先天综合判断观照下的深度增强学习:以AlphaGo Zero为例*

郑炳楠[1],贺威[2]

(1.南京大学;2.厦门大学)

摘　要:深度增强学习的出现引发了诸多关于人类思维与人工智能的思考。AlphaGo Zero的深度增强学习促使我们分析机器拥有先天综合知识的可能性。康德的先验感性综合理论对深度增强学习的观照体现在计算机的数学基础上,其感性纯直观中关于时间的部分与纯数学中的代数相对应,而二进制运算的过程体现了先天与经验的综合。先验知性综合理论对深度增强学习的观照体现于AlphaGo Zero围棋技能的获得过程。知性综合判断分为从简单到复杂的三个阶段,分别对应深度增强学习算法运行时由基础到综合的决策过程,从每个单位上的函数计算到函数之间组成的正负反馈,再到最终形成决策的通用智能,体现了这一程序在先天和经验上的综合性。然而,深度增强学习中很难存在先验理性。

关键词:先天综合判断;深度增强学习;AlphaGo Zero;人工智能

DOI:10.16397/j.cnki.1671-1165.2019001060

2017年10月19日,由谷歌的下属公司Deepmind研发的新版程序AlphaGo Zero的研究成果通过国际学术期刊《自然》揭开了面纱。AlphaGo Zero这一程序可以从空白状态开始,在没有任何人类棋谱输入的状态下自我对弈,自学围棋。并且,AlphaGo Zero刚出世不久就一举击败了AlphaGo,创下了惊人的记录,其能够迅速达到这个水平就在于采用了新的强化学习系统Reinforcement Learning,也就是Deep Reinforcement Learning(深度强化学习)。由于这一学习方法不需要人类输入,仅通过与自身比赛便在三天时间内达到了足以击败原版AlphaGo的水平,于是类比人类思维的产生和该程序的成功运行,学术界展开了关于"先验知识是否存在"的新一轮讨论。

康德在先天综合判断理论中将人的思维分为先验感性、先验知性和先验理性三个部分,并开创性地把先验与后天经验相结合,阐释了人形成知识的过程。基于AlphaGo Zero的性能,以在深度增强学习方面较有代表性的AlphaGo Zero为例,具体分析康德的先天综合判断理论,寻求对于人类思维及人工智能研究的新的思考方向成为可能,同时可以尝试解决深度增强学习中先验感性与先验知性如何存在、深度增

*　收稿日期:2018-11-12

作者简介:郑炳楠,南京大学哲学系、南京大学科学技术与社会研究所硕士研究生;贺威,厦门大学哲学系副教授,研究方向:科技哲学。

强学习中是否涉及先验理性等问题。

一、深度增强学习中的先验感性：计算机的数学基础

康德在先验感性学说中首先解释了，知识和外界对象借直观来发生直接关系，同时直观也是作为手段的思维的追求，而感性则是“通过我们被对象刺激的方式获得表象的能力”[1]。人对外界事物的认识要立足于受到的来自外界对象的刺激，也就是对象对表象能力发生了作用，即感觉。通过感觉所产生的现象是经验的，并且由于现象还没有被范畴规定，所以呈现杂多的特性。与感觉相应的现象的质料是后天给予人的，是经验的；反之有别于质料的形式就是先天蕴含在人的心灵之中的，是先验的、纯粹的，是感觉有序化的唯一途径。

当今科技水平下对于人工智能的研究正在致力于对人的思维进行更加精确的模仿、延伸和扩展，在这一方面，AlphaGo Zero通过深度增强学习这一优化项将人类给予的信息在程序运行时所占的比重降至更低，不需要人类棋谱的输入，形成了较为完善的闭合的决策体系。另一方面，AlphaGo Zero创造了在围棋这一专项领域击败人类思维的记录。所以，本文将深度增强学习这一程序放入AlphaGo Zero中进行说明，并分析其与人的思维同构的部分是可行的。AlphaGo Zero可以被视为存在先天综合判断中所提出的外界世界所给予的“经验”和其自身所先天拥有的部分，这一点以数学作为桥梁在感性直观的先验纯形式的部分得以体现。后文将通过AlphaGo Zero的部分计算原理对此加以详细说明，所涉及的数学公式、字母和符号均引用于David Silver和J. Schrittwieser等人在《自然》杂志上发表的 *Mastering the game of Go without human knowledge* 一文。

AlphaGo Zero需要处理的问题是围棋问题，即在棋盘的交叉点上落子的问题。围棋棋盘上总共有19×19=361个交叉点，每个点都有三种状态。在程序中，交叉点上的白色子用1表示，黑子用-1表示，无子用0来表示。那么，用$\vec{s}$表现此时棋盘的状态，即棋盘的状态向量记为$\vec{s}$，此时用公式[2]可表示为：

$$\vec{s}=\underbrace{(1,0,-1,\cdots)}_{361}\quad 。\tag{1-1}$$

那么在状态$\vec{s}$下，当可以落子时，下一步的落子向量$\vec{a}$便将是一个361维的向量[2]，即

$$\vec{a}=(0,\cdots,0,1,0,\cdots)\quad 。\tag{1-2}$$

这一程序运转的最终目的就在于在给定的任意一个状态$\vec{s}$下，寻找最优的$\vec{a}$的解，以在棋盘上取得最大的地盘。所以可以在康德的先天综合判断观照下以AlphaGo Zero作为例证，从围棋问题出发，寻找深度增强学习拥有先天综合知识的可能性。

在康德的理论中，感性纯直观是感性直观的纯先验形式，只包含时间和空间。在纯数学领域，时间与算学相对应，空间与几何学相对应。在《未来形而上学讨论》中，

康德提出:“几何学是根据空间的纯直观的;算学是在时间里把单位一个又一个地加起来,用这一办法做成数的概念。”[3]即算学和几何学都是时间和空间这一感性纯直观的表象,但同时也作为算学和几何学的形式,构成了数学命题中的先验部分。

AlphaGo Zero是在AlphaGo基础上进行了改良的计算机程序,其特色的深度增强学习更是程序运行的一个重要部分。而计算机程序本质上是通过指令的顺序,使计算机能按所要求的功能进行精确记述的逻辑方法,是以计算机为客观基础的数学逻辑方法,这一点从对于AlphaGo Zero所需要解决的问题阐释中也不难看出。计算机系统建立于二进制的数学系统之上,以补码的形式储存数据,用1和0来表示“开”和“关”,比如用1、-1和0表示白子、黑子和无子三种交叉点的状态。这些都属于算学,且是在算学的基础上展开的,所以在深度增强学习这一程序运行时,时间在整合了数字的运算等质料之后使二进制数学体系得以成型,并因此维持了计算机系统的运作。时间规范的是二进制运算中的最高原理、先验部分,其形成的算学的先验性公理是算学得以可能的条件,例如数学中的一致性等;同时二进制算法又是计算机的本质构成,所以这一先验部分也就成为了计算机程序运行的必要条件。计算机程序运行无法脱离其单独存在,而这种公理也蕴含在程序之中以彰显其作用。

综上所述,AlphaGo Zero程序将接收到的外部信息和刺激全部转化为以2为基数的数字,然后进行运算、储存等操作,其全部机能均建立于二进制体系之上。这体现了时间这一感性纯直观与纯数学之间的关系,以及若它在深度增强学习中存在则其存在方式和作用方式。

先验感性论探讨的是纯数学何以可能,实际讨论了数学如何把先验与经验在自身中结合的问题。康德认为数学判断本身是综合的,“人们由于看到数学家的推论全都是依据矛盾律进行的……,于是就使自己相信,数学原理也是出于矛盾而被承认的;他们在这里是弄错了”[4],数学中既有先验的部分,也包含经验的作用。例如在算学中1、2、3三个数字本身是有其先验基础的,但是1+2=3这一命题事实上无法单纯通过1+2推导出3的。1+2只是代表了两个数字的结合,而并没有指出其他的任何信息;其所得的结果也没有指明一定是3。之所以会得出1+2=3这个命题,一定是经历了经验的参与,也就是不能仅通过数字概念本身,而是要通过至少其中一个的直观来接近这个命题。

AlphaGo Zero的深度增强学习是一种运算体系,主要涉及的是代数的部分,所以将以代数的部分为切入口进行分析。与纯数学的推演方式同理,二进制体系作为数学运算体系中的一个部分,其运算也是需要依赖经验的。一方面,二进制其运算本身符合纯数学原理,其在接收外界信息转化为二进制数字和将二进制数字信号进行传递和再显示的过程当中都经过了二进制的简单运算,这些运算所得的结果不是数字概念本身所携带的先天必然,例如十进制中的2其自身并没有包含在二进制中用10

表示的内涵，而是依靠了经验得出；另一方面，计算机程序运行中必不可少的就是数据的储存，这一部分涉及的补码是一种范围更大且附加了更多运算条件的运算方式，例如要表示正数所对应的二进制负数就要经过将正数转化成二进制、所有位取反、加1这三个步骤。如在AlphaGo Zero程序中表示交叉点上的黑子状态则需要用到负数。-1的表示仍相对简单，若要表示-5则需要首先将其正数5转化为二进制数（00000101），所有位取反后得到二进制数（11111010），再加1之后得到-5的二进制表示(11111011)。数字越大，计算的过程越复杂，就越能得出数学系统中的经验性。

所以在AlphaGo Zero的深度增强学习中，数学基础是先天综合的，同时其先验部分对信息整合的过程也可以对应人的思维过程中的感性综合的部分——人通过感官接受到的关于外部对象的感觉经过感性纯直观时间和空间的整理形成感性直观。

二、深度增强学习中的先验知性：围棋技能的获得

（一）领悟直观综合与函数计算

人的心灵接受了表象后，便通过这一表象认识对象。这个对象被给予我们之后，还需要在表象和对象的关系中被再思维，才能有概念，再由概念组成知识。先验知性论的“A版演绎”从人的知性综合能力开始，认为感性直观所得到的表象是一种呈现于心灵的杂多，而知性则将其比较、分类、联结和整理，形成知识，完成综合。《康德〈纯粹理性批判〉指要》提出：“康德除考察人的感性之外，主要是研究了各种高级认识能力，即理性、判断力和知性，并认为只有知性的先天认识原理（由范畴体现的规律）才是‘构成性’的。”[5]

第一阶段是领悟直观的综合，即在心灵的杂多中产生出直观的统一体的过程。这种综合活动是非经验的，且遵从内感官的形式条件——时间，所以这种综合也是必要的，失去了领悟直观的综合则杂多无法成为一个统一体。

这一部分可以从组成深度增强学习的深度学习作为出发点进行分析。深度学习就是深度表征，可以理解为一个由很多函数组成的梯度，特定梯度最终指向一个既定的目标。当起始被给出，程序开始运行，通过函数计算等方式学习出有意义的向量以完成任务。这个过程中，原始的、无意义的起始信号进入到函数之中，首先成为二进制的数据，随后在某个函数中得到筛选、联结等初步的整理，使其成为适用于AlphaGo Zero这一软件的有价值信息。AlphaGo Zero应用了一个新的深度神经网络f_θ，其中的参数θ会通过训练不断得到调整。在网络输入之后，深度神经网络将整理出现在棋盘状态的向量$\vec{s}$以及包括现在的棋盘状态在内的8步历史落子记录。这些得到了初步整理的信息可以进一步进入深度学习网络，最终构成深度神经网络f_θ的预测，例如其网络输出的包含落子概率和评估值在内的一个函数，可表示为[2]：

$$f_\theta(\vec{s}) = (P, v) \quad 。 \tag{2-1}$$

其中$P_a = P_r(\vec{a}|\vec{s})$,$v$的阈值在[1,-1]之间。或者这些数据也可以参与强化学习的部分得到反馈。它们相比于原始输入更加系统、复杂,但也更具有统一性。

在AlphaGo Zero的深度增强体系中,负责深入学习的函数在逻辑上是先于原始输入而存在的。即函数在最初对数据进行的转化和处理是相对在先的,其形成并没有人输入棋谱数据的干涉,即网络输入并不能对函数本身以及函数之间的关系产生影响或是改变。

(二)想象再现综合与正负反馈

在第二个阶段中,康德认为,杂多经过第一阶段把握综合的活动得到了整理,随后那些经常共同出现或相继出现的表象之间会产生一种联结,能使心灵在一定条件下从一个对象向一个对象过渡。想象再现综合也具有先天性。

这一部分可以以深度强化学习中的强化学习部分来加以阐述。进行强化学习是为了使决策更加优化以达到最佳结果,且学习结果会影响到下一次的输入,所以本质上是一个闭环的系统。一个表示行为或状态函数在强化学习中会被评价,在评估这个函数的状态或者行为的好坏及其程度之后,得出正反馈或者负反馈,而如果目的是持续得到某种反馈,则代理函数将选择持续某种行为或动作。这种学习的逻辑模式实质上是一种联结。它将函数的行为或状态与某种正反馈或是负反馈联结在一起,进而使函数的行为或状态产生了某种倾向。AlphaGo Zero所应用的是自对弈强化学习算法,它在运行时,会在每一个现有的状态$\vec{s}$中,参照深度神经网络f_θ计算预测出的结果,执行蒙特卡洛搜索树算法,也就是MCTS搜索。MCTS搜索输出是每一个状态$\vec{s}$下,棋盘上不同交叉点所对应的落子概率π,且π是一个向量,其数学表示为[2]:

$$\pi_i = \frac{\overbrace{(a, b, c\cdots)}}{361} \quad 。 \tag{2-2}$$

其中$(a, b, c\cdots)$表示的是每一个交叉点上的落子概率。不同于最初运用随机来进行模拟的形式,AlphaGo Zero 1.0中使用的蒙特卡洛搜索树算法开始采用局面函数辅助策略函数作为落子的参考进行模拟。在模型中,每个状态$\vec{s}$下的落子选择都会对应三个结果数值:先验概率$P(\vec{s},\vec{a})$,访问次数$N(\vec{s},\vec{a})$和行动价值$Q(\vec{s},\vec{a})$。每一次的棋局模拟都以根状态作为开始,每次落子的结果均要进入最大化上限置信区间,其过程可表示为[2]:

$$Q(\vec{s},\vec{a}) + \mathrm{U}(\vec{s},\vec{a}) \quad 。$$

其中,

$$U(\vec{s},\vec{a}) \propto \frac{P(\vec{s},\vec{a})}{1 + N(\vec{s},\vec{a})} \quad 。 \tag{2-3}$$

直到分出胜负，棋局结束。局终之后，新的先验概率和评估值会被计算出来。例如在某一情况下，某一落子位置位于最大化上限置信区间，那么同一情况下选择这个位置落子的概率也会大大上升，下一次棋局模拟时会继续趋向于这种选择。

这种强化学习中的正负反馈联结也是由函数组成的，是不受外界数据干扰的，而且通过这一算法产生的结果组成了AlphaGo Zero的决策过程，导致了程序的最终决策结果，也就是成为了形成数据输出的基础。因此强化学习的特征也能够使深度增强学习在先天综合判断中被观照得以可能。

（三）概念认知综合与通用智能

概念的最终作用是给杂多一种统一性，即对某一事物产生稳定的认识、形成一种统一的表象。概念认知的综合实现在领悟直观综合和想象再现综合的基础上，使有序的、相联系的表象杂多形成稳定统一的表象。对于AlphaGo Zero而言，深度增强学习的运行也达到了类似的结果，即使数据不断趋于有条理，对落子的判断更加优化，力求达到最高的胜率，最终形成一套完善的对弈方法。

深度增强学习本质上提取了传统的深度学习和强化学习的优势，使用MCTS搜索进行自行对弈，在强化学习算法的迭代过程中训练深度学习网络，从而达到自主学习围棋的训练目的。最初，神经网络是随机初始化的θ_0，每一次自对弈都对应着一次迭代（$i \geqslant 1$）。当MCTS搜索至随机步数t时，$\pi_t = \alpha\theta_{i-1}(s_t)$，其数据被存储为$\vec{s}_t$、$\pi_t$、$z_t$。$z_t$表示的是$t$步的获胜者，即$z_t = \pm r_T$，其中$r_T$是在$T$步遇到双方都选择跳过、MCTS搜索的评估值低于投降线或棋盘没有交叉点能够落子的情况下，AlphaGo Zero程序根据胜负得到奖励$r_T \in \{-1,+1\}$。MCTS搜索在第t步使用的是前一次反馈的深度学习神经网络$f_{\theta_{i-1}}$，并以π_t的联合分布为根据对被存储的数据$(\vec{s},\pi,z)$进行采样以训练网络参数θ_i并进行再落子的决策。在此，深度学习算法与加强学习算法互相提供运行数据，形成了进一步整理信息、提高统一性和条理性的基础。同时深度学习神经网络与MCTS搜索并行训练参数θ，$f_{\theta_i}(\vec{s}) = (P,v)$将最大化$P_t$和$\pi_t$的相似度，最小化$v_t$与$z_t$的误差，与二者相关的损失函数[2]如下：

$$l = (z-v)^2 - \pi^T \log(P) + c\| \theta\|^2 \quad 。 \tag{2-4}$$

其中，c是防止过拟合运算中应用的系数。以上步骤完毕就意味着在AlphaGo Zero中一个训练步骤周期的完成。这些步骤的不断累加将导致数据持续的迭代，意味着联结的不断生成，也就使参数θ得到深入的训练，实现整个数据库的最高完善。即“训练一个强化学习策略网络，通过优化自对弈的最终结果来改进策略。它将调整策略来达到赢得游戏的目标，而不只是最大限度提高预测精度”[6]。

综上，深度增强学习的各个部分及其具体运行过程都与知性综合的各个阶段具

有相似性。深度增强学习这一算法系统自身,究其运行机制和基本原理而言,相对外界数据既有其先天性也有后天的经验性。这个程序既不能不依靠原始输入凭空开始运转,也相比于之前的程序尽可能放去掉了人类的"帮助"。在运算过程中,深度学习网络和强化学习的算法体系彼此支持,完成算法网络对信息进行较为高度的整理、规范,体现了由低到高综合的过程。

(四)深度增强学习中是否涉及先验理性

人的思维由感性到达知性,止步于理性,而理性之上是人的思维无法思考的范畴。理性的概念是超越经验的,体现出推理能力,不能被经验达到和给予,但是理性先天地包含着某些出自纯粹理性的知识和推论出的概念的起源。理性以先验知性综合产生的知识作为质料,对知性知识进行的是体系的统一。

先验灵魂说认为灵魂是一种实体,于是在这一命题中实体成为了一种用来规定灵魂的范畴,而脱离了先验知性综合的层面。这种实体性和单纯性所指向的是范导原理,而作为范导原理的蓝图无法被具体地表现,不是直接指向思维主体的属性。在AlphaGo Zero中,深度增强学习只是一种算法,其虽然具有稳定的、一致的运算系统和程序框架,但是基于当今的科学研究水平,这种程序之上并没有像人一般的更高层次的、独立的理智。计算机运算虽然在某些情况下能够比人类更快更准确地处理庞大的数据,但至今其一直作为人类的工具在发挥着作用,是人类思维的延伸,是一种人的创造物,所以一定要对其进行更高层次的抽象的话,最终结果也会回到人类自身。而即便在人的立场上,康德也认为,纯粹理性是在心灵之上的,不直接导致心灵的具体属性。

纯粹理性的第二个范导原理关于一般世界的概念。理性宇宙论将应当作用于经验的先验形式和感性直观的纯形式等用于判断世界的整体性质,造成了宇宙论中的二律背反。这一矛盾显示,对于存在理性的可思维的自然而言,纯粹理性只是一种范导性原则,不具体规定现实中自然的具体属性和条件。而自在的自然并不受到理性的干涉和影响。深度增强学习作为一种计算机程序,在第三次科技革命的基础上得到发展,同时又引领着第四次科技革命浪潮,是人类最新研发出的科技成果。它虽然是人的创新产物,但利用的是自然中的条件和自然的运行规律、原则。深度增强学习实质上是自然众多条件和原则的一种具体的组合方式,若将其以自然作为方向进行更高度的抽象,则其程序作为软件而存在的内核将涉及基础数学,而作为硬件基础存在的部分将涉及基础物理学,也就是均指向了科学的最基本理论。但显然,康德认为理性是在这些具体理论的之上存在的更为范导性的原则,所以显然纯粹理性在深度增强学习中也是难以立足的。

纯粹理性的第三个思辨对象是上帝的理性概念。康德认为"上帝"这一理念是从外部事物和人的思维当中概括出来的先验理想,没有客观现实性。根本上,理性将世

界的一切联结进行了系统的体系化，不包含任何建构性的原则，所以纯粹理性不是上帝，无所谓是否存在。深度增强学习是整个世界的一个组成部分，那么将这一理论延伸至深度增强学习之中，则意味着深度增强学习是受到了纯粹理性体系化的具体事物，纯粹理性是使我们认识到深度增强学习系统是如此的最高范导。人根据自己的自主意识，在理论科学和实际应用科学发展的基础上研发了AlphaGo Zero这一程序，在科技发展历程中有其必然性。这也支撑了对神学批判的这一部分的观点，即康德主要指出了纯粹理性与上帝的本质区别，肯定了人的思维的能动性和自主意识，批判了理性神学理论。因此在深度增强学习系统中同样可以将“上帝存在”这一命题看作伪命题。

总结以上三个思辨话题可以看到，康德所认为的纯粹理性不是传统形而上学概念下的，而只是一种范导原则。这种原则体现了比先验知性综合更高度的系统统一性，并不是为超验知识提供建构性的条件。康德对于传统形而上学理性概念的批判态度，但是在创建新的形而上学体系的过程中又走入了误区，最终没能完全建立起来，这导致要从全新的角度看待纯粹理性部分的问题较为困难。但是将深度综合学习置于先天综合判断的理性综合之中，触发了一些值得人类在未来的人工智能研究等领域中值得进一步探讨的问题。

其一，理性是人区别于世间万物的特质，是人思考这个世界，对感官的杂多进行整理的范导性原则。因此人工智能的研究虽然创造了诸如AlphaGo Zero打败人类棋手此类的优异成果，但仍然与人类智能之间存在鸿沟，于是人的理性也成为了长久以来人工智能研究和发展的标杆与方向。在日后的研究中或许可以思考，人工智能的发展最终是否能够赋予其“理性”或类似“理性”的规范？这种规范是否可能是建立在其自身之上的？其二，基础数学和基础物理学的理论研究最终可以指向哲学的部分，那么当下人工智能的研究也不妨重视基础物理和数学理论的研究进展，一方面为自身的硬件软件升级做基础，一方面或许会产生的哲学观点的更新也有可能为人工智能日后的研究打开突破口。其三，对神学的批判其实人类应将目光投射于自身，认可自身思维的力量，进一步推进科学技术的进展。深度增强学习的自主学习功能体现了这种以往在人工智能中较为少见的自主性，是这一观点的有益延伸。在日后的研究中，探索人工智能的自主学习、自主更新等也不失为一种可行的方向。

三、结语

本文以时下人工智能技术代表性新成果AlphaGo Zero的深度增强学习作为突破口，将人类的前沿科技与康德的经典哲学理论先天综合判断进行比较和综合分析，一方面从新的视角审视人工智能技术的发展，力求寻找人工智能发展问题的新思考方式；另一方面也在某种程度上为先天综合判断提供了佐证。

总之,软件程序在本质和函数运算系统的构成等方面呈现出了与感性综合和知性综合类似的特征;并且,虽然计算机与人脑有本质的不同,但它高度自主的学习过程在各个阶段仍然呈现出了先验与综合的结合。然而,康德自身并没有完成新的形而上学体系的建立,本文中涉及的深度增强学习在这一部分也持有批判的观点,并期待包括AlphaGo系列在内的人工智能研究能够在这些问题当中取得新的进展。

参考文献:

[1]康德.纯粹理性批判(注释本)[M].李秋零,译.北京:中国人民大学出版社,2011:52.
[2]David Silver, J Schrittwieser. Mastering the game of Go without human knowledge[J]. Nature, 2017, 550: 354-359.
[3]康德.未来形而上学导论[M].庞景仁,译.北京:商务印书馆,1978:42.
[4]康德.纯粹理性批判[M].邓晓芒,译.北京:人民出版社,2004:11.
[5]杨祖陶,邓晓芒.康德《纯粹理性批判》指要[M].北京:人民出版社,2001.
[6]David Silver, Aja Huang. Mastering the game of Go with deep neural networks and tree search [J]. Nature,2016, 529: 484-489.

(责任编辑 朱凯)

Deep Reinforcement Learning under Innate Comprehensive Judgment: A Case Study of AlphaGo Zero

Zheng Bingnan[1], He Wei[2]
(1. Nanjing University; 2. Xiamen University)

Abstract: The emergence of deep reinforcement learning has led to many analogical reflections on human thinking and artificial intelligence. AlphaGo Zero is a good example in the discussion that whether computers have innate comprehensive knowledge. The transcendental perception in deep reinforcement learning is reflected in the mathematics of computers, and the part about time corresponds to algebra in pure mathematics, and the binary computing process reflects the synthesis of innate and experience. The transcendental intellectuality in deep reinforcement learning is reflected in the process where AlphaGo Zero acquires Go skills. Kant classified the transcendental intellectuality into three stages from simple to complex which corresponds to the decision-making process from the basic to the comprehensive. The function calculations on each unit, positive and negative feedback between functions and general intelligence that ultimately results in decision-making embody the comprehensiveness of this procedure, both innate and empirical. However, it is difficult to have prior rationality in deep reinforcement learning.

Key words: innate comprehensive judgment; deep reinforcement learning; AlphaGo Zero; artificial intelligence