人工智能：模型与算法

人工智能发展与挑战

吴飞

浙江大学计算机学院

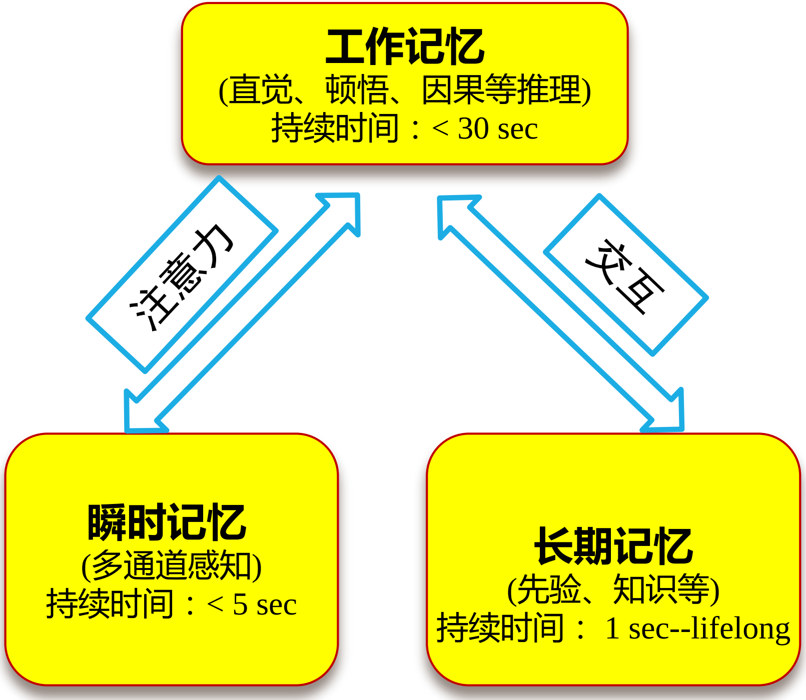
提纲

1、 记忆驱动的智能计算

2、 可计算社会学

3、若干挑战

记忆是人类认知的基石

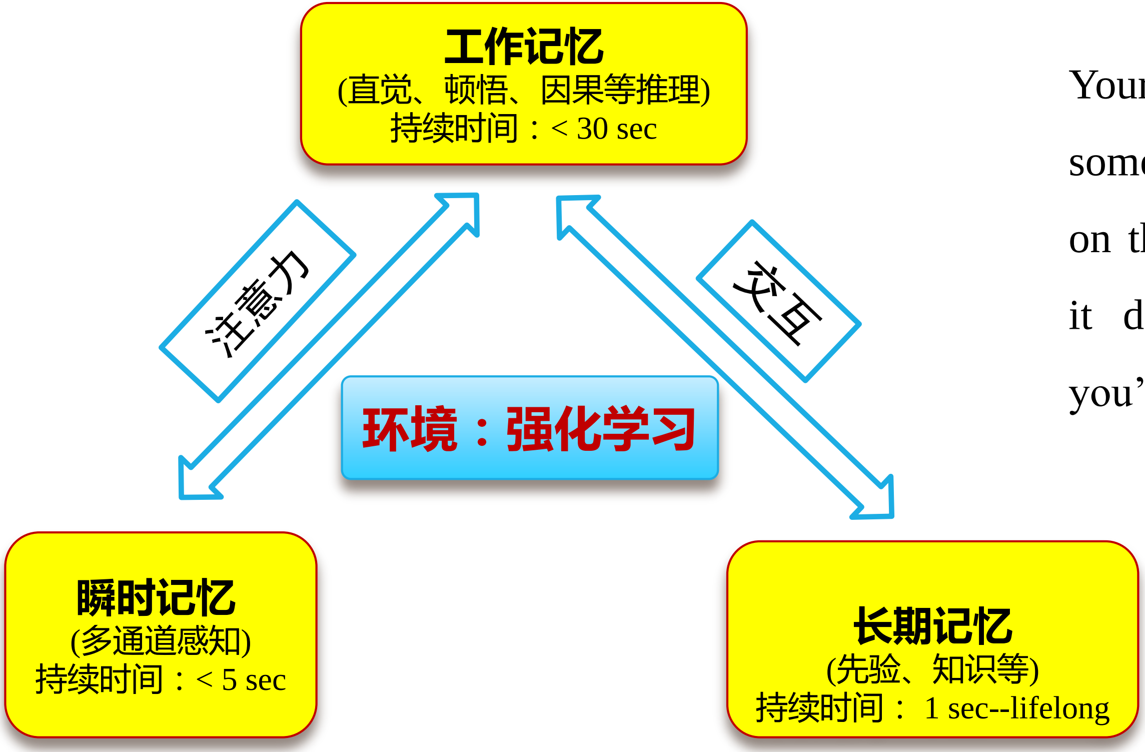


知之在人者谓之**知**	**知觉：**人所固有认识外界客观事物本能，如视觉、听觉和触觉等能力
知有所合谓之**智**	**智慧：**知觉对外界事物的认知
所以能之在人者为之**能**	**本能：**人身上所具用来处置事物能力
能有所合谓之**能**	**智能：**对外界所产生的认知和决策

《荀子．正名》

Alan Baddeley, Working memory: looking back and looking forward, Nature Reviews Neuroscience 4, 829–839, 2003

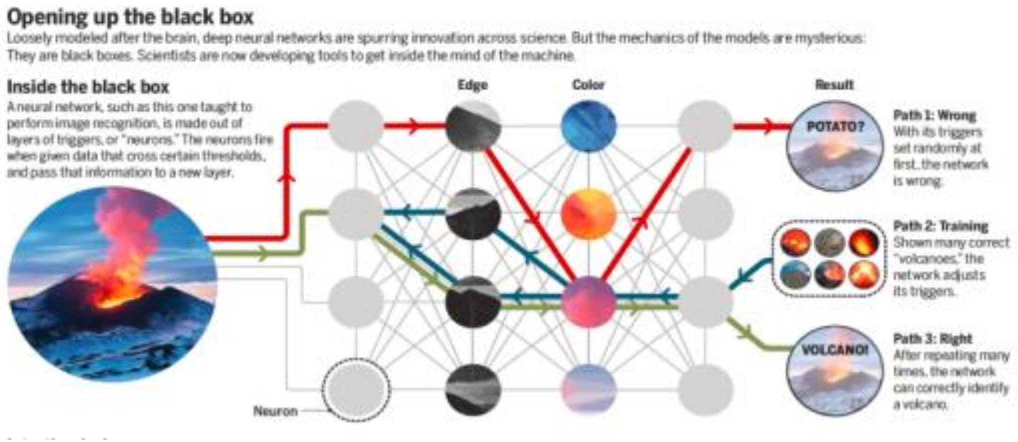
记忆单元之间及其与环境的交互是提升智能能力的重要途径



Your ability to remember something doesn’t just depend on the strength of the memory, it depends on the state that you’re in

Human-level control through deep reinforcement learning, ***Nature***, 518:529–533, 2015

模型中的记忆(internal memory)



Opening up the black box

Loosely modeled after the brain, deep neural networks are spurring innovation across science. But the mechanics of the models are mysterious: They are black boxes. Scientists are now developing tools to get inside the mind of the machine.

Inside the black box

A neural network, such as this one taught to perform image recognition, is made out of layers of triggers, or "neurons." The neurons fire when given data that cross certain thresholds, and pass that information to a new layer.

Path 1: Wrong

With its triggers set randomly at first, the network is wrong.

Path 2: Training

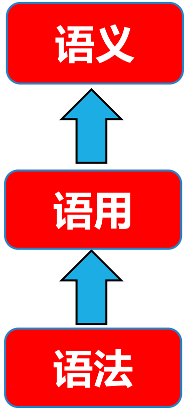
Shown many correct "volcanoes," the network adjusts its triggers.

Path 3: Right

After repeating many times, the network can correctly identify a volcano.

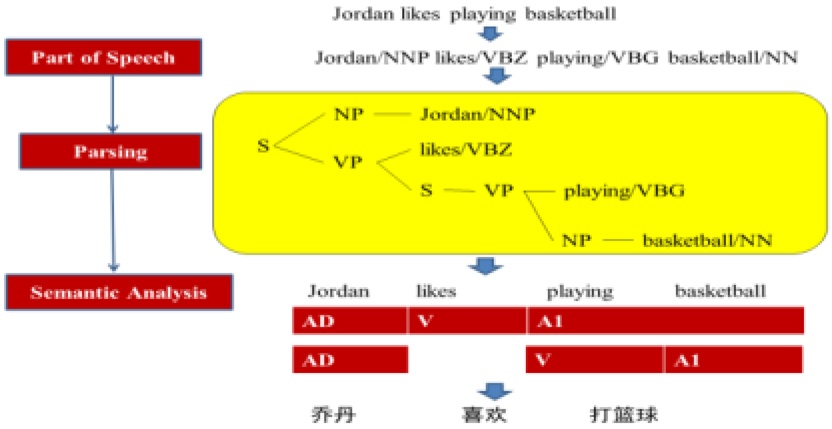
- 神经网络中的赫布理论（Hebbian theory）：突触可塑性原理，即突触前神经元向突触后神经元的持续重复的刺激可以导致突触传递效能的增加
- 神经元间权值信息既执行计算任务、又承载记忆功能。
- 记忆被新知识覆盖导致多任务难以实行、模型泛化能力不强、结果解释性弱

从分段学习到“端到端”学习：以自然语言翻译为例

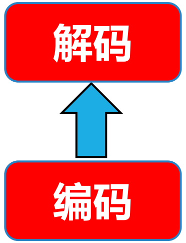

语义

语用

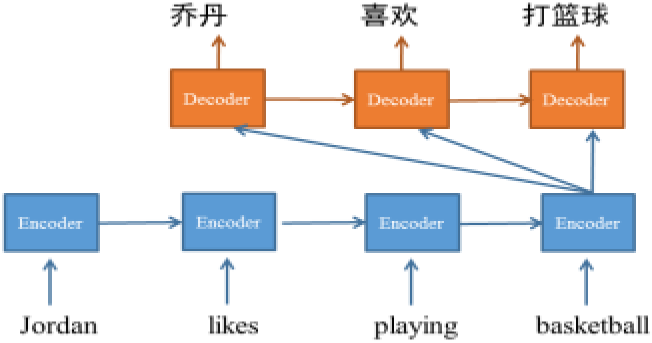
语法



- Conditional random field
- Hidden Markov Model

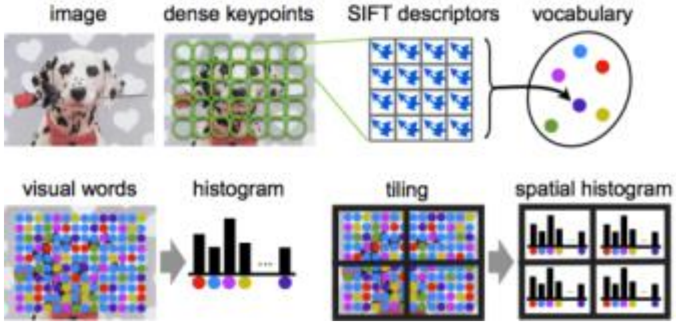

解码

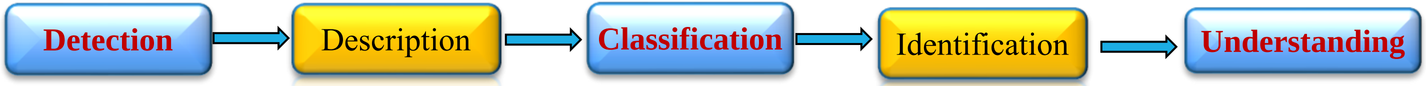
编码

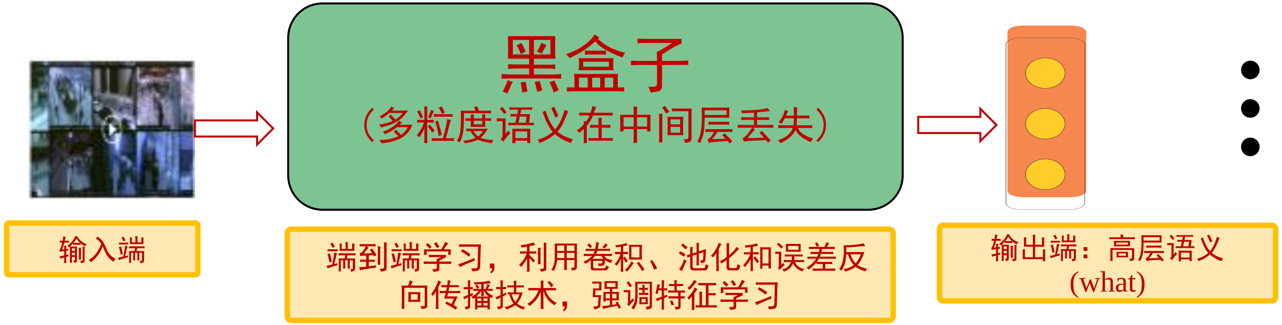


- word2vec
- Paragraph2vec
- Node2vec
- path2vec

从分段学习到“端到端”学习：以视觉分类和理解为例




Detection → Description → Classification → Identification → Understanding


黑盒子

（多粒度语义在中间层丢失）

输入端

端到端学习，利用卷积、池化和误差反向传播技术，强调特征学习

输出端：高层语义 (what)

- Two stream model
- Region-CNN
- Mask R-CNN

从分段学习到“端到端”学习

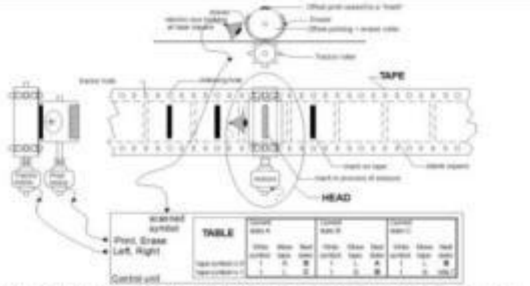
分段学习

每个阶段可灵活引入先验、经验与知识，但并不知所引入信息的合理性

端到端学习

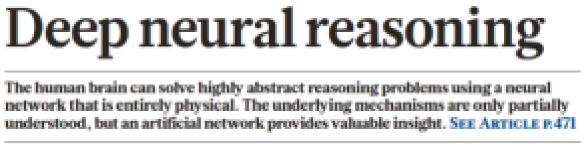
数据说话（你见或不见我，我就在那里，不悲不喜），但缺乏了人类语言可表述的“interpretability”

从图灵机到神经图灵机：利用外在记忆体中的知识



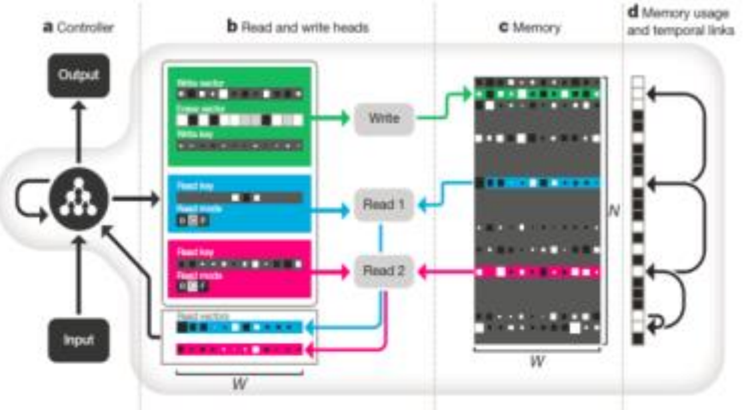
A fanciful mechanical Turing machine's TAPE and HEAD. The TABLE instructions might be on another "read only" tape, or perhaps on punch-cards. Usually a "finite state machine" is the model for the TABLE.

A.M.Turing, On Computable Numbers with an Application to the Entscheidungsproblem, *Proceedings of the London Mathematical Society*, Ser. 2, Vol. 42, 1937



Deep neural reasoning

The human brain can solve highly abstract reasoning problems using a neural network that is entirely physical. The underlying mechanisms are only partially understood, but an artificial network provides valuable insight. SEE ARTICLE P.471



Alex Graves, et al., Hybrid computing using a neural network with dynamic external memory, *Nature* 538, 471–476,2016

利用外在记忆体中的知识进行可计算推理

MEMORY NETWORKS

Jason Weston, Sumit Chopra & Antoine Bordes
Facebook AI Research
770 Broadway
New York, USA
{jase,spchopra,abordes}@fb.com

ABSTRACT

We describe a new class of learning models called *memory networks*. Memory networks reason with inference components combined with a long-term memory component; they learn how to use these jointly. The long-term memory can be read and written to, with the goal of using it for prediction. We investigate these models in the context of question answering (QA) where the long-term memory effectively acts as a (dynamic) knowledge base, and the output is a textual response. We evaluate them on a large-scale QA task, and a smaller, but more complex, toy task generated from a simulated world. In the latter, we show the reasoning power of such models by chaining multiple supporting sentences to answer questions that require understanding the intension of verbs.

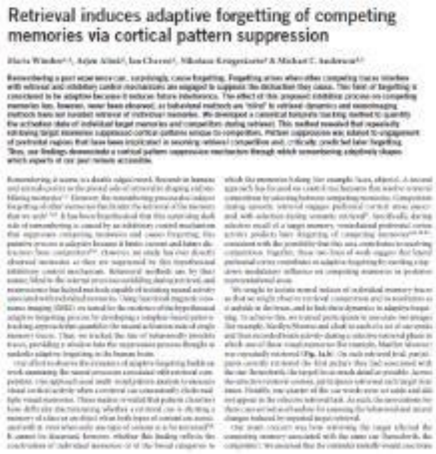
1 INTRODUCTION

Most machine learning models lack an easy way to read and write to part of a (potentially very large) long-term memory component, and to combine this seamlessly with inference. Hence, they do not take advantage of one of the great assets of a modern day computer. For example, consider the task of being told a set of facts or a story, and then having to answer questions on that subject. In principle this could be achieved by a language modeler such as a recurrent neural network (RNN) (Mikolov et al., 2010; Hochreiter & Schmidhuber, 1997) as these models are trained to predict the next (set of) word(s) to output after having read a stream of words. However, their memory (encoded by hidden states and weights) is typically too small, and is not compartmentalized enough to accurately remember facts from the past (knowledge is compressed into dense vectors). RNNs are known to have difficulty in performing memorization, for example the simple copying task of outputting the same input sequence they have just read (Zaremba & Sutskever, 2014). The situation is similar for other tasks, e.g., in the vision and audio domains a long term memory is required to watch a movie and answer questions about it.

In this work, we introduce a class of models called memory networks that attempt to rectify this problem. The central idea is to combine the successful learning strategies developed in the machine learning literature for inference with a memory component that can be read and written to. The model is then trained to learn how to operate effectively with the memory component. We introduce the general framework in Section 2, and present a specific implementation in the text domain for the task of question answering in Section 3. We discuss related work in Section 4, describe our experiments in 5, and finally conclude in Section 6.

A neural algorithm for a fundamental computing problem

[illegible]



Retrieval induces adaptive forgetting of competing memories via cortical pattern suppression

[illegible]

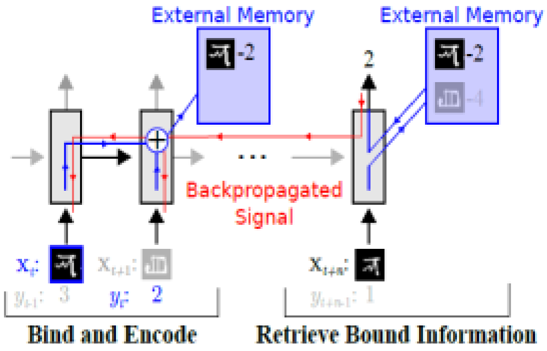
弦外之音、画外之意：
利用外在记忆体的深度神经推理

记忆的激活：
哈希索引的相似度搜索

记忆的可塑性维持、用尽废退：
强化学习中记忆的形成、巩固和遗忘

- Jason Weston(facebook), et al., Memory Networks, arXiv:1410.3916
- Retrieval induces adaptive forgetting of competing memories via cortical pattern suppression, Nature Neuroscience, 18, pages582–589 (2015)
- A neural algorithm for a fundamental computing problem, Science , 358, Issue 6364,793-796, 2017

外在记忆体中的知识的不同利用方式



one shot learning

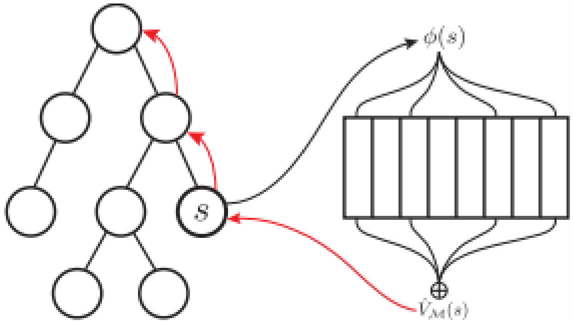


Figure 1: A brief illustration of M-MCTS. When a le state s is searched, the feature representation $\phi(s)$ generated, which is then used to *query* the memo., based value approximation $\hat{V}_{\mathcal{M}}(s)$. $\hat{V}_{\mathcal{M}}(s)$ is used to update s and all its ancestors according to equation (9), as indicated by the red arrows in the figure.

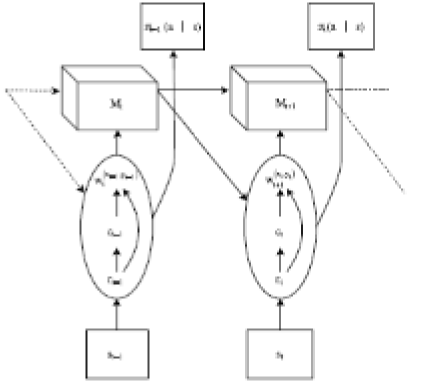


Figure 1: A visualization of two time steps of the neural map.

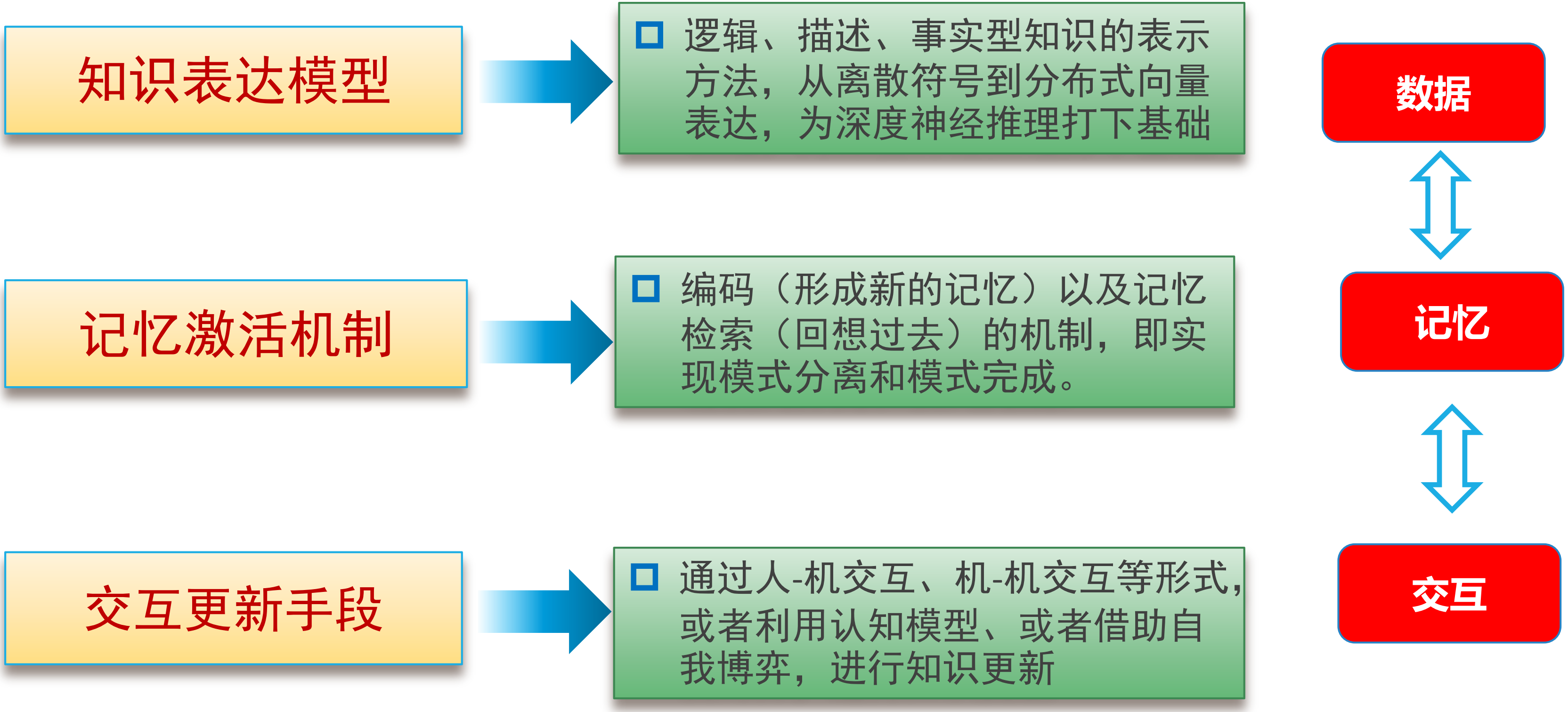
Structured memory

- Adam Santoro,et al., Meta-Learning with Memory-Augmented Neural Networks
- Chenjun Xiao, et al., Memory-Augmented Monte Carlo Tree Search
- Emilio Parisotto, et al., Neural Map:Structured Memory for deep reinforcement learning

有效利用当前数据、已有知识和未知交互

Shallow models	Deep models	备注
Language model	Neural language model	● 不只是单纯追求将浅层模型拓展到深层模型。 ● 更为重要的是，在这个转变过程中，巧妙融合数据、知识和交互经验，多种手段和方法的综合利用。
Bayesian Learning	Bayesian deep learning	
Turing Machine	Neural Turing Machine	
Reinforcement Learning	Deep Reinforcement Learning	
Generative Model	Deep Generative Model	
X	Deep or Neural + X	

有效利用当前数据、已有知识和未知交互的挑战



知识表达模型

□ 逻辑、描述、事实型知识的表示方法，从离散符号到分布式向量表达，为深度神经推理打下基础

记忆激活机制

□ 编码（形成新的记忆）以及记忆检索（回想过去）的机制，即实现模式分离和模式完成。

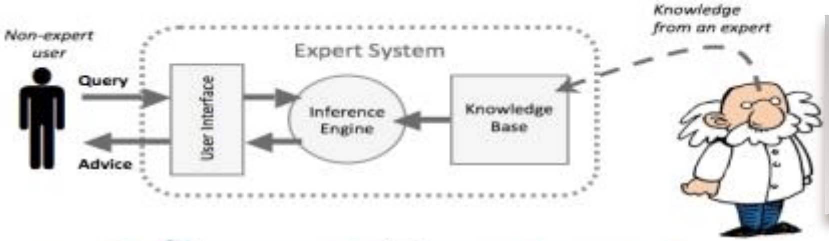
交互更新手段

□ 通过人-机交互、机-机交互等形式，或者利用认知模型、或者借助自我博弈，进行知识更新

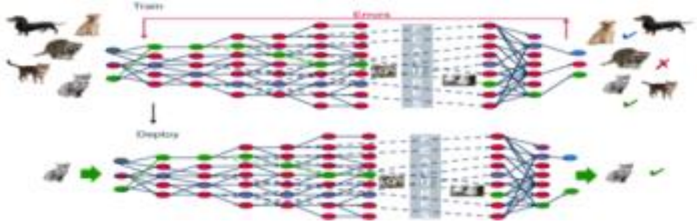
对推理过程逐渐松绑，使推理逐步走向对思维广泛模拟：**跨媒体综合推理**

跨媒体综合推理

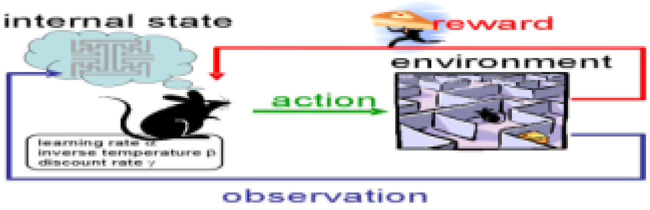
知识推导、数据驱动和交互反馈等学习方法缺乏像人类一样能接受多通道输入、综合利用多种推理模式进行协同推理的计算框架。为此，跨媒体综合推理思想被提出。



模拟高级思维活动，侧重知识表示，但难以拓展

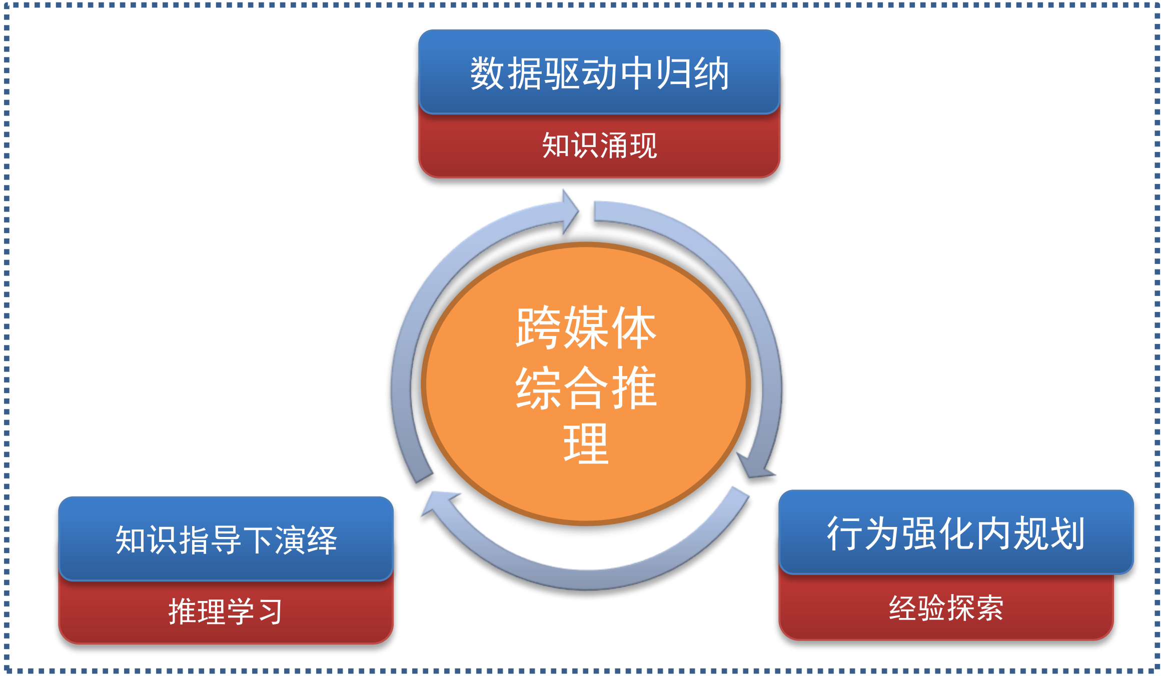


从大量实例中学习，侧重感知能力，但过程难以理解与干预



模拟低级的认知过程，侧重自我适应，极大的依赖于策略

跨媒体综合推理

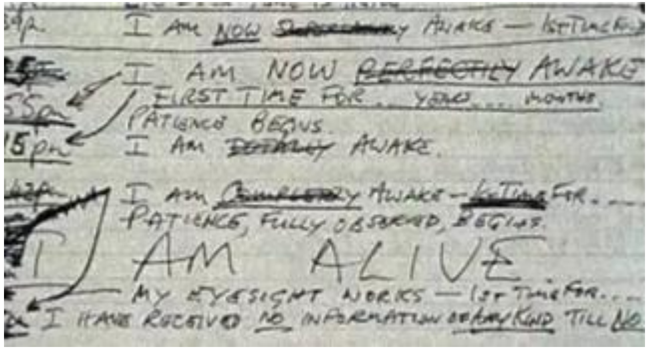


有机协调“知识指导下的演绎”、“数据驱动中的归纳”和“行为强化内的规划”等理论模型和方法手段，建立知识、数据和反馈于一体的人工智能理论和模型

仅有短期记忆的人生

英国指挥家Clive Wearing：

- 因 herpes simplex virus（单纯疱疹病毒）侵蚀大脑hippocampus（海马体）而患上anterograde amnesia （顺行性遗忘症）
- 海马体是将短期记忆传递成长期记忆的重要器官



8:?p I AM NOW ~~SUPERFICIALL~~Y AWAKE — 1st TIME

I AM NOW ~~PERFECTLY~~ AWAKE

FIRST TIME FOR .. YEARS ... MONTHS

55pm PATIENCE BEGINS.

15pm I AM ~~DEFINITELY~~ AWAKE.

I AM ~~COMPLETELY~~ AWAKE — ~~1st TIME~~ FOR. . . .

PATIENCE, FULLY OBSERVED, BEGINS.

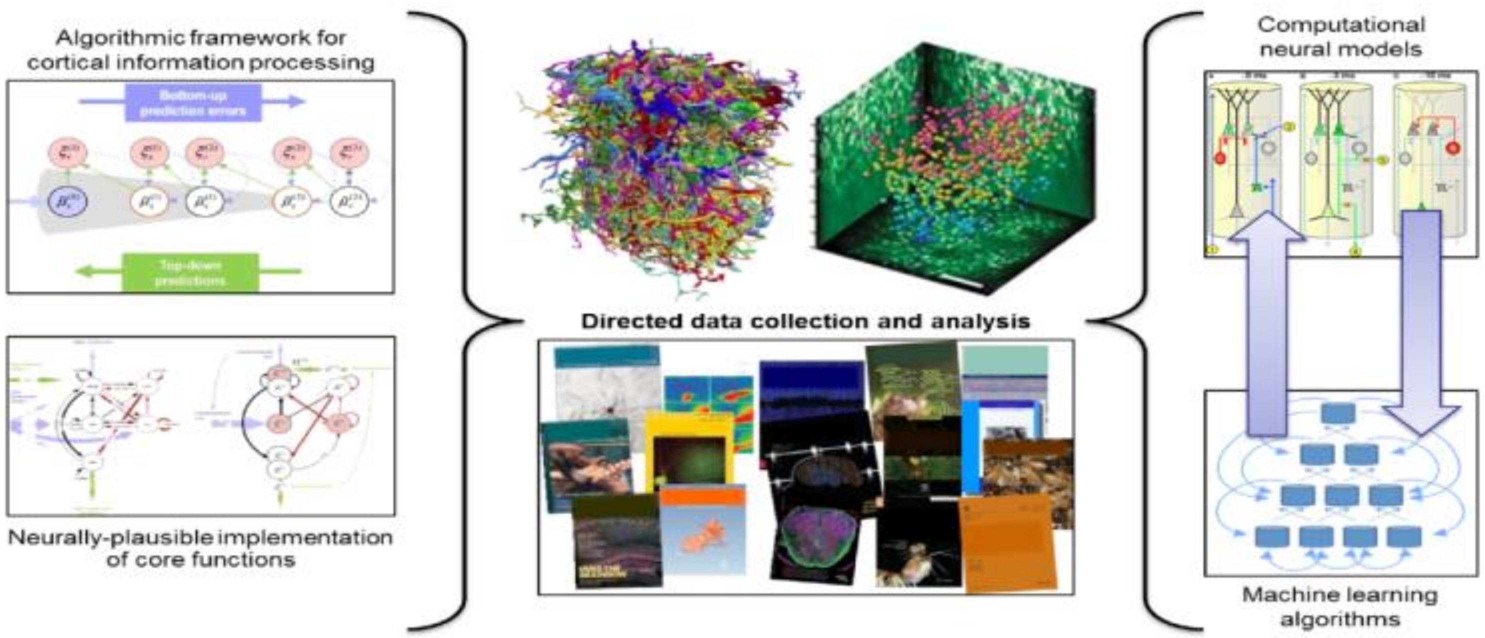
I AM ALIVE

MY EYESIGHT WORKS — 1st TIME FOR. . . .

I HAVE RECEIVED NO INFORMATION OF ANY KIND TILL NO

他的记忆，和金鱼一样，只要“七秒”就会消失。他的生命是一段又一段的空白，没有过去，没有未来。

要证明现在致力于构建的通用智能是可行的，人类大脑是现有唯一证据，因此把神经科学作为新算法的灵感来源是有意义的。


Algorithmic framework for cortical information processing
Bottom-up prediction errors
Top-down predictions
Neurally-plausible implementation of core functions
Directed data collection and analysis
Computational neural models
Machine learning algorithms

Neuroscience-Inspired Artificial Intelligence, Neuron, Volume 95, Issue 2, p245–258, 19 July 2017

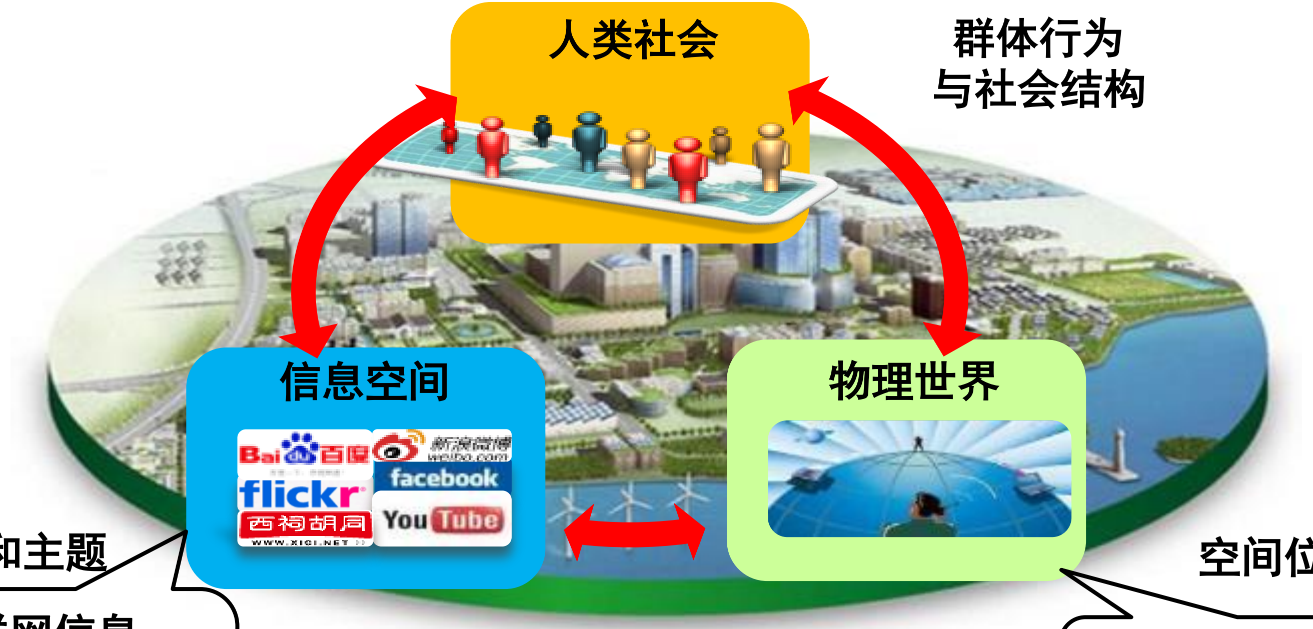
提纲

1、 记忆驱动的智能计算

2、 可计算社会学

3、若干挑战

人类进入信息空间(Cyberspace)、物理世界 (Physical)和人类社会(Society)三元空间融合时代(CPS)



人类社会

群体行为
与社会结构

信息空间

物理世界

事件和主题

空间位置信息

互联网信息

百度每天搜索请求
100亿；处理数据量
100PB

空间位置信息

每年增加1亿GPS设备

三元空间大数据相互映照、彼此补充：计算社会的数据基础

SOCIAL MEDIA

Self-reflection, online

Some scientists might not like the persona they see when they look online. But they can do something about it.

BY VIRGINIA GEWIN

Michael Tobis has spent years tackling the controversies surrounding climate-related science on his blog 'Only In It For The Gold'. An Earth systems modeller at the University of Texas Institute for Geophysics in Austin, Tobis admits that he may now be more prominent as a blogger than as a published scientist. Although that might not appeal to all researchers, scientists have good reason to take online media seriously. "Scientists have an online persona — whether they like it or not," says Tobis.

Online media offer researchers unique ways to express their interests and goals, foster collaborations and garner invitations and opportunities. But even scientists who don't blog or tweet have an online presence that evolves apace — with or without their intervention. Auto-generated profiles from citation databases, Wikipedia entries, even photos from college can, by virtue of a simple

Scientists can also now engage with the public in new and innovative ways, as demonstrated by a researcher who was contacted about his ancestry after publishing his genome on the Internet (see *Nature* 468, 880–881; 2010). "Even if you never pay attention to the online world and don't want anything to do with it, it's bleeding into your real life," says Liz Neeley, the Seattle-based assistant director of science outreach at the Communication Partnership for Science and the Sea, an organization that helps scientists to engage with the public.

It is important, and fairly easy, for ▸

31 MARCH 2011 | VOL 471 | NATURE | 667

North Carolina recently by attempting to demystify it — and to therefore dispel the notion that it was "a mysterious alien creation born to suck out our brains". The humorous style was intentional. "The public has a very narrow view of how scientists act, look and behave, and I wanted a blog that helped dispel the staid stereotype," says McClain. It worked. His blog gets, on average, 1,000–2,000 hits a day, a lot for an independent blog site. "If people are entertained, they come back for more," he says.

BLOGGERS BEWARE

Scientists actively cultivating a web persona should, however, tread carefully — it is difficult to remove something once it has been unleashed to the blogosphere. "The Internet is forever. People should keep in mind that when you blog or tweet something, it becomes Google-able," says Goldman, who has had guest blogs on *The Guardian* and *Scientific American*'s 'Mind Matters' blog.

And not all blogs have the same aims. Dennis Meredith, author of *Explaining Research: How to Reach Key Audiences to*

those instances, Nielsen Hayden opts to 'disemvowel', or remove the vowels from, rude posts. Bloggers have every right to remove unhelpful comments, she says. They should also read the comments left on their blog, says Nielsen Hayden. "You can't throw a party and not attend it." But that takes time — and that's where Twitter's swift tempo can trump a blog.

For some scientists, Twitter serves as a filter whereby trusted tweeters post links to studies, articles or conference presents. Those that might pique a scientists' interest or even help to advance research aims. "It's the meta-information that Twitter carries that is so important," says Paul Filmer, programme director for the Inter-American Institute for Global Change Research, based in Arlington, Virginia, part of the National Science Foundation (NSF), who admonishes the NSF and Voyager2 Twitter feeds. Other scientists may fail to see the point of messages that cannot be longer than Twitter's limit of 140 characters.

"It's easy to mock the idea of Twitter," says Tobis. "With its silly name and teenagers reporting on what they had for lunch. But

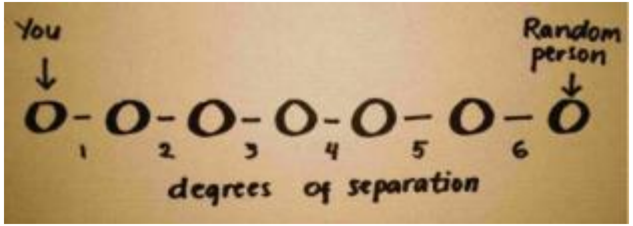
Virginia Gewin is a freelance journalist based in Portland, Oregon.

….The 'virtual' world and the 'real' world complement each ….

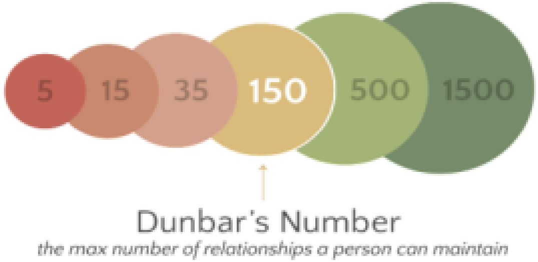
互联网数据是对现实世界的 "自我映照"，与现实世界相互补充。

Viginia Gewin, Self-reflection, *Nature* , 471,667-669,2011, March

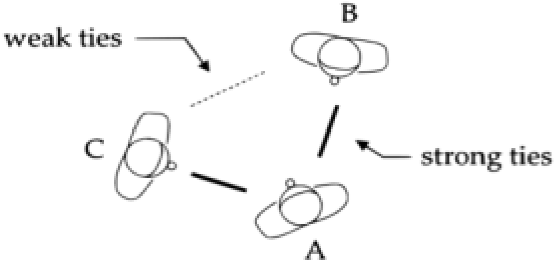
社会学研究：观察与实验



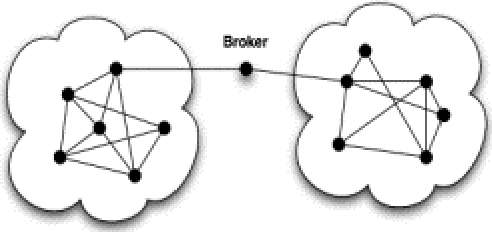
六度分割理论
Stanley Milgram. The small world problem. Psychology Today, 2(1):60–67, 1967. Cited by 6967 (as of May 2016)



邓巴数字即150定律：人类智力所允许稳定网络社交人数150人，Facebook社区用户平均好友人数是120人
Robin I. M. Dunbar. Neocortex size as a constraint on group size in primates. *Journal of Human Evolution* 22 (6): 469–493. Cited by 1393 (as of May 2016)



弱链接优势
Granovetter, Mark S. The strength of weak ties. *American journal of sociology* 78.6 (1973): 1360-1380. Cited by 43186 (as of April 2017)



结构洞：位置比关系更重要
Ronald S. Burt. *Structural Holes: The Social Structure of Competition*. Cambridge: Harvard University Press. 1995. Cited by 20207 (as of May 2016)

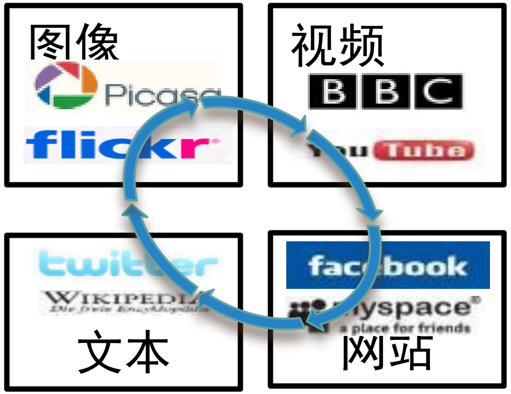
社会学研究：观察与实验

时间 （年）	人　　物	事　　件
1793	E üler	七桥问题
1959	Erdǒs和R ényi	随机图理论
1967	Milgram	小世界实验（六度空间）
1973	Granovetter	弱连接的强度
1998	Watts和Strogatz	小世界模型
1999	Barab ási和Albert	无尺度网络

迄今为止最大规模小世界验证：2011年，Facebook与意大利米兰大学公布了关于Facebook中用户之间六度分离验证的结果。2007年以来，Facebook上两个用户之间的平均距离仅为4.74，任何两个用户之间间隔不超过5度的概率99.6%。注意2011年5月数据包括了Facebook上大约7.21亿活跃用户以及69亿朋友关系链接。

Lars Backstrom, et. al., *Four Degrees of Separation*, http://arxiv.org/abs/1111.4570

三元空间大数据相互映照、彼此补充：计算社会的数据基础



图像

视频

文本

网站

COMPUTER SCIENCE

Mining Our Reality

Tom M. Mitchell

Something important is changing in how we as a society use computers to mine data. In the past decade, machine-learning algorithms have helped to analyze historical data, often revealing trends and patterns too subtle for humans to detect. Examples include mining credit card data to discover activity patterns that suggest fraud, and mining scientific data to discover new empirical laws (1, 2). Researchers are beginning to apply these algorithms to real-time data that record personal activities, conversations, and movements (3–8) in an attempt to improve

Machine Learning Department, Carnegie Mellon University, Pittsburgh, PA 15213, USA. E-mail: tom.mitchell@cs.cmu.edu

Meanwhile, new algorithms aim to address privacy concerns arising from data sharing and aggregation (9, 10).

To appreciate both the po… vacy implications of real-ti… consider the data available ju… company, based on your p… those of millions of other ind… going about their daily lives … phone—a device that contain… tioning System (GPS) sensor… within a few meters, an acc… detects when you are walking … a microphone that detects both … and background noises, a camera that records

e-mail and text message. The potential benefits of mining such data are various; examples include reducing traffic congestion and pollu-

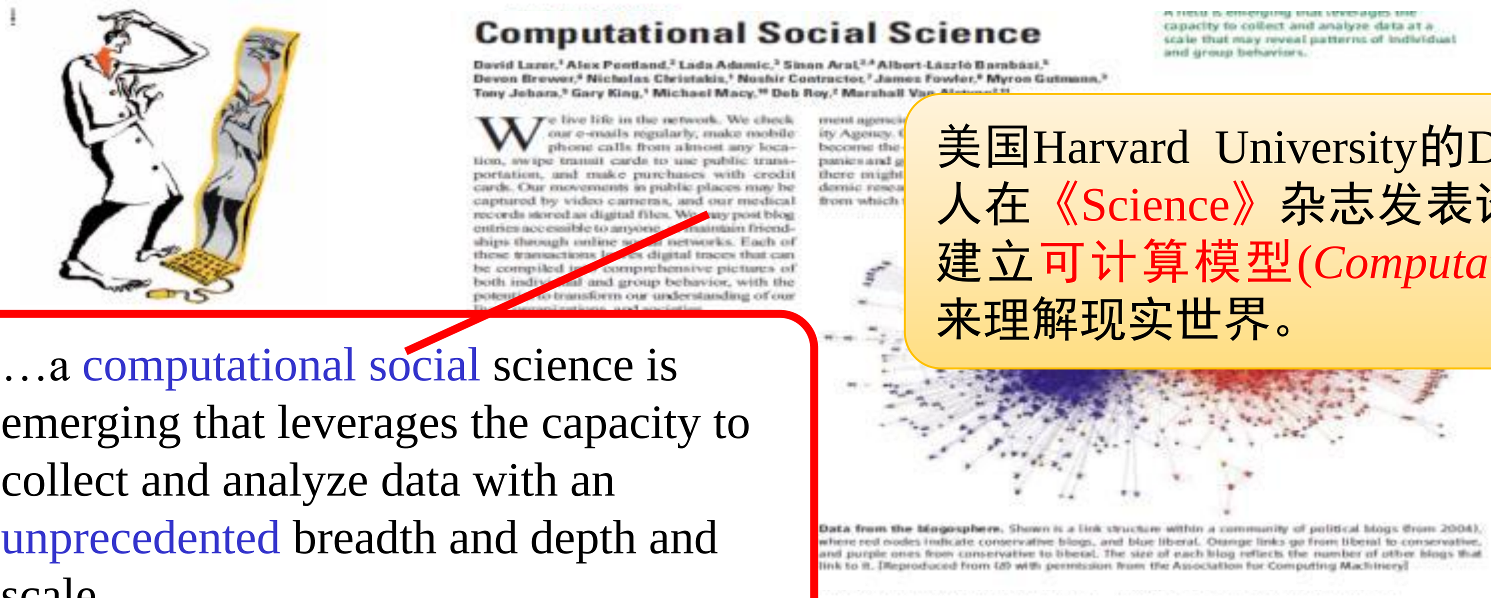
through services such as Google Maps (11).

美国CMU的Tom Mitchell教授在《Science》发表名为“*Mining our Reality*”的文章指出：….从相互关联的网络海量数据中寻找“蛛丝马迹”，更有利于理解现实世界…

….use computers to mine data…. machine learning algorithms have helped to analyze historical data, often revealing trends and patterns too subtle for humans to detect.

Tom M., Mitchell, Mining our Reality, *Science*, 326,2009,1644-1645

可计算社会学：应用计算模型和方法来理解和预测现实世界



A field is emerging that leverages the capacity to collect and analyze data at a scale that may reveal patterns of individual and group behaviors.

Computational Social Science

David Lazer,[1] Alex Pentland,[2] Lada Adamic,[3] Sinan Aral,[2,4] Albert-László Barabási,[5] Devon Brewer,[6] Nicholas Christakis,[1] Noshir Contractor,[7] James Fowler,[8] Myron Gutmann,[3] Tony Jebara,[9] Gary King,[1] Michael Macy,[10] Deb Roy,[2] Marshall Van Alstyne[11]

We live life in the network. We check our e-mails regularly, make mobile phone calls from almost any location, swipe transit cards to use public transportation, and make purchases with credit cards. Our movements in public places may be captured by video cameras, and our medical records stored as digital files. We may post blog entries accessible to anyone, or maintain friendships through online social networks. Each of these transactions leaves digital traces that can be compiled into comprehensive pictures of both individual and group behavior, with the potential to transform our understanding of our lives, organizations, and societies.

ment agenci
ity Agency. C
become the
panies and g
there might
demic resea
from which t

Data from the blogosphere. Shown is a link structure within a community of political blogs (from 2004), where red nodes indicate conservative blogs, and blue liberal. Orange links go from liberal to conservative, and purple ones from conservative to liberal. The size of each blog reflects the number of other blogs that link to it. [Reproduced from (8) with permission from the Association for Computing Machinery]

美国Harvard University的David Lazer等人在《Science》杂志发表论文也认为需建立可计算模型(*Computational Model*)来理解现实世界。

…a computational social science is emerging that leverages the capacity to collect and analyze data with an unprecedented breadth and depth and scale

David Lazer, *et. al.*, Computational Social Science, *Science*, 323,721-723, 2009

计算思维成为可计算社会学研究的思维方式

在人工智能、大数据、互联网等支撑下所形成的计算思维与实验思维和理论思维同等重要，人工智能正成为一种通识教育，渗透进入其他知识技术教育之中。

Viewpoint | Jeannette M. Wing

Computational Thinking

It represents a universally applicable attitude and skill set everyone, not just computer scientists, would be eager to learn and use.

Wing, Jeanette M., Computational thinking, *Communications of the ACM*,49 (3): 33,2006

计算思维是运用计算机科学的基础概念进行问题求解、系统设计、以及人类行为理解等涵盖计算机科学之广度的一系列思维活动

思维方式	机制	手段
实验思维	实验⟶观察⟶发现、推断与总结	**观察与归纳**
理论思维	假设、预设⟶定义/性质/定理⟶证明	**推理与演绎**
计算思维	**设计，构造与计算**	**设计与构造**

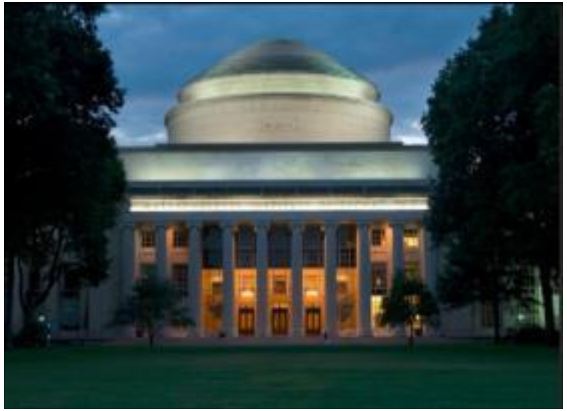
MIT通过学科交叉重塑人才培养模式：计算机和人工智能为引擎

Massachusetts Institute of Technology

MIT reshapes itself to shape the future

MIT于2018年10月16日宣布，投资十亿美元建设新的计算学院，致力于将计算机技术以及人工智能纳入所有研究领域，同时寻求改变计算和人工智能相关的公共政策和道德研究教育。

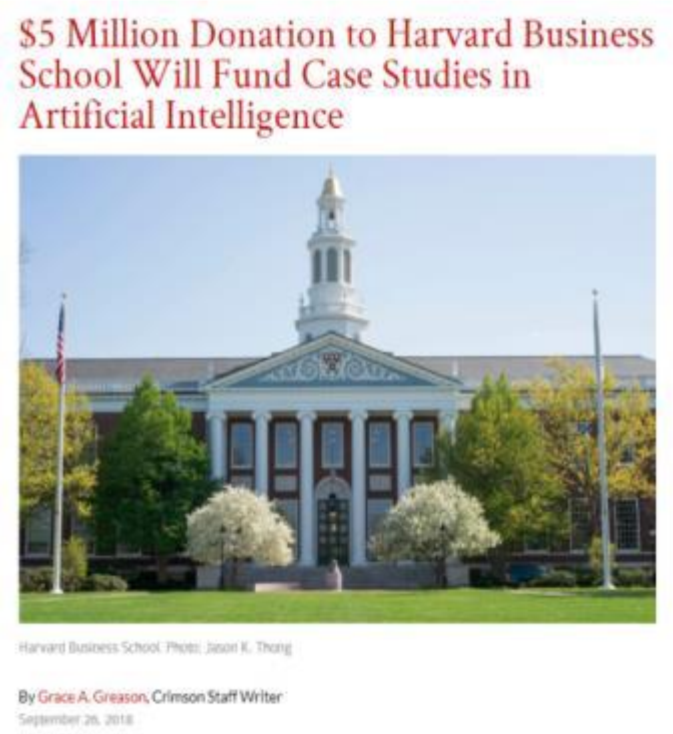
make AI part of every graduate’s education
AI-focused College



- 2019年秋季学期正式启动新学院
- 将计算机和人工智能力量带到 MIT所有学习领域
- 创造 50 个跨越新学院和其它学系的新教师职位
- 为 MIT五个学院建设公共合作结构，用于计算机和人工智能领域的教育、研究和创新
- 改变关于计算机和人工智能公共政策和道德方面教育和研究

哈佛大学商学院在案例分析中重视人工智能

HARVARD | BUSINESS | SCHOOL

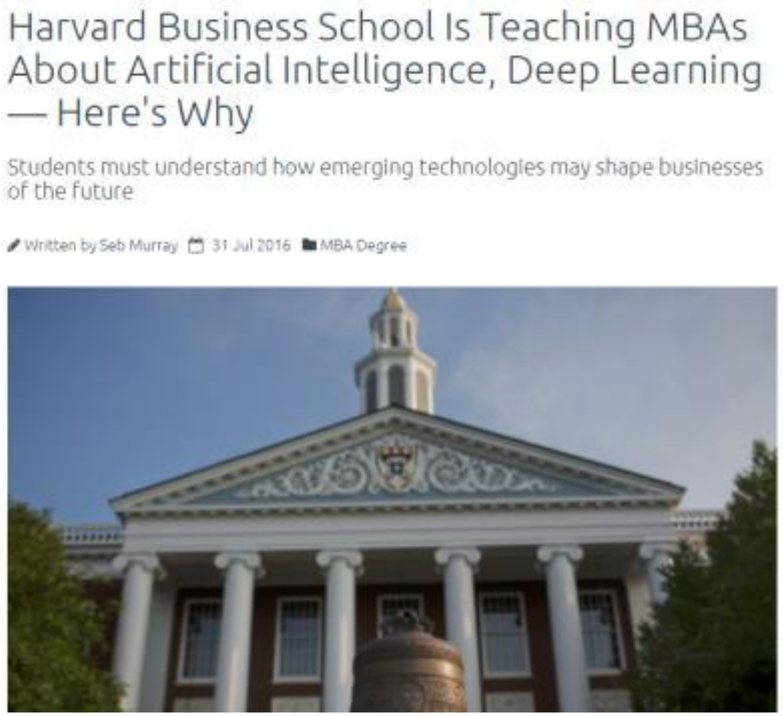


$5 Million Donation to Harvard Business School Will Fund Case Studies in Artificial Intelligence

Harvard Business School Photo: Jason K. Thong

By Grace A. Greason, Crimson Staff Writer

September 26, 2018



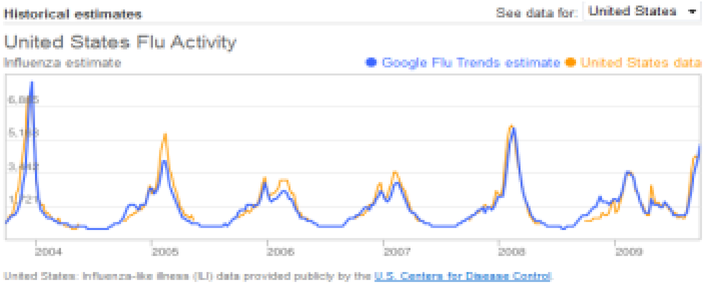
Harvard Business School Is Teaching MBAs About Artificial Intelligence, Deep Learning — Here's Why

Students must understand how emerging technologies may shape businesses of the future

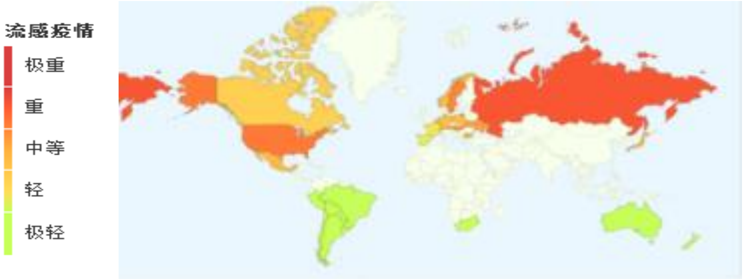
Written by Seb Murray 31 Jul 2016 MBA Degree

谷歌预测流感方法中的人工智能

□ Google推出 “谷歌流感趋势”项目，通过分析搜索单词来预测流感的爆发。与美国疾病控制和预防中心提供的报告对比，其对追踪疾病的精确率达到97%-98%



United States: Influenza-like illness (ILI) data provided publicly by the U.S. Centers for Disease Control



Google预测的流感爆发与实际情况对比

Google预测的流感分布图

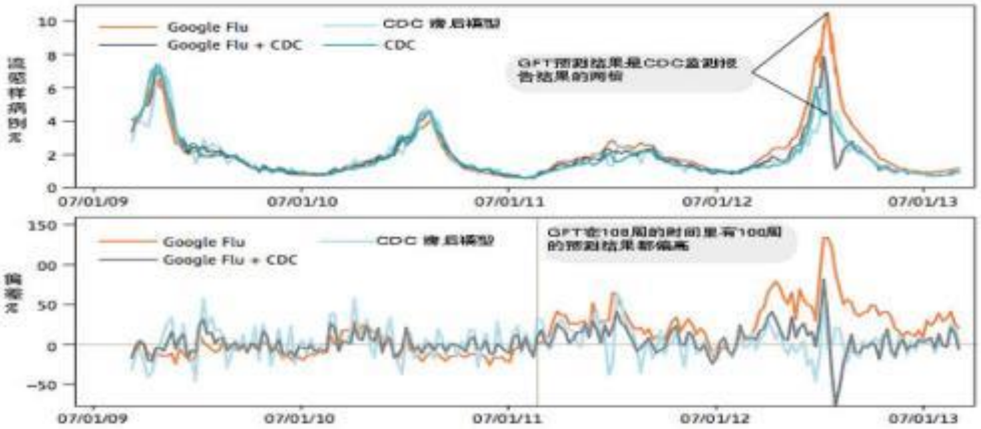
Detecting influenza epidemics using search engine query data, *Nature* 457, 1012-1014 (19 February 2009)

谷歌预测流感方法中的人工智能

- □ 谷歌流感预测（Google Flu Trends，GFT）所预测流感样病例门诊数超过了与美国疾病预防控制中心（Centers for Disease Control and Prevention，CDC）根据全美各实验室监测报告得出的预测结果的两倍。

- □ “大数据傲慢”(Big Data Hubris): 认为大数据可以完全取代传统的数据收集方法，而非作为后者的补充

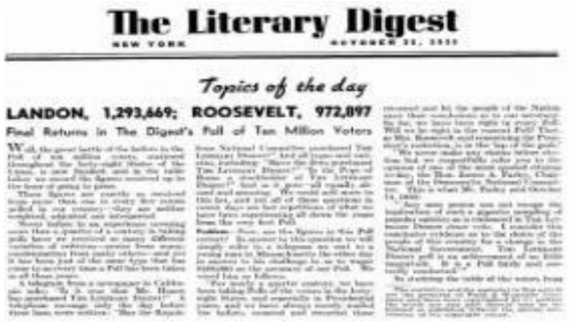
GFT的平均绝对偏差为0.486，CDC滞后模型的平均绝对偏差为0.311，GFT与CDC相结合的平均绝对偏差为0.232



The Parable of Google Flu: Traps in Big Data Analysis, *Science* 14 Mar 2014

小数据 Versus 大数据

- □ 1936年《文学文摘》（The Literary Digest）预测总统选举：民主党人艾尔弗雷德 兰登（Alfred Landon）与时任总统富兰克林 罗斯福（Franklin Roosevelt）
- □ 《文学文摘》给自己的订户邮寄了1000万份调查问卷，根据回收到的240万份回执，预测艾尔弗雷德 兰登将会以55比41的优势击败富兰克林 罗斯福赢得大选。
- □ 美国数学家、抽样调查方法创始人、民意调查的组织者乔治 盖洛普（George Gallup）仅通过3000人问卷调查，得出了准确得多的预测结果。
- □ 良好的“小数据”可战胜“含噪和偏差”的大数据，小数据可做到“以小见大”、“一叶知秋”。数据质量对于机器学习预测结果影响甚大。



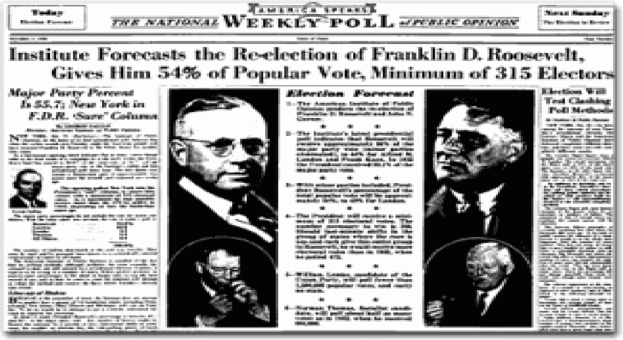
The Literary Digest

NEW YORK

Topics of the day

LANDON, 1,293,669; ROOSEVELT, 972,897

Final Returns in The Digest's Poll of Ten Million Voters



Today

AMERICA SPEAKS

THE NATIONAL WEEKLY POLL OF PUBLIC OPINION

Next Sunday

Institute Forecasts the Re-election of Franklin D. Roosevelt, Gives Him 54% of Popular Vote, Minimum of 315 Electors

Major Party Percent Is 55.7; New York in F.D.R. 'Sure' Column

Election Forecast

Election Will Test Clashing Poll Methods

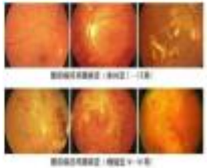
人工智能与学科交叉中存在的“数据之学科鸿沟”

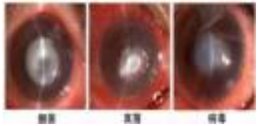
医学数据

法学数据

心理学数据

人文社科数据





算法

模型

算力

数据化知识 ≠ 数据驱动

浙江大学学者用7亿条通话记录回答：移民融入成功了吗

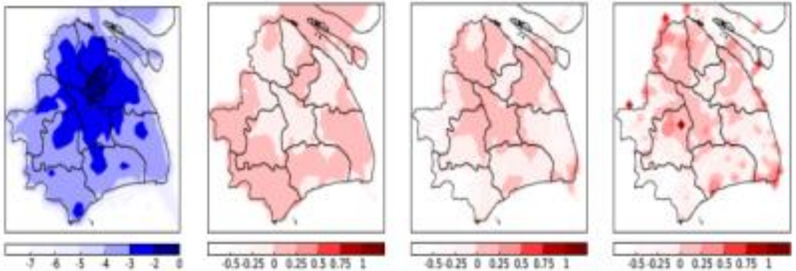
通过对5400万用户的6.98亿条通话记录的学习，人工智能建立起了一项看似天马行空的关联：通话记录与城市移民状态

- 两周通话记录：是走还是留？群越多越杂，越能融入、参谋城市规划

计算社会学：通话记录“洞悉”城市移民状态

求是印象

科学封面（第21期）——计算社会学：通话记录“洞悉”城市移民状态



(a) Log overall average probability.

(b) Log odds ratio for locals.

(c) Log odds ratio for settled migrants.

(d) Log odds ratio for new migrants.

上海市居民地域分布

Yang Yang（杨洋）, Chenhao Tan, Zongtao Liu, Fei Wu, and Yueting Zhuang. Urban Dreams of Migrants: A Case Study of Migrant Integration in Shanghai. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence(AAAI)*, 2018

提纲

1、 记忆驱动的智能计算

2、 可计算社会学

3、若干挑战

第三波人工智能：“third wave” AI

“first wave” (rule-based)

符号主义人工智能为手段

以语言和可描述信息处理为主

“second wave”
(statistical-learning-based)

以模型假设的机器学习为手段

在数据建模基础上、从数据中学习模式

“third wave” AI
theory and applications

以自适应和推理为核心目标

human-like communication and reasoning capabilities

第三波人工智能：“third wave” AI

“first wave” (rule-based)

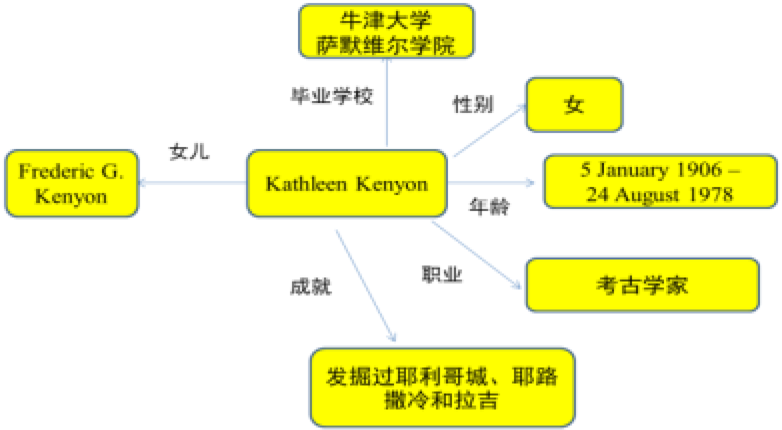
符号主义人工智能为手段
以语言和可描述信息处理为主

IBM“沃森”的推理

主持人问：Kathleen Kenyon's excavation of this city mentioned in Joshua showed the walls had been repaired 17 times.（Kathleen Kenyon对这个在《《圣经•约书亚记》》中提到的城市的发掘表明，该城的城墙曾被修复17次）
沃森回答：What is Jericho （耶利哥城是什么？）

答案排序：**耶利哥（Jericho）97%**、耶路撒冷（Jerusalem）42%、拉吉（Lachish）7%

- Kathleen
- Kenyon
- Kathleen Kenyon
- Joshua



第三波人工智能：“third wave” AI

“second wave”
(statistical-learning-based)

以模型假设的机器学习为手段

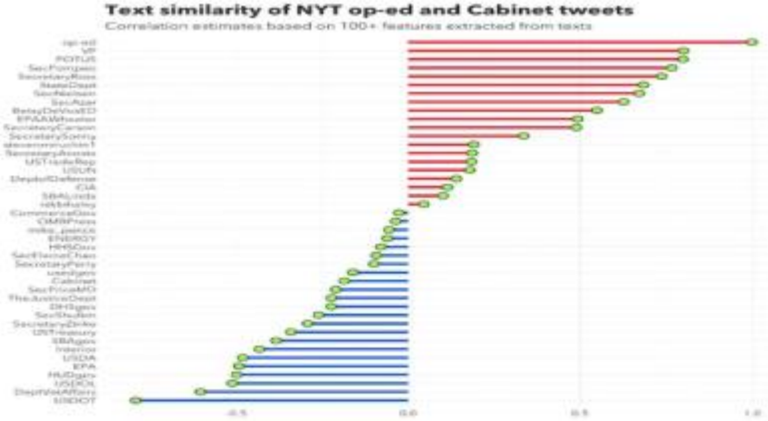
在数据建模基础上、从数据中学习模式

白宫版“潜伏”，数据分析锁定副总统

Opinion

I Am Part of the Resistance Inside the Trump Administration

I work for the president but like-minded colleagues and I have vowed to thwart parts of his agenda and his worst inclinations.



Text similarity of NYT op-ed and Cabinet tweets

Correlation estimates based on 100+ features extracted from texts

第三波人工智能：“third wave” AI

可观测性问题	What if we see A (what is?)	$P(y\|A)$
决策行动问题	What if we do A (what if?)	$P(y\|do(A))$ （如果采取A行为，则B真）
反事实问题 (Counterfactual)	What if we did things differently	(why?) $P(y'\|A)$ （如果A为真，则B将不同）
Options: with what probability		

关联(association)：
直接可从数据中计算得到的统计相关

介入(intervention):
无法直接从观测数据就能得到关系，如“某个商品涨价会产生什么结果”

反事实(counterfactual)：
某个事情已经发生了，则在相同环境中，这个事情不发生会带来怎样的新结果

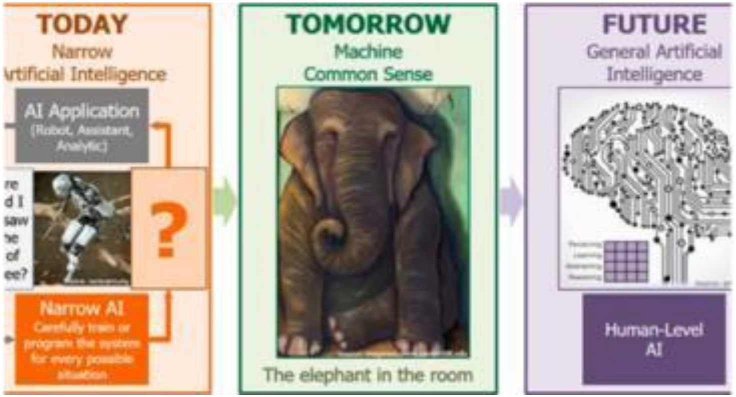
挑战 (1)：人类常识知识支持下顿悟与直觉

2 billion dollars over the next five years

Machine Common Sense (MCS)

Department of Defense
Defense Advanced Research Projects Agency
Information Innovation Office
Oct 19, 2018

DARPA is soliciting innovative research proposals in the area of machine common sense to enable Artificial Intelligence (AI) applications to understand new situations, monitor the reasonableness of their actions, communicate more effectively with people, and transfer learning to new domains. Proposed research should investigate innovative approaches that enable revolutionary advances in science, devices, or systems. Specifically excluded is research that primarily results in evolutionary improvements to the existing state of practice.



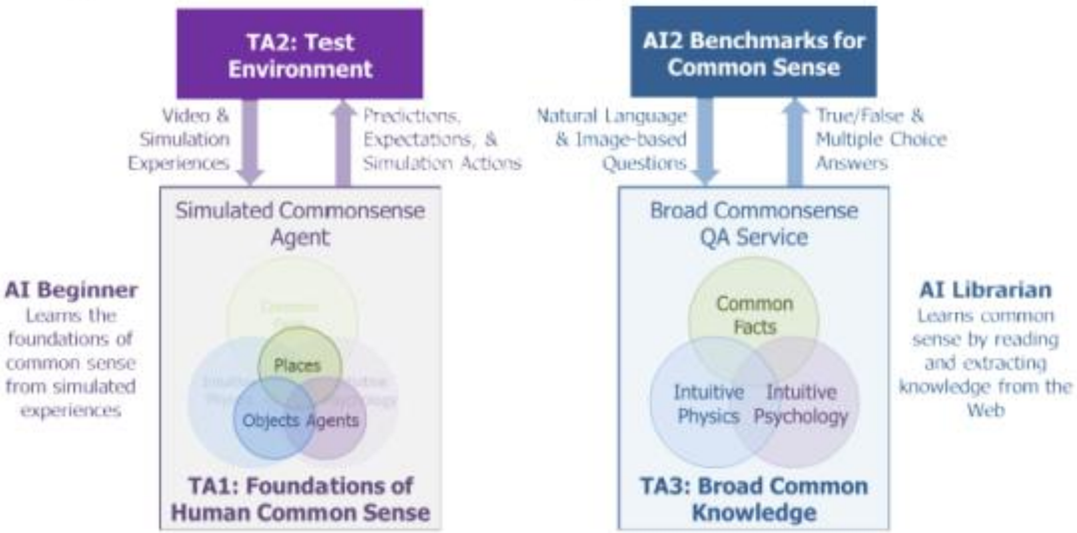
见一叶落，而知岁之将暮；审堂下之阴，而知日月之行，阴阳之变；见瓶水之冰，而知天下之寒，鱼鳖之藏也

《淮南子说山训》

挑战(1)：人类常识知识支持下顿悟与直觉

Machine Common Sense (MCS)

- **Foundations of Human Common Sense**: a service that learns from experience, like a child, to construct computational models that mimic the core domains of cognition for objects (intuitive physics), places (spatial navigation), and agents (intentional actors). These models will be evaluated against the cognitive development milestones as evidenced in developmental psychology experiments with human children from 0-18 months old.
- **Broad Common Knowledge:** a service that learns from reading the Web, like a research librarian, to construct a commonsense knowledge repository capable of answering natural language and image-based questions about commonsense phenomena. This service will mimic the general knowledge of an average American adult in 2018, and be evaluated against the Allen Institute for Artificial Intelligence (AI2) Common Sense Benchmarks.


TA2: Test Environment
Video & Simulation Experiences
Predictions, Expectations, & Simulation Actions
Simulated Commonsense Agent
Places
Objects
Agents
TA1: Foundations of Human Common Sense
AI Beginner
Learns the foundations of common sense from simulated experiences
AI2 Benchmarks for Common Sense
Natural Language & Image-based Questions
True/False & Multiple Choice Answers
Broad Commonsense QA Service
Common Facts
Intuitive Physics
Intuitive Psychology
TA3: Broad Common Knowledge
AI Librarian
Learns common sense by reading and extracting knowledge from the Web

常识是人工智能的黑洞

Common sense is the dark matter of artificial intelligence

挑战(1)：人类常识知识支持下顿悟与直觉

常识知识的获取

自然语言、视觉和听觉等

常识知识的记忆

存储与激活等

常识知识的利用

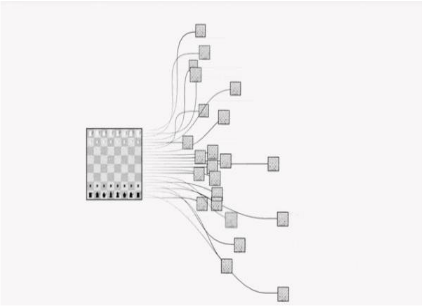
顿悟与直觉推理

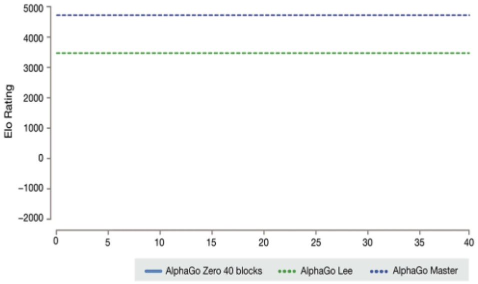
- 威廉和他的三岁宝贝儿子谁身高更高
- 针扎入胡萝卜，它会留下小孔，请问这里的“它”指的是什么

挑战 (2)：元学习(meta-learning)理论：“利用与探索”的 多巴胺神经计算机制

多巴胺（dopamine, 神经传导物质，快乐信号，2000年诺贝尔医学奖）在人脑学习过程中“周旋”于已得回报和预期回报

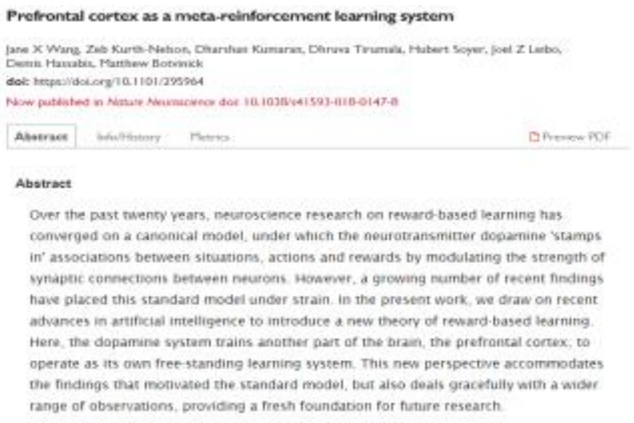
starting *tabula rasa*
(一张白纸绘蓝图)





AlphaGo Zero: 深度学习、蒙特卡洛树搜索、强化学习，40天训练、2900万次自我对弈，89:11击败AlphaGo Master

多巴胺奖励信号不仅仅是对权重进行调整，它还负责编码、传递抽象任务与规则结构的重要信息，从而使快速任务适应成为可能。



Prefrontal cortex as a meta-reinforcement learning system

Jane X Wang, Zeb Kurth-Nelson, Dharshan Kumaran, Dhruva Tirumala, Hubert Soyer, Joel Z Leibo, Demis Hassabis, Matthew Botvinick

doi: https://doi.org/10.1101/295964

Now published in *Nature Neuroscience* doi: 10.1038/s41593-018-0147-8

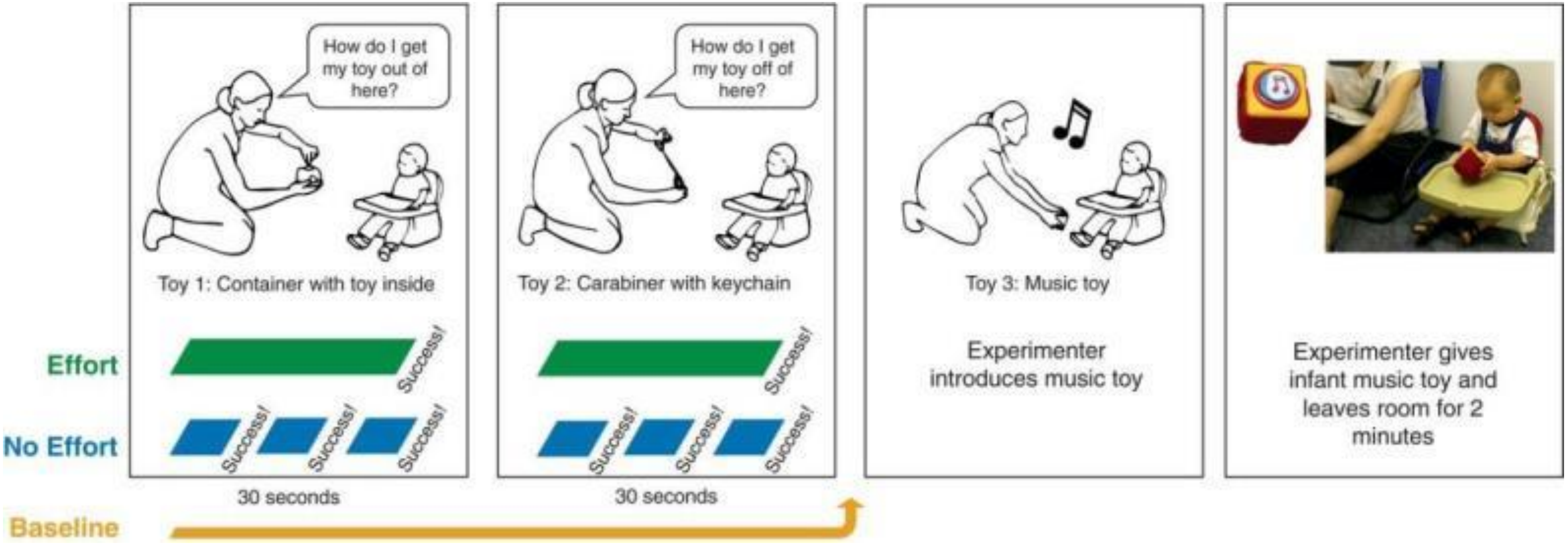
Abstract Info/History Metrics Preview PDF

Abstract

Over the past twenty years, neuroscience research on reward-based learning has converged on a canonical model, under which the neurotransmitter dopamine 'stamps in' associations between situations, actions and rewards by modulating the strength of synaptic connections between neurons. However, a growing number of recent findings have placed this standard model under strain. In the present work, we draw on recent advances in artificial intelligence to introduce a new theory of reward-based learning. Here, the dopamine system trains another part of the brain, the prefrontal cortex, to operate as its own free-standing learning system. This new perspective accommodates the findings that motivated the standard model, but also deals gracefully with a wider range of observations, providing a fresh foundation for future research.

Prefrontal cortex as a meta-reinforcement learning system, Nature Neuroscience,2018

挑战(2)：元学习(meta-learning)理论：“利用与探索”的 多巴胺神经计算机制


How do I get my toy out of here?
Toy 1: Container with toy inside
How do I get my toy off of here?
Toy 2: Carabiner with keychain
Toy 3: Music toy
Experimenter introduces music toy
Experimenter gives infant music toy and leaves room for 2 minutes
Effort
No Effort
Baseline
Success!
Success!
Success!
Success!
Success!
Success!
Success!
Success!
30 seconds
30 seconds

通过观察成人努力实现目标的行为和反应，婴幼儿的坚持性提高，并且努力去完成一个新的和困难的任务

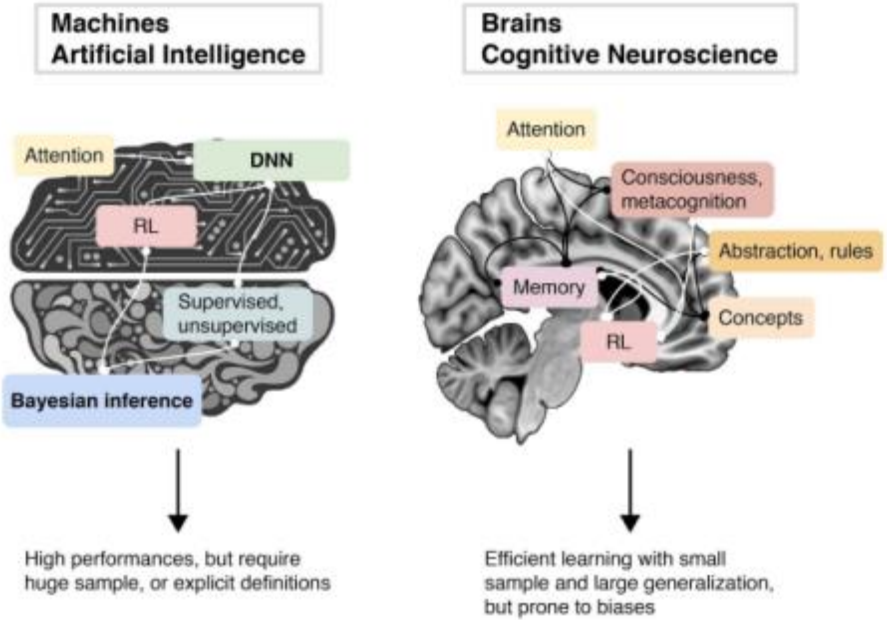
Julia A. Leonard, Yuna Lee, Laura E. Schulz. (2017, September 29). Infants make more attempts to achieve a goal when they see adults persist, ***Science***, 10.1126/science.aan2317.

挑战(2)：元学习(meta-learning)理论：“利用与探索”的 多巴胺神经计算机制

神经启发 认知可解 计算可行

- 基于结构和模型的学习是如何发生在脑中的
- 为什么多巴胺本身就编码模型信息
- 前额叶皮质的神经元是如何调节学习信号的

道之运行轨迹为元

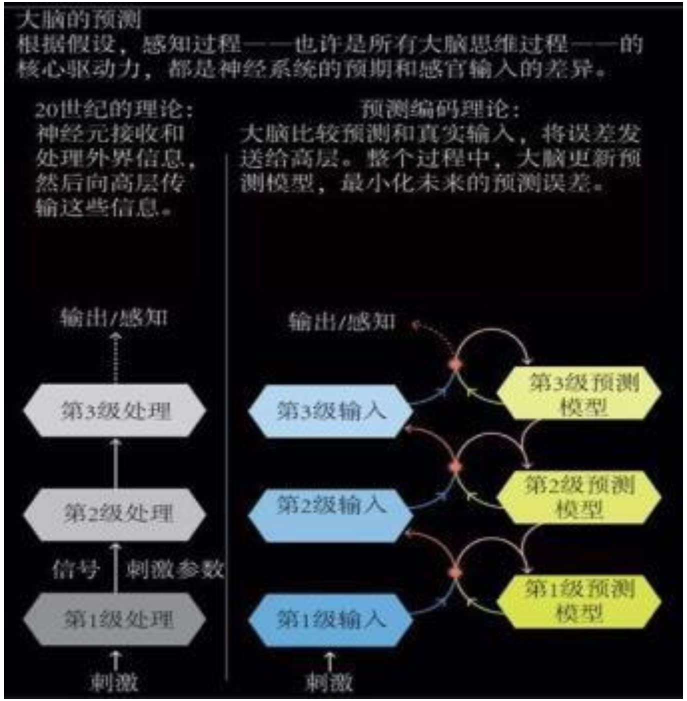

Machines
Artificial Intelligence
Attention
DNN
RL
Supervised, unsupervised
Bayesian inference
High performances, but require huge sample, or explicit definitions
Brains
Cognitive Neuroscience
Attention
Consciousness, metacognition
Abstraction, rules
Memory
Concepts
RL
Efficient learning with small sample and large generalization, but prone to biases

General schematic, solutions to complex problems in artificial intelligence and nature (brains). Higher cognitive functions continuously interact between them and with reinforcement learning to drive generalization and learning from small sample

挑战(3)：大脑的贝叶斯模型和因果推理

认知科学：大脑将外界信息计算得到概率赋予其所主动构建的“世界假设”，并加以调整优化。即比较“现有经验”和“未来期望” 。

- 大脑如何在神经回路尺度上进行模型构建和推断，将自底向上与自上而下相互结合
- 贝叶斯模型如何整合感知、认知、理性和意识



大脑的预测

根据假设，感知过程——也许是所有大脑思维过程——的核心驱动力，都是神经系统的预期和感官输入的差异。

20世纪的理论：神经元接收和处理外界信息，然后向高层传输这些信息。

预测编码理论：大脑比较预测和真实输入，将误差发送给高层。整个过程中，大脑更新预测模型，最小化未来的预测误差。

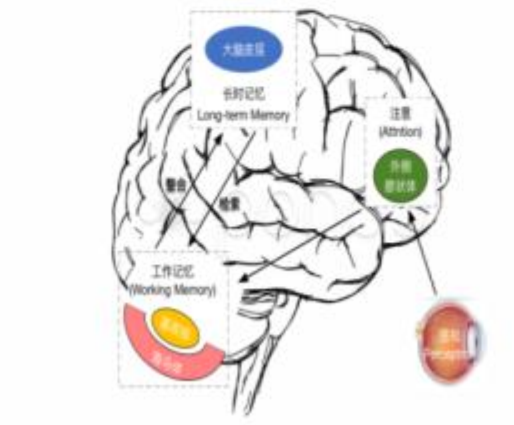
挑战(3)：大脑的贝叶斯模型和因果推理

认知科学：大脑不是在被动等待外界输入信息，而是主动构建尚未“显性描述”的世界的假设。

似然概率　　**先验概率**

⇩　　⇩

$$P(\text{疾病}|\text{症状}) = \frac{P(\text{症状}|\text{疾病}) \times P(\text{疾病})}{P(\text{症状})}$$



挑战(4)：控制论与人工智能和博弈对抗

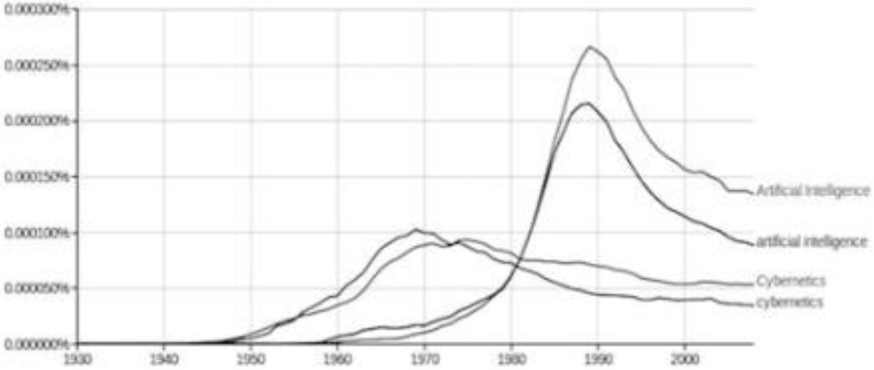
- Interesting fact: John von Neumann
  - Father of computer science + Father of game theory
- Game theory deeply interacts with computer science
  - Theoretical computer science: algorithmic game theory
  - AI: multiagent systems, game playing, human-robot interaction, machine learning, internet advertising
  - Web and Internet: internet economics, sharing economics
  - Distributed system: Blockchain

计算博弈策略

(智能体对决)

博弈机制设计

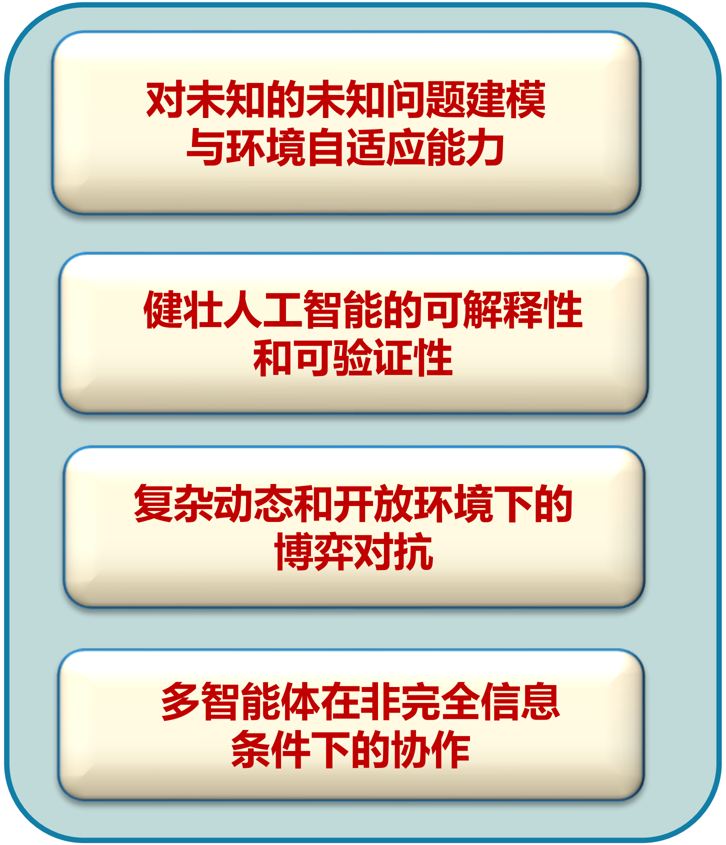
(拍卖、城市资源分配)



“人工智能”与“控制论”词频对比
（引自《人工智能简史》）

挑战(4)：控制论与人工智能和博弈对抗

- 健壮人工智能：
  - 自主无人系统在新环境中须通用性和自适应性要求极高
  - 极端复杂条件对人工智能“健壮性”要求极高
  - 人工智能体系的综合程度史无前例（云、网、端、算法、模型、人）
  - 人是“智能回路”中最重要因素



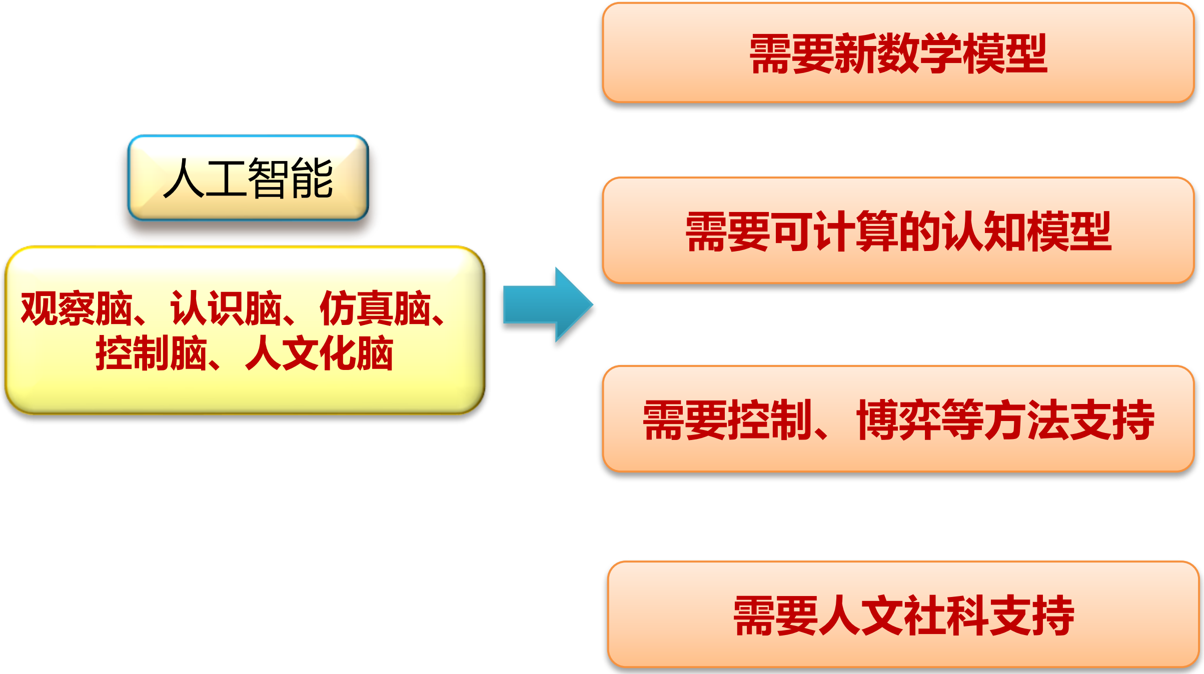
对未知的未知问题建模与环境自适应能力

健壮人工智能的可解释性和可验证性

复杂动态和开放环境下的博弈对抗

多智能体在非完全信息条件下的协作

多学科交叉促进人工智能发展



人工智能

观察脑、认识脑、仿真脑、控制脑、人文化脑

需要新数学模型

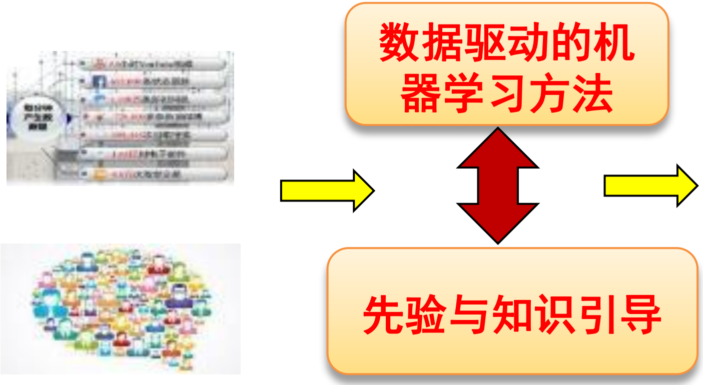
需要可计算的认知模型

需要控制、博弈等方法支持

需要人文社科支持

- 常识知识
- 元学习
- 贝叶斯大脑
- 因果推理
- 控制、博弈对抗、AI

可解释、更通用和适应性强的人工智能
数据利用、知识引导与能力学习



从浅层计算到深度神经推理

从单纯依赖于数据驱动的模型到数据驱动与知识引导相互结合

从领域任务驱动智能到更为通用条件下的强人工智能（从经验中学习）

Yueting Zhuang, Fei Wu, Chun Chen, Yunhe Pan, Challenges and Opportunities: From Big Data to Knowledge in AI 2.0, *Frontiers of Information Technology & Electronic Engineering*, 2017,18(1):3-14

“我们必须知道，我们必将知道”----大卫 希尔伯特 （David Hilbert）

谢谢