朱松纯: 走向通用人工智能——从大数据到大任务

原创 CSDN App AI科技大本营 2019-11-01



出品 | AI科技大本营 (ID:rgznai100)

10 月 31 日,由北京智源人工智能研究院主办的 2019 北京智源大会在国家会议中心开幕,本次大会围绕人工智能基础研究现状及面临的机遇和挑战、人工智能技术未来发展的核心方向等话题,回归 AI 技术本身,开展深入研讨,探索人工智能前沿科技发展趋势。

在第一天的主论坛上,美国加州大学洛杉矶分校的朱松纯教授带来了《走向通用人工智能:从大数据到大任务》的主题演讲,并提出了任务是智能中心的观点。

以下是朱松纯演讲内容实录, AI科技大本营 (ID:rgznai100) 整理:

关于通用人工智能,每个人都有自己的想法,有人认为不可能,有人认为马上就要到来,而且会很可怕。不管哪一种观点,我们都要扎扎实实地去研究问题,今天我给大家分享一个从大数据到大任务的思路。



第一,人工智能的两种范式之争:大数据VS 大任务;第二,智能的中心核心是任务,每时每刻都被各种各样的任务驱动;第三,如何通过构建一个大任务的平台,来研究通用人工智能。

两种AI范式之争:"大数据"对"大任务"



首先是两个范式之争。如果你问大部分人工智能的研究者,他们会告诉你AI = B+C+D,也就是人工智能等于大数据加上算力加上深度学习,这是被普遍接受的观点,但是我一直在反对这个观点。几年前我反对把深度学习作为人工智能之解时,很多人还很不满,但是今天很多人已经同意这个观点了。

目前以大数据驱动的人工智能,在产业落地的过程中遇到了很多问题,这个问题其实一早就可以预想到。比如只能做特定的、人类事先定义的任务,而不能做通用任务,或者自己定义任务。第二,每项任务需要大量的数据,成本非常高,而且模型不具有解释性,知识表达与人不同。



人,以及后来的 MIT 实验室主任等,在中国湖北莲花山标数据。当时大数据刚刚兴起,我们雄心勃勃地想标数据,制定了200 多页的数据标注手册,包括莲花如何分解,花蕊、花瓣以及生活环境等。

标了几年之后,我发现这里面有问题。一些学生刚开始问我如何标注时,我还能回答,但是后来我就答不出来了。于是我发现这条路走不通,所以2009年左右,我就转型去做认知科学。



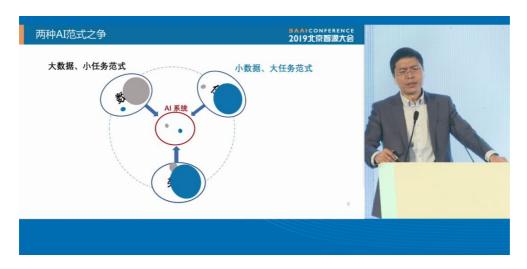
我之前写过一篇文章,介绍了两种人工智能的模式。

一种称之为"鹦鹉范式",鹦鹉可以与人类对话,但是不理解你在说什么。比如你说林黛玉,它也说林黛玉,但是它并不知道林黛玉是什么。

还有一种是"乌鸦范式"。乌鸦找到核桃之后,会把核桃扔在路上,让车去压,压碎了再吃。但是因为路上车太多乌鸦吃不到核桃,于是乌鸦把核桃扔到斑马线上,因为这里有红绿灯,绿灯亮时车都停住了,它就可以去吃。这个例子是非常惊人的,因为乌鸦既没有大数据,也没有监督学习,却完全可以自主地研究其中的因果关系,然后利用资源完成任务,而且功耗非常小,小于1瓦,这给了我们很大的启发。



但是乌鸦能做到的不仅是这些,它们还知道如何利用工具,我相信直到今天,很多机器人的规划能力都远达不到乌鸦对物理的理解。

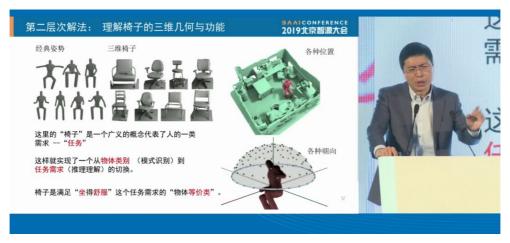


假设我们要定义一个人工智能系统,我可以认为任何一个动物、机器都是 AI 系统,它往往决定了三个要素:第一是构架,你的脑袋里缺一块,就永远进化不到一个层次,人的智能90%多都是天生的;第二是环境数据;第三是任务。



第一种层次的解法是用大数据给一个任务,比如人脸识别。给一个构架,比如深度学习有多少层,这就是目前通用的 大数据系统。我认为人类走了另外一条路,就是有很小量的数据,但是有很大量的任务,不过人类有很高级的构架, 这就成了另外一个系统。

举个例子,怎么教计算机学会识别椅子?用大数据的办法很简单,也非常暴力。就是搜索大量的例子,用人工标注。你需要搜集各种材质、各种摄像头角度的椅子,穷举后拿去训练,训练之后系统记住这些特征。但是艺术家总是设计新的椅子出来,总是有特例,机器总是搞不清楚,所以不可泛化,不可解释什么是椅子,这就是它的核心问题所在。在自动驾驶、视频监控等领域都存在这样的问题,即不能穷举所有的例子。



第二种层次的解法是什么呢?假设我们要理解椅子,定义什么是椅子。首先获得人的经典坐姿,之后用坐姿拟合图像,各种位置、各种朝向、各种姿态的坐姿,能够坐舒服的就是椅子,这就是一个任务。



椅子变成了一个任务的等价类,不管什么物体能够坐,坐得舒服的就是椅子。这代表了一种想象,我要想象这把椅子怎么坐,这和深度学习不同: regression是统计,拿特征去拟合, simulation 是用我的身体去想象,这是两者根本的区别。

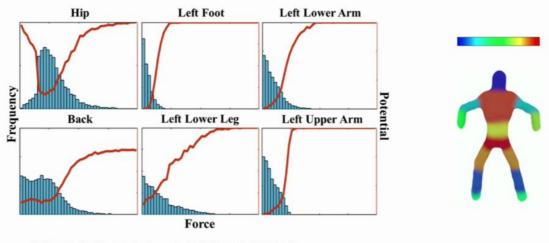


第三个层次,什么是坐得舒服?首先这个椅子要坐得稳,物理稳定是神经系统里非常敏感的感知。

我做了一个简单的实验,把各种各样的椅子放在办公室和实验室里,看学生进来时先坐哪儿、后坐哪儿,这时人会有一个偏好。

另外一件我们感同身受的事是,我们可以感觉这个人是如何用力,如何受力的,这都可以映射,这是人类非常强的能力。映射过来之后,我就知道你如何坐得舒服,从中学到你的价值观。

下图是从少量样本学习得到的关于'坐椅子'的价值函数,由此可以解释人类的各种行为。



横轴是各部位受力大小, 红线纵轴是价值的负数。

Zhu, et al. Inferring Forces and Learning Human Utilities from Video, CVPR 2016.

理解这点之后,就不需要任何数据了。我知道椅子是为了让人坐得舒服,从根本出发,我根本不需要数据。这就是我说的小数据、大任务。

以任务为中心的智能



以任务为中心的智能是大家经过漫长的时间之后才体会到的。

第一层的表达是以图像为中心。我们看到了一张图像,然后以图像为中心抽取各种特征,理解各个部件。第二层变成了以场景和物体为中心,即几何表达。

第一层是深度学习,即把图像看成图像,没有几何、物理和功能的理解。第二层开始以三维的几何来表达场景,比如 三维场景的桌子、椅子等。

最后,我认为真正的表达是以任务为中心的 task-centered representation,因为我们先有任务,才有世界上各种各样的物体,这些物体都是被设计出来是为了满足人的需求和完成某种任务。以任务为中心的表达,比如功能、因果、价值观、物理、社会常识,我们统称为physical and social common sense,这些常识是小孩在 18 个月之前获取的。

以任务为中心的智能 例子: 开瓶

3AAICONFERENCE 2019北京智源大会

理解因果、按照任务理解物体, 而不是分类



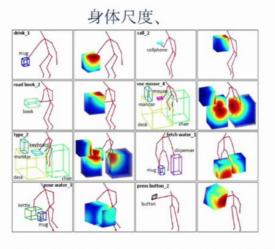
什么是以任务为中心的表达?我们考虑世界不是从某个物体类别的角度出发。比如开酒瓶,并不是只有开瓶器能开酒瓶,任何东西都能开酒瓶,砸核桃也隐藏着一个物理原理。我们知道物理原理之后,并不需要固定的东西开酒瓶,只要能完成开酒瓶任务都可以。我认为这种就是通用的人工智能。

以任务为中心的智能

3AAICONFERENCE 2019北京智源大会

17

人脑的很多知识是按照任务来组织的, 分两个尺度: 汉字有大量以人的身体部件为边旁部首





4D Human-Object Interaction (P. Wei et al 2013). 4d hand-object interaction (Xie et al. ICRA 2018)

这个问题怎么解?最近的研究发现,人脑里大部分的知识表达并不是按照物理的类别来组织的,比如椅子是一类,桌子是一类,车子是一类,而是按照如何使用它们来组织,大概可以分成两个尺度:身体的尺度和手的尺度。

其实这一点不新鲜,因为我们汉字中超过 1/3 的汉字都是与人有关的偏旁部首。我们之所以发明汉字,是因为汉字里有一些任务,任务就是要把人放进去,一起来考虑,比如手、脚、耳朵、身体等。

还是以砸核桃为例:物理常识与任务的泛化

3AAICONFERENCE 2019北京智源大会

学习: 从一个例子中琢磨出模型



测试:泛化推广与创新



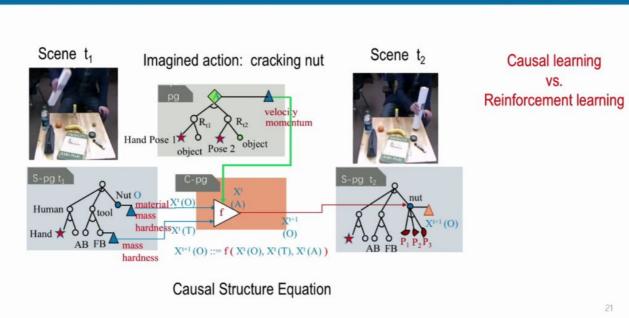
19

还是以砸核桃为例,乌鸦用车轮碾核桃,如果在一个新的环境中,工具被拿走,但是你还是能完成砸核桃的任务,这 是小学教育很重要的事情,就是教会你具有基本的常识。

砸核桃虽然只有一个例子,但是人要从上万个选择中选择一个方案,在这个过程中有大量的simulation。我看到了核桃和几种工具,大脑会快速地想怎么做,可能你的大脑出现了上万种选择迅速。这是大量计算的过程,但不是深度学习的计算,而是一种 simulation 的计算。

时空、因果联合解译与推理

3AAICONFERENCE 2019北京智源大会



还有时空因果的推理,现在这种情况要达到各种各样的目标,中间有各种各样的动作,然后形成一个因果方程,在物理上统一。

另外一点是 Causal Learning 和 Reinforcement Learing。RL现在也非常火,但是据神经科学研究者称,RL 是老鼠这类的低等动物用的,要用大量的例子反复实验。而人使用的是因果学习,只需要有两三个例子。

砸核桃的任务转换到机器人不能只是简单的转换,必须是一个物理因果的等价,要推理物理的功能。

任务学习: 从人到机器人

3AAICONFERENCE 2019北京智源大会

机器人不能简单模仿人的动作,根据自身物理功能,规划任务和运动

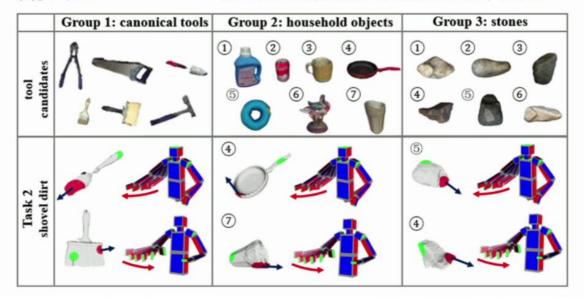


22

这是一个从一两个例子中学习的过程。一个比较聪明的人可以从几个简单的选择里就悟出坐椅子的价值观,从一个简单的砸核桃的动作悟出本质。一旦悟出来还要数据做什么呢?所以,这是一个核心问题。

任务: 铲土

与基于死记硬背的模式识别方法形成鲜明对比



以铲土为例,如果让你用工具铲土,你就会想象如何去铲。如果没有工具,用家里的工具同样你也能铲土。经过自动计算,机器的第一个选择是锅,第二个选择是杯子。

当人类或猿人走过石器时代时,神经系统已经学习了工具和物理知识,他理解的是本质。

如何定义任务?

3AAICONFERENCE 2019北京智源大会

任务是以符合因果的方式改变场景中的流态

流态Fluents: 时变的物理、内心状态, 社会关系等.

物理流态



社会流态

Internal fluents (force, pain, ...) Social relations







各种流态组成的复合数学空间 (calculus of fluents) 就定义了任务的空间。
-- 空间的任何两点 (A, B) 就对应一个任务。

24

现在回到如何定义任务?图像有多少个像素可以很清楚地定义,但是任务如何定义呢?定义任务是以符合因果的方式,改变场景中的流态。"流态"是牛顿发明的一个词,包括时变的物理状态、内心状态、社会关系等,可以简单分类为物理流态和社会流态。

前面讲到了物理任务: 改变物体的状态. 现在介绍社会常识与任务 2019北京智源大会

AI要与人交流,必须先知道很多社会流态:

- · 人在什么时候知道了什么 (belief)
- 人有什么目标与意图 (goals and intents)
- · 人下一步可能怎么做 (plans)



12月大的正常幼儿开始指东西

R. Saxe and N. Kanwisher (2006) Reported special cortical areas for the Theory of Minds in neuroimaging experiments.

如果定义了这些原子的任务空间(atomic space),就可以组合产生复合的数学空间,这就是任务。这个事情说清楚,人工智能的问题就解决了一大半。

目前人工智能之所以遇到了很大的困难,是因为大家说不清楚到底要做什么任务。任务定义不清楚,是很多产品卖不出去,或者卖出去后被投诉的原因。比如扫地机器人产品定义不清楚应该吸和不该吸的东西有哪些,给机器的任务本身就没有清晰的定义。监控也一样,到底什么人该抓,什么人不该抓,或者什么样的环境都无法准确定义。

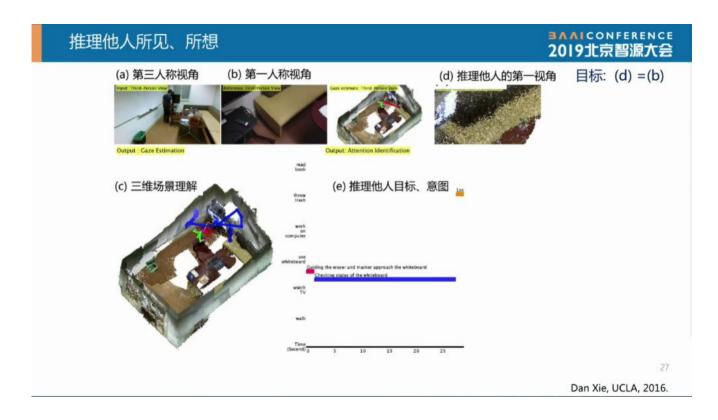
前面谈到了一些基本的物理任务和常识,物理常识是目前人工智能面临的主要障碍。比如自然语言理解,自然语言最多也是符号对符号,比如什么叫"玩水",如果没有三维数据的体验,没有物理常识,其实很难搞懂这个词的含义,所以物理常识是关键。

中国有一则寓言故事"盲人辨日"。一个从未见过太阳的盲人,无法解释什么是太阳,这就是自然语言的尴尬。自然语言必须和认知科学、计算机视觉、机器人联系在一起,否则没法研究清楚,这是我的观点。

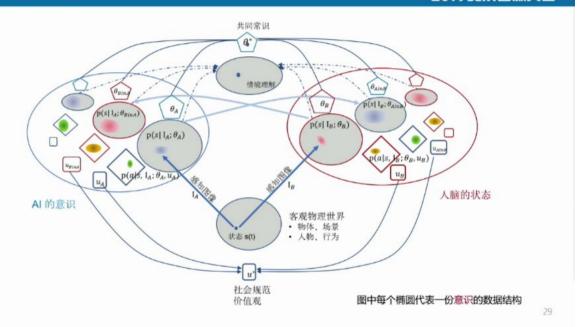
25



另外一种智能是社会的常识和任务。人类幼儿 12 个月之后开始指东西,他认识一个东西,但认为你不知道,所以他指给你看,这是一种非常强的智能。要实现这种智能,首先要有视角的转换,即推理他人所见所想,这是智能基本的东西。人类在对话时要有context,要知道上下文是什么,共同知道哪些东西。



一个人从第三视角看,是他真正看到的东西。然后我们从第三人称想他看到了什么,这是计算机推理,相当于我大概 知道你在看什么,你突然问我问题时我就知道如何回答。这其中还有我们达成的共识,我知道你知道,你也知道我知 道,这样才形成了一个共同的任务。

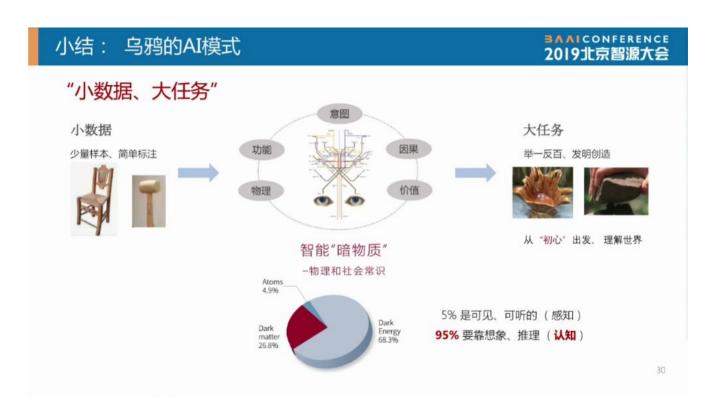


人的认知构架是什么?要形成对话,对话和语言是很重要的问题。每个客观的世界用一个圆圈表示,其中每一个点都代表一个状态。红色是我脑中所想,蓝色是机器人看到的。

首先,机器人看到的是不完整、不确定的世界,人类看到的是共享的世界,因为看问题的角度一样。我们彼此从对方的角度看问题,双方互相了解对方看到的东西。有了共同的东西,才能建立模型。

决策函数,即我知道你在这种状态下应该如何做,我应该如何做。价值函数,即我大概知道你应该怎么做,以及我认为你会怎么做,以及你认为我怎么做,有共同的情境和知识,就会有共同的价值观。最后通过交流的过程,我们达成了共识。

Christopher Manning 讲到人和人之间的通信只有 10 个 bit, 很慢, 跟 5G 相比差太远了, 但是通信很快, 为什么? 就是因为我们有这些东西。



我总结一下 AI 的乌鸦模式。

用少量的例子,但是有功能、因果、价值观等以后,就可以举一反百,从初心出发来理解这个世界。我称之为**智能暗物质**。你看见了这把椅子,想象身体如何去坐,这叫做暗物质。其中95%都是认知推理,只有把这 95% 搞定了,才能去理解剩下的 5%,否则就只能穷举所有的情况。

例:基于物理认知+社会认知的人机协作

3AAICONFERENCE 2019北京智源大会



31

这是一个简单的演示,机器人如何与人类进行交互。机器人看到人进来了以后,它要理解人的意图,知道意图之后可以帮人打开冰箱,知道要把食物放到冰箱里。在整个过程中不仅有语言的交流,还有动作、表情的交流,使得双方能够达成一个共识,即猜测你的意图是什么,这是一个基本的方式。

人工智能系统原型:集成视觉、认知、语言、学习、机器人等功能 2019北京智源大会



这是我们刚做出来的桌面机器人,它能够进行三维场景的重建。三维场景重建可以使用一些基本的计算机视觉方法。 然后想象在这个场景中人能够做什么,才能定义家具的用处。

这其中包括 top-down 的 inference,但永远无法识别小物体,必须要通过场景 context。这是一个非常致命的问题,现在所有的深度学习都没有 top-down,只有 bottom-up。

人工智能系统原型: 集成视觉、认知、语言、学习、机器人 等功能 2019北京智源大会

AI六大领域的统一与融合

• 计算视觉: 物体识别、属性理解、三维重建、场景理解、行为分析

• 认知科学: 功能用途、物理属性、因果判断、社交意向、脑中有脑

• 语言对话: 语义解译、对话意图、模式切换、语境落地、共享情景

• 机器学习: 实例归纳、因果模型、价值获取、本质洞察、规律学习

• 机器人学: 任务规划、物理推导、因果推断、镜像学习、社交礼仪

• 博弈伦理: 价值函数、博弈平衡、社会关系、伦理规范、道德标准

32

这时我们要做一个统一的系统,融通六大领域,即计算机视觉、认知科学、语言对话、机器学习、机器人学习等,融合了才能有一个像样的计算机系统。

如何构建"大任务"训练与测试平台?

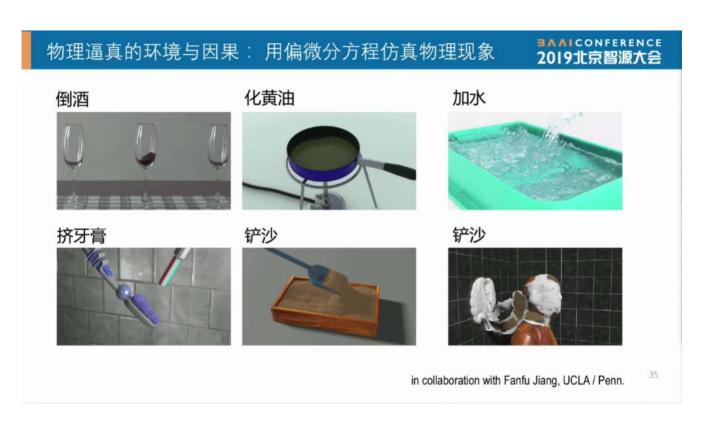
3AAICONFERENCE 2019北京智源大会

3、构建"大任务"训练与测试平台:

-- 如何训练出一只"乌鸦"— 走向通用的人工智能?



当然,只在一个物理场景中训练是不够的。第一步要根据人的需求,生成大量的数据库中的三维物体。这是生成的各种例子,生成以后可以在各种环境中测试。现在做的大数据拟合,大家都可以测试,这是其中一种玩法。

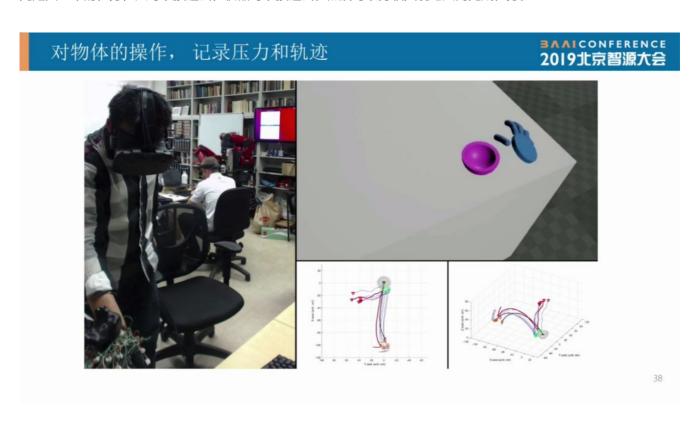


我的玩法不同,智能系统来了以后,我用一个全新的系统给你看,看你能不能完成各种各样的任务,而不是事先规定 的任务。

在这个系统里面必须要做物理逼真的东西,比如可以倒酒、倒水、挤牙膏、玩沙子、玩水,这是非常费劲的,我们已经做了好多年。



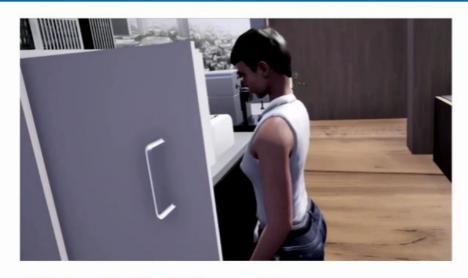
先定义基本的任务,人可以接进去,机器可以接进去,然后可以身临其境地共同完成任务。



我们做了一个手套,它可以精细化地记录一些感知和运动行为。我进入了以后就知道有一些基本的操作,就是 Learing from demonstration。

AI智能体:完成"大任务" -- 鲜榨果汁

3AAICONFERENCE 2019北京智源大会



训练出一只自主的智能体 --- "乌鸦"

40

这是一个虚拟的机器人智能体,让它完成鲜榨果汁。比如说它要先找橘子,然后把橘子切了,再到榨果汁的地方去压,它要有训练的过程。比如做饭、做面条,这都是非常大的任务。一般的日常生活中,你们看不起的东西越难。



人可以在里面与机器进行交互,也就是人机协作完成一件事情。

Al 智能体: 任务学习与泛化

■AAICONFERENCE 2019北京智源大会

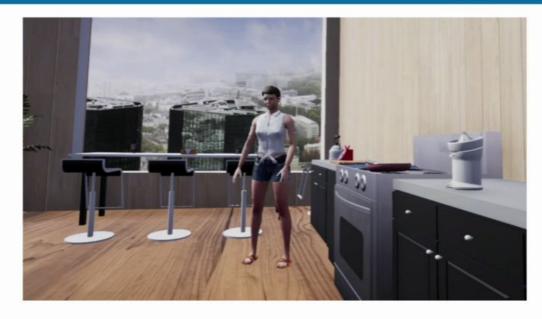
Demonstrator teaches learner how to crack nut



最后人也可以教机器人,比如说这个男的演示怎么砸核桃,这个动作背后其实有很大的工程量。



机器人会去想,怎样在一个新环境下完成砸核桃,整个推理的过程都可以在这个地方完成。



43

最重要的事情,我可以随时把机器叫停,然后问它,你现在知道什么,你知道我在干什么,或者叫它解释,打算做什么,为什么要做,这就是Explainable AI。

智能体的核心就是把自然语言对话、计算机视觉、机器人等结合在一起,就像教小孩一样,怎么样用小数据教会他。

总结

第一,"大数据"对"大任务"两种范式之争,99%的人都是押注大数据,但是10年前我就押注大任务。

第二,我认为任务是智能中心。我们有task-orientated操作系统、编程语言和体系结构。

第三,如何构建大任务的平台。我要能在里面训练出"乌鸦",通过自主的任务去驱动它。这里面有很多数学问题、理论问题和工程问题要解决,中美必须合作来解决这个问题。

(*本文为 AI科技大本营原创文章, 转载请微信联系 1092722531)



2019 中国大数据技术大会 (BDTC) 再度来袭!豪华主席阵容及百位技术专家齐聚,15 场精选专题技术和行业论坛,超强干货+技术剖析+行业实践立体解读,深入解析热门技术在行业中的实践落地。5 折票倒计时1 天!



推荐阅读

- 元学习的崛起
- "Python之父"从Dropbox退休
- 通俗易懂: 图卷积神经网络入门详解
- 最近程序员频繁被抓,如何避免面向监狱编程
- 惊! 又一家知名企业被查
- 利用 50 行 Python 代码构建一个在线文本生成器
- 技术干货 | 详谈AI芯片分类和关键技术
- 重磅 | 边缘计算核心技术辨析
- 回首30年存储演进之路 再看新数据时代阿里云存储如何革故鼎新?
- Wi-Fi 6到底有什么特别?
- 百度工程师深度好文, 横评3大企业级区块链技术平台

(水点的每个"在看", 我都认真当成了AI