数据驱动与知识引导相结合：*self-thinking*

浙江大学人工智能研究所

吴飞

http://person.zju.edu.cn/wufei

2017年4月21日

提纲

- **三种智能学习的模式**
- 从 AlphaGo 到 Libratus
- 数据驱动与知识引导的思考
- 总结

三种智能学习的模式

命题、谓词等

核心：知识符号表示下的逻辑推理

监督、非监督和半监督等

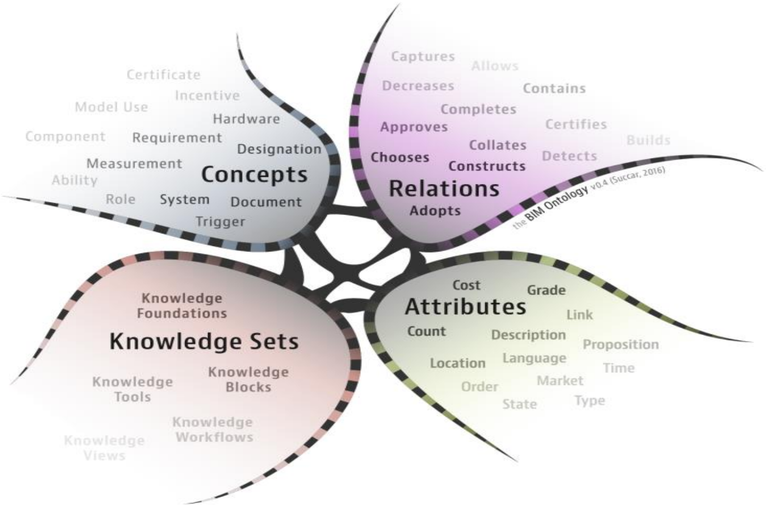
核心：任务导向型的识别、分类

从经验中学习

- Trail-and-Error
- Exploit-Exploration

用规则教 → 用数据学 → 用问题引导

三种智能学习的模式



学新知识
（逻辑推理引擎）

学模式
（假设空间与先验）

学策略与决策
（试错式自我学习）

用规则教 → 用数据学 → 用问题引导

三种智能学习的模式


The New York Times

How Many Computers to Identify a Cat? 16,000

Jim Wilson/The New York Times

An image of a cat that a neural network taught itself to recognize.

By JOHN MARKOFF

Published: June 25, 2012

学新知识
（逻辑推理引擎）

学模式
（假设空间与先验）

学策略与决策
（试错式自我学习）

用规则教 → 用数据学 → 用问题引导

三种智能学习的模式

Science | DOI:10.1145/2949662 Marina Krakovsky

Reinforcement Renaissance

The power of deep neural networks has sparked renewed interest in reinforcement learning, with applications to games, robotics, and beyond.

EACH TIME DEEPMIND has announced an amazing accomplishment in game-playing computers in recent months, people have taken notice.

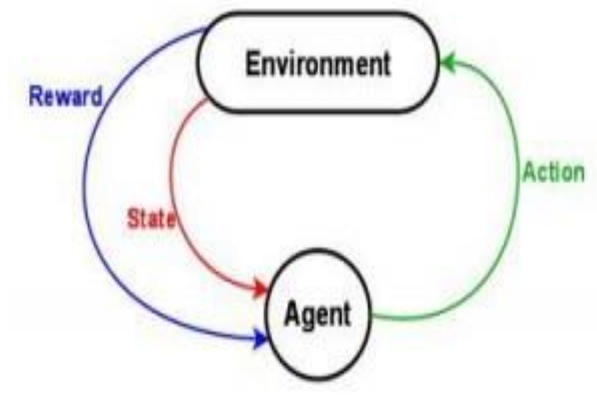
First, the Google-owned, London-based artificial intelligence (AI) research center wowed the world with a computer program that had taught itself to play nearly 50 different 1980s-era Atari games—from Pong and Breakout to Pac-Man, Space Invaders, Boxing, and more—using as input nothing but pixel positions and game scores, performing at or above the human level in more than half these varied games. Then, this January, DeepMind researchers impressed experts with a feat in the realm of strategy games: AlphaGo, their Go-playing program, beat the European champion in the ancient board game, which poses a much tougher AI challenge than chess. Less than two months later, AlphaGo scored an even greater victory: it won 4 games in a best-of-5 series against the best Go player in the world, surprising the champion himself.

The idea that a computer can learn to play such complex games from scratch and achieve a proficient level elicits gee-whiz reactions from the general public, and DeepMind's triumphs have heightened academic and commercial interest in the AI field behind DeepMind's methods: a blend of deep neural networks and reinforcement learning called "deep reinforcement learning."

"Reinforcement learning is a model of learning where you're not given a solution—you have to discover it by trial and error," explains Sridhar Mahadevan, a professor at the University of Massachusetts Amherst, a longtime center of research into reinforcement learning.

IMAGE BY MAXUSER

12 COMMUNICATIONS OF THE ACM | AUGUST 2016 | VOL. 59 | NO. 8



Reinforcement Renaissance, Communications of the ACM, 2016,59(8):12-14

学新知识 （逻辑推理引擎）	学模式 （假设空间与先验）	学策略与决策 （试错式自我学习）
用规则教	用数据学	用问题引导

三种智能学习的模式

学新知识
（逻辑推理引擎）

学模式
（假设空间与先验）

学策略与决策
（试错式自我学习）

用规则教 → 用数据学 → 用问题引导

从知识到数据、从数据到能力

三种智能学习的模式

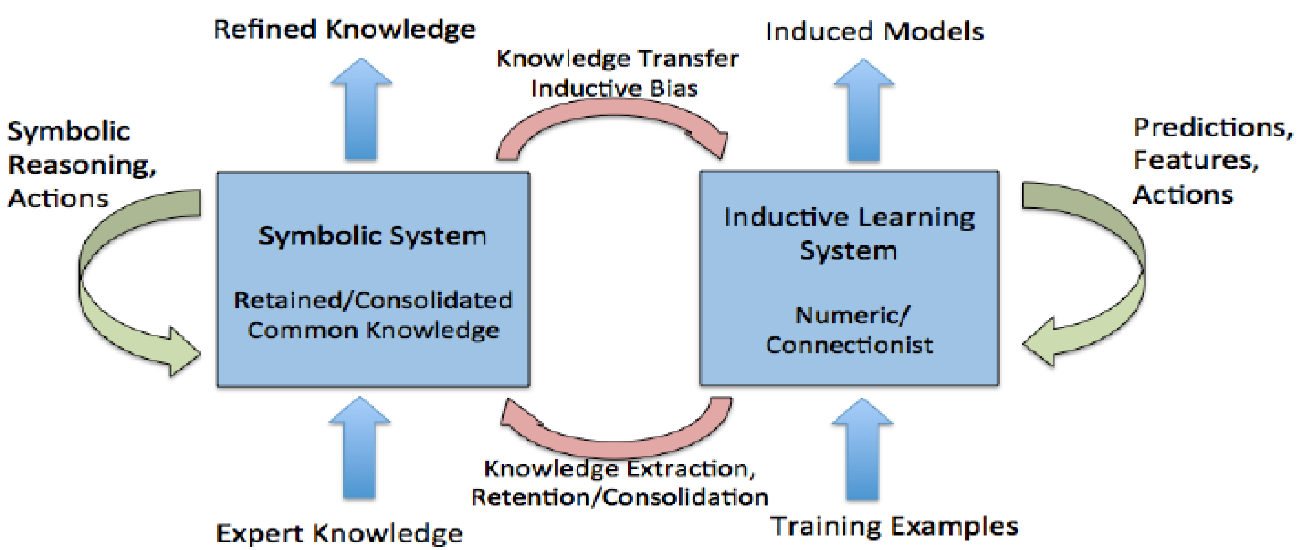
学新知识
（逻辑推理引擎）

用规则教

优势：与人类逻辑推理相似，解释性强

不足：知识库和推理规则天生会不完善（手工构造而存在偏见bias）

趋势：其中一种趋势是演绎方法与归纳手段的结合(如，neural–symbolic integration)



三种智能学习的模式

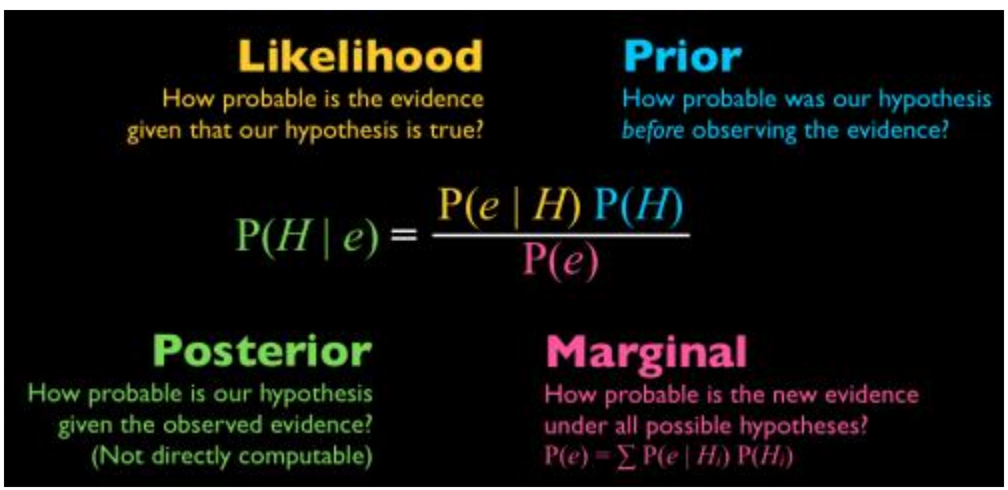
学模式
（假设空间与先验）

用数据学

优势：直接从大（或小）数据中学习

不足：数据依赖性强、解释能力弱

趋势：其中一种趋势是将数据驱动与知识引导相结合，广泛结合逻辑、先验和知识以及数据



Bayes对后验概率解释：
先验概率
假设空间：似然函数

三种智能学习的模式

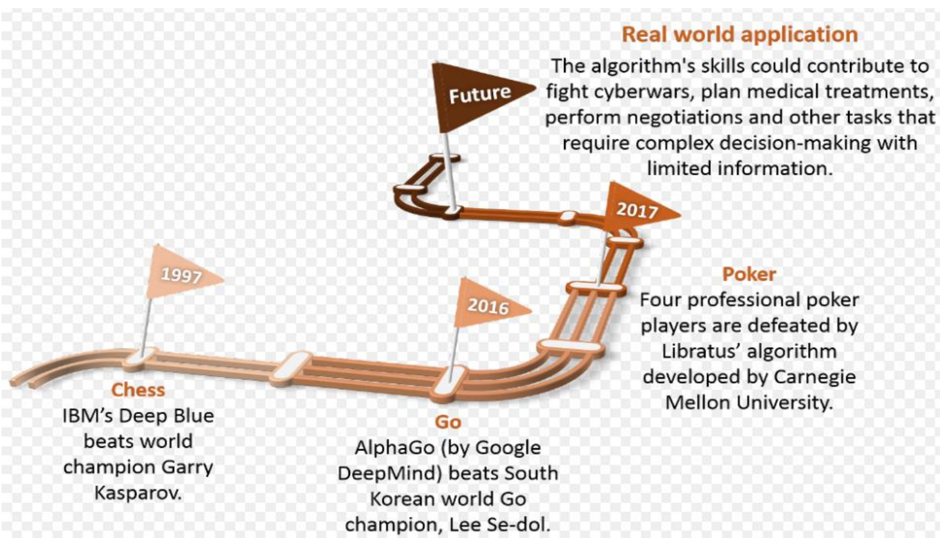
学策略与决策
（试错式自我学习）

用问题引导

优势：从经验中进行能力的持续学习(甚至因为具有学习能力而可以self-play形式学)

不足：因非穷举式搜索（因空间巨大或信息不完全）而需更好策略 (如采样和模拟等)

趋势：从完全信息下的博弈决策向非完全信息的博弈决策迈进

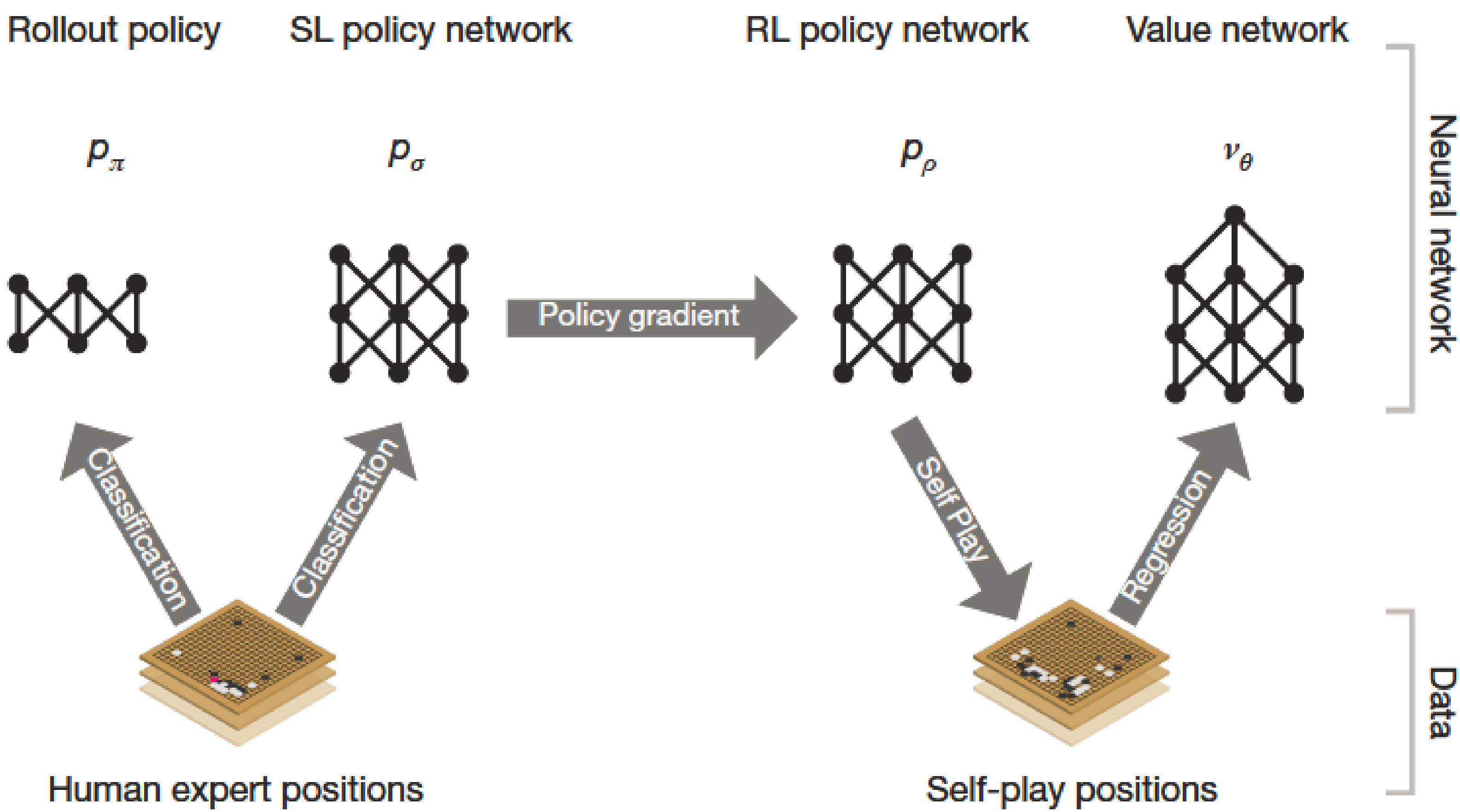


Playing at a normal human pace, one can't beat a computer program for limit Texas Hold'em with statistical significance in a lifetime.

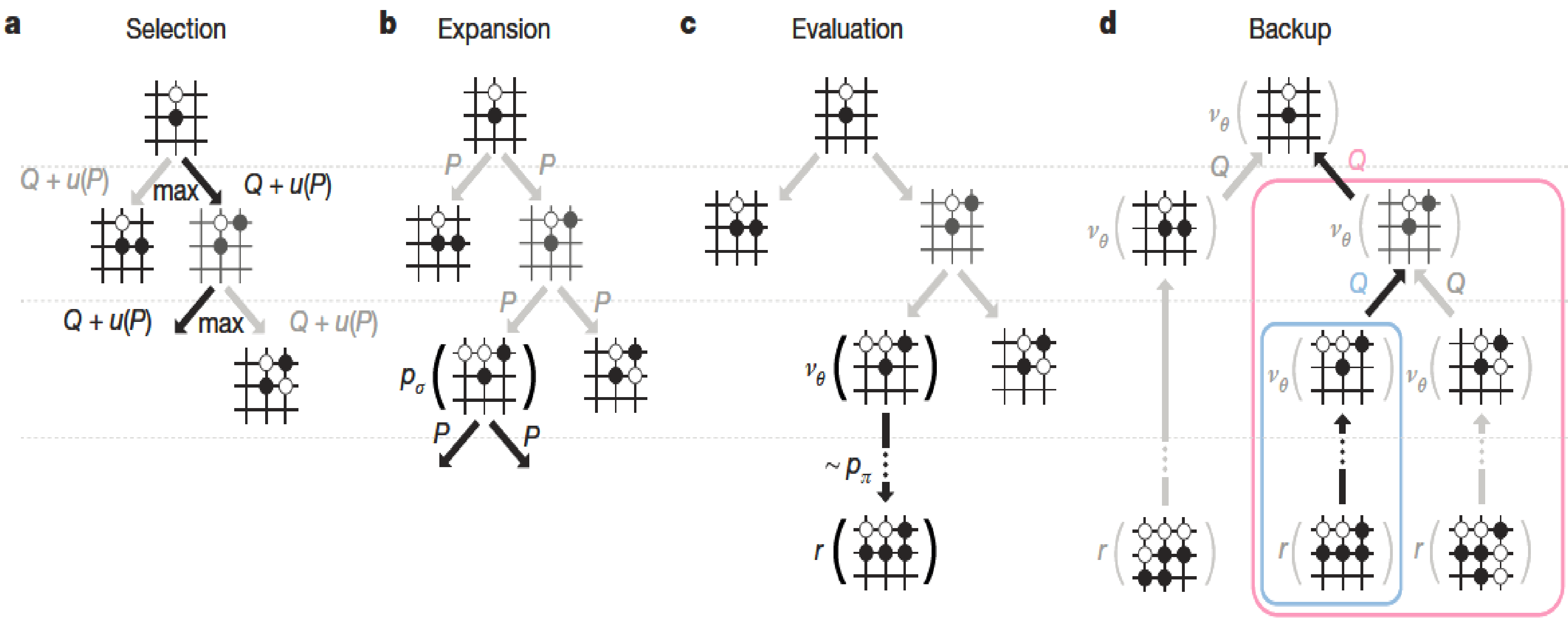
提纲

- 三种智能学习的模式
- **从 *AlphaGo* 到 *Libratus***
- 数据驱动与知识引导的思考
- 总结

AlphaGo：完全信息下的博弈



AlphaGo：利用(exploitation)与探索(exploration)



AlphaGo综合考虑了三个因素进行决策： $\underset{a}{\operatorname{argmax}}\left(Q(s_t, a) + \frac{P(s,a)}{1+N(s,a)}\right)$

- action value Q （争取胜利）
- 人类棋手先验P（不乱下棋）
- 访问频率少 N （偏好奇招）

AlphaGo：利用(exploitation)与探索(exploration)

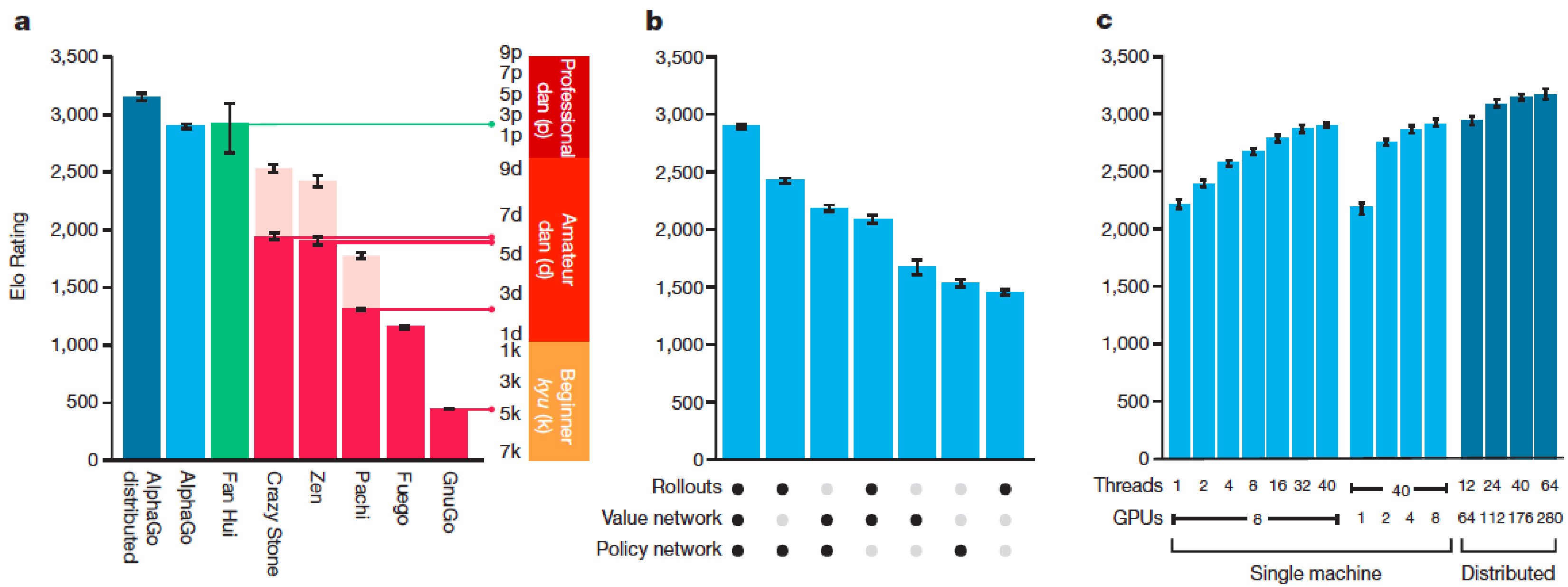


Figure 4 | Tournament evaluation of AlphaGo. a, Results of a tournament between different Go programs (see Extended Data Tables 6–11). Each program used approximately 5 s computation time per move. To provide a greater challenge to AlphaGo, some programs (pale upper bars) were given four handicap stones (that is, free moves at the start of every game) against all opponents. Programs were evaluated on an Elo scale[37]: a 230 point gap corresponds to a 79% probability of winning, which roughly corresponds to one amateur *dan* rank advantage on KGS[38]; an approximate correspondence to human ranks is also shown, horizontal lines show KGS ranks achieved online by that program. Games against the human European champion Fan Hui were also included; these games used longer time controls. 95% confidence intervals are shown. **b,** Performance of AlphaGo, on a single machine, for different combinations of components. The version solely using the policy network does not perform any search. **c,** Scalability study of MCTS in AlphaGo with search threads and GPUs, using asynchronous search (light blue) or distributed search (dark blue), for 2 s per move.

注：腾讯AI Lab研发的围棋程序“绝艺”于2017年3月参加日本最具传统和权威的“UEC杯”计算机围棋大赛，从30个参赛程序中脱颖而出，以11战全胜的战绩夺得冠军

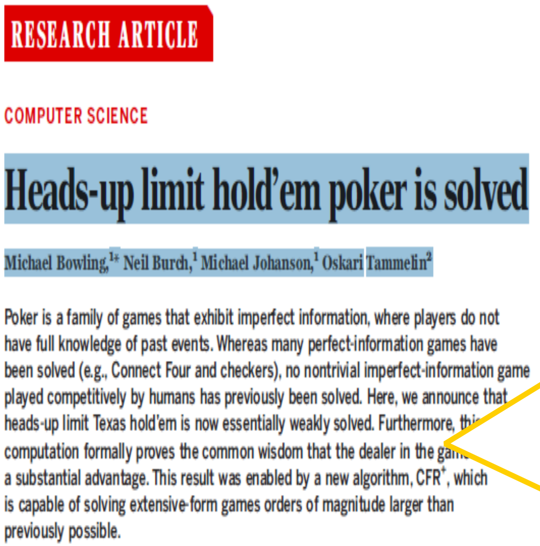
从完全信息博弈到非完全信息博弈

与AlphaGo是在规则已知下的博弈情况不同，现实社会中诸多行为（如新经济运行、环境变化、产业布局、网络空间安全等）决策是非完全信息的博弈，即在信息未能全面掌握条件下进行推理和决策。

面临的巨大挑战（以扑克为例）

- Hidden information (other players' cards)
- Uncertainty about future events
- Deceptive strategies needed in a good player
- Very large game tree

从完全信息博弈到非完全信息博弈


RESEARCH ARTICLE

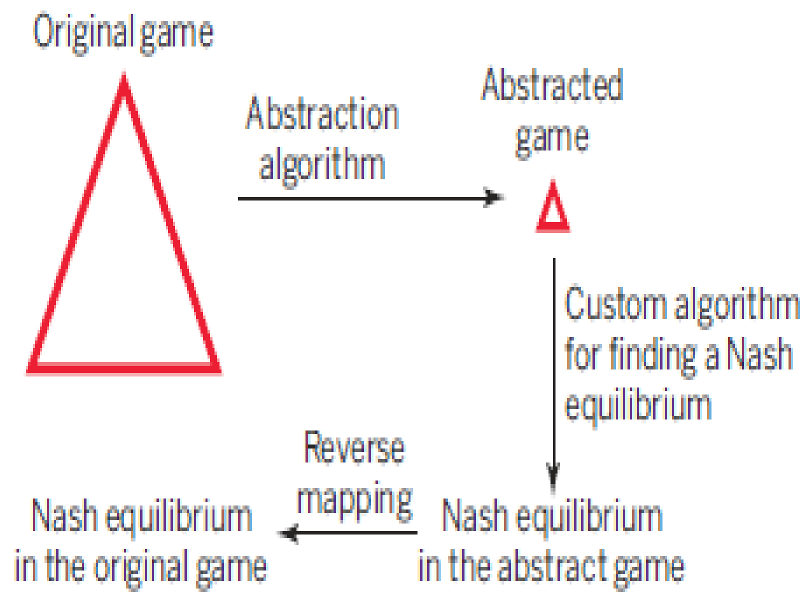
COMPUTER SCIENCE

Heads-up limit hold'em poker is solved

Michael Bowling,[1*] Neil Burch,[1] Michael Johanson,[1] Oskari Tammelin[2]

Poker is a family of games that exhibit imperfect information, where players do not have full knowledge of past events. Whereas many perfect-information games have been solved (e.g., Connect Four and checkers), no nontrivial imperfect-information game played competitively by humans has previously been solved. Here, we announce that heads-up limit Texas hold'em is now essentially weakly solved. Furthermore, thi computation formally proves the common wisdom that the dealer in the gam a substantial advantage. This result was enabled by a new algorithm, CFR+, which is capable of solving extensive-form games orders of magnitude larger than previously possible.

we announce that heads-up limit Texas hold'em is now **essentially weakly solved**...This result was enabled by a new algorithm, CFR+ (**Counterfactual regret minimization**)...



Leading approach to imperfect-information games

Michael Bowling, Neil Burch, Michael Johanson, Oskari Tammelin, Heads-up limit hold'em poker is solved, *Science*, 347(6218):145-149,2015

从完全信息博弈到非完全信息博弈

Libratus 是一个无限德州扑克的人工智能程序，使用了很多策略，如诈唬。借助匹兹堡超算中心的Bridges超级计算机，训练成本是1500万core-hour，比赛期间耗用1000多万core-hour。未用深度学习。

三大模块

- 模块一：Nash equilibrium approximation before competition（Monte-Carlo Sampling for Counterfactual regret minimization）
- 模块二：Endgame solving during competition
- 模块三：Continual self-improvement after competition

Noam Brown, Tuomas Sandholm, Safe and Nested Endgame Solving for Imperfect-Information Games, Workshop on Computer Poker and Imperfect Information at the AAAI Conference on Artificial Intelligence (AAAI), 2017

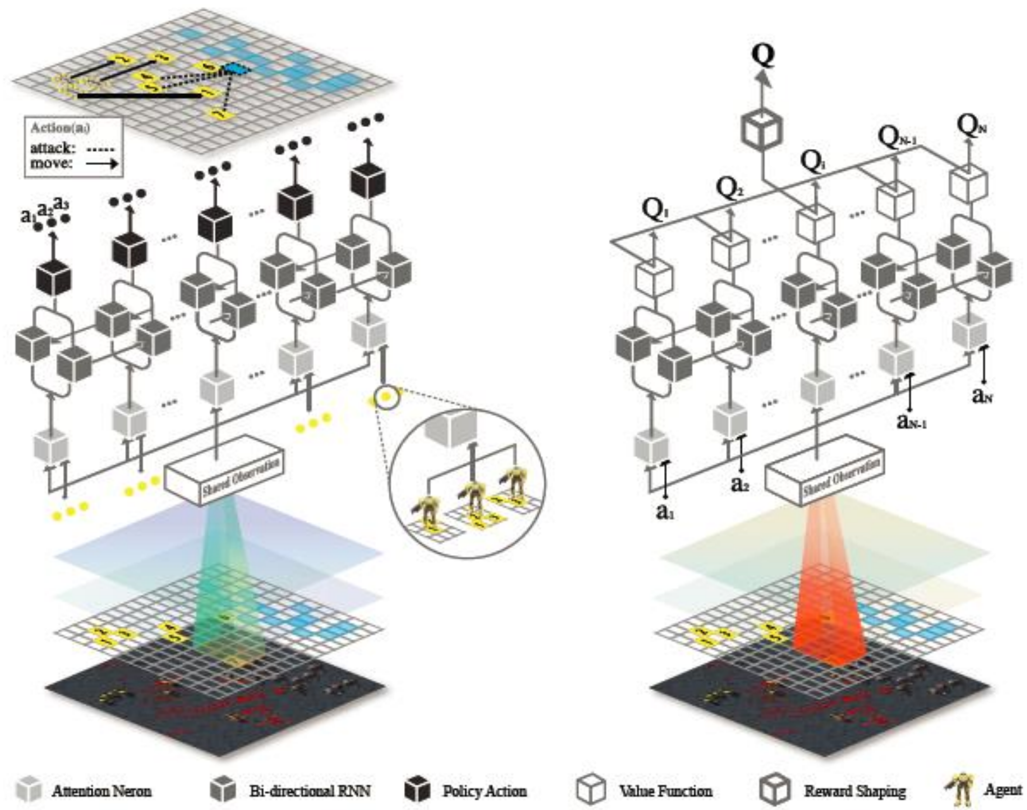
从完全信息博弈到非完全信息博弈

Multiagent Bidirectionally-Coordinated Nets for Learning to Play StarCraft Combat Games

Peng Peng[†], Quan Yuan[†], Ying Wen[‡], Yaodong Yang[‡], Zhenkun Tang[†], Haitao Long[†], Jun Wang[‡] *
[†]Alibaba Group, [‡]University College London

Abstract

Real-world artificial intelligence (AI) applications often require multiple agents to work in a collaborative effort. Efficient learning for intra-agent communication and coordination is an indispensable step towards general AI. In this paper, we take StarCraft combat game as the test scenario, where the task is to coordinate multiple agents as a team to defeat their enemies. To maintain a scalable yet effective communication protocol, we introduce a multiagent bidirectionally-coordinated network (BiCNet ['bɪknet]) with a vectorised extension of actor-critic formulation. We show that BiCNet can handle different types of combats under diverse terrains with arbitrary numbers of AI agents for both sides. Our analysis demonstrates that without any supervisions such as human demonstrations or labelled data, BiCNet could learn various types of coordination strategies that is similar to these of experienced game players. Moreover, BiCNet is easily adaptable to the tasks with heterogeneous agents. In our experiments, we evaluate our approach against multiple baselines under different scenarios; it shows state-of-the-art performance, and possesses potential values for large-scale real-world applications.



(a) Multiagent policy networks with grouping (b) Multiagent Q networks with reward shaping

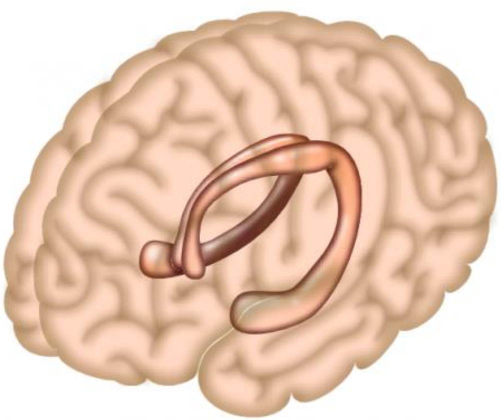
Figure 1: Bidirectionally-Coordinated Net (BiCNet).

提纲

- 三种智能学习的模式
- 从 AlphaGo 到 Libratus
- **数据驱动与知识引导的思考**
- 总结

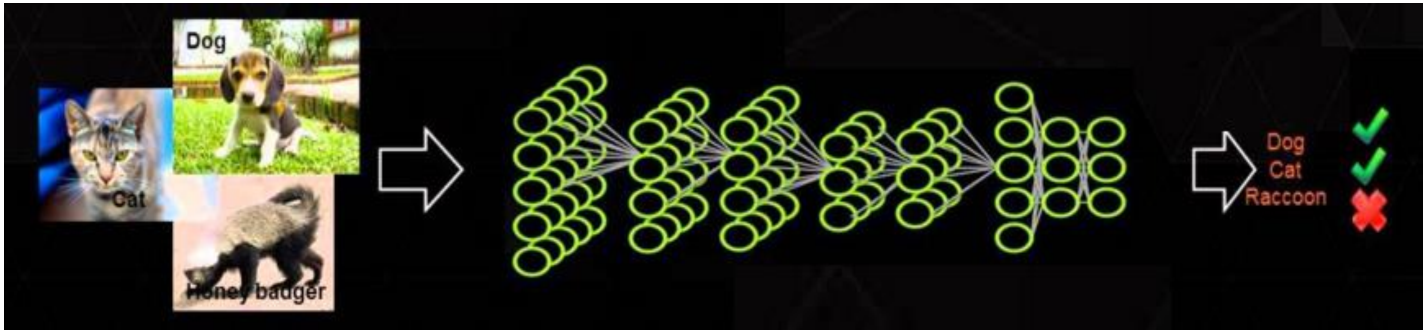
数据驱动与知识引导

- 知识与先验从何而来：
  - 人类大脑中的海马体：先验和知识的储存体

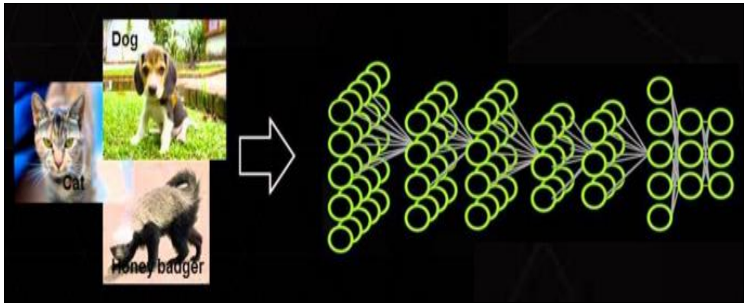


如何将知识引入

□ 以端到端学习为例：



深度学习

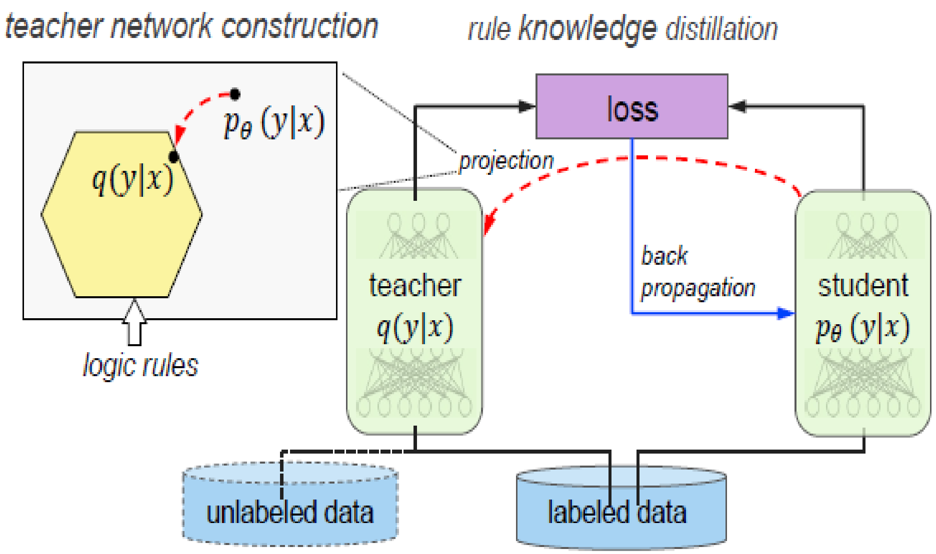


- 条件随机场
- 长短时记忆模型
- 随机森林

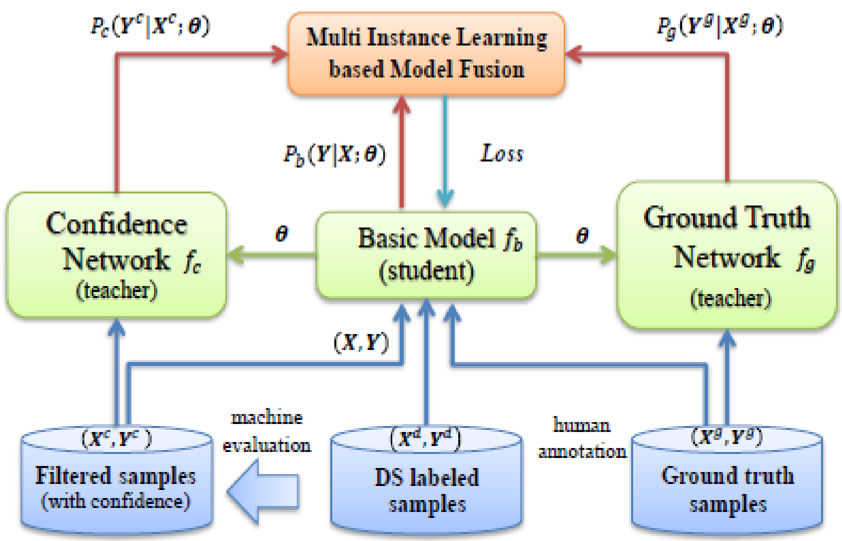
深度学习 + 推理模型

如何将知识引入

□ 以端到端学习为例：



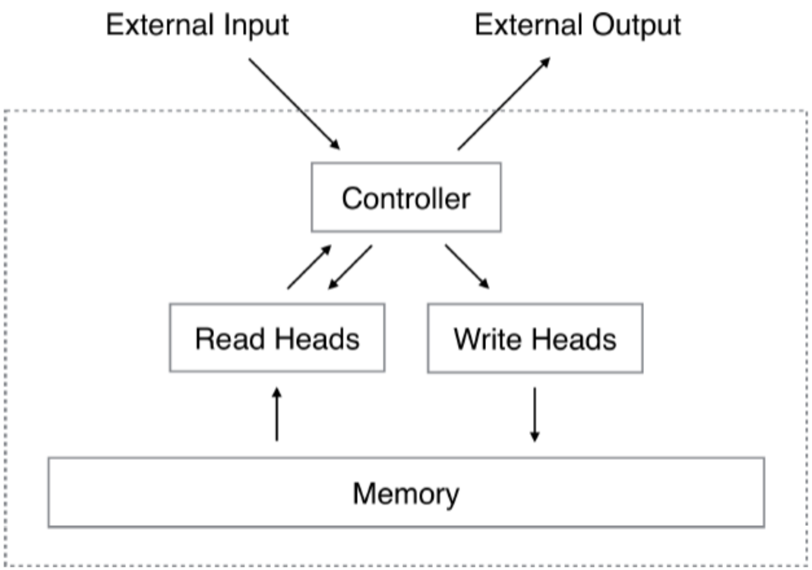
Harnessing Deep Neural Networks
with Logic Rules



ENCORE: External Neural Constraints
Regularized Distant Supervision for
Relation Extraction

Learning with external memory

- Neural Turing Machines



◆ Reading

$$\mathbf{r}_t \longleftarrow \sum_i w_t(i)\mathbf{M}_t(i),$$

$$\sum_i w_t(i) = 1, \qquad 0 \leq w_t(i) \leq 1, \forall i.$$

◆ Writing

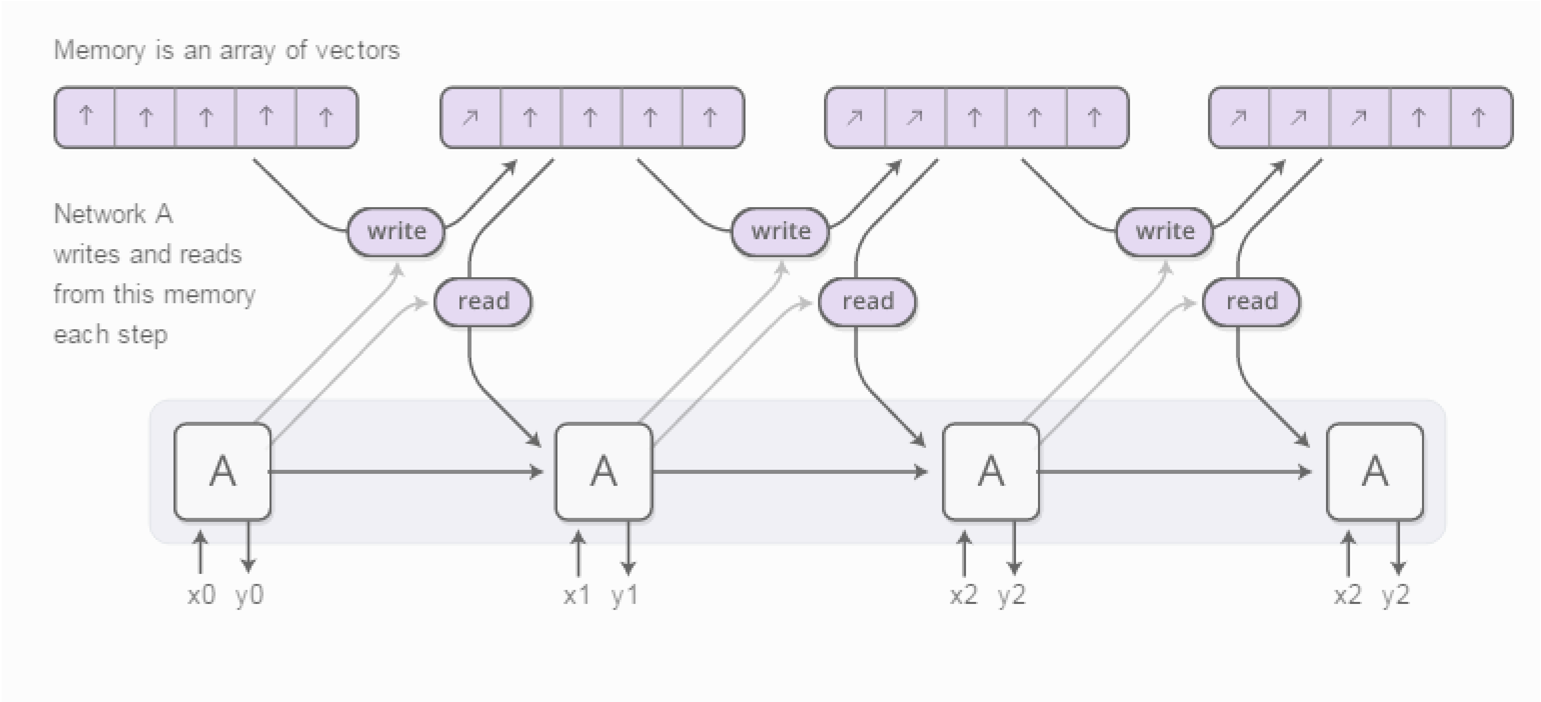
$$\tilde{\mathbf{M}}_t(i) \longleftarrow \mathbf{M}_{t-1}(i)\left[\mathbf{1} - w_t(i)\mathbf{e}_t\right],$$

$$\mathbf{M}_t(i) \longleftarrow \tilde{\mathbf{M}}_t(i) + w_t(i)\,\mathbf{a}_t.$$

Graves A, Wayne G, Danihelka I. Neural Turing Machines,arXiv preprint arXiv:1410.5401, 2014, DeepMind
J. Weston, S. Chopra, A. Bordes. Memory Networks. ICLR 2015 (and arXiv:1410.3916, Facebook AI)

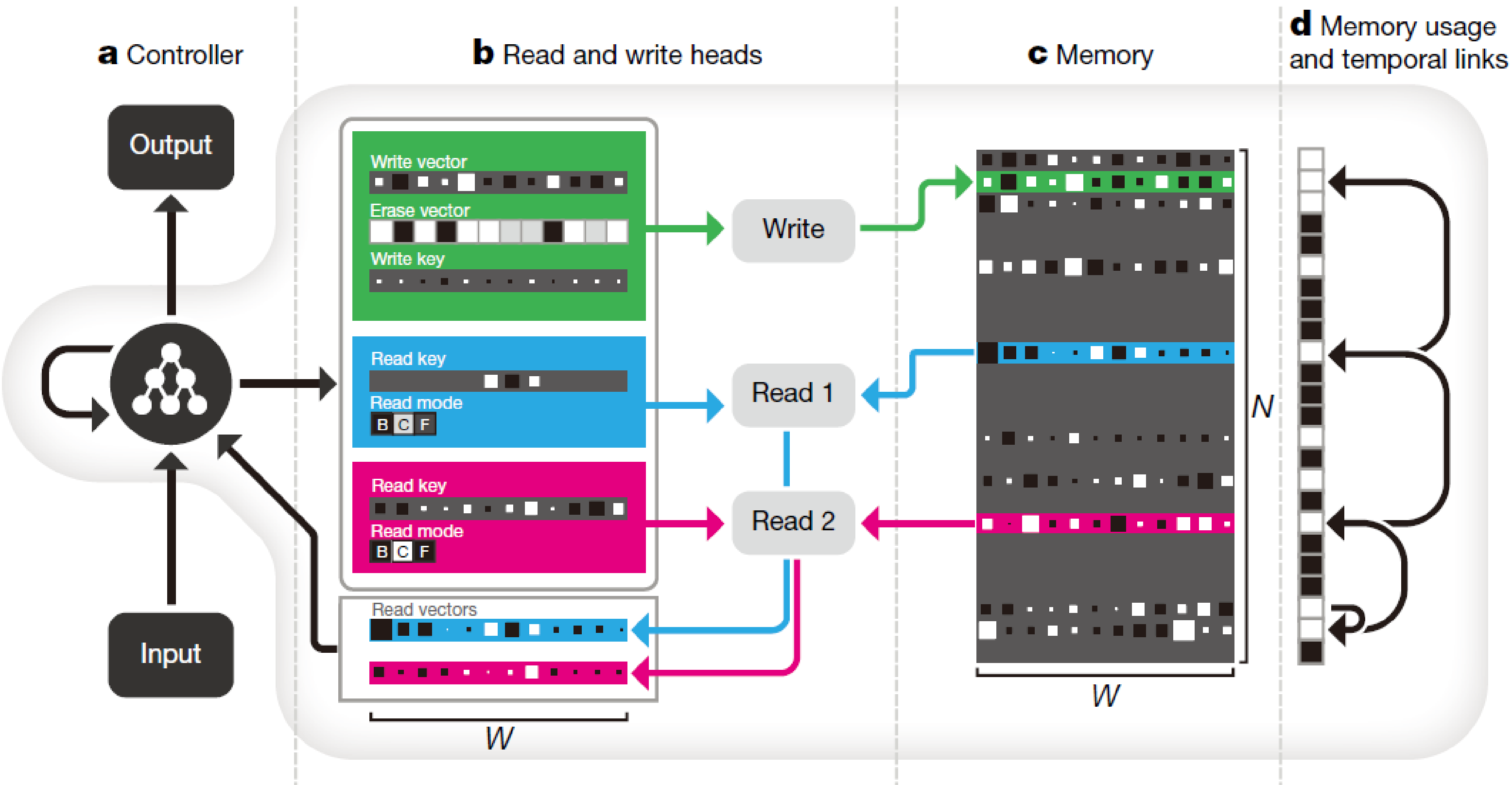
Learning with external memory

- Neural Turing Machines



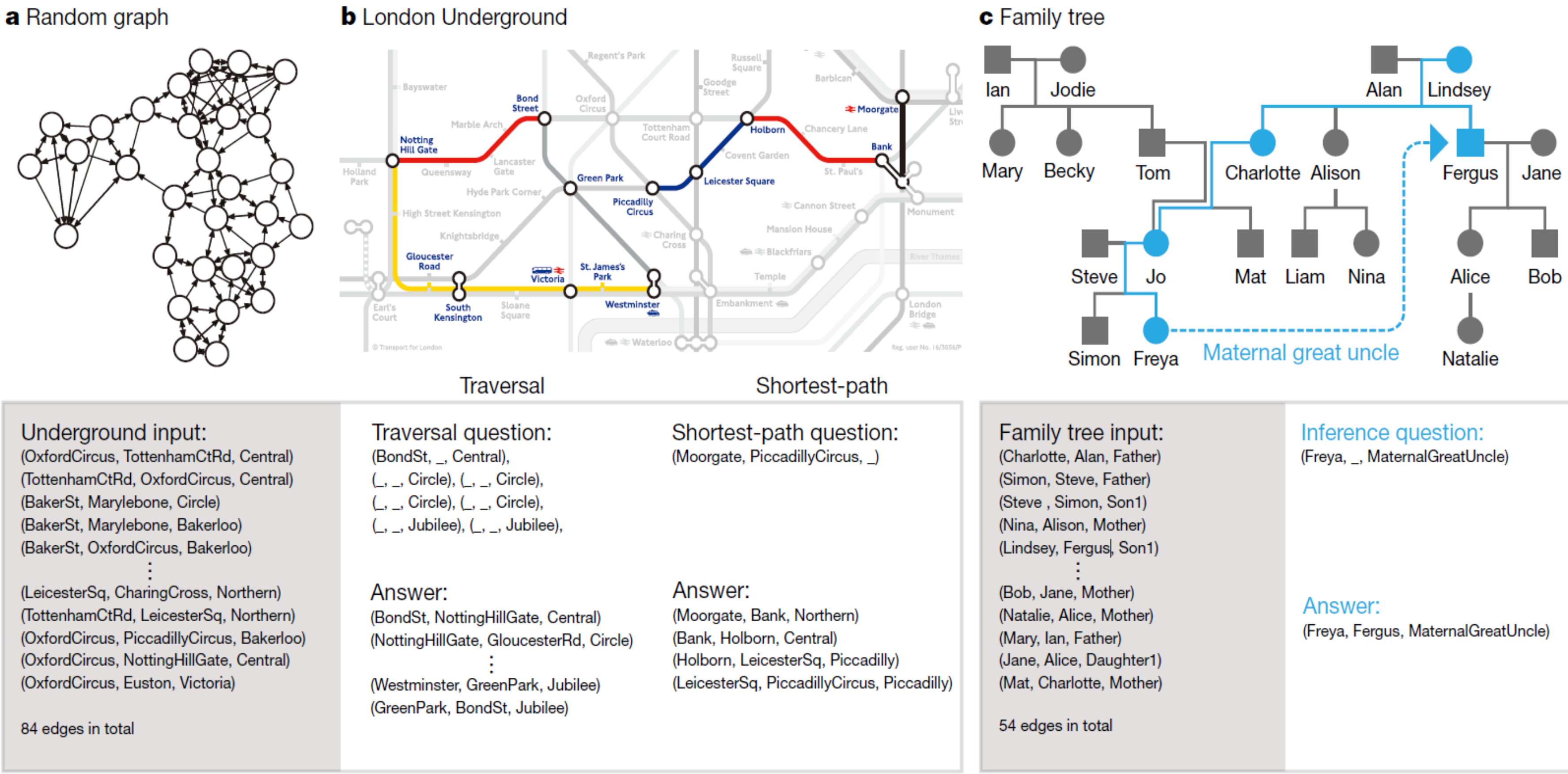
Learning with external memory

- □ 可微分神经计算机(differentiable neural computer，DNC)：
  - □ An achievement that has potential implications for the neural–symbolic integration problem(神经网络-符号计算的统一)
  - □ Deep neural reasoning and one-shot learning



Graves, Alex, et al. "Hybrid computing using a neural network with dynamic external memory." Nature 538.7626 (2016): 471-476.

Learning with external memory



Deep neural reasoning

（连续空间与离散空间相互协调）

Learning with external memory

Learning of Basic Algorithms using Reasoning, Attention, Memory (RAM)

Methods include adding stacks and addressable memory to RNNs:

- “Neural Net Architectures for Temporal Sequence Processing.” M. Mozer.
- “Neural Turing Machines” A. Graves, G. Wayne, I. Danihelka.
- “Inferring Algorithmic Patterns with Stack Augmented Recurrent Nets.” A. Joulin, T. Mikolov.
- “Learning to Transduce with Unbounded Memory” E. Grefenstette et al.
- “Neural Programmer-Interpreters” S. Reed, N. de Freitas.
- “Reinforcement Learning Turing Machine.” W. Zaremba and I. Sutskever.
- “Learning Simple Algorithms from Examples” W. Zaremba, T. Mikolov, A. Joulin, R. Fergus.
- “The Neural GPU and the Neural RAM machine” I. Sutskever.

Gives 'memory' to AI

DeepMind crafted an algorithm that lets a neural network '**remember**' past knowledge and learn more effectively. The approach is similar to how your own mind works, and might even provide insights into the functioning of human minds.

Much like real synapses, which tend to preserve connections between neurons when they've been useful in the past, the algorithm (known as **Elastic Weight Consideration**) decides how important a given connection is to its associated task

Overcoming catastrophic forgetting in neural networks

James Kirkpatrick[a,1], Razvan Pascanu[a], Neil Rabinowitz[a], Joel Veness[a], Guillaume Desjardins[a], Andrei A. Rusu[a], Kieran Milan[a], John Quan[a], Tiago Ramalho[a], Agnieszka Grabska-Barwinska[a], Demis Hassabis[a], Claudia Clopath[b], Dharshan Kumaran[a], and Raia Hadsell[a]

[a]DeepMind, London EC4 5TW, United Kingdom; and [b]Bioengineering Department, Imperial College London, London SW7 2AZ, United Kingdom

Edited by James L. McClelland, Stanford University, Stanford, CA, and approved February 13, 2017 (received for review July 19, 2016)

The ability to learn tasks in a sequential fashion is crucial to the development of artificial intelligence. Until now neural networks have not been capable of this and it has been widely thought that catastrophic forgetting is an inevitable feature of connectionist models. We show that it is possible to overcome this limitation and train networks that can maintain expertise on tasks that they have not experienced for a long time. Our approach remembers old tasks by selectively slowing down learning on the weights important for those tasks. We demonstrate our approach is scalable and effective by solving a set of classification tasks based on a hand-written digit dataset and by learning several Atari 2600 games sequentially.

synaptic consolidation | artificial intelligence | stability plasticity | continual learning | deep learning

Achieving artificial general intelligence requires that agents are able to learn and remember many different tasks (1). This is particularly difficult in real-world settings: The sequence of tasks may not be explicitly labeled, tasks may switch unpredictably, and any individual task may not recur for long time intervals. Critically, therefore, intelligent agents must demonstrate a capacity for continual learning: that is, the ability to learn consecutive tasks without forgetting how to perform previously trained tasks.

Continual learning poses particular challenges for artificial neural networks due to the tendency for knowledge of the previously learned task(s) (e.g., task *A*) to be abruptly lost as information relevant to the current task (e.g., task *B*) is incorporated. This phenomenon, termed catastrophic forgetting (2–6), occurs specifically when the network is trained sequentially on multiple tasks because the weights in the network that are important for task *A* are changed to meet the objectives of task *B*. Whereas recent advances in machine learning and in particular deep neural networks have resulted in impressive gains in performance across a variety of domains (e.g., refs. 7 and 8), little progress has been made in achieving continual learning. Current approaches have typically ensured that data from all tasks are simultaneously available during training. By interleaving data from multiple tasks during learning, forgetting does not occur because the weights of the network can be jointly

In marked contrast to artificial neural networks, humans and other animals appear to be able to learn in a continual fashion (11). Recent evidence suggests that the mammalian brain may avoid catastrophic forgetting by protecting previously acquired knowledge in neocortical circuits (11–14). When a mouse acquires a new skill, a proportion of excitatory synapses are strengthened; this manifests as an increase in the volume of individual dendritic spines of neurons (13). Critically, these enlarged dendritic spines persist despite the subsequent learning of other tasks, accounting for retention of performance several months later (13). When these spines are selectively "erased," the corresponding skill is forgotten (11, 12). This provides causal evidence that neural mechanisms supporting the protection of these strengthened synapses are critical to retention of task performance. These experimental findings—together with neurobiological models such as the cascade model (15, 16)—suggest that continual learning in the neocortex relies on task-specific synaptic consolidation, whereby knowledge is durably encoded by rendering a proportion of synapses less plastic and therefore stable over long timescales.

In this work, we demonstrate that task-specific synaptic consolidation offers a unique solution to the continual-learning problem for artificial intelligence. We develop an algorithm analogous to synaptic consolidation for artificial neural networks, which we refer to as elastic weight consolidation (EWC). This algorithm slows down learning on certain weights based on how important they are to previously seen tasks. We show how EWC can be used in supervised learning and reinforcement learning problems to train several tasks sequentially without forgetting older ones, in marked contrast to previous deep-learning techniques.

Significance

Deep neural networks are currently the most successful machine-learning technique for solving a variety of tasks, including language translation, image classification, and image generation. One weakness of such models is that, unlike humans, they are unable to learn multiple tasks sequentially. In this work we propose a practical solution to train such models sequentially by protecting the weights important for previous tasks. This approach, inspired by

NEUROSCIENCE

APPLIED MATHEMATICS

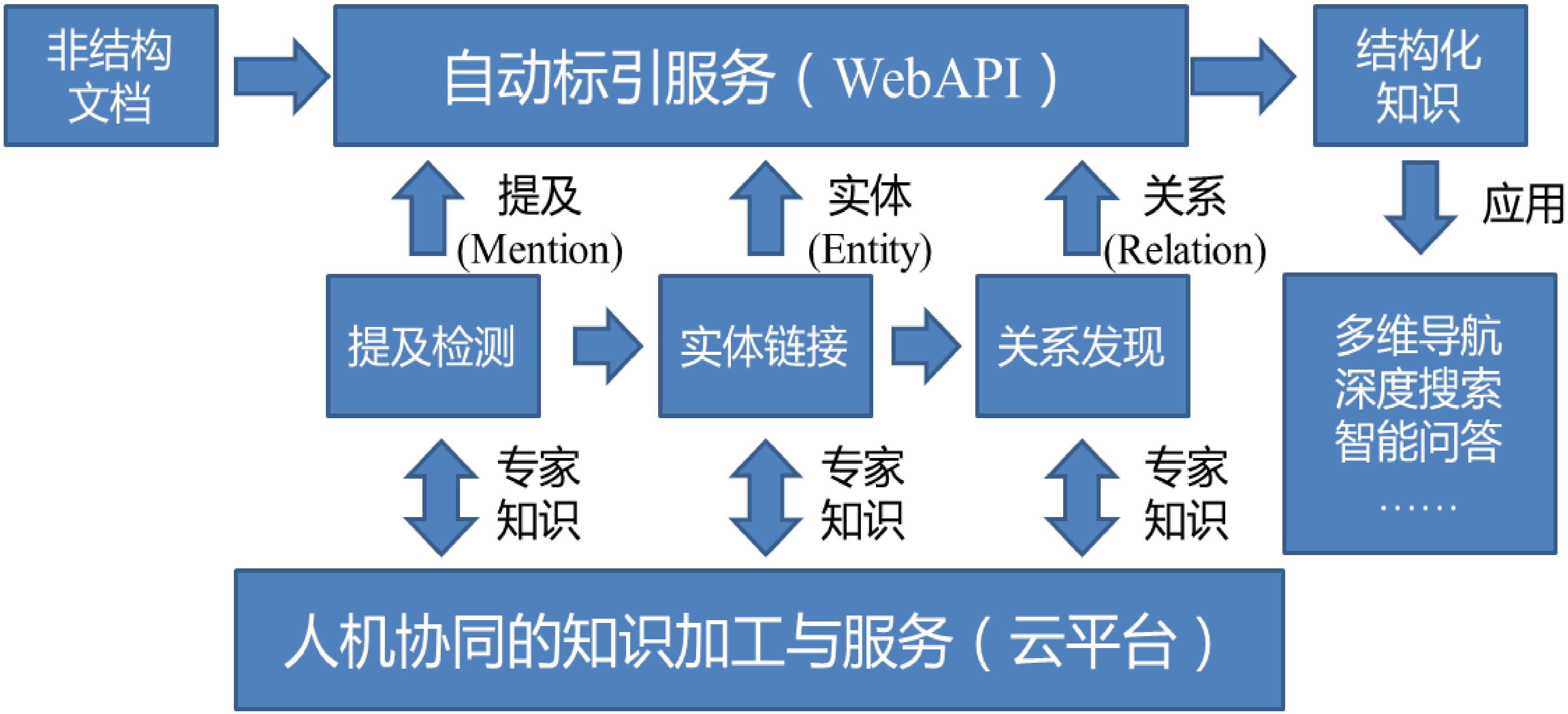
James Kirkpatrick, Razvan Pascanu, et al., Overcoming catastrophic forgetting in neural network, PNAS, http://www.pnas.org/cgi/doi/10.1073/pnas.1611835114

知识计算引擎: KS-Studio

4,765,271	102,136	27,454	560,711
Wikipedia实体数	ChEBI实体数	Mesh实体数	API调用次数

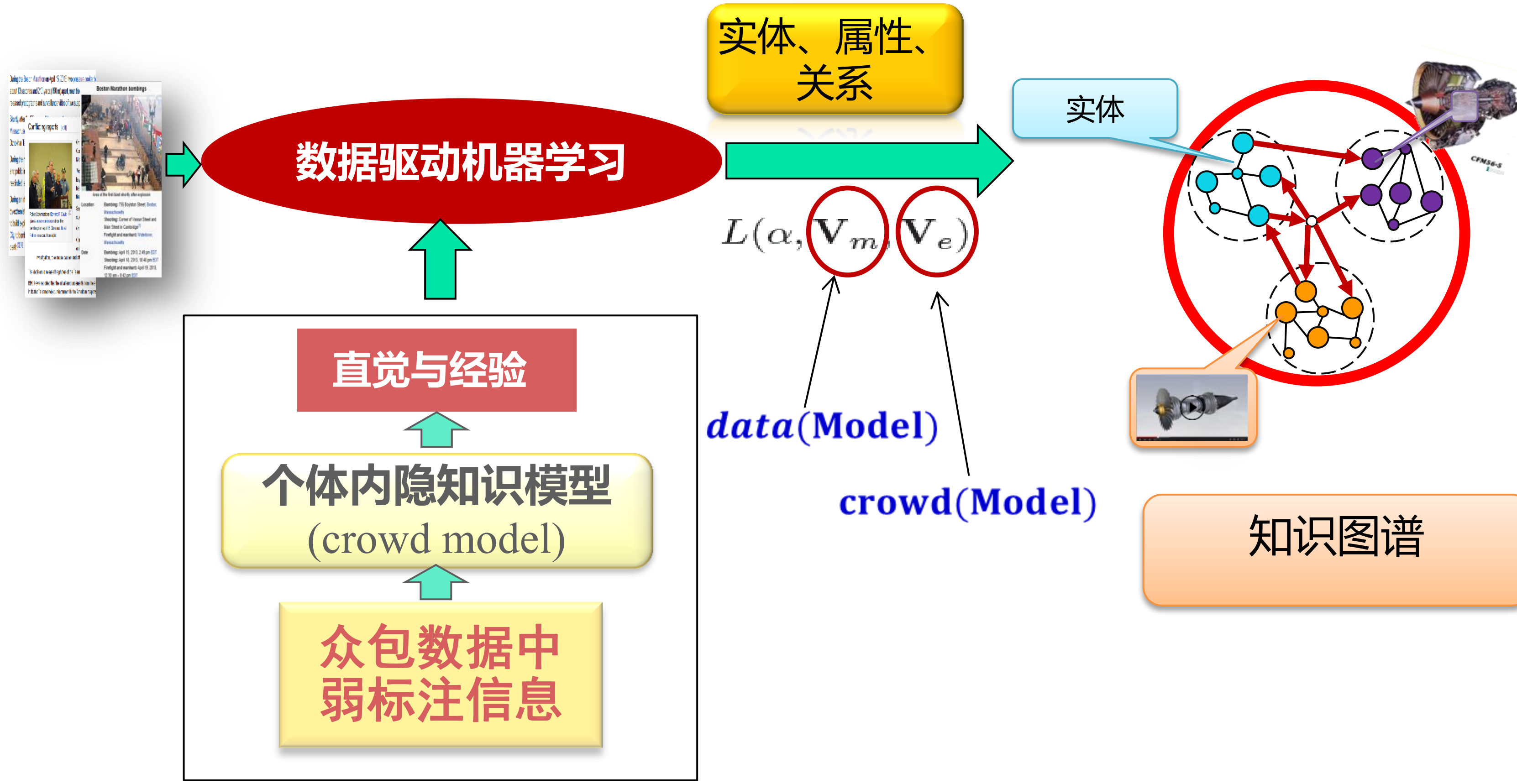
http://www.ksstudio.org/

知识计算引擎: KS-Studio



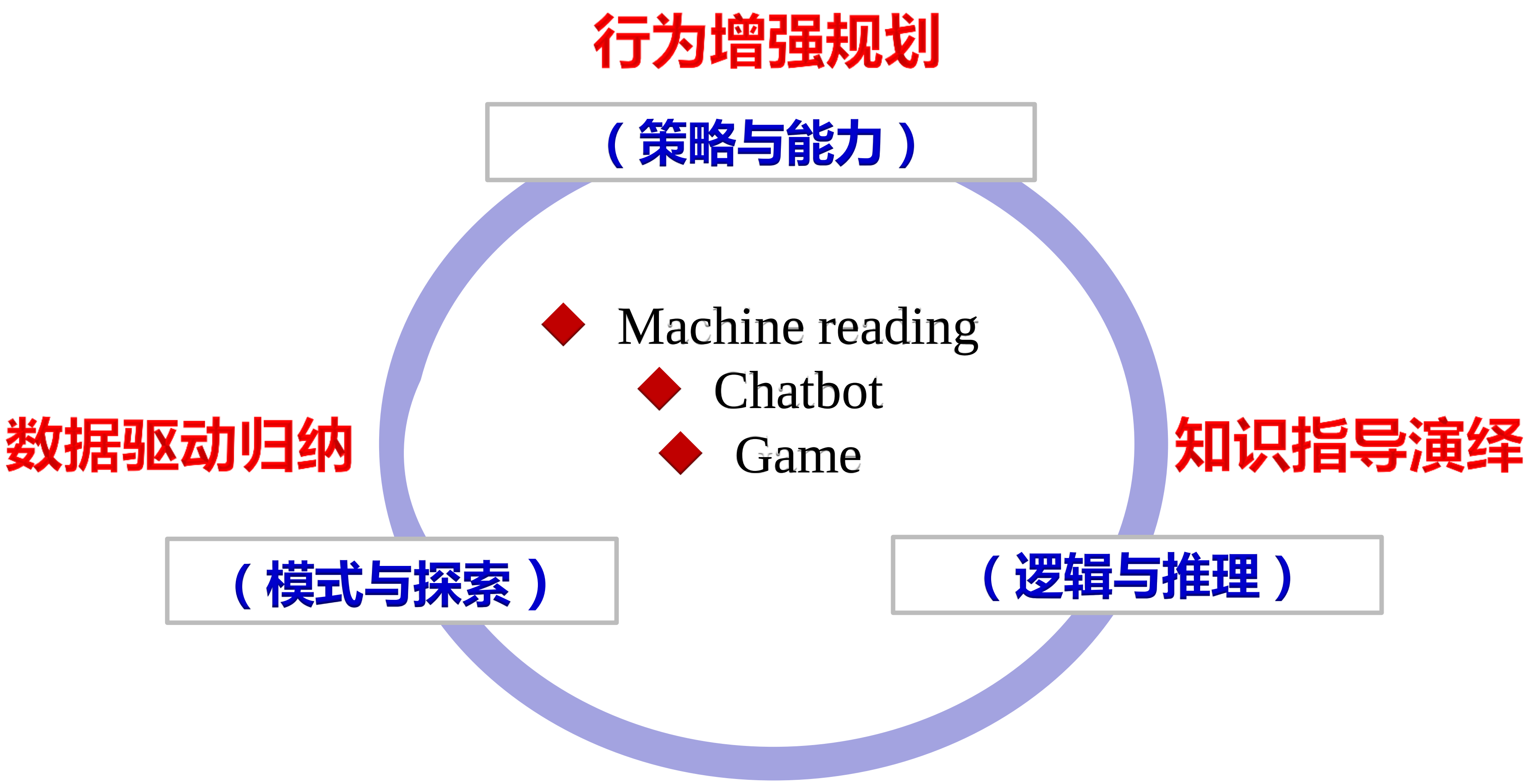
知识计算引擎技术框架

知识计算引擎: KS-Studio



在数据驱动机器学习中引入"众包数据"或"知识规则"，拓展单纯数据驱动的概念识别手段，建立解释性强的人工智能方法。

总结


行为增强规划
（策略与能力）
数据驱动归纳
（模式与探索）
知识指导演绎
（逻辑与推理）
Machine reading
Chatbot
Game

总结

实现可解释、更鲁棒和更通用的人工智能
（数据利用、知识引导与能力学习）

数据驱动的机器学习方法

先验与知识引导

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

谢谢大家

Email: *wufei@cs.zju.edu.cn*