

论智能^{*)}

史俊^{1,2} 史忠植¹

(中科院计算所智能信息处理重点实验室 北京100080)¹ (中科院研究生院计算机学部 北京100039)²

摘要 信息化的必然趋势是智能化,智能革命将开创人类后文明史。图灵测试、物理符号假设等对智能研究的发展起了积极的推动作用。掌上电脑的先驱者霍金斯的新书 On Intelligence(论智能)闯进了人工智能的领域并使该领域萌生新机。新书提出了对人脑的新理解,将可能导致创建真正的智能机器,关键是机器智能需要神经科学。智能科学主张智能的研究不仅要在功能上模拟自然智能,而且要进行机理的研究,提出智能研究的新思路。

关键词 智能,神经科学,智能科学,图灵测试,物理符号假设

On Intelligence

SHI Jun^{1,2} SHI Zhong-Zhi¹

(Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080)¹ (Graduate School of Chinese Academy Sciences, Beijing 100039)²

Abstract The inexorable trend of Informatization is Intelligentization. Intelligent revolution will initiate post civilization era. Turing test, physical symbolic hypothesis and soon play important roles to push research on intelligence. Jeff Hawkins who is palm pilot fame rode into artificial intelligence and shot up the place. His new book titled On Intelligence proposes that a new understanding of the brain will lead to the creation of truly intelligent machine. The key issue is that machine intelligence meets neuroscience. Intelligence science claims that research on intelligence not only simulate natural intelligence in function, but also should research on mechanism, which is a new idea for intelligence research.

Keywords Intelligence, Neuroscience, Intelligence science, Turing test, Physical symbolic hypothesis

1 引言

人类从磨擦起火把机械能转变为热能,瓦特发明蒸汽机,又把热能转变为机械能,这是人类历史上两次最伟大的创举,创造了人类前文明史。智能革命将开创人类后文明史。与能量革命实现能量的转换与利用不同,智能革命实现智能的转换与利用,即人把自己的智能赋予机器,智能机把人的智能转换为机器智能,并放大人的智能;人又把机器智能转换为人的智能,加以利用。如果说蒸汽机魔术般地创造了工业社会,那么智能机也一定能奇迹般地创造出智能社会^[4]。

智能的概念是什么,它从哪里来?从某种程度上来看,我们是由一些按预定的方式运作着的物质组成——最低限度地定义一台机器——这已经很清楚,机器可以有智能。如果我们试图要拼凑智能机器现已存在的证据,为什么不足够理智地去制造一台不是基于碳的人脑?

人工智能从1956年提出以来取得了很大的进展和成功,赋予机器某种智能。无论这种机器展示的是何种智能,它们有可能都不同于人类的智能。类似地有,人造飞行器不能像鸟儿一样飞翔,船的推进器也不能模仿鱼在水中航行。

我们想要实现的人造智能似乎是无边无际的。例如,当程序员编写一段计算机程序的时候,他可能会产生很多错误。为什么不让一台有智能的机器无论在何时都可以写出完美的程

序?这些机器还可以写诗、故事、书籍、还有音乐。如果给定它们化学规则,它们还可以产生制造新药的化学公式,它们可以解决 Fermat 最新的定理,思考物理界重大的一统理论,告诉我们谁会藏在草木繁盛的山上。这是人类智力多么大的扩展!

对人工智能的研究使人们学到了很多有关智能方面的知识。智能不是约束计算机按有利的方式来执行和阐明任务,而是脱离原有的目标产生新的智能。也许有一天我们最终会达到目标,但是伴随着时间的流逝,这个目标好像离我们越来越远。

2004年10月,掌上电脑的先驱者霍金斯(Jeff Hawkins)闯进了人工智能的领域并使该领域萌生新机。他新书封面的标题是《On Intelligence》,提出了对大脑新的理解,这将导致产生真正的智能机器^[1]。匆匆看一下封面将立刻使自从1977年以来没有受到多少刺激的神经元苏醒,让霍金斯来使它们再次开始是非常有趣的。十几年来,霍金斯就一直是神经系统科学狂热的爱好者,但是直到今日,才投入足够的时间对该主题进行认真的思考,并且将他的观点付诸笔墨。

除了提供对人工智能和神经系统科学领域的回顾以外,霍金斯还描述了他自己有关大脑皮层如何产生智能和智能机器的本性等观点,这将使我们发展神经连通性的模式和操作(也许是他自己)并找出适合实现这种模式的硬件的途径成为

*)基金项目:国家自然科学基金(No. 60435010, 90104021);国家重大基础研究项目(973)(No. 2003CB317000)。史俊 硕士研究生,主要研究方向为智能科学,图像理解,视觉信息挖掘,模式识别。史忠植 研究员,博士导师,主要研究领域为智能科学,分布智能,知识工程,机器学习,认知信息学。

可能。

大脑皮层是一个极其复杂的系统。认知神经科学和脑科学的研究成果表明,脑的感知部分,包括视觉、听觉、运动等,不仅具有输入/输出通道的功能,而且直接与思维活动有关。一个智能系统,不仅要由推理机和知识库组成,而且要研究感知通道的作用,从机理上模拟人脑的功能^[5]。

2 图灵测试

英国数学家图灵(A. M. Turing)于1950年提出了一个测试标准,来判断电脑能否被认为是“能思考”。这个测试被称为图灵测试,现在已被多数人承认。

要分辨一个想法是“自创”的思想还是精心设计的“模仿”是非常难的,任何自创思想的证据都可以被否决。图灵试图解决长久以来关于如何定义思考的哲学争论,他提出一个虽然主观但可操作的标准:如果一台电脑行为(act)、反应(react)和互相作用(interact)都和有意意识的个体一样,那么它就应该被认为是有意意识的。消除人类中心的偏见,图灵设计了一种“模仿游戏”,即现在说的图灵测试:远处的人类测试者在一段规定的时间内,根据两个实体对他提出的各种问题的反应来判断是人类还是电脑。通过一系列这样的测试,从电脑被误判断为人的几率就可以测出电脑智能的成功程度。

图灵测试只是说明了人类思维的一个方面,即使通过了图灵测试也不意味着电脑能够思考。

3 物理符号系统

1976年纽威尔和司马贺提出了物理符号系统假设^[6],说明物理符号系统的本质。用物理符号系统主要是强调所研究的对象是一个具体的物质系统,如计算机的构造系统、人的神经系统、大脑神经元等。所谓符号就是模式;任何一个模式,只要它能和其它模式相区别,它就是一个符号。不同的英文字母就是不同的符号。对符号进行操作就是对符号进行比较,即找出哪几个是相同的符号,哪几个是不同的符号。物理符号系统的基本任务和功能就是辨认相同的符号和区分不同的符号。

物理符号系统主要假设内容如下:

(1) 物理符号系统假设:物理系统表现智能行为的必要和充分条件是它是一个物理符号系统。

(2) 必要性意味着表现智能的任何物理系统将是一个物理符号系统的示例。

(3) 充分性意味着任何物理符号系统都可以进一步组织表现智能行为。

(4) 智能行为就是人类所具有的那种智能;在某些物理限制下,实际上所发生的适合系统目的和适应环境要求的行为。

由此可见,在物理符号系统假设中,既然人具有智能,他就是个物理符号系统。人类能够观察、认识外界事物、接受智力测验、通过考试等,这些都是人的智能的表现。人所以能够表现出智能,就是基于他的信息加工过程。这是由物理符号系统的假设得出的第一个推论。第二个推论是,既然计算机是一个物理符号系统,它就一定能表现出智能,这是人工智能的基本条件。第三个推论是,既然人是一个物理符号系统,计算机也是一个物理符号系统,那么我们就用计算机来模拟人的活动。我们可以用计算机在形式上来描述人的活动过程,或者建立一个理论来说明人的活动过程。

纽威尔和司马贺的物理符号系统假设架起了精神世界和

物质世界的桥梁,为符号智能的发展奠定了理论基础。但是,物理符号系统假设也只是从功能上论述实现机器智能的可能途径。

4 机器智能需要神经科学

霍金斯从小心地区分智能和智能行为相互关联的地方入手,对著名的图灵测试进行解析。智能无论如何也不能使智能的主体认为它们也是智能的。图灵的提议给了深远的洞察力,但是随着时间的流逝,霍金斯逐渐地认识到这是平庸的并且毫无用处的,更多的是逃避而不是有用的定义^[2]。

霍金斯叙述了塞尔(John Searle)的“中国屋”的试验,这个试验在房间内放一个智能主体,他可以在房间的任何地点不假思索地遵循机械的指令,这将导致他正确地回答一系列问题(这些问题是用该主体不能理解的中文写成的)最终走出房间。

塞尔指出,中文输出的接受者推断在房间内的主体一定理解中文,可能,甚至展示发人深省的洞察力。但是主体自己什么事也不做,——机器也做不了——他只是简单地遵循由某些人写成的程序罢了。

塞尔说他自己不知道智能是什么,但是这个试验表明计算机并不具有智能。塞尔的实验清楚地说明图灵测试很容易被愚弄。我们要对智能做更多的努力而不是仅仅进行一些主观的判断。

另外一些人认为主体、指令、还有房间的结合共同反映了智能——尽量避免我们不得不面对的对智能并不了解很多的事实,尽管我们在许多科学分支上面作了长期的大量工作。

霍金斯相信智能是大量群集的神经元涌现的行为,用基于记忆的世界模式产生连续不断的对未来事件的一系列预测。他认为时间是大脑做什么和怎样做最至关重要部分。他相信大脑的行为有三个至关重要的组成部分:大脑工作于输入的时间序列流,涉及许多的反馈(有证据表明神经网络以这种方式在大脑中构成),真实网络的层次结构所具有的模式,看起来对它们的功能十分重要。

大脑本质上是独立于我们肢体和感觉的智能设备,他提醒我,大脑是通过感觉觉察并且思考外部世界的唯一途径。大脑的每一个部分都不断地传递着对外部世界实时度量的相关模式。我们考虑视觉,听觉还有触觉,它们是十分不同的,但是都将表达的模式序列传送到我们的大脑。人类是千百万年来进化的产物,我们的大脑和感觉以及运动控制紧紧结合为一体。我们很容易掉进图灵的陷阱,认为“如果看上去是智能的,那他就是智能的”。

霍金斯提到了克勒(Helen Keller)这个例子。尽管她又聋又瞎,但是她学会了语言并且成为一名优秀的作家。是的,我们的大脑利用拍电报的方式帮助我们产生经由视觉和听觉等传感器传输到大脑的感觉。但是智能很明显地可以在没有那些感觉的情况下产生。

霍金斯建议把神经皮层的神经网络当作分布式存储的模式序列,以不变的形式联想访问和存储,并且分层次地进行安排。这种模型给了我们巨大的感觉。它可以帮助我们解释为什么那么多聪明的人都会按照类似的思维模式进行思考。他们可以看到,在看起来完全不同的事务底下的模式。当他们在无意识的情况下进行关联检索的时候,有许多表面上看来令人非常吃惊的相关性。按照霍金斯的预料,这些人可能存储了专

(下转封四)

(上接第 110 页)

一的不变的表达概念方式,并在必要的时候采用它形成大量的令人震惊的引人注目的想法。我们经常感觉到,只有少数的想法是计算机学科独一无二的。例如,我们现在学习的大部分仅仅是已经储存在我们脑海中那些概念的新名词,只是我们是否把它们联系起来。

为了帮助理解他有关神经网络存储的序列模式,霍金斯建议读者考虑那些偶然在他们身上发生的事,并且设想把这些故事告诉他们的朋友。

这个训练揭示了故事必须被当作一个从头到尾的序列来讲述。你可以从中间开始,但是不能反向地讲述故事。每一个状态“唤醒”下一个状态。这也是那些能够凭借记忆演奏长而且复杂音乐的音乐家们的真理。但是你让他们从中间部分开始,他们将陷入困境。

霍金斯也指出无论我们是朗诵一篇存储在记忆中的故事,还是将它打出来,或写出来,我们都是按同一个不变的方式把它们从记忆的仓库里取出来。那些信息与大脑的其他功能一起,帮助我们完成握手,适当地动嘴说话、朗诵,以及移动手臂和胳膊来写作等事情。

依照霍金斯的理论,有意性——一个智能的特点——好像很适合他所提出的所有神经网络的基本功能就是形成连续的预测,并且正确地存储通过感官所获得的世界观这一模式。

5 智能研究的新思路

2003年史忠植发表了“展望智能科学”一文。他指出“智能科学研究智能的基本理论和实现技术,是由脑科学、认知科学、人工智能等学科构成的交叉学科。脑科学从分子水平、细胞水平、行为水平研究人脑智能机理,建立脑模型,揭示人脑的本质。认知科学是研究人类感知、学习、记忆、思维、意识等人脑心智活动过程的科学。人工智能研究用人工的方法和技术,模仿、延伸和扩展人的智能,实现机器智能。三门学科共同研究,探索智能科学的新概念、新理论、新方法,必将在21世纪共创辉煌”^[3]。史忠植的专著《智能科学》将系统地论述智能科学的内容,讨论智能研究的新思路。

学习是基本的认知活动,是经验与知识的积累过程,也是对外部事物前后关联地把握和理解以便改善系统行为的性能的过程。学习的神经生物学基础是神经细胞之间的联系结构

突触的可塑性变化,已成为当代神经科学中一个十分活跃的研究领域。突触可塑性条件即在突触前纤维与相联的突后细胞同时兴奋时,突触的连接加强。

人类进化过程中,语言的使用使大脑两半球功能分化。语言是最复杂、最有系统、而应用又最广的符号系统。语言符号不仅表示具体的事物、状态或动作,而且也表示抽象的概念。汉语以其独特的词法和句法体系、文字系统和语音声调系统而显著区别于印欧语言,具有音、形、义紧密结合的独特风格。从神经、认知和计算三个层次上研究汉语,给予我们开启智能之门极好的机遇。

工作记忆系统能同时储存和加工信息,这和短时记忆概念仅强调储存功能是不同的。工作记忆分成三个子成分,分别是中枢执行系统、视空初步加工系统和语音环路。大量行为研究和神经心理学上的许多证据表明了三个子成分的存在,有关工作记忆的结构和作用形式的认识也在不断地丰富和完善。人们发现工作记忆与语言理解能力、注意及推理等联系紧密,工作记忆蕴藏智能的玄机。

意识也许是人类大脑最大的奥秘和最高的成就之一。研究意识可以将觉知(awareness)和非觉知作为切入点,找到神经相关物在脑活动中的区别。

21世纪,人类将在智能科学领域取得突破性进展,这是极大的机遇和挑战。我们要开拓创新,研究人类智能的本质和规律,探索人类的智力如何由物质产生和人脑信息处理的过程。在攀登智能科学高峰崎岖曲折的道路上,不畏艰险,不怕困难,对人类文明进程作出我们的贡献。

参 考 文 献

- 1 Hawkins J, Blakeslee S. On Intelligence. Times Books, 2004
- 2 Colwell B. Machine Intelligence meets neuroscience. Computer, 38(1), 2005, 12~15. 史俊译. 机器智能需要神经科学. 中国计算机学会通讯
- 3 Shi Zhongzhi. Perspectives On Cognitive Informatics. IEEE ICCI'03, 2003: 129~136
- 4 童天湘. 智能化是信息化发展的必然趋势. 见:智能科学技术基础理论重大问题研讨会, 2004. <http://www.intsci.ac.cn/>
- 5 史忠植. 展望智能科学. 科学中国人, 2003(8): 47~49
- 6 史忠植. 智能科学. 清华大学出版社, 2005

计算机科学

(1974年1月创刊)

第32卷第6期(月刊)

2005年6月25日出版

ISSN 1002-137X
CN50-1075/TP

定价: 25.00元 国外定价: 5美元

邮发代号: 78-68

发行范围: 国内外公开

主管单位: 国家科学技术部

主办单位: 国家科技部西南信息中心

编辑出版: 《计算机科学》杂志社

重庆市渝中区胜利路132号 邮政编码: 400013

电话: (023) 63500828 E-mail: jsjcx@swic.ac.cn

网址: www.jsjcx.com

社 长: 牟炳林

主 编: 彭 丹

副 主 编: 朱宗元

主编助理: 徐书令

印刷者: 重庆科情印务有限公司

总发行处: 重 庆 市 邮 政 局

订购处: 全 国 各 地 邮 政 局

国外总发行: 中国国际图书贸易总公司(北京399信箱)

国外代号: 6210-MO