计算能力也有较高要求，这一点很多控制系统因为计算能力限制和实时性要求无法满足；再者，深度学习算法缺乏对所作决策的可解释性和可信性刻画，这也给控制算法设计也带来了挑战；另外，现有深度学习算法难以保证鲁棒性，在极小的噪声下就可能失效，这严重影响了将其应用于往往对系统运行的稳定可靠要求严苛的控制系统中^[14]。

2.2 人在 AI 赋能控制的闭环中

关于 AI 技术的未来发展，大多数从业人员相信在可以预见的未来我们只可能拥有弱人工智能（或专用人工智能，Artificial Narrow Intelligence，指在某一特定领域达到或超过人的能力的人工智能），而不能实现强人工智能（或称通用人工智能，Artificial General Intelligence，指在所有领域达到人的能力的人工智能）^[17]。

这样，依照应用领域不同，AI 赋能的控制就存在着两个发展可能：如果应用领域是弱人工智能可以完美解决的，那么 AI 赋能的控制就事实上退化到了传统的以自动化为目标的控制系统的范畴；与此相对应的，如果应用领域是

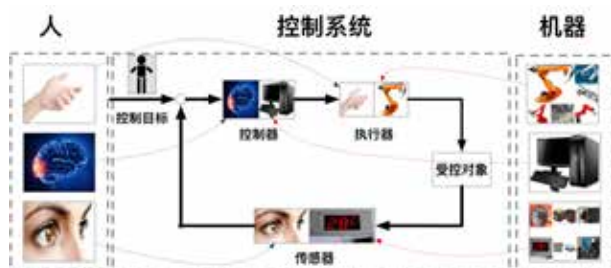


图2 控制系统的组件可由人或机器担当

弱人工智能无法完美解决的，那么 AI 赋能的控制就势必要与人共存。

有理由相信 AI 将在绝大多数领域中得到应用，而弱人工智能能够完美解决的专业领域是较为受限的，这意味着未来人将存在于大量 AI 赋能的控制系统的闭环中。从控制和自动化的发展历史上看，这是一个本质性的变化：控制系统的研究一开始的初衷是希望将越来越多的功能自动化，从而让机器取代人的劳动，但现在随着 AI 在控制中的渗透，我们反而因为 AI 所带来的自主性要将人请回控制的闭环中，见图 2。

引发这一悖论的原因在于，AI 赋能的控制极大地拓展了控制系统适用的范围，使得控制系统可以应用于更不确定的环境、更复杂的动态、更多变的目标等条件下，而这原本就是人绝对控制的领域。因此，AI 赋能的控制因为能力提升得以进入这些拓展了的应用中，但仍需要人在系统的闭环中承担某种职责，而传统领域则仍是无人自动化的。

人在 AI 赋能的控制闭环中的概念引发了人机融合模式的改变。传统上研究人与机器交互的领域，如人机界面、人因工程等，都不曾将机器视为拥有与人类类似的智能和自主性的对象，因此在概念和方法上都不适用。另外，人在 AI 赋能的控制闭环中更多涉及的是在理论和算法层面的问题，

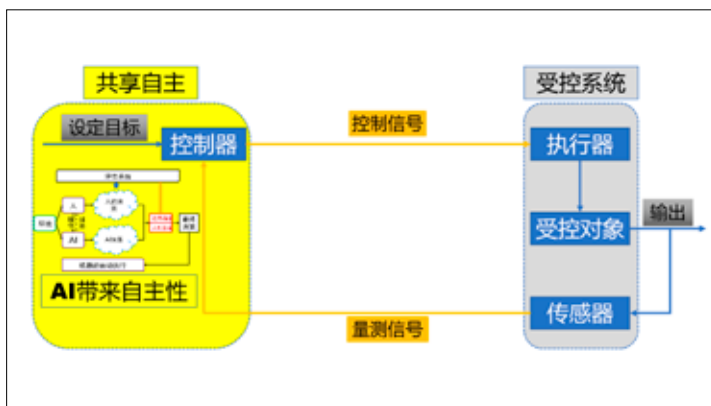
这也有别于其他人机融合的途径,比如类脑智能通过工程方法实现与人类大脑具有相同结构的人造大脑^[18],脑机接口通过物理连接人类大脑和外部机器,利用人类大脑的智能操控外界^[15]。

2.3 人机共享自主的控制

人和具有自主性的 AI 共同存在于一个控制系统中,产生了一个我们未曾见过的控制图景,即,系统的控制权需要在人和 AI 之间分配和协调。

解决这一问题的难点主要在于:1)在任何工程系统中,人都是一个不受欢迎的因素,因为它不标准、难量化、鲁棒性差。因此在大多数工程系统中人都游离于系统之外,但在这里的控制图景中,人是系统闭环的关键组成部分;2)基于深度学习的 AI 本身具有难解释、可信度差、鲁棒性差等问题,对其也难以操控。

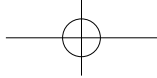
解决上述人和 AI 共享系统控制的问题,人机共享自主(Shared Autonomy)是一个可能的控制范式。共享自主的概念久已有之,在不同场景下也称为可信自主(Trusted Autonomy)、可调节自主(Adjustable Autonomy)等,它研究如何“理解和设计在两种各自具有一定程度自主性的实体间的交互空间(understanding and designing the interaction space between two entities each of which exhibits a level of autonomy)^[16]”;特别是,共享控制(Shared control)作为共享自主在控制系统领域的具体化,按照由初创了 IEEE SMC Technical Committee on



Shared Control 的三位专家在共同撰写的文章中给出的定义：“在共享控制中，人与机器在感知—执行循环中和谐交互来执行动态任务，而在理想环境下人或机器都可以独立执行这些动态任务 (In shared control, human(s) and robot(s) are interacting congruently in a perception-action cycle to perform a dynamic task that either the human or the robot could execute individually under ideal circumstances)^[17]。”这一定义强调了人和机器在共享控制的框架中具有某种平等的地位，而在一般控制系统中并非如此。

尽管因为概念上的相类,共享自主/共享控制据信是解决人在 AI 赋能控制中问题的有效控制范式,但传统共享自主/共享控制的概念和方法并不能简单迁移到我们面临的问题中,原因在于:首先,传统共享自主/共享控制的研究框架自身并不完善,特别是更为相关的共享控制的概念直到 2018 年才开始给出尝试性的定义^[17];其次,传统共享自主/共享控制的研究框架也缺少对 AI 赋能控制的针对性,更多是研究人与传统机器人之间的交互问题。这两方面的原因使得我们需要研究新形式的共享控制框架以应对当前问题。

知名学术机构和学者已经在包含 AI 因素的人机共享控制方面展开了相关的研究。比如,斯坦福大学研究人员提出了一种人机共享自主钢琴教学策略,通过在人和 AI 之间共享演示视频的选择权提升人跟随视频学习弹奏钢琴的学习效果^[18];伯克利人工智能实验室 (Berkeley Artificial Intelligence Research



BAIR)提出了一种人机共享自主操控无人机的策略,通过由人和机器负责各自擅长的任务,达到了单独人或无人机自身都无法达到的操控效果^[19];麻省理工学院人类中心AI实验室(Human-Centered AI at MIT)则提出了一种人和AI系统的仲裁机制,可用于提升特斯拉自动驾驶的安全性^[20]。

现有研究基本上还是一事一议性质的,缺乏系统性和完整性,也缺乏对该领域中若干关键基础性问题的研究。比如,人和AI各自的自主性边界如何定义和确定?是否存在通用的人和AI的介入和切换规则?等等。总的来讲,这一领域的研究正在萌芽,长路刚启,任重道远。

3. 结束语

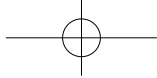
本文从技术变革影响控制系统结构变迁的角度,探讨了两种不同于传统控制系统的控制架构,即信息传输需经由数据通信网络从而需要在共享信息下进行控制的网络化控制系统,和控制决策需要在人和AI之间进行共享的人机共享自主控制。文章主要阐述问题的来源和其重要性,而不多涉及相应的技术方面,感兴趣的读者可从相关参考文献中做进一步了解。

我们认为,不管是共享信息的控制还是共享自主的控制,都是控制在具体的技术变革下的适应,具有鲜明的时代特点。研究该类问题,一方面扩展了控制理论和工程本身的内涵,另一方面也是技术发展带来的对控制研究的变革性要求,具有基本的重要性和紧迫性。

因学识所限,上述讨论一定是极为浅薄和失于片面的,不值一哂;不揣浅陋而仍作此“皇皇大论”的原因,固有科学普及和宣传的考虑,更多是为抛砖引玉,冀图就教于方家,解我等之惑耳。

参考文献

- [1] Aström K J A, Kumar P R. Control: A perspective[J]. Automatica, Elsevier Ltd, 2014, 50(1): 3–43.
- [2] Baillieul J, Samad T. Encyclopedia of Systems and Control[M]. J. Baillieul, T. Samad. London: Springer, 2015.
- [3] Murray R M, Astrom K J, Boyd S P, 等. Future directions in control in an information-rich world[J]. IEEE Control Systems magazine, 2003, 23(2): 20–33.
- [4] 黄琳. 控制理论发展过程的启示[J]. 系统工程理论与实践, 1990, 1(1): 71–77.
- [5] 席裕庚. 控制科学漫谈——学科的思考[J]. 系统与控制纵横, 2015: 1–6.
- [6] 程代展, 赵寅. 关于控制论的几点蠡测[J]. 中国科学院院刊, 2012, 2: 005.
- [7] Zhao Y–B, Sun X–M, Zhang J, 等. Networked Control Systems: The Communication Basics and Control Methodologies[J]. Mathematical Problems in Engineering, 2015, 2015(1): 1–9.
- [8] Hespanha J P, Naghshtabrizi P, Xu Y. A Survey of Recent Results in Networked Control Systems[J]. Proceedings of the IEEE, 2007, 95(1): 138–162.
- [9] Zhang L, Gao H, Kaynak O. Network-Induced Constraints in Networked Control Systems: A Survey[J]. IEEE Transactions on Industrial Informatics, 2013, 9(1): 403–416.
- [10] Park P, Ergen S C, Fischione C, 等. Wireless Network Design for Control Systems: A Survey[J]. IEEE Communications Surveys & Tutorials, 2017, 20(2):978–1013.
- [11] Al-Dabbagh A W, Chen T. Design considerations for wireless networked control systems[J]. IEEE Transactions on Industrial Electronics, 2016, 63(9): 5547–5557.
- [12] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553): 436–444.
- [13] 国务院. 国务院关于印发新一代人工智能发展规划的通知[R]. : 1–27.
- [14] M ü ller V C. Fundamental Issues of Artificial Intelligence[M]. V.C. M ü ller. Cham: Springer, 2016.



- [15] Brown M F, Brown A A. The promise of cyborg intelligence[J]. *Learning & Behavior*, 2016, 45(1): 5–6.
- [16] Abbass H A, Petraki E, Merrick K, 等. Trusted Autonomy and Cognitive Cyber Symbiosis: Open Challenges.[J]. *Cognitive computation*, 2016, 8(3): 385–408.
- [17] Abbink D A, Carlson T, Mulder M, 等. A Topology of Shared Control Systems—Finding Common Ground in Diversity[J]. *IEEE Transactions on Human–Machine Systems*, 2018, 48(5): 509–525.
- [18] Zhou S, Mu T, Goel K, 等. Shared Autonomy for an Interactive AI System[A]. New York, USA: ACM Press, 2018: 20–22.
- [19] Reddy S, Dragan A D, Levine S. Shared Autonomy via Deep Reinforcement Learning[J]. ar, 2018, cs.LG.
- [20] Fridman L, Ding L, Jenik B, 等. Arguing Machines: Human Supervision of Black Box AI Systems That Make Life–Critical Decisions[A]. *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019: 1–9.

