

智能系统的广义智能定性评价之研究^{*})

刘东 尹怡欣 涂序彦

(北京科技大学信息工程学院 北京 100083)

摘要 本文专门研究智能系统的智能特性和智能水平评价,旨在促进其健康发展。采用的主要方法是:在图灵测试的基础上,借鉴了人的智商测试方法,运用了广义人工智能的广义思想,提出了广义智能评价理论体系。该理论体系包括了一个关于广义智能评价的定义、特指系统的性能和特指系统技术的两个广义智能标准、评价产品或系统智能时使用的三个广义规范,提出了广义智商算法、广义智能评价的定性和定量评价方法及评价规则,它留下的广义空间可包容不同类型的智能系统作为其评价对象。在实施评价的过程中,采用共享法、仿真法和人机结合等测试方法,举例说明了对典型智能控制系统的智能水平评价和对智能移动机器人自识别特定对象的智能特性评价结果。

关键词 智能控制,智能特性,定性评价,广义智能测试和评价

Research on Qualitative Evaluation of Generalized Intelligence in Intelligent System

LIU Dong YIN Yi-Xin TU Xu-Yan

(Information Engineering School, Beijing University of Science and Technology, Beijing 100083)

Abstract In this paper, we research exclusively on the intellectual characterization and level of intelligent systems in order to promote the healthy development of intelligence evaluation. The main method we use is on the base of Turing test; bring forward the evaluation theoretical system of generalized intelligence while considering the methods to test human intelligence and applying the generalized conception of artificial intelligence. This system is composed of a definition of general intelligence evaluation, two generalized intelligence standards of capability and technique of a certain system, three generalized criteria when evaluating the intelligence of products or systems, putting forward the calculation method of generalized intelligent, the qualitative and quantitative evaluation methods and rules of generalized intelligence. It concludes different intelligent systems as its goals to evaluate. In the process of evaluation, we apply the method of sharing, imitation and combination of human and computer etc. and we illustrate the evaluation result of intelligence characteristics of typical intelligent control system and self recognition of cyber.

Keywords Intelligence control, Intelligence characteristics, Qualitative evaluation, Generalized intelligence test and evaluation

长期以来,人们对“智能”的解释众说纷纭:有的人把含有电脑 CPU 的系统称作智能系统,有的人把某些自动化的系统称作智能系统,有的人把带有信息采集和处理功能的系统称作智能系统。于是,各种智能仪表、智能器械、智能设备、智能装置、智能产品、智能电器、智能武器、智能系统纷纷涌入市场。在这其中,确有真正“智能”的存在,但也有某种程度上对“智能”误解的存在。

我们认为,电脑化、自动化、信息化不等同于智能化,它们只是智能化的基础、条件和手段。电脑化是使传统的产品或系统含有电脑,电脑可以用来实现计算等功能。自动化是指产品或系统在没有或很少有人参与的情况下,按着人的要求,采取自动控制或自动运行的方式,完成其承担的任务。信息化则体现在系统信息的采集、传输、处理及应用上。在电脑化、自动化和信息化当中不一定都含有智能的成分。

然而,什么是智能化,目前还需要研究,才能澄清尚存的一些模糊概念,正确地认识“智能”和正确地评价智能系统。我们认为,智能化的产品或系统应从两个方面来理解:一是指

它具有了某些仿人或拟人的智能特性,二是指它采用了人工智能的理论、方法和技术。

因此,针对智能系统的智能进行评价研究,对澄清其概念,促进其发展是十分必要的。这不仅是学术的需要,也是应用的需要。

1 前人的工作

前人在对机器的智能测试和对人的智商测试方面做过两项有益的工作:一是针对机器的图灵测试,二是针对人的智商测试(IQ)。

图灵测试是采用了人工智能的理论和方法,对机器智能进行的测试。1950年,美国数学家图灵提出了这个测试标准,具体的做法是通过图灵设计的著名“模仿游戏试验”,将被测试的机器和陪测人分别隔离在两间屋子里,开始相互提问。如果陪测人在不知情的条件下,在相当长的时间里,分不清与他交谈的对象是人还是机器,那么这台机器就通过了图灵测试,于是认为该机器在这个问题上具有了与人同样的智能。

^{*})国家自然科学基金(基金编号:60374032)、北京市教委科技发展计划面上项目(项目编号:KM200511417011)。刘东 博士生,教授,主研方向:人工智能、智能控制;尹怡欣 教授,博导,主研方向:人工智能、人工生命、智能控制;涂序彦 教授,博导,中国人工智能学会荣誉理事长,全球华人智能控制与智能化大会主席,世界专家系统大会远东区主席,主研方向:人工智能、人工生命、大系统控制。

现在,著名的图灵测试已被大多数人承认。

IQ测试是采用分类测试、数理统计的方法,对人的智商进行测试,上世纪初,由德国心理学家W·斯特思最早提出。50年代后期,美国心理学家D·韦克斯勒又提出离差智商(IQD),具体做法是将数学、语言、认知等各类测试题目汇编成测试量表,让被测儿童或成人答题,正确回答的题目越多,被测人的积分越高。通过对某个国家或某个地区进行的大量测试数据进行统计,得出该国家或地区的标准值。在实施个体测试时,将被测人的积分与标准值进行比较计算,从而计算出被测人的离差智商值,并得出其智商所处的高低程度。例如,韦氏量表测试的结论分为:离差智商值在130以上者为非常优秀;120~129为优秀;110~119为中上(聪明);90~109为中等;80~89为中下(迟钝);70~79为临界状态;70以下为智力缺陷。IQ测试提出至今已在美国、法国等许多发达国家中使用。在我国各地区,尤以对儿童IQ测试的应用较为普遍;对成人的IQ测试也有所研究和应用,例如某些单位在招聘时就使用了IQ测试。

除此之外,人们针对系统的稳定性、快速性、可观性、能测性、鲁棒性等静态和动态性能测试方面进行了许多的研究和应用。

上述种种测试方法对我们研究智能系统的智能评价很有启发,但不能直接套用。其主要原因如下:一是智能系统不是人,不能将对人的IQ量表用于机器和系统的测试。二是除图灵测试外的以上各种系统测试方法都不是针对机器智能或系统智能的测试,故不在本文的讨论范围之内。三是图灵测试没有明确地把智能特性区分开来进行测试。四是图灵等测试不能识别出机器或系统中是否采用了人工智能的理论、方法和技术。五是图灵测试不进行智能水平的测试,即它不把机器或系统的智能划分出等级。因此,我们要在前人工作的基础上,进行新的、专门的研究,提出针对智能系统智能特性与智能水平的测试及评价的理论、方法和技术,解决尚存的问题,把智能系统的智能评价研究与应用向前推进。

2 广义智能评价的理论体系

智能系统是将以前的传统系统与人工智能相结合的产物,例如智能控制系统、智能管理系统等^[1,2]。著名的美国斯坦福大学人工智能研究中心尼尔逊教授对人工智能的定义是:“人工智能是关于知识的学科—怎样表示知识以及怎样获得知识并使用知识的科学。”美国麻省理工学院的温斯顿教授认为:“人工智能就是研究如何使计算机去做过去只有人才能做的智能工作。”即人工智能是研究人类智能活动的规律,构造具有一定智能的人工系统,研究如何让计算机去完成以往需要人的智力才能胜任的工作,也就是研究如何应用计算机的软硬件来模拟人类某些智能行为的基本理论、方法和技术。北京科技大学的涂序彦及尹怡欣先生先后提出和论述了广义人工智能和人工生命及应用的思想^[3]。我们认为,智能系统是指那些含有拟人的智能特性、应用了人工智能的理论、方法和技术系统。

然而,我们如何实现对智能系统的智能进行测试和评价呢?为此,我们提出了智能系统的广义智能评价理论体系,包括广义智能评价的一个定义、二个标准、三个规范。

定义 广义智能评价是专门针对机器或系统进行的智能测试和评价。它在图灵测试的基础上,借鉴人的智商测试方法,运用广义算子模型进行系统描述,从性能和技术两个标准

出发对系统进行测试,按照广义智商算法对系统智能进行评价,从而实现对系统智能特性的定性评价和对系统智能水平的定量评价。

标准1 特指系统的性能标准。如果一个系统具有自学习、自适应、自组织、自寻优、自镇定、自整定、自识别、自规划、自协调、自修复、自繁殖、自联想、自推理、自感知、自认识、自诊断等智能特性之一,那么该系统就是一个智能系统。

标准2 特指系统的技术标准。如果系统采用了专家系统、人工神经网络等人工智能的技术方法之一,那么该系统就是一个智能系统。

广义规范1 如果系统具有的智能特性多,AND、采用的人工智能技术方法多 THEN,其智能水平就高。

广义规范2 如果系统具有的智能特性较多 AND、采用的人工智能技术方法较多 THEN,其智能水平就较高。

广义规范3 如果系统仅具有智能特性 OR、仅采用人工智能技术方法 THEN,系统也有智能,只是智能的等级不同而已。

符合规范1和2者称为“双标准”产品或系统。符合规范3者称为“单标准”产品或系统^[4]。例如,具有自识别、自适应等特性、采用人工神经网络等技术的自主智能汽车就为“双标准”智能系统。再如,以前的自适应、自校正控制系统中没有使用后来的人工智能技术方法,但系统具有了自适应的智能特性;传统的PID控制系统中不具有自学习特性,但可能具有PID参数自整定的智能特性,它们都属“单标准”系统。

因此,广义智能评价理论体系在前人对智能测试的相关理论和方法的基础上,突出了对智能特性的区分和对智能水平的测度,包括了对人工智能技术方法运用状况的考察,留下了广义的空间来容纳对目前已有的和将来可能有的智能系统进行智能测度和评价。由于广义算子模型本身就可具有多层次性、分布性等特点,因此系统的多层智能、多智体、分布智能和网络智能的评价等都可以包容在广义智能评价的体系框架之中。

3 广义智商算法与广义智能评价方法

3.1 广义智商算法

广义智商算法在对系统的智能水平进行定量评价时使用。使用广义智商算法之前,必须先获得系统个体在项目测试时的各项积分。广义智商算法如下。

计算个体得分F值:

$$F = \sum W_i V_i \quad (1)$$

式(1)中,F为系统个体的总积分, W_i 为权值, V_i 为各项测试积分。

计算产品或系统的广义智商IQ_G:

$$IQ_G = M + \frac{F-B}{S} \quad (2)$$

式(2)中,M为中等智商值,对于“双标准”的产品或系统,取M=100,“单标准”的取M=80。

式(2)中,B为团体平均值:

$$B = \frac{1}{N} \cdot \sum_{j=1}^N F_j \quad (3)$$

其中,N为参加测试的个体数量, F_j 为第j个个体的测试积分。

式(2)中,S为同一类产品或系统测试得分的标准差,由下式计算:

$$S = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^N (F_i - B)^2} \quad (4)$$

3.2 广义智能评价方法

广义智能评价方法分为针对系统智能特性的定性评价方法和针对智能水平的定量评价方法。

3.2.1 广义智能评价的定性评价方法

广义智能评价的定性评价方法为“预、测、评”三步法。具体含义如下。

(1)预:对某被测产品或系统进行智能特征描述、评价模型搭建等预先准备。

(2)测:按照广义智能评价标准和规范³,进行智能特性和技术特性的有无测试。

(3)评:根据测试结果,做出被测产品或系统有无智能的定性评价。

3.2.2 广义智能评价的定量评价方法

广义智能评价的定量评价方法为“预、测、算、评”四步法。具体含义如下。

(1)预:对某被测产品或系统进行智能特征描述、评价模型建立等预先准备。

(2)测:按照广义智能评价标准和规范,对相应测试项目进行测试,同时记录积分。

(3)算:根据测试得到的积分,按照广义智商算法,计算出被测产品或系统的广义智商(IQ_G)值。

(4)评:根据算出的 IQ_G 值,按照广义智商定量评价规则,评价出被测产品或系统的智能水平所在的等级。其中,广义智商定量评价规则如下:

BEGIN

IF IQ_G 值低于中等智商 M THEN 其智能水平为 * 级

IF IQ_G 值等于或接近中等智商 M THEN 其智能水平为 * * 级

IF IQ_G 值高于中等智商 M THEN 其智能水平为 * * * 级

IF IQ_G 值很高于中等智商 M THEN 其智能水平为 * * * * 级

产品或系统得到的“*”越多,其智能水平越高。

END

对于“低于、等于或接近、高于、很高于”的界定由专家经验确定。

4 评价的实施

无论进行定性评价还是进行定量评价,都要经过相关的项目测试。常用的测试方法有黑盒法、白盒法、灰盒法及经验法。测试结果的获得方法有共享法、仿真法、人工测试法、人机结合测试法、自动测试法及以上方法的相互混合法。其中,共享法是指对于已有产品或系统鉴定结论中包含的智能相关参数值,评价时可采用直接使用。

本课题组在文[5]中采用共享法和仿真法介绍了针对两个智能控制系统智能水平进行定量评价的智商计算举例,在文[6,7]中也进行了相关讨论。此外,本课题组还通过对智能移动机器人对特定对象自识别特性的测试进行了广义智能定性评价的应用。

智能移动机器人可以具有自学习、自识别、自适应等多种智能特性。本课题在实验室中以中科院自动化所高创中心研制开发的自主移动机器人平台为实验平台,对自识别特性进行了实验设计及实施评价。实验室中共有两台机器人,其中机器人 A 配备了自识别的软硬件,机器人 B 没有配备。实验得出的定性评价结论是:机器人 A 具有了自识别特定对象的智能特性,如图 1 所示。



图 1 机器人 A 的自识别实验

结论与展望 综上所述,得出以下几点:

(1)广义智能评价理论体系及评价方法的提出是本课题组就智能系统智能水平评价研究的阶段性结果,欲继续用于指导和开展对智能系统的智能特性和技术特性的定性评价和系统智能水平的定量评价。

(2)广义智能评价在图灵测试的基础上,借鉴了人类智商测试的方法,发扬了广义人工智能的广义思想,提出了广义智能评价的定义、标准、规范、方法及广义定量评价中用到的广义智商算法和评价规则。

(3)广义智能评价的广义特征,使得它可包容不同类型的智能系统为其评价对象,这样做的思考是,力求有益于当前智能系统的发展和智能评价的起步。

(4)本课题组运用共享法、仿真法和人机结合等方法对特定智能控制系统的智能水平和智能移动机器人的自识别特性进行了广义智能评价。

(5)广义智能评价有待于不断设计和积累针对各类智能系统的测试用例。

参考文献

- 涂序彦. 智能控制的理论、方法与技术. 见:第二届全国智能控制专家讨论会论文集(上), 清华大学, 1994. 27~34
- 涂序彦. 智能控制及其应用. 自动化学报, 1977(1)
- 尹怡欣, 涂序彦. 人工生命及应用. 北京:北京邮电大学出版社, 2004. 88~95
- Wang Wei, Xiao Du, Jun Ping, et al. The Study on the Criterion and Standard of Intelligent System, Oct. 2004. 27~30
- 刘东, 尹怡欣, 涂序彦. 一种智能控制系统智能水平的评价方法. 中南大学学报, 2005(专辑). 13~16
- Liu Dong, Yin Yixin, Dong Jie, et al. Calculating Method of Evaluating the Intelligence Quality of Intelligent Control Systems. In: 2006 IEEE International Symposium on Intelligent Control, Munich, 2006
- Liu Dng, Yin Yixin, Tu Xuyan. A System Of Evaluating The Intelligence Quality of ICS. In: Third International Conference on Impulsive Dynamical Systems and Applications, Qing Dao: Ocean Univ of China, 2006