**P1：各位同学大家好，我是邓浩，软件学院预聘助理教授。**

我的办公室在济事楼408R，欢迎同学们随时来办公室串门。

这是我的电子邮箱。

欢迎大家加我联系方式。

**P2：参加这门课的教学，对我来说是一个很好的学习机会。**

我目前在负责一些物联网与智慧城市的项目研发工作，我们其中很多研究任务也涉及到将人工智能技术用在城市管理、城市规划问题中。因此参加这门课的教学，对我的科研项目工作也具有很好的启发和帮助。

因此，也希望这门课提到的一些内容可以给大家未来解决科研、工作、生活问题，带来一些启发。

最后，欢迎同学们能在学有余力的情况下，参加到学院、学校的各类科研项目过程中。尤其欢迎大家参加到我们的科研项目里来，为未来的研究生阶段打基础。

在本部分内容——“人工智能的一些基本概念”介绍之前，我们还是先进行一下引导性介绍。

**P3：**

我们知道，全球经济曾经是由工业化推动的。

然而今天，它是由知识和信息所推动的。主要的技术进步和平台转变，加速了这种转变：

1990年代是互联网。2000年代带来了云计算的时代。2010年代产生了智能手机的普遍性。曾经的新兴平台继续扩大了知识的获取，改变了人们沟通、创造和消费内容的方式。

我们现在开始预感：我们正处于下一个平台转变的边缘。知识和信息经济将由人工智能来定义。

而且这种预测，已经在这两年越发地清晰：比如我们看到了这两年生成式AI开始大火，而大模型（chatGPT, sora）更是火上加火。有人开始调侃：现在这个时代是人类在沉迷娱乐，机器在学习。

除了chatGPT外，还有很多其他的好玩的生成式AI应用。比如：

1、~~解决任何问题的 ChatGPT~~

2、专为写文案而生的 Jasper AI

3、生成真人视频的 Synthesia

4、解决法律问题的 Do Not Pay

5、创作艺术作品的 Dall-E-2

6、自动发布社交媒体的 Repurpose IO

7、帮你写论文的 Jenni AI

8、写故事 Tome

9、会议记录 Fireflies

10、生成真人演讲的 Murf

11、时间管理大师 Timelt

12、聊天机器人 Chatbot Live

13、Sora

这个图是大模型技术的发展脉络图。大模型技术目前主要包括了开发者可以私人访问的模型（如GPT等），以及开源模型比如Meta的开源模型OPT等。

**P4：**

回到ChatGPT。

经过一年的发展，很明显，我们的人工智能发展在大模型领域已与openai等公司的技术差距水平并没有缩小。可能几年前也有专家已考虑到这个问题。因为这个问题不只是基础理论、算法的问题，也涉及到整个半导体领域受限带来的算力瓶颈等等，所以这个问题也是可能预见到的。

当然这也不只是我们感到压力，openai的对手们也是感受到了压力。谷歌的市值一度跌到了千亿美金。有同学理解谷歌（可控核聚变/AI重写排序算法，速度快70%：DeepMind AlphaDev）受影响的原因吗？

---

**我们的焦虑和压力不像谷歌和亚马逊**，并不局限于一个团体，更多是基于对国家间竞争感受到的压力。相信大家已经开始逐渐感受到国家间竞争对我们日常生活的影响，包括对我们同学们未来出国留学的影响。

也有电子行业的资深从业者认为：中国科技全面领先。他认为科技的核心能力，是材料、能源、信息，这些能力，通过基础建设，会导致社会效率高、能力强、成本低。因此从近十年中国在相关方向的发展，中国所有和美国同时起步的领域都超过美国。而在这个基础上形成制造能力，基础建设能力，也就远超美国。目前全世界140个国家和地区是中国第一大贸易伙伴。

---【补充解释】

*社会上流行着这样一种想法，美国虽然现在制造业能力不行了，但是它的基础科研能力还是强大的。*

*持这种想法的，或者没有从事过生产实践，也做不了科研。*

*全世界科研水平怎么起来的？一切的科研都是建立在生产实践上。近代西方为什么科技水平很强大？因为欧洲是上百个小国，盎格鲁.撒克逊人、高卢人、雅利安人、维京人、斯拉夫人打了上千年，有非常高的杀人的要求，为什么近代西方有大科学家，牛顿、诺贝尔、居里、爱因斯坦，他们都是为了研究杀人武器，牛顿研究炮弹打出去落在什么地方，才发现了万有引力。二战后，美国的技术突飞猛进，就是和苏联展开军备竞赛，花大量资金用于航天航空，互联网怎么来的，这也是军事项目，为了防止通信网络被整个端了，所以搞出一个没有中心的网络，各个终端是连在一起，互不隶属，这就是互联网。*

*一切的科研都是寄托在市场需求之上，不受市场影响，自己爱好搞科研，你做梦吧。*

*今天中国和美国比，核电、智能电网、特高压、风力发电、光伏发电、新能源电池、自动驾驶、智能交通管理、高铁、铁路运输、地铁建设、盾构机、大型工程机械、铁路建设、桥梁建设、地下通道建设、造船、港口建设、稀土提炼、钢铁生产、水泥生产、新型舰艇制造、电磁弹射、超高速导弹、通信网络技术、5G技术都是全面领先的，可以说基础设施、生产制造能力中国远远领先，今天世界上最伟大的工程一定是中国做的，就是苹果、特斯拉没有中国，也一定要完蛋，如果不是中国生产，就不会有这样的效率，就不会有这么低的成本，就无法生产出来，及时交货，也不会在价格上还有竞争力。*

*所有这些能力，一方面对于技术，包括底层技术提出更高的要求，同时大量资金的积累，就大大加快技术的研发能力。一个具体的例子，手机我们希望玻璃摔不碎，这就是一个非常底层的技术，华为就率先投入研发，做出了昆仑破晓玻璃，把强度提升了10倍。这个技术还有可能用在众多的产品上，甚至有一天也可以用在飞机上，苹果呢就没有搞出来，它还是很容易摔碎。好在美国还有苹果，如果苹果都没有了，哪个大学有那么多钱，没事搞个高强度玻璃？*

*从来没有脱离产业的基础研究，基础研究都是随着产业发展，产业给基础领域提出更高要求，产业为科技提供大量资金，产业带动基础领域发展。*

*美国产业完蛋了，哪有钱来投入，产业都不领先了，科研会知道往哪里研究？*

*这两年中国和美国的技术水平还没有完全拉开，只有我这种产业一线，对美国企业情况很了解的，才会觉得美国科研水平要完蛋了，一般人感受不到，但是过几年，我估计2030年，一般人就会有明显的感受了。这就和我推动5G时，很多普通人，一听5G，觉得我们不需要，我们感觉不到5G的作用。其实中国5G基站230万个了，占全世界60%以上，中国5G套餐用户超过11亿了，中国的物联网用户数已经超过手机1.6亿了，中国的移动物联网连接占全球的70%。中国正在建设起来强大的物联网体系。*

*有人说，我不知道，你没有感觉到。你当然没有感觉到，你是一个远离社会实践、老了、距科技发展很远的社会下层，你没感觉到难道不是很正常？*

**没有产业进步，没有技术发展，没有生产制造，就是没有科学研究**！

大家觉得这个观点如何？

----

**我建议的是：请大家忽视观点，而应重视逻辑和事实。**我觉得项老师的话讲得很好，我也觉得很多专家的反方面的话讲的很好。因为他们都有很好的逻辑。而其中的事实确实极为复杂的，我没有太多能力掌握全部。所以我不会给大家一个明确的观点建议。但是共性地，我们有共识的地方在于：我需要为发展国家的科技水平努力。最近，中共中央政治局在2月21日下午就加强基础研究进行第三次集体学习。中共中央总书记习近平在主持学习时强调，加强基础研究，是实现高水平科技自立自强的迫切要求，是建设世界科技强国的必由之路。

**P5：**

同学们在学习的过程中也要留意。我们真正要学习知识，不能完全偷懒，只希望别人给一个适合自己的观点。或者自己要意识到：自己是以何种形式应用观点的。自己没有验证过的观点可以采纳，但是要多留一个心眼。因为，瑞达利欧提出：“观点很廉价，几乎所有人都愿意和你分享观点。许多人会把观点表述为事实。你要区分观点和事实。”

**在科研领域也有类似的道理：缺乏执行能力的新思路是很廉价的……**

很不巧，作为一门导论课程，我们可能会涉及很多观点，请大家在听课的时候小心。不要轻易地同意我的观点。

也因此，为了加强基础知识和尽量贴近事实层面，我们这次课会匆匆介绍一下人工智能涉及到的一些基本概念。这些概念中当然包括了很多观点。

如果大家需要理解这些概念的基本知识，请大家课后阅读《人工智能：一种现代方法这本书》。

**P6：**

在开始一些基本概念之前，我们首先讨论一个根本性的观点问题：当我们说人工智能的时候我们究竟在说什么？

**我们来看这个图。**

这个图是人工智能专家弗吉尼亚·迪戈纳姆（Virginia Dignum）对人工智能的理解。

她把人工智能系统需要具备的功能分成了三个方面：学习、推理、交互。

这三个方面看起来挺清晰明了。

我们前边的课程已经讲过什么是人工智能。我想随机问2位同学，你们认为的人工智能是什么？

***\*\*\*2个学生回答\*\*\****

很难准确回答。首先我会困惑：迪戈纳姆提到的学习是什么、推理是什么、交互是什么？

**1、学习是什么呢？**

维基百科上把学习定义为通过**阅读、听讲、研究、观察、理解、探索、实验、实践**等手段获得知识或技能的过程，是一种使个体可以得到持续变化（**知识和技能，方法与过程，情感与价值的改善和升华**）的行为方式。

***\*\*\*板书\*\*\****

**这样看来，相比于人类而言，人工智能系统的哪些学习手段比较厉害？*\*\*\*2个学生回答\*\*\****

好像人工智能系统的研究（快曾经）、探索、**听讲（曾经）**能力比较弱。

人工智能系统的阅读、理解、观察、实验、实践方面，跟人比起来各有千秋。

随着chatgpt等的出现，听讲能力比较弱可能开始已经成为过去时。

**然后，在“知识和技能，方法与过程，情感与价值的改善和升华”这三方面，我们现在的人工智能系统比较容易实现哪个？**

好像人工智能系统比较容易获得知识和技能，方法与过程，是不是？

**最后，我想问的是：我们的人工智能算法，如果模仿这些学习过程，会不会让算法更有效？**

我觉得应该是这样。我们很多科学家的研究思路，确实也就是从人类学习启发而来，比如强化学习，它本身就是启发于人的学习。

https://zh.wikipedia.org/wiki/%E5%AD%A6%E4%B9%A0

**2、推理是什么呢？**

维基百科上将推理定义为：“使用理智，从某些前提产生结论”的行动。

那么，需要问一下，我们的人工智能系统实现的推理，符合这个定义吗？

https://zh.wikipedia.org/wiki/%E6%8E%A8%E7%90%86

**3、交互是什么呢？**

维基百科上将交互定义为：让2个或多个物体相互产生影响的行为。

我的问题是：大家觉得我们人工智能系统在交互上做的如何？不只是人工智能系统与人的交互，我们的系统与系统交互又做得如何？

https://en.wikipedia.org/wiki/Interaction

***\*\*\*学生回答\*\*\****

**对于这些问题，**我推荐大家看参考书《人工智能：一种现代的方法》。

这本书里，作者认为理性的智能体Agent被认为是实现人工智能的核心。

作者将人工智能涉及的6个领域“**问题求解，知识与推理，合乎逻辑的行动，不确定知识与推理，学习以及通信、感知与行动**”都统一到了智能体Agent这一框架下，形成了一个相互联系的整体。

目前《人工智能：一种现代的方法》已经更新到第四版，我们课程确定参考书的时候第四版还没正式发行。新的一版，作者已经把对人工智能新发展的理解加入了。他们把近十年AI 进展，尤其是深度学习所带来的影响纳入整体框架后给出的最新阐释，体现了两位大师对人工智能趋势和学科体系发展的洞见。

**P7:**

**我们再来看看近年来的一些著名的人工智能技术发展案例（提问，问学生的了解情况）。**

首先是2016年-2017年李世石、柯洁与AlphaGo的世纪大战。

2019年OpenAI 2:0击败了Dota2世界冠军OG。

然后是2020年AlphaFold2实现了高精度的蛋白结构预测。这个事件被认为是结构生物学历史上的一个革命性突破。

最后是2021年下半年到现在的生成式AI的大爆发。21年下半年DeepMinde在nature发文，提出一种机器学习模型，可以帮助发现数学定理。22年基于扩散模型的AI作画，大受追捧。以及持续到现在，ChatGPT形成了广泛影响。

**这些案例给了我们对人工智能无限的憧憬。让我个人在22年年中一度对人工智能技术落地应用的相对悲观，到现在再次形成积极展望：相信在经济形势逐渐承压的情况下，软件工程相关行业仍然会是未来发展重点。这对今年毕业同学的就业来说也许是好事。**

记得去前年我在课上还在说：很多时候我们面临的人工智能仍然像人工智障。那么今年我可能不敢这么说了，就好像大家也不再怎么提“图灵测试”了。在22年的时候，已经有人提出：“图灵测试的实质就是要让人机交互在限定时间内做到真假莫辨。玩过GPT3的同学们都清楚，其实这一点已经做到了。从这个角度看，图灵测试已经过时了。”

不过，即使图灵测试真的已经过时，人工智能与人类智能还是有很大的不同。为此，强化学习之父还写了一本书：AI研究的苦涩教训

**究其原因：人工智能与人类智能的根基是不一样的，很多方面其实没有可比性，甚至也不必要强求可比性，即便在20年前我们人类的计数能力可能就比不过电脑。**

当然我们发展人工智能很多时候又在追求可比性。不然为什么需要在智能前加人工两个字？

**那么人工智能与人类智能的根基不一样在什么地方呢？**

我认为：一种是**碳基生命**，一种是**硅基系统**。

我们常见的硅基系统，比如计算机，都是建立在半导体器件的基础上的。通过半导体在导电和不导电之间切换，实现0和1信号的模拟。

实现这一模拟的关键器件叫三极管。

我们就是在这样的物理材料和电子器件上，实现我们现在看到的各种人工智能技术的。

**而我们人类的智能是怎么实现的呢？**

我不知道，我知道反正不会是通过半导体三极管实现的。

我们人体内各种信号的工作原理跟半导体三极管的工作原理应该差距也挺大。

所以人类智能与人工智能体现出不具有可比性的智慧，我觉得不需要惊讶。

我觉得需要惊讶的是：为什么人工制造的三极管集群，开始具备了人类不太理解的、不太能解释的一些智慧。比如《人工智能：现代方法》的作者Stuart Russell 教授就说：ChatGPT 真正令人印象深刻的是它的泛化能力，它能够在其与用户进行的对话和此前读过的文本中找到相似之处并进行适当的转换，所以它的回答看起来很智能。但是，我们不知道模型是如何做到这一点的，我们也不知道这种泛化能力的边界在哪里，我们不知道这种泛化是如何在电路中实现的。如果我们真的知道了，那确实可以说是人工智能的进步，因为我们能够把它作为一个基础，我们能够基于 ChatGPT 开发其他系统。但现阶段而言，一切都还是谜。我们所谓往前走的唯一方法是——模型不 work？好吧，我们再给它更多数据，把模型再做大一点。

**这一切的不理解，就要从深度神经网络开始。**

在深度神经网络之前，我们的人工智能都是在我们人类的数学逻辑之下的，包括我们目前在机器学习中涉及的大部分算法。

但是深度神经网络出现之后，人类开始不太能解释人工系统的这种智慧表现。

在我们现在所用的高精度图像识别之前，很多科学家已经研究了很多年模式识别，但是性能始终比不上用人的眼睛来识别。

于是，早些年人工智能技术遇到了很长的瓶颈期，搞得很多研究者也挺绝望。

我的导师也是那时候从人工智能领域转到了无线通信领域，赶上了无线通信的大发展，不过后来错过了人工智能的大发展。

后来，2000年前后，经过**燕·勒存（Y. LeCun）、杰弗里·辛顿（G. Hinton）**的开创性工作后，深度学习模型和算法开始发挥威力，并掀起了人工智能的研究浪潮。

***\*\*\*板书\*\*\****

**但是为什么深度学习的模型和算法带来了传统模式识别技术没有的高精度图像识别性能？我们的科学家们现在还不能彻底证明。**

这就好比法拉第最开始发现电磁感应现象的时候一样，人类对于经典电磁学——麦克斯韦理论是无知的，也解释不清楚法拉第电磁感应定律。

后来，经过麦克斯韦、爱因斯坦等人构建好电磁学理论的数学模型后，人类才开始深刻理解各种电磁现象，并稳步进入了现代信息社会，实现了我们看到的各种人工智能技术。——杨振宁

**其实，我觉得同样值得惊讶的是：人类距离能解释 “人类智能” 还非常遥远。**

我们可以对比一下人脑和电脑的能耗水平，人脑运行速率比电脑慢几百万倍，一天的能耗在**20瓦左右**，但是可以轻松做到很多不可思议的事。

AlphaGo的能耗是多少？大约是人类的**37万倍**。这个角度看，人类是不是很厉害？

~~所以从这个视角出发，国内外很多科研团队也在打造另一个科研概念：类脑计算。~~

~~https://cloud.tencent.com/developer/news/410629~~

**P8：**

**闲扯了很久，我们看看其他专家对于人工智能的观点。可能梁老师上节课已经讲过，我们在回顾一下。**

目前人工智能主要包括了4类定义：

**1、像人一样思考**

霍格兰德认为 “人工智能就是要打造有头脑的机器”。

贝尔曼认为 “人工智能是使得决策、问题求解、学习等活动自动化”。

**2、理性思考**

切尔尼克和麦克德蒙认为 “人工智能是采用计算模型实现人的理性思维”。

温斯顿认为 “人工智能是通过计算实现感知、推理和行动”。

**3、像人一样行动**

库茨维尔认为 “人工智能是创造一种机器，这种机器可以完成需要人类智力才能完成的任务”。

里奇和奈特认为 “人工智能是研究如何使计算机能做到 ‘那些目前人比计算机更擅长的事情’ ”。

**4、理性行动**

切尔尼克和麦克德蒙认为 “人工智能主要研究智能体的设计，实现智能计算”。

尼尔松认为 “人工智能需要实现人工设备做出智能行为”。

“像人一样的思考和行动”通常被认为是强人工智能，“合理的思考和行动”通常被认为是弱人工智能。

**讲到这里的时候，我们应该已经忘了，前边人工智能专家弗吉尼亚·迪戈纳姆提到的人工智能系统应该具备三个功能：学习、推理、交互。**我觉得，我们也没必要记住各个专家的定义。

重要的是：我们自己怎么理解人工智能，我们想做什么样的人工智能系统。

我们是想做像人一样思考的系统，还是做理性思考的系统？这应该不一样的。

我们是想做做像人一样行动的系统，还是做理性行动的系统？这应该也不一样。

**像人一样思考或行动，不代表总是理性思考或行动。**所以你要设计的系统是要像人一样，还是理性，是有本质区别的。

**思考和行动也是有本质区别。**你设计的系统是只提供决策建议和指令，还是重点关注机器输出，或人和机器交互，或机器和机器交互？

这里边，就包含了直觉经验和逻辑理性的区别。

**P9：**

**再来看看我们日常使用的人工智能技术。比如：抖音、头条、滴滴、翻译器。**

好像挺平平无奇，但是他们都在我们提到的4类人工智能技术里。

另一方面，以5年一个维度看，我们的图像识别问题、自然语言处理问题都因为这几年人工智能技术提高了不少，甚至可以说是巨大。

**综合人工智能技术的4种分类与我们身边的这些人工智能应用案例，**我们再来考虑下什么是我们自己理解的人工智能？

是实现**人类**逻辑+**机器**逻辑吗？

是实现**人类直觉**+**机器逻辑**吗？

是实现**人类逻辑**+**机器直觉**吗？

是实现**人类**直觉+**机器**直觉吗？

而且，什么是直觉（压缩即智能）。什么是逻辑？——强化学习里就需要用到机器直觉式的试错

对应的，什么是机器直觉，什么是机器逻辑？这些问题，以后大家继续做人工智能相关科研的时候，可以继续思考，并丰富他们的内涵。

https://zh.wikipedia.org/zh-cn/%E7%9B%B4%E8%A6%BA\_(%E7%9F%A5%E8%AD%98%E8%AB%96)

https://zh.wikipedia.org/wiki/%E9%80%BB%E8%BE%91

**到这里，我推荐一本书《思考，快与慢》。**有的同学应该已经读过，没读过的同学有时间、有兴趣也可以翻翻。人工智能系统是否也可以，或者已经可以实现快思考或慢思考？

* **再回到人工智能来。**无论未来人工智能如何演变，内涵如何变化，现在我们都可以将它涉及的几个关键概念，按照右下图这样划分逻辑。

他们是一个由大到小的包含关系。所以，人工智能不等于机器学习，也不等于深度学习，

**P10：**

我们再来看看一些专家提到的三个人工智能技术特征：

第一个特征是：**人工智能应该由⼈类设计，为⼈类服务**，本质为计算，基础为数据；这个特征里边，我有点怀疑“**人工智能应该由⼈类设计，为⼈类服务**”这个说法还能成立多久，也许未来机器智能与人工智能已经很难区分。正如图灵测试可能已经过时一样。

第二个特征是：人工智能应该能感知环境，能产⽣反应，能与⼈交互，能与⼈互补；说到这里，上了用户交互技术的同学是不是感觉很熟悉。

第三个特征是：人工智能应该有适应特性，有学习能⼒，有演化迭代，有连接扩展。这个特征大家怎么理解？

**P11：**

此外，通常来说，人工智能被划分为了三个流派。

1.符号主义

符号主义又称逻辑主义、⼼理学派或计算机学派。起源于数理逻辑。

其主要观点是**利用物理符号系统**及**有限合理性原理**来实现人工智能。具体来讲，符号主义认为人类思维的基本单元是符号，而基于符号的一系列运算就构成了认知的过程，所以人和计算机都可以被看成具备逻辑推理能力的符号系统，换句话说，计算机可以通过各种符号运算来模拟人的“智能”。

因为这种学派对于AI的解释和人们的认知是比较相近的，可以较容易地为大家所接受，所以可以说它在AI历史中的很长一段时间都处于主导地位，其代表人物是纽威尔、肖、西蒙和尼尔森。符号主义在不同历史时期都有些代表性的成果，例如：

1. ~~Allen等人发明的“逻辑理论家”程序，可以证明出《自然哲学的数字原理》(Principia Mathematica)中的38条数学定理(后来可以证明全部52条定理)，而且某些解法甚至比人类数学家提供的方案更为巧妙~~
2. Allen和Simon等人提出了通用问题解决器(General Problem Solver)推理架构以及启发式搜索思路，影响相当深远(比如AlphaGO就借鉴了这一思想)。
3. 专家系统。专家系统对20世纪AI的繁荣起到了非常重要的推动作用，理论上来讲它也属于符号主义的研究成果。
4. ~~知识库和知识图谱。专家系统的主要难点在于：知识的获取构建以及推理引擎的实现。所以学者们围绕这些困难点发展了不少理论，比如反向链(Backward Chaining)推理、Rate算法等。~~

2.联结（连接）主义

连接主义又称仿⽣学派或⽣理学派，这是因为它的其中一个**研究重点在于人脑的运行机制**，然后将研究结果应用到人工智能的分析中。由于这种学科间的交叉关系，我们有时候会发现研究人工智能的科学家可能同时也会是脑神经科学家，又或者是心理学家。连接主义理论的创始人Edward Lee Thorndike就是一名心理学家(教育心理学的奠基人)，他从动物的实验研究中得到了启发，然后提出了连接主义的理论基础。

**连接主义主要原理为神经网络及神经网络间的连接机制与学习算法**。连接主义其后的发展出现了多个里程碑事件，包括：

1. 1943年的人工神经元
2. 1957年的感知机
3. 多层感知机和反向传播算法
4. 1995年后期出现SVM
5. 2010年后的深度神经网络

当时间步入21世纪的第二个十年后，情况又逐渐发生了变化。**以深度神经网络为代表的人工智能以及它的各种应用场景**，以迅雷不及掩耳之势抢占了人们工作生活的方方面面，造成了目前“人人谈AI”的局面。

https://zhuanlan.zhihu.com/p/368123491

3.⾏为主义

⾏为主义又称进化主义或控制论学派。起源于控制论。

⾏为主义的原理是**控制论及感知—动作型控制系统**。**强化学习的灵感就来自于行为主义**。强化学习是让一个智能体（Agent）不断的采取不同的行动（Action）， 改变自己的状态（State），和环境（Enviroment）进行交互，从而获得不同的奖励（Reward）,我们只需要设计出合适的奖励（Reward）规则，智能体就能在不断的试错中习得合适的策略。

控制论思想早在20世纪40~50年代就成为时代思潮的重要部分，影响了早期的人工智能工作者。**维纳(Wiener)和麦克洛克(McCulloch)**等人提出的控制论和自组织系统以及**钱学森**等人提出的工程控制论和生物控制论，影响了许多领域。控制论把神经系统的工作原理与信息理论、控制理论、逻辑以及计算机联系起来。早期的研究工作重点是模拟人在控制过程中的智能行为和作用，如对自寻优、自适应、自镇定、自组织和自学习等控制论系统的研究，并进行“控制论动物”的研制。到20世纪60~70年代，上述这些控制论系统的研究取得一定进展，播下智能控制和智能机器人的种子，并在20世纪80年代诞生了智能控制和智能机器人系统。行为主义是20世纪末才以人工智能新学派的面孔出现的，引起许多人的兴趣。这一学派的代表作者首推布鲁克斯(Brooks)的六足行走机器人，它被看作是新一代的“控制论动物”，是一个基于感知-动作模式模拟昆虫行为的控制系统。**——以及系统科学**

**P12：**

好了，进行了上述的引导性介绍。我们现在开始快速浏览一遍人工智能的一些基本概念。首先是智能体。

按照我们参考书的观点：理性Agent智能体是我们人工智能方法的核心。同学们以后在学习人工智能的时候要关注Agent智能体的本质是什么，Agent有哪些类别，Agent智能体面临了哪些环境等等。以及了解环境如何影响Agent智能体的设计，哪些Agent智能体会比其他Agent智能体在环境中表现更出色。

**P13：**

智能体Agent的本质是什么呢？从概念上讲：智能体是通过传感器感知所处环境，并通过执⾏器对该环境产⽣作⽤的东西（事物）。

这个图就示例性地说明智能体、环境、传感器、执行器关联关系。**智能体可以分为从简单到复杂的类型：**

被动智能体或无目标智能体（如简单模拟中的物体）

具有简单目标的智能体（如鸟群中的鸟）

认知主体（具有复杂的计算）

具体地，一些普遍意义上的智能体包括：软件、机器人，人类或人类团队。对于机器人智能体，它的传感器一般包括：摄像头、测距仪等，执行器包括各种马达等。大家认为，人作为智能体，他的传感器是什么，执行器是什么？

**P14：**

大模型智能体（问同学们的理解）。

**P15：**

对应地，智能体的数学定义是：

**智能体可以通过智能体函数来实现“给定感知序列——到行动的映射”。智能体函数的抽象数学表⽰为：**



其中，f就是这个映射函数，P\*是感知序列，A是行动。

最后，智能体函数是通过智能体程序实现的。智能体函数是一个抽象的数学表示。智能体程序是智能体的一个具体表现。

**P16：**

那么，人工智能方法的核心——理性智能体是什么？

定义：**对每⼀个可能的感知序列**，根据已知的感知序列提供的证据和智能体具有的先验知识，理性智能体应该选择**期望能使其性能度量最⼤化的⾏动**。

**注意：理性（Rational）与全知（Omniscience）是有区别的**

1. 理性 ≠ 全知

⼀个**全知的智能体明确地知道它的⾏动产⽣的实际结果并且做出相应的动作**，全知者在现实中是不可能的。

理性并不要求全知。理性的选择只依赖于到当时为⽌的感知序列。

2. 理性 ≠ 完美

理性是使期望的性能最⼤化 完美是使实际的性能最⼤化

对智能体⽽⾔，完美是不太合理的要求。无论是ChatGPT还是人。

**P17：**

判断一个智能体，是否是理性智能体，我们需要视情况而定：

首先，我们需要说明**性能度量**是什么？

其次，需要清楚**智能体对环境**了解多少？

智能体**具有哪些传感器和执行器**

若智能体的执行结果与其预期性一样，那么就可以认为智能体确实是理性。

以吸尘器机器人为例，一旦清除了所有灰尘，该智能体将会毫无必要地反复来回；**如果性能度量考虑对每个动作罚1分**，那么智能体的表现就会很差。

在确定所有方格都干净的情况下，一个更好的智能体不会做任何事情。如果干净的方格可能再次变脏，智能体应该偶尔检查，并在必要时重新清理。

如果环境的地理信息是未知的，**智能体则需要对其进行探索**（explore）。

**P18：**

下面我们介绍一下人工智能中的“面向问题求解的搜索”。

**P19:**

智能体中最简单的是**反射型智能体**。这类智能体类Agent 存有在何种状态下可采取何种行动的直接映射表，它们的行为就取决于这种映射。有些环境中 Agent 的这种映射表可能非常大，导致占用很多存储空间或者查表消耗的时间太长而无法学习，Agent 在这样的环境中难以运转。另一方面，基于目标的 Agent 会考虑将要采取的行动及行动的可能后果，即与目标还有多远。

这样就带来了一个问题：我们的智能体要面向目标搜索或者求解。这尤其是“**基于目标的智能体**”关注的。我们称之为称为问题求解 Agent(problem-solvingAgent)。

**要讨论问题求解，首先要讨论的是对问题及其解的精确定义**。问题求解中“问题”的形式化描述可以分为5个部分，**分别为初始状态，⾏动，转移模型，⽬标测试和路径耗散**。

**要进行问题求解，需要用到一些求解问题的通用的搜索算法**。主要包括2类：⾸先是⽆信息的搜索算法。

【**⽆信息搜索**】也称为盲⽬搜索。是指除了问题定义中提供的状态信息外没有任何其他信息。搜索算法要做的是⽣成后继状态并区分⽬标状态与⾮⽬标状态，⽆信息搜索⼀般分为以下⼏类：宽度优先搜索，⼀致代价搜索，深度优先搜索，深度受限搜索，迭代加深的深度优先搜索，双向搜索

其次是，【**有信息搜索算法**】。

如果能找到⼀种搜索⽅法，能够充分利⽤待求解问题⾃⾝的某些特性信息，以指导搜索朝着最有利于问题求解的⽅向发展，那么搜索的效率就会⼤⼤提⾼。在这⾥就要⽤到启发式搜索了，也称为有信息搜索策略。

关于问题求解中的搜索算法，大家还可以关注：连续空间中的局部搜索，使用不确定动作的搜索，基于博弈的对抗搜索，以及基于“约束满足问题”的相关搜索方法。对人工智能感兴趣的同学，希望课后多看我们的参考书。**我们现在很多研究生的主要还落在改模型的层面，这其实限制了我们形成类似ChatGPT这样的突破性成果的能力**。要想以后参与到突破性的工作中，请重视对基础知识的学习。

**P20：**

下面我们介绍面向“知识、推理与规划”的逻辑智能体。

**P21：**

我们容易理解：人类通过对世界的认识来帮助他们做事。具体来说，人类智能不是靠反射机制，而是靠知识在人脑的内部表示，来进行推理和决策。**人类的这种智能，属于基于知识的智能体**。

我们这一部分会强调“**将逻辑作为支持基于知识的Agent的通用表示**”。这样，我们的智能体可以通过对信息的组合和再组合，最终去适应各种用途。

我们要理解基于知识的智能体。先讲一个故事：

你的室友回到了寝室，他/她⾝上湿透了。你知道以下⼏件事：

1. 你的室友⾝上湿了

2. 如果你的室友⾝上湿了，要么是因为下⾬，要么是洒⽔车

3. 如果你的室友是被洒⽔车弄湿了，那么洒⽔车⼀定正在运⾏

4. 如果你的室友是被⾬淋湿了，那么他/她⼀定没带伞

5. 伞不在伞架上

6. 如果伞不在伞架上，那么要么是你拿了伞，要么是你室友拿了伞

7. 你没有拿伞

**思考**

1. 你能推断出洒⽔车是否在运⾏吗？

2. ⼈⼯智能能推断出洒⽔车是否在运⾏吗？

**P22：**

这个故事其实对应了一个知识库：

1. 室友湿了

2. 室友湿了=> (室友是被⾬淋湿的 OR 室友是被洒⽔车淋湿的)

3. 室友是被洒⽔车淋湿的=>洒⽔车在运⾏

4. 室友是被⾬淋湿的=>NOT(室友拿了伞)

5. 伞不见了

6. 伞不见了=> (你拿了伞 OR 室友拿了伞)

7. NOT(你拿了伞)

而基于知识的智能体的核心部件就是这个知识库。知识库=**由形式语言组成的句子的集合**。这里的句子我们也称为语句，是一个技术术语。它与英语、中文等其他自然语言相关，但不完全等价。

知识库里的语句用知识表示语言表达，表示了关于世界的某些断言。如果某语句是直接给定而不是推导得到的时候，我们就把它称为**公理**。

**P23：**

其次，逻辑是什么。

**逻辑是⼀个⽤来处理事实的形式化系统，⽬的是得到正确的结论。事实可以由我们刚才提到的知识库来支撑**。

我们知道，知识库是由语句构成的。一般，我们需要根据表示语言的语法来表达这些语句。

**语法**是构造知识库有效句⼦的规则~~：例如, x + 2 ≥ y 是有效的算术语句, ≥x2y + 就不是；~~

逻辑则必须定义语言的语义（也就是定义那些由语法规范后的语句的含义）。

**语义**是句⼦的“含义”,定义了每个语句在每个可能世界的真值，或者逻辑语句和真实世界之间的关系，具体⽽⾔，语义决定了句⼦的真假，~~例如，x + 2 ≥ y~~

~~为真时可以取x = 5 且y = 7~~

当需要精确描述逻辑时，我们就需要建立**模型**来代替“可能世界”。可能世界可以被认为是（潜在的）真实环境，智能体可能在也可能不在其中，而模型是数学抽象，对于每个相关的语句，每个模型都有固定的针织（真或假）。

**P24:**

命题逻辑是一种虽然简单但很强大的逻辑。命题是能判断真假的陈述句。

~~下面我们介绍一种虽然简单但很强大的被称为命题逻辑的逻辑。~~

~~命题是能判断真假的陈述句。这种陈述句的判断只有两种可能：⼀种是 正确的判断，⼀种是错误的判断。称判断为正确的命题的真值（或 值）为真，称判断为错误的命题的真值（或值）为假。~~

~~因此又可称命题是具有唯⼀真值的陈述句或判断结果惟⼀的陈述句。~~

~~于是，命题的真值就是判断的结果，真值的取值在真与假中⼆者取⼀，真命题就是真值为真的命题，假命题则是真值为假的命题。~~

**P25：**

~~我们看看，命题的两个分类~~

~~一类是简单命题（也叫原⼦命题）：这是由简单陈述句构成的命题~~

~~一类是复合命题：这是由简单命题⽤连接词联结⽽成的命题~~

~~这个表格是复合命题中连接词的真值表。连接词的运算规则可以用真值表总结，真值表指定了复合命题在“其组成部分的真值赋值”后如何计算真值。具体细节请参考参考书。~~

**P26**：

下面介绍一阶逻辑。

**P27：**

**命题逻辑作为表示语言，可以阐述逻辑和基于知识的智能体。但是命题逻辑是一种表达能力很弱的语言，无法以简洁的方式表示复杂环境的知识**。而一阶逻辑，则有丰富的表达能力，可以表示大量的常识知识。

~~一阶逻辑的研究内容主要是从命题逻辑种的组成部分（比如个体词、谓词、量词等）间的逻辑关系，总结出正确的推理形式和规则。~~

命题逻辑和一阶逻辑的最根本区别在于每种语言对现实本质的假设是不同的。**数学上来说，命题逻辑假定了世界中的事实要么成立要么不成立，每个事实只能处于真或假的两种状态。一阶逻辑的假设更多，也就是世界由对象构成，对象之间的某种关系或者成立或者不成立**。一阶逻辑的对应形式模型比命题逻辑的更加复杂。

**P28：**

~~要想使用一阶逻辑，首先需要明确它的基本要点，⼀阶逻辑的元素有：对象、关系、函数、等价性；一阶逻辑语法的基本元素有：常量、谓词、函数、变量、连接符、等词、量词；一阶逻辑的原⼦语句和术语分别以这两种方式表达；一阶逻辑的复杂语句是由原⼦语句通过连接词和量词连接组成的；右图展示了一阶逻辑的语法。~~

~~在人工智能的“知识、推理与规划”方面，同学们还可以通过参考书了解下：规划与决策方法，知识表示相关内容。对自动驾驶等感兴趣的同学，应该关注规划与决策方法部分；对知识图谱感兴趣的同学，应该看看知识工程这一部分。~~

**P29：**

下面介绍不确定的量化。

**P30：**

智能体的环境可能是部分可观察或不确定的，这个时候智能体需要处理不确定性。但是，智能体在不确定性环境下，可能很难预测经过一系列动作后自己所处的状态或位置。

对此，**一些传统的解决办法是基于“对所有可能世界状态组成的集合的表示”（这被称为信念状态）和“生成应急措施”来实现对不确定性问题的求解**。但是这些方法真正用于创建Agent程序时仍有很多缺陷：

* 当解释观察到的部分信息时，逻辑 Agent 必须考虑每一种逻辑上可能的解释。这导致信念状态的表示无法忍受地庞大而复杂。
* 一个处理所有可能意外情况的正确的应急规划必须考虑每一种可能情况(即使是可能性很小的情况)，可能会变得非常庞大。
* 有时，没有可以保证达到 Agent 的目标的规划，然而Agent 仍需要有所行动。Agent必须能够对这些不能确保达到目标的行动规划的好坏做出对比。

~~以自动驾驶的出租车Agent为例。出租车Agent的目标是将乘客按时送到机场。Agent做了一个规划 At:在飞机起飞 90分钟前出发，并以合理的速度驶向机场。即使机场仅仅~~

~~只有5英里远，一个逻辑 Agent 也无法确定地得到这样的结论:“规划At将让我们及时到~~

~~达机场”，然而它可以做出这样的弱一些的结论:“规划At将让我们及时到达机场，只要车~~

~~不抛锚，汽油不耗尽，不遇到任何交通事故，桥上也没有交通事故，飞机不会提前起飞，而且没有陨星砸到我的车，……”。~~

~~可以看到，这些条件没有一个是能够演绎的，所以也无法推断这个规划能否成功。这就是一个限制问题(qualification problem，见第7章)。对于限制问题，目前为止我们还没发现直正的解。~~

**P31：**

对于这样的情况，我们需要一些方法来处理不确定性。不确定环境下，智能体的知识提供相关语句的信念度，**处理信念度的主要⼯具是概率理论**。概率提供了⼀种⽅法以概括由我们的惰性和⽆知产⽣的不确定性~~：例如可以说规划A25将使我们及时到达机场的概率(可能性)是0.04。~~

**概率论是⽤于处理信念度的理论，效⽤理论是对偏好进⾏表⽰和推理，是有效性的度量**，进而我们可以得出**决策理论 = 概率理论 + 效⽤理论**。决策理论的基本思想为⼀个智能体是理性的，当且仅当它选择能产⽣最⾼期望效用的⾏动；期望效用是行动的所有可能结果的平均。

**P32：**

概率是关于什么的？

像逻辑断言一样，概率断言时关于可能世界的断言。逻辑断言考虑的是要严格排除哪些可能世界（排除所有那些断言不成立的世界），而概率断言考虑的是各种可能世界的可能性有多大。

当某个概率断言在没有完全确定概率模型的情况下给出了基本的概率模型的限制，**这样的概率是无条件概率或者先验概率**。它们是指不知道其他信息的情况下对命题的信念度。

然而，大多数情况下，我们会有一些已经为我们已知的信息，通常称为证据。比如我们已经看到一个色子是5，此时我们还在等待第二个色子停止翻滚。这种情况下我们不再关心得到相同数字5的先验概率，而是更关心再第一个色子是5的前提下两个色子结果相同的概率。**这样的概率称为条件概率或后验概率**。

**P33：**

**在不确定性的量化这一部分，大家课后可以重点关注链式法则**。最近几年深度学习席卷了机器学习多个领域，计算机视觉，自然语言处理，推荐系统等等，成为各个领域当中的最佳方法。深度学习或者说神经网络当中最重要的一个要素是反向传播算法（Back propagation）。反向传播算法与数学当中求导链式法则有非常密切的关系，当前的流行的网络结构，无不遵循这个法则，比如计算视觉当中的LeNet、AlexNet、GoogLeNet、VGG、ResNet，还有其它的各种网络。**有同学能帮我解释下反向传播算法吗？**

我记得，新加坡国立大学校长青年教授，北京潞晨科技有限公司董事长尤洋老师（2021年4月入选亚洲福布斯30岁以下精英榜）在社交媒体上讨论过：他认为深度学习中的反向传播就是复合函数求导中（链式法则）。

大家课后学习后，请思考下是否认可他的观点。

https://zhuanlan.zhihu.com/p/37082439

**P34：**

**贝叶斯规则**大家有时间也要课余学习。这是多数进⾏概率推理的⼈⼯智能系统的基础。

**P35：**

下面介绍概率推理。

**P36：**

刚才我们讲到：**决策理论 = 概率理论 + 效⽤理论**

在参考书上概率理论的相关章节中，作者强调了独立性和条件独立关系对于简化世界的概率表示的重要性。为此同学们可以关注**一种以贝叶斯网络的形式明确表示这些关系的系统化方法**，最终我们希望通过使用这类方法捕获不确定性知识。

我们要了解到：**在概率理论中，完全概率分布能够回答关于问题域的任何问题, 但是存在这样的两个问题**。

首先是随着变量数目增多会增大到不可操作的程度.

其次为每个可能世界逐个指定概率是不可能的

**P37：**

那么为了解决这两个问题, 我们就可以采取贝叶斯网络。

**贝叶斯网络是一种使用简单局部分布, 也就是条件概率, 来描述复杂联合分布的技术**. 贝叶斯网络是一个有向无环图，也称之为图模型. 它是用于表示变量之间的依赖关系. 理论上来说, 它可以表示任何完全联合概率分布。

什么是有向无环图？图由顶点和连接这些顶点的边所构成。每条边都带有从一个顶点指向另一个顶点的方向的图为有向图。

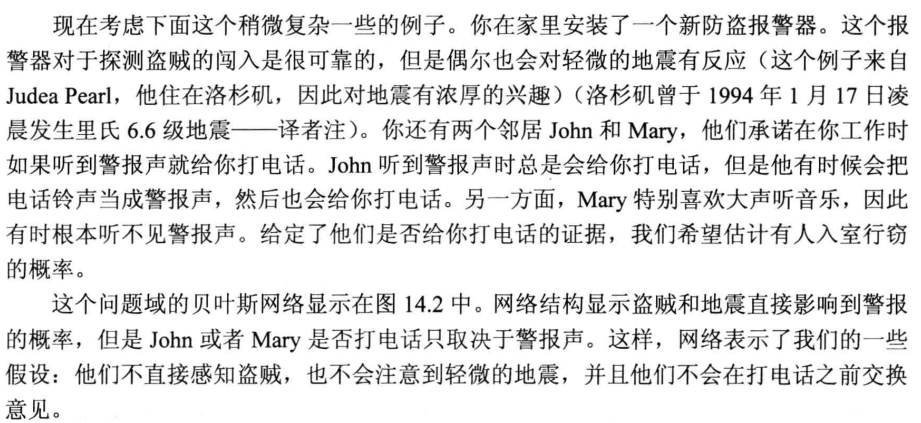
在图论中，如果一个有向图无法从某个顶点出发经过若干条边回到该点，则这个图是一个有向无环图（DAG图）。

https://zh.wikipedia.org/zh-hans/%E6%9C%89%E5%90%91%E6%97%A0%E7%8E%AF%E5%9B%BE

https://cloud.tencent.com/developer/article/1732855

**P38：**

~~这里我们介绍一个贝叶斯网络的例子.~~

~~~~

~~图中这个贝叶斯网络可以看到：地震和盗贼可能会引发家用报警器, 听到了报警器的声音之后, 邻居John和Mary可能会给我打电话. 图中的一个节点对应一个随机变量, 有向边表示的是连接的变量间的直接影响, 通过一个条件概率分布可以量化父结点对于该结点的影响 , 这就构成了一个条件概率表. 这样的网络拓扑反应了因果关系。~~

**P39：**

~~在对于贝叶斯网络有个基本认识之后, 我们需要对于贝叶斯网络的语义有一个初步的了解. 贝叶斯网络是完全联合概率的一种表示. 完全联合概率分布由局部条件概率分布的乘积进行定义. 如下面的公式所示.~~

~~举个例子来说, 比如在刚才的例子中, 两个人都给你打电话的概率可以表示成这样的形式, 那么我们就可以计算出两个邻居都给你打电话的概率是0.00063。~~

**P40：**

~~然后我们介绍一下构造贝叶斯网络的方法. 首先, 我们需要构造贝叶斯网络中节点, 我们需要选择一组排好序的随机变量. 然后, 我们构造贝叶斯网络中的边, 遍历全部的节点, 我们执行这样的步骤:~~

~~首先从这个节点之前的节点选择父节点的最小集合, 然后在父节点和 之间插入一条边, 最后我们可以得到一个条件概率表.~~

~~构造贝叶斯网络过程利用到了乘法法则和链式法则。具体细节，大家有兴趣，课后看书。~~

**P41：**

基于贝叶斯网络，我们就可以进行推理。一种方法是通过枚举来进行推理，也就是在贝叶斯网络中通过计算条件概率的乘积再求和，再来回答需要推理的问题。

比如，我们在报警器问题中，“为了知道两个邻居都给你打电话下，出现盗贼的概率”，可以遍历贝叶斯网络，将每条路径上的值相乘, 并且再“+”号处求和。

可以得到: 在两个邻居都给你打电话的条件下，出现盗贼的概率大约是28.4%。

**将贝叶斯网络中每条路径上的值相乘, 并且再“+”号处求和，这种做法可以采用“归一化指数函数，或Softmax函数”的形式，用在深度学习分类任务中。这种Softmax函数也叫贝叶斯分类器**。

谁能来讲讲自己对Softmax函数的理解？这也是个很重要的概念，大家后续估计都会在其他课程或毕设，或科研中掌握。

**P42：**

需要注意：

我们最终得到的网络可能有单连通网络和多联通网络两种形式.

对于单连通网络: 1. 网络中任意两个结点之间顶多只有一条(无向)路径 2. 精确推理的时间和空间复杂度与网络规模呈线性关系

对于多连通网络 1. 变量消元算法在最坏情况下具有指数量级的时间和空间复杂度。 2. **贝叶斯网络推理是#P难题（严格难于NP完全问题）**

对于多连通网络，大家要小心。需要用算法去降低其复杂度。比如用聚类算法。这是未来大家可能会重点研究或涉及的算法之一。

在不确定知识与推理这部分，除了我们提到的不确定性量化、概率推理，“时间上的概率推理”一章中的隐马尔可夫模型、卡尔曼滤波器等都应该有所了解。隐马尔可夫模型虽然不如基于深度学习模型效果，但是在语音识别领域已经得到广泛使用。

还需要有所了解的是效用理论、效用函数、决策理论、价值迭代、策略迭代等。推理不是我们的最终目标，我们最终目标是制定决策。而这些章节的内容则是为大家提供决策方面的知识。

**P43：**

下面介绍样例学习。样例学习又叫例中学(learning from worked examples),是学习者从对样例的研习中习得样例中蕴含的知识、技能的过程

**P44：**

这一部分的内容聚焦于一类实际应用非常广泛的问题: 我们希望一个智能体可以从一组“输入-输出”中学习能够预测新输入相对应的输出的函数。

为什么我们希望智能体能够学习？有同学回答一下吗？

如果Agent的设计能够被改进，为什么设计者不以“改进”为思路，自己进行编程？这主要包括了三个理由：

首先，设计者不能预测Agent可能身处的环境，比如迷宫机器人需要通过学习新迷宫的布局才可以走新迷宫，如果走新迷宫都需要设计者重新编程，那就效率太低了。其二，不能预期随时间推移可能出现的所有变化，比如股票预测问题，需要智能体自适应地预测。其三，有的时候，人类程序员也不能编写出完成任务的程序，除非使用学习类算法。

但是，要让智能体具备学习能力, 需要依赖四个改进因素：

1. 改进哪个部件（实际上Agent任何部件都可以通过从数据中进行学习获得改进，这是我们很常用的研究范式）

2. agent具备什么样的先验知识

3. 如何表示数据和组件

4. 学习中可用的反馈是什么.

**其中对于学习可用的反馈主要有三种, 分别是无监督学习, 监督学习和强化学习**.

对于无监督学习, 我们没有来自外部环境的直接反馈（输出值），我们需要自行组织学习的输入, 这需要我们挖掘数据中的隐含规律.

对于监督学习.对每一个输入模式都有一个正确的目标输出, 那么外部环境的反馈需要告知agent的期望输出是什么，或当前输出是否正确, 通过这样的反馈机制 我们可以从一组输入－输出的实例数据集中学习出一个函数, 或者说是一组模型参数. 那么当新的数据到来时，可以根据这个函数预测结果.

对于强化学习, 我们主要的反馈信息是奖励和惩罚组成的序列, 以及奖惩之前的哪些动作需要为结果负主要责任

大家是否还听过弱监督、自监督等概念？了解的同学请介绍下。

**P45：**

监督学习是一种典型的样例学习。

我们再介绍一下监督学习的基本概念, 对于监督学习, 它的任务是：给定由N个“输入-输出”对样例组成的训练集。其中，每个y由未知函数y=f(x) 生成。

**监督学习的任务**就是发现一个函数h ,它逼近真实函数f 。 其中, x和y可以是任何形式的值~~, 比如在分类任务中, y就是一些离散的值, 比如我们需要对图片分类, 识别这是猫还是狗还是两者都不是。~~

~~在回归任务中, 我们则需要去预测比如明天的气温具体是多少度.~~

在监督学习中, 如果我们假设的函数 h能够完美拟合某些数据, 我们称 h与f是一致的.

不同的假设h可能对数据的拟合程度不一样, 复杂程度也不同, 如下图所示: 虽然每个函数都完美拟合了数据点, 但是显然复杂的程度是不同的。

我们的假设函数h对应了一个假说空间H。这时就面临了一个基本问题：如何从假说空间中的多个一致假说进行选择？

一个答案是：选择与数据一致的**最简单**假说。这个原则被称为，奥卡姆剃刀规律（不要人为地将事情复杂化；如无必要，勿增实体，勿增实体）。有谁能给我们介绍下奥卡姆剃刀定律在生活中应用的一些例子（软件工程里也许特别适用）。

不过**定义简单**并不是一件容易的事。

**P46：**

一种常见（也是最简单的）的监督学习是决策树。~~这里我就不细讲了。~~

~~决策树将从一组训练数据中学习到的函数被表示为一棵树 决策树通过把实例从根节点排列到某个叶子节点来分类实例： 树上的每一个节点指定了对实例的某个属性的测试 节点的每一个后继分支对应于该属性的一个可能值。 叶子节点即为实例所属的分类. 举个例子来说, 我们要决定是否在一个餐厅里面等餐, 我们这里列出全部的属性, 比如我们是否有一个替代的选项? 等待的区域是否舒适, 现在是不是周五或者周六, 我们是否饥饿等等一共十个属性, 这些属性可能会影响我们最终的决策. 我们则可以通过这些属性构造出一棵决策树,比如最右边的路径可以表示为右上角的公式的形式。~~

**P47：**

~~那么如何构造决策树呢. 假设, 对于刚刚提到的这些属性, 我们收集到了一些样例, 其中, 有每个属性具体的值,以及对应 的最终的结果. 在构造决策树的时候, 总是优先选择重要的属性, 从而将问题分解为更小的问题, 以便于进行递归求解. 在递归中考虑四种情况: 1. 如果剩下的都是正例或者负例, 则可以输出对应的结果 2. 如果剩下的既有正样本又有负样本, 则选择最好的属性继续分裂 3. 如果没有剩下任何样例, 则对于这个属性的组合没有观察到任何样例, 可以返回父节点中 最多的分类 4. 如果剩下的既有正样本又有负样本, 且没有属性可用了, 则返回剩余样本中最多的分类.~~

**P48：**

另一类曾经被广受重视的监督学习是人工神经网络。

如图所示, 这是一个简单的神经元的简单数学模型。 包括输入, 输入函数, 激活函数以及输出, 通过输入函数求和, 以及激活函数激活, 最终可以得到输出. 粗略地说，当输入的线性组合超过某个（硬或软）阈值时，神经元“点火”。一个神经网络是由一组单元连接而成：网络的性质由它的拓扑和神经元的性质确 定。**与我们现在认识到的生物神经元工作原理类似**（右图）。

在以后我们的群体智能部分介绍中，还会讲到它。

**P49：**

人工神经网络模型一直受到深层参数训练问题困扰。**这个问题，直到人工神经网络模型提出几十年后才有研究者提出了反向传播算法来解决**。

具体来说，为了学习到正确的参数, 我们需要从输出层反向传播误差, 在这过程中不断的更新权重以及传播误差, 不断重复这个过程, 最终可以拟合出一个适用于当前问题的函数.

输出层和隐藏层的权重的更新方式类似, 除了Δ值的计算略有不同.代表的是网络的输出与目标输出之间的差值。

总体的，反向传播过程可以归纳为:

1. 利用观测到的误差值，计算输出单元的∆值

2. 从输出层开始，对网络中的每一层重复下面的步骤，直到到达最前面的隐层 (a)向前面的层反向传播∆值 (b)更新两层之间的权值。

样例学习我们就介绍到这里。在样例学习中，支持向量机等知识大家也需要关注。

除了样例学习，大家也可以看看如何将学习的逻辑公式化这些内容。无监督学习和强化学习同样很重要，在以后的课程中我们会再次讲到。

**P50：**

下面介绍自然语言处理。

**P51：**

人类因为具有语言的能力而区别于其他物种。虽然黑猩猩、海豚等动物也能掌握数百个符号构成的词汇，但是**只有人类能够适用离散的符号，可靠地传递大量信息**。

随着人工智能的发展，人类开始想要智能体能够处理自然语言。因为，我们希望智能体能与人类交流，也希望智能体能从人类的自然语言中获取信息。

图灵测试的提出，也是基于语言的（有同学能说出图灵测试的内容吗？图灵提出，**测试者与被测试者（一个人和一台机器）隔开的情况下，通过一些装置（如键盘）向被测试者随意提问。 进行多次测试后，如果机器让平均每个参与者做出超过30%的误判，那么这台机器就通过了测试，并被认为具有人类智能**）。在这个定义下，大家感觉图灵测试是否已经过时？

---

目前，互联网上已有超过万亿数量的信息网页，几乎所有的这些页面都是用自然语言描述的。Agent想要获取知识，就需要理解（至少部分理解）人们所使用的具有歧义的、杂乱无章的语言。

对于这个需求，我们考虑的就是一个**“信息查找任务”**的问题。这又涉及了三个子问题：**文本分类、信息检索和信息抽取**。

解决这些问题，有个共同的需要，**就是要采用语言模型**。这个模型是用来预测语言表达的概率分布情况的。

所有的形式语言，如Java和Python等编程语言，都有精确确定义的语言模型。在Python语言中，“print(2+2)” 是一段合法的程序，而“2)+(2 print”则不是。合法的程序的数目虽然是无限的, 但是我们可以通过语法来 描述.

但是在自然语言中却有一些不同，它不能描述为一个确定的语句的集合. 一般认为"Not to do"是符合语法的, "To not do"是错误的, 但是这并不是绝对性的错误. 因此, 通过概率分布来定义自然语言模型是更加有效的.

自然语言也存在歧义. 同一句话在不同的语境中会有不同的含义, 对一个句子的解释应该使用多个意义上的概率分布

自然语言的模型是对自然语言的近似. 由于自然语言在不断的发展中, 不断的有新的单词和语句被制造出来, 所以只能做到近似, 而不能完全的模拟

**P52：**

这三个是常用的语言模型。我不再细讲。

*首先介绍n元字符模型. 长度为n的序列我们称之为n元组的. N个字符序列上的概率分布就成为n元模型. 假设我们有一个包含100 个字符的三元模型, 那么他就有100万个参数, 显然非常的复杂, 因此对于参数的估计我们可以通过从语料库中统计得到. 得到 了n元模型, 一个最简单的任务就是语言识别, 计算机可以很轻松的识别出来对应的短文本是英语还是德 语. 同时, 这个模型还可以完成诸如拼写纠错, 体裁分类, 命名实体识别等任务.*

*但是对于一个句子“The program issues an http request” 其中http不在英语的语料库中, 则很直接采用n元 字符模型不能够识别出来这是英语. 因此需要对于调整低频计数的概率, 这个过程就要平滑. 经典的平滑方 法有: 拉普拉斯平滑, 回退模型, 线性插值平滑等.*

*n元单词模型显然比n元字符模型复杂的多, 原因在于单词的集合远远比字符的集合大, 一般一个语言的符 号不会超过100个, 但是单词的数量往往不可计数. 同时还有未知的单词和数字和邮箱等特殊的组合. 那么这么复杂的模型到底能带来什么好处呢, 在一个小的文本中建立单词模型, 我们可以从123元的单词 模型的随机取样.显然三元和二元的表达远比一元的效果更好*

**P53：**

语言模型之后，现在我们详细介绍文本分类任务。

**给定某个文本，我们需要判断它属于预定义类别集合中的哪个类别**。

语言识别 和体裁分类都是文本分类的例子，情感分析（判断电影或产品评论是积极的还是消极的）和垃圾邮件检 测（判断电子邮件信息是垃圾邮件或非垃圾邮件)也是文本分类。

~~我们可以将垃圾邮件检测看作一种监督 学习问题。训练集很容易得到：正例（垃圾邮件）在垃圾邮件文件夹里，负例(非垃圾邮件)在收件箱里。 下面是一些摘录样本, 其中spam是垃圾邮件, ham是正常邮件. 从这里面我们可以发现"for cheap"和"you can buy"等n元词序列很想吃垃圾邮件的特征 还有"yo,u d-eserve"等也可以是垃圾邮件的特征.~~

**对于文本分类任务, 我们有两种方法 一种是基于语言模型的方法**, 通过训练得到一个计算的垃圾邮件的n元语言模型以及正常邮件的n元语言模型. 然后使用贝叶斯规则对其分类. **另外一种是机器学习的方法**, 通过将n元组作为特征,使用机器学习算法进行分类任务, 只要确定了垃圾邮 件的特征集, 就可以使用任务监督学习的技术, 完成分类

**P54：**

信息检索(Information retrieval, IR)的任务是寻找与用户的信息需求相关的文档。

网上的搜索引擎就是 一个众所周知的信息检索系统的例子~~。用户将类似[AI book]的查询信息输入到搜索引擎，就能得到相关 的网页的列表~~。

一个信息检索系统具有如下特征：

1. 文档集合。每个系统都必须确定其需要处理的文档：一个段落文本、一页文本还是多页文本。

2. 使用查询语言描述的查询。查询描述了用户想知道的内容。~~查询语言可以是一个单词列表，如[AI book];可以是一个必须连续出现的单词短语，如[“AI book”]~~

3. 结果集合。该集合是文档集合的子集，包含了IR系统判断的与查询相关的那部分文档。

4. 结果集合的展示。结果集合可以简单地用有序的文档标题列表来展示，也可以采取复杂的展示方法。当前，信息检索系统使用一种简单的基于词袋的语言模型，它在处理大规模文本语料时，在召回率和准确率上 也有好的表现。在万维网语料上，链接分析算法能够提升性能。

**信息检索也是我们用户交互技术中的重点内容**。

**P55：**

信息抽取是一个通过浏览文本获取特定类别的对象以及对象之间的关系的过程。

**其典型的任务包括，从网页中抽取地址实例信息，获取街名、城市名、州名以及邮政编码等数据库字段的内容；从天气报道中抽取暴风雨信息，获取温度、风速以及降雨量等数据库字段的内容**。在受限的领域内，信息抽取可以达到高的准确率。而在更一般的领域内，则需要更复杂的语言模型和更复杂的学习技术。

目前自然语言处理技术大放光彩。除了机器翻译、语音识别外，还有哪些应用？

**P56：**

下面介绍感知（智能）。一种说法，我们正从感知智能时代迈入认知智能时代。有同学听说过吗？能否讲讲其中区别。

**P57：**

好的，我们看看人工智能中的感知任务是什么。

**感知是通过解释传感器的响应而为Agent提供它们所处的世界的信息**。

传感器作为Agent程序的输入，提供环境测量信息。传感器可以是一个简单的开关（比如输出一个比特表示开或者关），也可以是复杂如人眼一样的部件。

对于机器人来说，**有些机器人可以进行主动传感**，也就是说它们发射出一个信号，比如雷达信号或超声波，然后感知从环境中反射回来的信号。**因此可以看到，人工智能研究其实通常面临跨学科的交叉知识的学习，尤其是物理知识。像去年很火的扩散模型，思路就来源于热力学，而有些学习模型更是直接结合物理模型构建的**。

这里, 我们将重点讨论一种感知形态：视觉感知

对应视觉感知，主要包括三种图像处理技术, 分别是：**边缘检测**,**纹理分析**以及**光流计算**, 从而支持agent进行**图像分割**。

**P58:**

1、边缘检测

边缘是图像中的**直线或曲线段**，**穿过边缘的图像亮度有“显著的”变化**。

边缘检测的目标是根据大量的、成兆字节的图像数据进行抽象，形成更紧凑、更抽象的表示方式（**这个任务很好地体现了：压缩即智能**）。

边缘检测可以获得图像中的各种轮廓，在我们的项目中也有实际的应用。

**P59：**

2、纹理分析与光流计算

纹理, **它指的是在表面空间上重复出现的、能够通过视觉感觉到的模式**。

纹理的实例包括建筑物上窗户的 模式、汗衫上的针脚、美洲豹皮肤上的花斑等. 在不同的光照下, 纹理一般是不变的, 这使得纹理成为了识别物理目标的重要依据。对应了纹理分析技术。

光流则**描述的是一个视频序列。图像中的物体在运动或是我们的镜头在相对物体运动时，由此会引起的图像 中的明显的运动，我们称之为光流**。 如图所示, 最右边图是对应一帧到另一帧像素的位移。计算这个光流通常是一个重要任务。

**P60：**

图像分割是指**基于像素点的相似性将图像分解成若干区域的过程**。

其基本思想如下：每个图像像素都可以关 联某些视觉特性，诸如亮度、色彩和纹理。在一个物体中，或者是它的单独一部分中，这些属性的变化 相对非常小，而穿过物体之间的边界时，典型情况下这些属性中的一个或多个会出现较大的变化。

**有两 种方法可用于图像分割**，一种主要致力于检测这些区域的边界，而另一种则致力于检测出区域本身。 如图所示(a)是原始图像, (b)是边缘轮廓(c)和(d)则是对应不同精细程度的区域分割,(c)更加精细,(d)更加粗糙。

大部分项目中图像分割与边缘检测是同时用到的。最终目标是完成图像分割。

**P61：**

这是一个典型的应用研究：基于外观的目标识别。

外观指的是一个物体看上去的情况。一些物体类（比如说棒球）在外观上会变化很小：在大部分的情形下 这类物体看起来基本一样。

基于此，我们可以计算一些描述包含这些物体的图像的特征，然后据此训练 出分类器，对物体进行分类和识别。

**其他类别的物体—如房子，它们的变化一般很大。如一所房子可能具有不同的大小，颜色或者形状，甚 至在不同的角度来看它也可能是不同的**。一个理念是物体一般由些局部部分组成，而物体的变化主要是 这些局部部分相互之间的移动导致的。我们可以通过检测局部的特征，来检测整体是否存在， 而不必关心各个部分的位置。

在识别的过程中, 还存在这种各样的问题, 如遮挡, 不同的朝向, 形变, 透视等等。这些问题会导致图像的特征被影响. 但是无论物体怎么变化, 总会在图像中表现出某些固定的结构.

不过，我们现在的物体识别算法已经可以达到非常好的效果, 如图所示:

这种进步，是前几年吸引大量人才涌入人工智能领域的关键。包括现在还有大量的研究人员在研究计算机视觉相关问题。像感知中的三维重建、基于结构的物体识别等，我们学院的张林老师、梁爽老师就在做长期深入研究。

**P62：**

大家学习人工智能也不要忽视机器人学，以及机器人学所研究的硬件对象（**具身智能**）。其次，正如我们基于参考书的基础概念这部分教学内容开始所提到的，大家也不要忽视人工智能的发展伦理、数理基础，以及未来动态。

**人工智能的发展伦理方面：**要关注信息安全、隐私保护、机器人权利、公平与偏见等议题。

**人工智能的数理基础方面：**要注意到一些数学异议，比如图灵和哥德尔证明了某些数学问题原则上是不能用形式系统解决的（**哥德尔不完全性定理**）。甚至，也要注意到一些物理原理（**海森堡测不准原理**）：一个微观粒子的某些物理量（如位置和动量，或方位角与动量矩，还有时间和能量等），不可能同时具有确定的数值，其中一个量越确定，另一个量的不确定程度就越大。薛定谔的猫就是讲这个问题。这些异议或原理，就是我们的科学边界，大家都看过三体吧。高水平的人工智能研究最终都要触碰科学边界（一个可能例子：就好比图灵测试已经被触碰）。另外大家也要广泛地关注存在于各个学科（自然科学方面）的第一性原理知识，比如能量守恒，要在未来深入的科研中将自己的研究与各类第一性原理反复印证。*大家思考下为什么我们把数学的定理说成异议，海森堡测不准原理说成是原理*？

**人工智能的未来动态方面：**要关注可解释性、安全性（比如智能体的安全保护）、如何衡量人工智能（图灵测试似乎已经过时）。

大家能否探讨一下：我们在沿着正确的方向前进吗？

目前《人工智能：现代方法》第四版中文版于近期问世，一篇Stuart Russell 教授的专访大家可以课后看看。最新第四版是两位作者把近十年 AI 进展，尤其是深度学习所带来的影响纳入整体框架后给出的最新阐释，体现了两位大师对人工智能趋势和学科体系发展的洞见。大家注意：**《人工智能：现代方法》第四版有一个重要的更新，那就是不再假设 AI 系统或智能体拥有固定的目标**。

对此作者是这样解释的：目标其实是一件非常复杂的事情。例如我说午饭想买个橙子，这可以是一个目标，对吧？在日常语境中，目标被视为某种可以被实现的东西，一旦实现了，事情就完结了。但在哲学与经济学定义的理性选择理论中，其实并不存在这样的目标，我们有的是对各种可能的未来的偏好或排序，每一种可能的未来都从现在一直延伸到时间的尽头，里面包含了宇宙中的所有。我想，这是对目标、对人类真正想要什么的一种更复杂、更深远的理解。

https://weibo.com/ttarticle/p/show?id=2309404871275821793925