记忆驱动的智能学习

吴　飞

浙江大学计算机学院

2018年4月22日

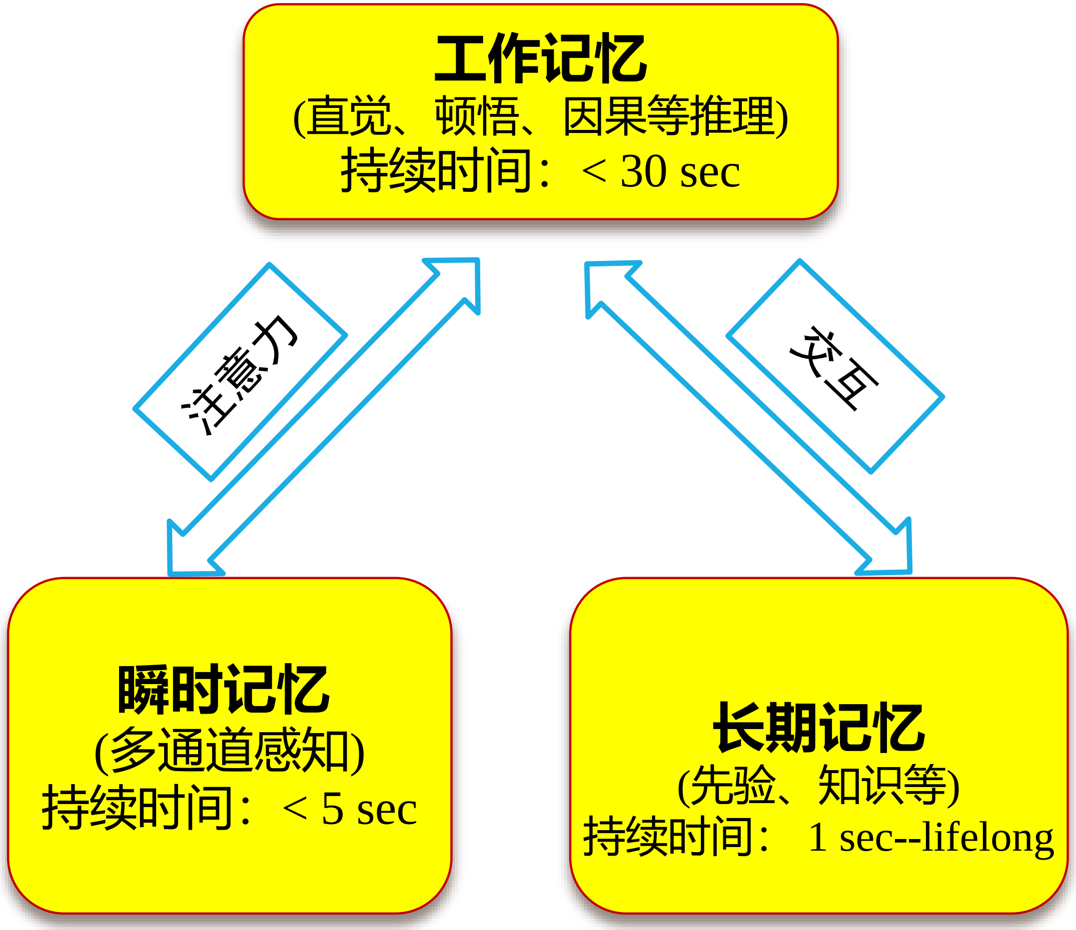
提纲

1、记忆是认知的基石

2、若干工作

3、总结

记忆是人类认知的基石

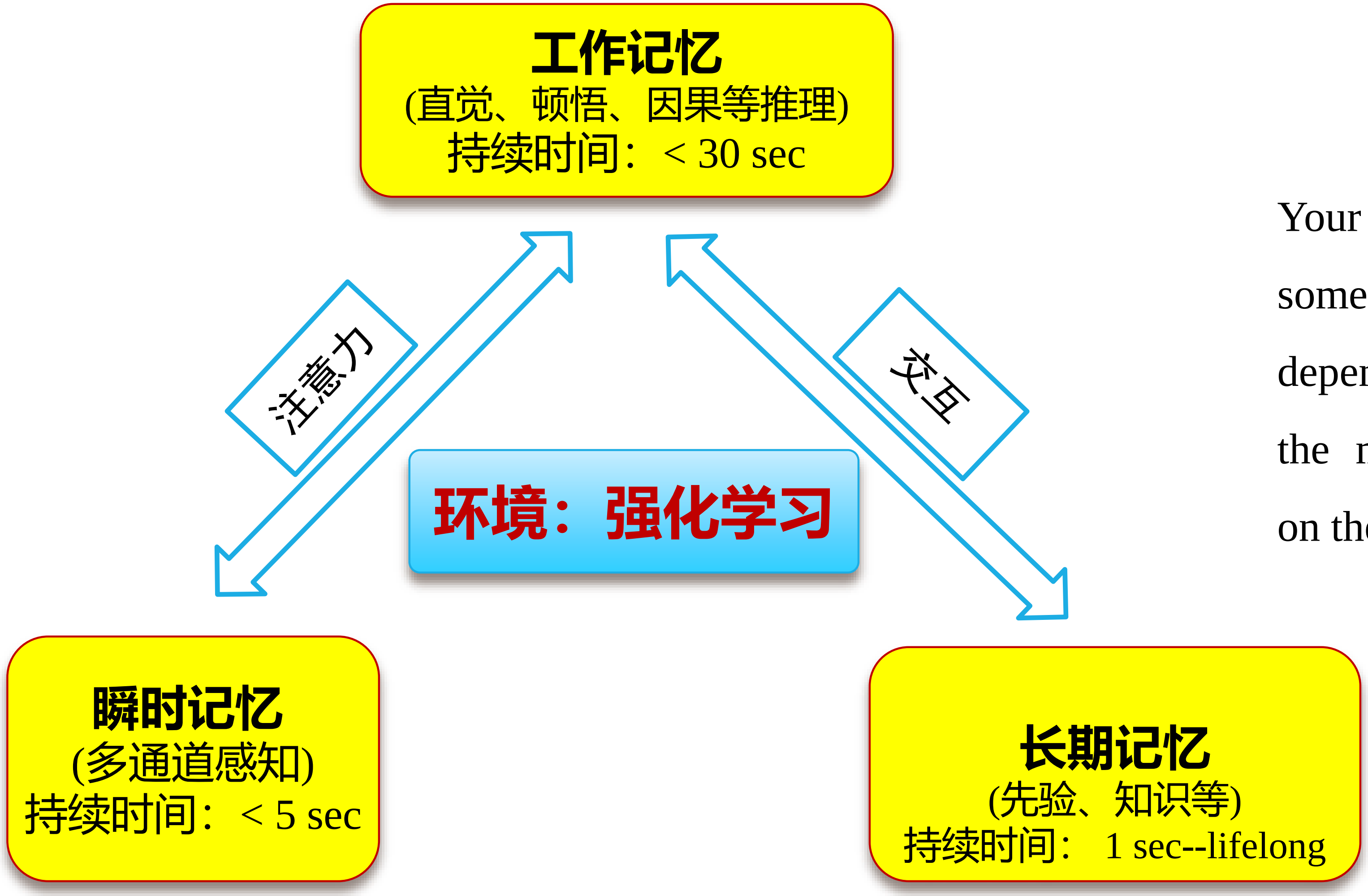


知之在人者谓之**知**	**知觉：**人所固有认识外界客观事物本能，如视觉、听觉和触觉等能力
知有所合谓之**智**	**智慧：**知觉对外界事物的认知
所以能之在人者为之**能**	**本能：**人身上所具用来处置事物能力
能有所合谓之**能**	**智能：**对外界所产生的认知和决策

《荀子．正名》

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

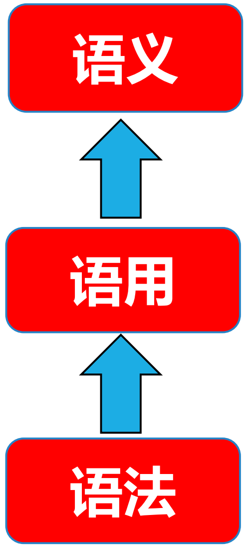
记忆单元之间及其与环境的交互是提升智能能力的重要途径

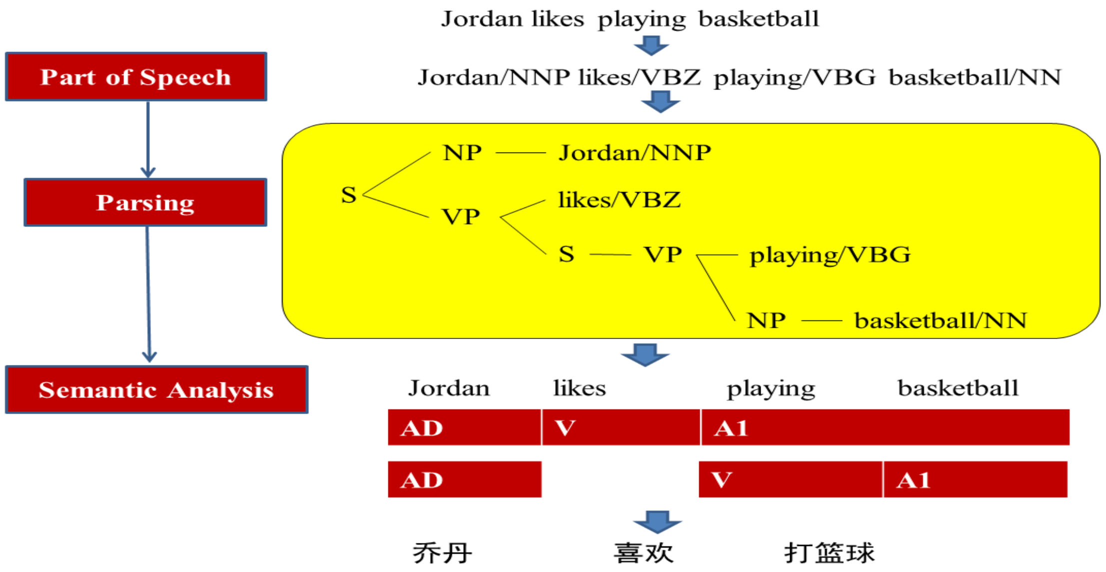


Your ability to remember something doesn’t just depend on the strength of the memory, it depends on the state that you’re in

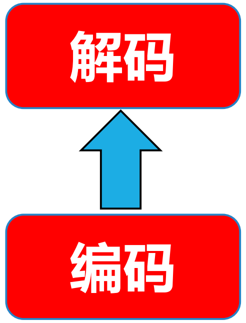
Human-level control through deep reinforcement learning, ***Nature***, 518:529–533, 2015

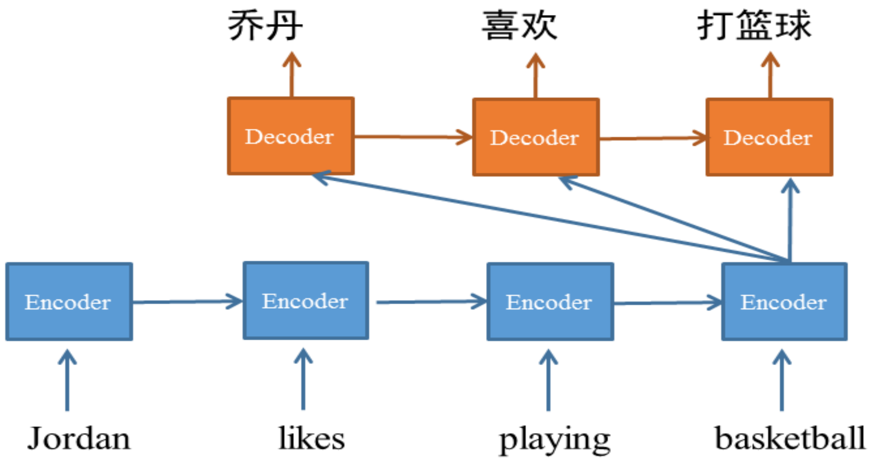
从分段学习到“端到端”学习：以自然语言翻译为例





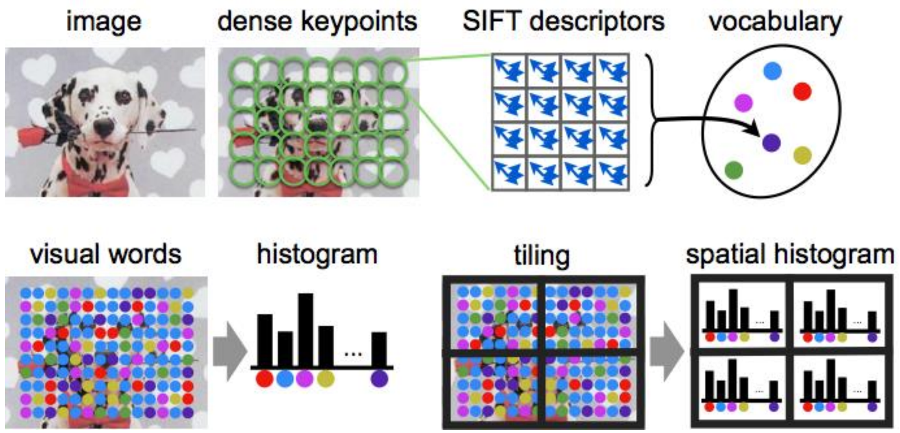
- Conditional random field
- Hidden Markov Model





- 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



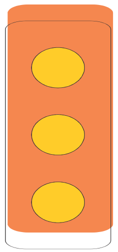
Detection → Description → **Classification** → Identification → **Understanding**

输入端

黑盒子

（多粒度语义在中间层丢失）

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

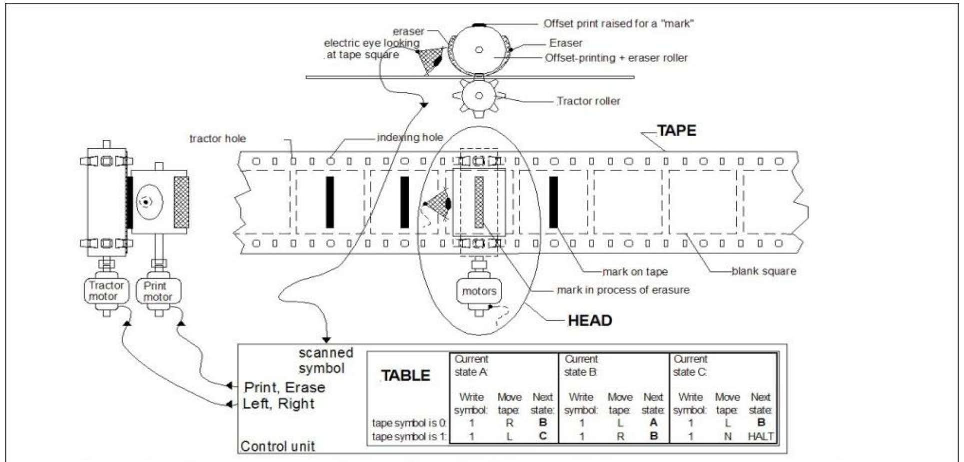
分段学习

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

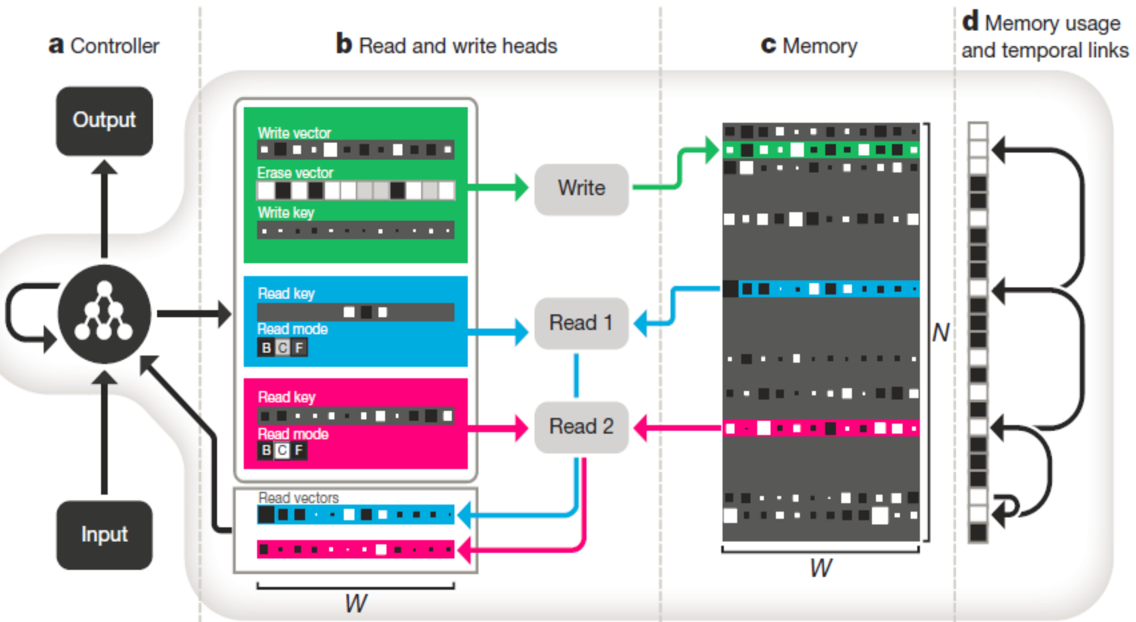


A fanciful mechanical Turing machine's TAPE and HEAD. The TABLE instructions might be on another "read only" tape, or perhaps on punch-cards. Usually a "finite state machine" is the model for the TABLE.

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

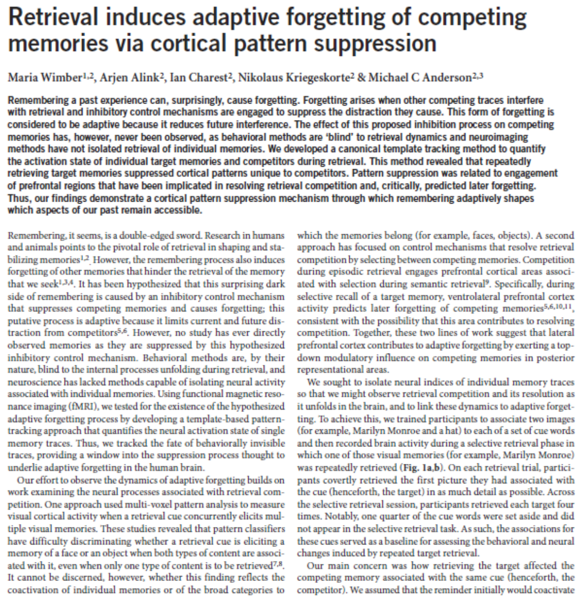
Most machine learning models lack an easy way to read and write to part of a (potentially very large) long-term memory component, and to combine this seamlessly with inference. Hence, they do not take advantage of one of the great assets of a modern day computer. For example, consider the task of being 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

Sanjoy Dasgupta,[1] Charles F. Stevens,[2,3] Saket Navlakha[2]*

Similarity search—for example, identifying similar images in a database or similar documents on the web—is a fundamental computing problem faced by large-scale information retrieval systems. We discovered that the fruit fly olfactory circuit solves this problem with a variant of a computer science algorithm (called locality-sensitive hashing). The fly circuit assigns similar neural activity patterns to similar odors, so that behaviors learned from one odor can be applied when a similar odor is experienced. The fly algorithm, however, uses three computational strategies that depart from traditional approaches. These strategies can be translated to improve the performance of computational similarity searches. This perspective helps illuminate the logic supporting an important sensory function and provides a conceptually new algorithm for solving a fundamental computational problem.



Retrieval induces adaptive forgetting of competing memories via cortical pattern suppression

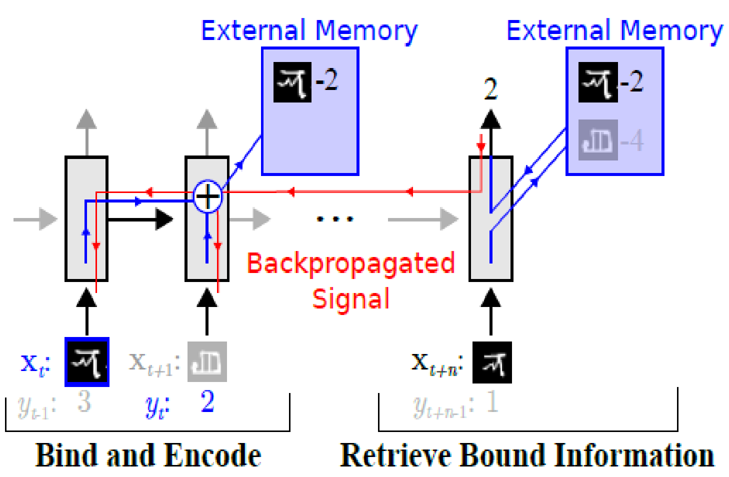
Maria Wimber[1,2], Arjen Alink[2], Ian Charest[2], Nikolaus Kriegeskorte[2] & Michael C Anderson[2,3]

Remembering a past experience can, surprisingly, cause forgetting. Forgetting arises when other competing traces interfere with retrieval and inhibitory control mechanisms are engaged to suppress the distraction they cause. This form of forgetting is considered to be adaptive because it reduces future interference. The effect of this proposed inhibition process on competing memories has, however, never been observed, as behavioral methods are 'blind' to retrieval dynamics and neuroimaging methods have not isolated retrieval of individual memories. We developed a canonical template tracking method to quantify the activation state of individual target memories and competitors during retrieval. This method revealed that repeatedly retrieving target memories suppressed cortical patterns unique to competitors. Pattern suppression was related to engagement of prefrontal regions that have been implicated in resolving retrieval competition and, critically, predicted later forgetting. Thus, our findings demonstrate a cortical pattern suppression mechanism through which remembering adaptively shapes which aspects of our past remain accessible.

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



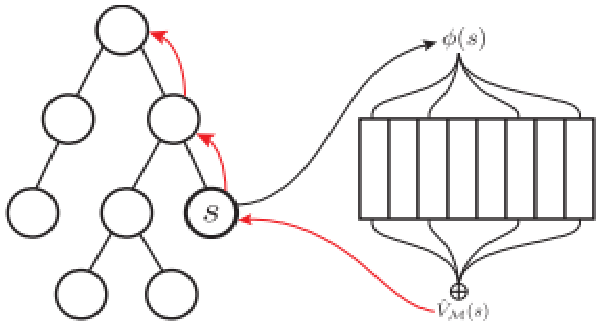


Figure 1: A brief illustration of M-MCTS. When a leaf state s is searched, the feature representation $\phi(s)$ is generated, which is then used to *query* the memory based value approximation $\hat{V}_{\mathcal{M}}(s)$. $\hat{V}_{\mathcal{M}}(s)$ is used to update s and all its ancestors according to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

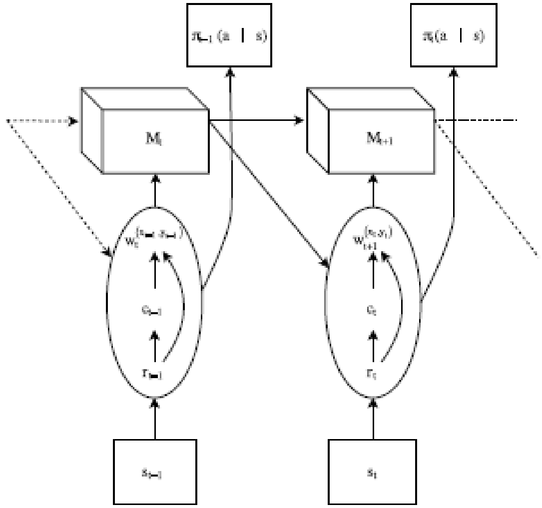


Figure 1: A visualization of two time steps of the neural map.

one shot learning

Monte Carlo Tree Search

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

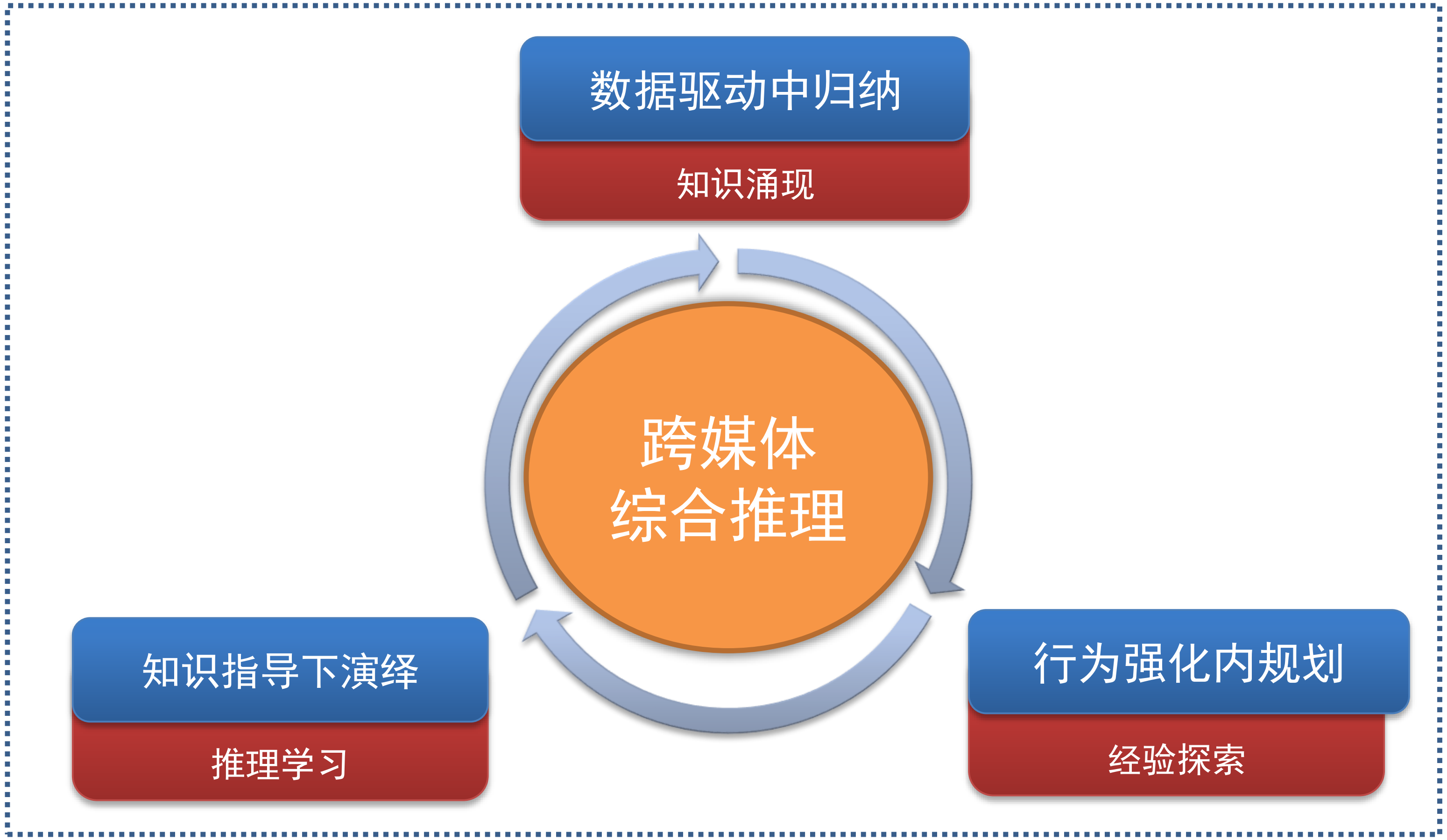
数据

记忆

交互

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

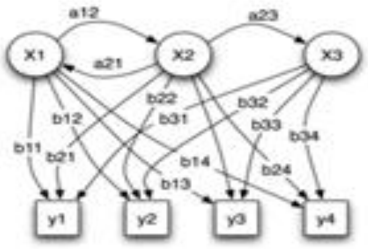


有机协调“知识指导下的演绎”、“数据驱动中的归纳”和“行为强化内的规划”等理论模型和方法手段，建立知识、数据和反馈于一体的人工智能理论和模型

跨媒体综合推理

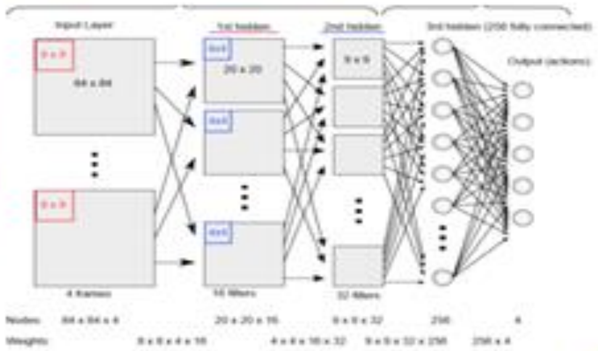
多个源头、多种数据、三向交互的综合推理机制

自顶向下



知识指导（推理）

自底向上



数据驱动（感知）

经验学习（自适应）

提纲

1、记忆是认知的基石

2、若干工作

3、总结

人工智能的基础科学问题

人工智能基础理论及应用(第194期双清论坛，2017.12.2-3,长沙)

1. 场景理解过程中的跨层次关联
2. 视觉知识表示和推理
3. 场景理解中的触类旁通能力的学习
4. 视觉与语言的交互机制
5. 脑机理认知
6. 类脑可计算模型
7. 计算架构与能力
8. 多源碎片化知识的表示
9. 多源碎片化知识推理与发现
10. 多源碎片化知识的适配
11. 人工智能的自动推理
12. 非合作博弈数学模型
13. 从记忆中提取答案
14. 无时不刻的预测
15. 基于常识的推理
16. 语言和知识
17. 对新问题如何提出解决方案
18. 开放动态环境中机器学习面临新的挑战：分布偏移、类别可增、属性变动、目标多样、环境变迁
19. 神经形态器件
20. 神经处理器
21. 神经计算机基础软件
22. 神经网络的记忆机制

194期双清论坛归纳整理后的科学问题

脑与认知

- 脑观测与脑认知
- 神经网络的记忆机制
- 类脑可计算
- 脑观测

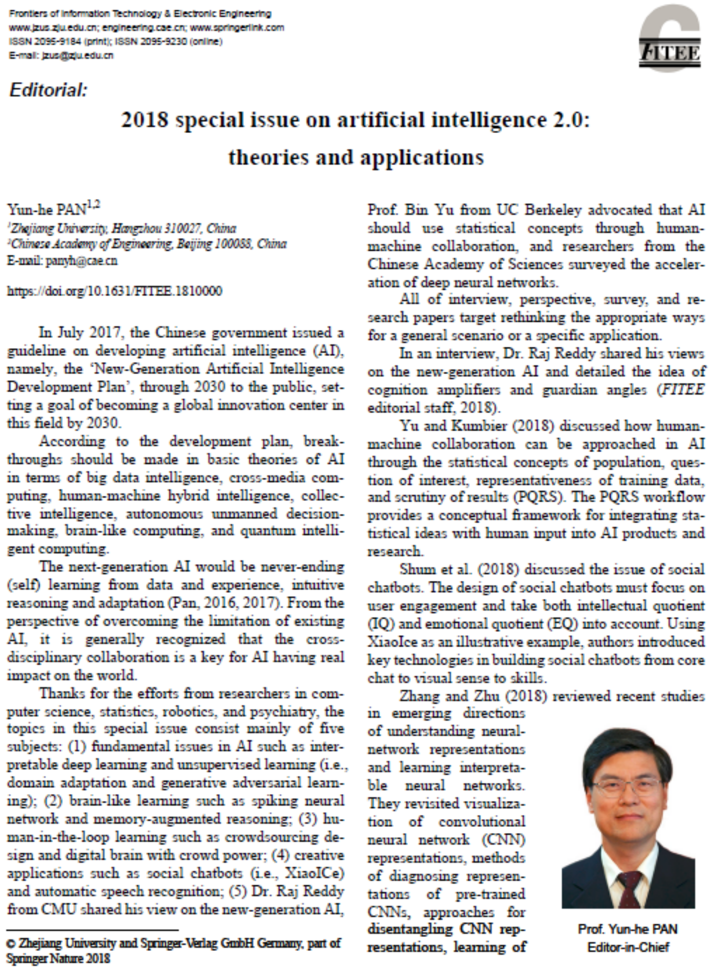
机器学习

- 知识表示与推理
- 动态环境下机器学习
- 记忆的学习方法
- 视觉、语言与学习

工具与平台

- 神经形态器件
- 神经处理器
- 神经计算机基础软件

2018人工智能2.0： 理论与应用



Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

FITEE

Editorial:

2018 special issue on artificial intelligence 2.0: theories and applications

Yun-he PAN[1,2]
[1]*Zhejiang University, Hangzhou 310027, China*
[2]*Chinese Academy of Engineering, Beijing 100088, China*
E-mail: panyh@cae.cn

https://doi.org/10.1631/FITEE.1810000

In July 2017, the Chinese government issued a guideline on developing artificial intelligence (AI), namely, the 'New-Generation Artificial Intelligence Development Plan', through 2030 to the public, setting a goal of becoming a global innovation center in this field by 2030.

According to the development plan, breakthroughs should be made in basic theories of AI in terms of big data intelligence, cross-media computing, human-machine hybrid intelligence, collective intelligence, autonomous unmanned decision-making, brain-like computing, and quantum intelligent computing.

The next-generation AI would be never-ending (self) learning from data and experience, intuitive reasoning and adaptation (Pan, 2016, 2017). From the perspective of overcoming the limitation of existing AI, it is generally recognized that the cross-disciplinary collaboration is a key for AI having real impact on the world.

Thanks for the efforts from researchers in computer science, statistics, robotics, and psychiatry, the topics in this special issue consist mainly of five subjects: (1) fundamental issues in AI such as interpretable deep learning and unsupervised learning (i.e., domain adaptation and generative adversarial learning); (2) brain-like learning such as spiking neural network and memory-augmented reasoning; (3) human-in-the-loop learning such as crowdsourcing design and digital brain with crowd power; (4) creative applications such as social chatbots (i.e., XiaoICe) and automatic speech recognition; (5) Dr. Raj Reddy from CMU shared his view on the new-generation AI, Prof. Bin Yu from UC Berkeley advocated that AI should use statistical concepts through human-machine collaboration, and researchers from the Chinese Academy of Sciences surveyed the acceleration of deep neural networks.

All of interview, perspective, survey, and research papers target rethinking the appropriate ways for a general scenario or a specific application.

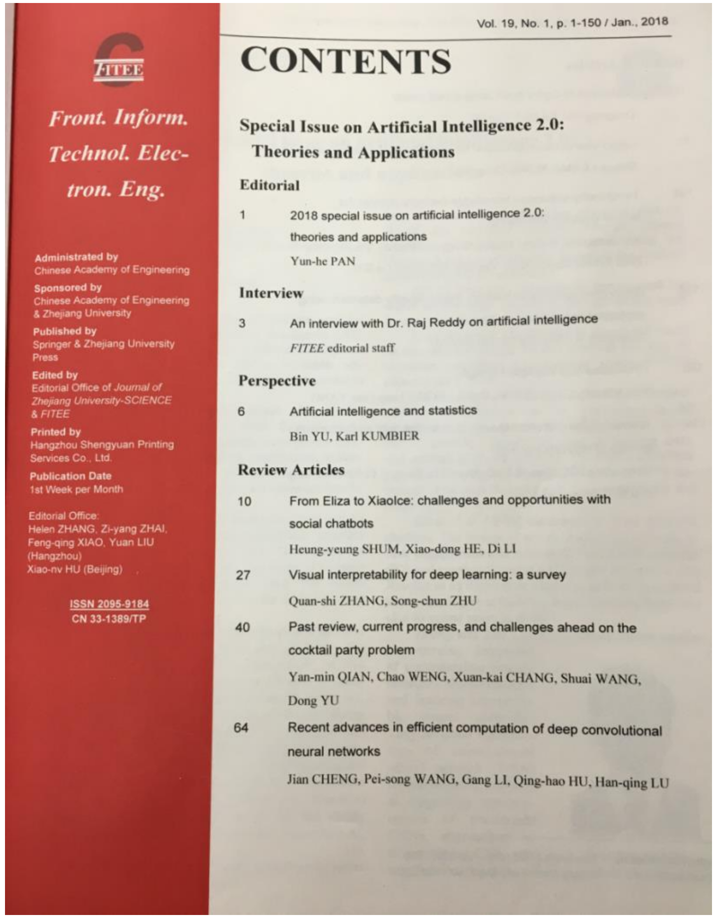
In an interview, Dr. Raj Reddy shared his views on the new-generation AI and detailed the idea of cognition amplifiers and guardian angles (*FITEE* editorial staff, 2018).

Yu and Kumbier (2018) discussed how human-machine collaboration can be approached in AI through the statistical concepts of population, question of interest, representativeness of training data, and scrutiny of results (PQRS). The PQRS workflow provides a conceptual framework for integrating statistical ideas with human input into AI products and research.

Shum et al. (2018) discussed the issue of social chatbots. The design of social chatbots must focus on user engagement and take both intellectual quotient (IQ) and emotional quotient (EQ) into account. Using XiaoIce as an illustrative example, authors introduced key technologies in building social chatbots from core chat to visual sense to skills.

Zhang and Zhu (2018) reviewed recent studies in emerging directions of understanding neural-network representations and learning interpretable neural networks. They revisited visualization of convolutional neural network (CNN) representations, methods of diagnosing representations of pre-trained CNNs, approaches for disentangling CNN representations, learning of

Prof. Yun-he PAN
Editor-in-Chief



FITEE

Front. Inform. Technol. Electron. Eng.

Administrated by
Chinese Academy of Engineering
Sponsored by
Chinese Academy of Engineering & Zhejiang University
Published by
Springer & Zhejiang University Press
Edited by
Editorial Office of *Journal of Zhejiang University-SCIENCE & FITEE*
Printed by
Hangzhou Shengyuan Printing Services Co., Ltd.
Publication Date
1st Week per Month

Editorial Office:
Helen ZHANG, Zi-yang ZHAI, Feng-qing XIAO, Yuan LIU (Hangzhou)
Xiao-nv HU (Beijing)

ISSN 2095-9184
CN 33-1389/TP

Vol. 19, No. 1, p. 1-150 / Jan., 2018

CONTENTS

Special Issue on Artificial Intelligence 2.0: Theories and Applications

Editorial

Interview

Perspective

Review Articles

- 人工智能基本理论问题，如可解释性深度学习和无监督学习
- 类脑学习，如脉冲神经网络和记忆增强推理
- 人在回路智能学习
- 创意智能应用，如社交聊天机器人（即小冰）和自动语音识别
- 卡耐基梅隆大学Raj Reddy博士分享了他对新一代人工智能的看法，加州大学伯克利分校郁彬教授主张在人机协作中使用统计概念以提升智能，中国科学院程健研究员等综述了深度神经网络加速方法。

时序增强的知识记忆网络在问答中的应用

Is ther a place where all sports bike enthusiasts meet? I've got a Triumph Daytona 675 and it gets lonely to drive with all the cars around.	Question
There may be in Villagio, I'm not sure.	Potential
Hello, best of luck, and congratulations on your Triumph Daytona 675.	Bad
Triumph Daytona 675 and lonely? Talk to Dani @ Al Fardan in Villagio. I know they organize rides on Fridays.	Good
Hey! So you are also into sports bike?	Bad
Villagio in the parking near the the carrefour gate on Fridays. I personally don't like riding with guys there because most of them are new to bikes. On the other hand, if you're looking for a real ride around, I will get my K6 in a month. Hit me up then and we'll go out with a couple of friends.	Good
Yup. Dani @ Al Fardan. He organizes rides on Fridays but not sure with the weather now... Ride Safe! Live to Ride, Ride to Live.	Good

Q-A问答中观点逐步凝练形成

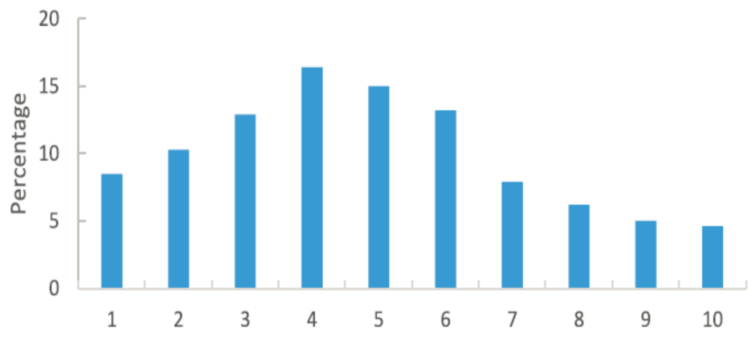


Fig. 2. The occurrence ratio of the best answer in the first tenth answers w.r.t a given question in our statistics on Baidu Zhidao data.

Q-A问答中用户之间复杂交互

能用众力，则无敌于天下矣；**能用众智，则无畏于圣人矣**(语出《三国志 吴志 孙权传》)

时序增强的知识记忆网络在问答中的应用

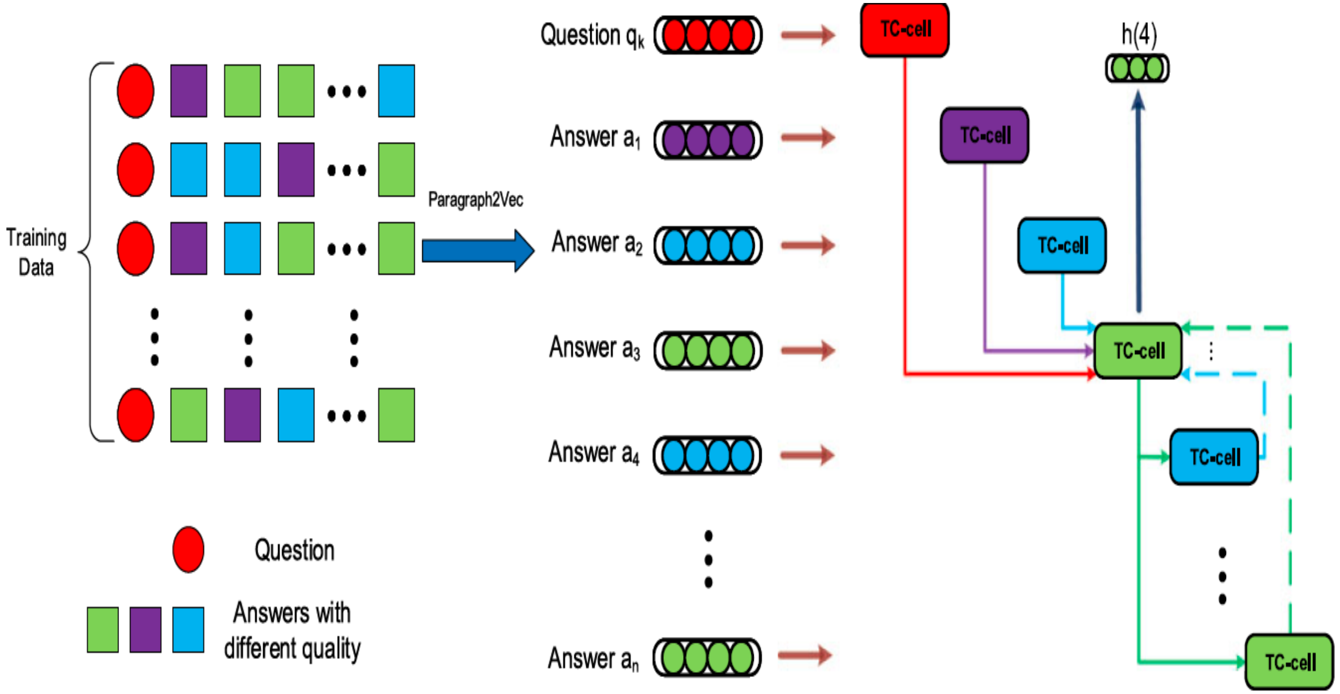


TABLE 4
The Precision Performance When the First Good Answer Occurring at the ith ($i = 1, 2, 3, 4, 5,$ or 6) Position is Identified as the *Good* Answer (%)

Method \ ith	1	2	3	4	5	6
SVM	78.16	50.60	32.89	13.16	9.38	7.41
CRF	77.54	50.23	33.84	13.63	9.56	7.68
TL	79.31	53.88	37.45	16.29	11.88	9.63
DBN	75.48	49.78	33.35	13.61	9.80	7.85
mDBN	76.08	50.87	34.64	14.19	10.32	8.28
CNN	76.32	51.80	35.79	15.14	11.04	9.10
R-CNN	77.03	52.21	36.02	15.24	11.20	9.30
QA-LSTM	78.10	51.54	34.27	14.39	10.40	8.64
MemNN	79.42	54.77	38.45	16.80	12.30	10.22
TC-LSTM	80.12	55.75	39.28	17.61	13.37	11.36

用户在凝练答案过程中不仅会理解问题本身，同时也会理解先前已有答案，之后在此基础上对自己所给出回答进行修正和补充，进而给出质量更高的答案。

时序增强的知识记忆网络在问答中的应用

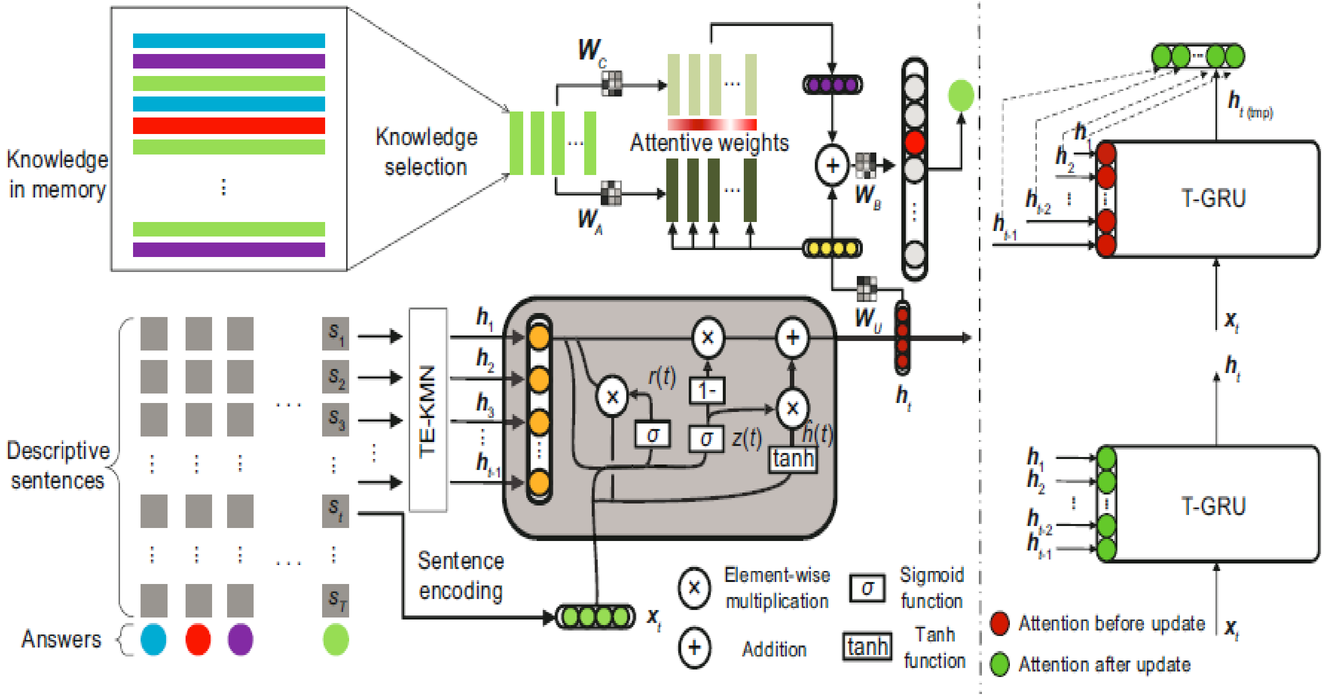
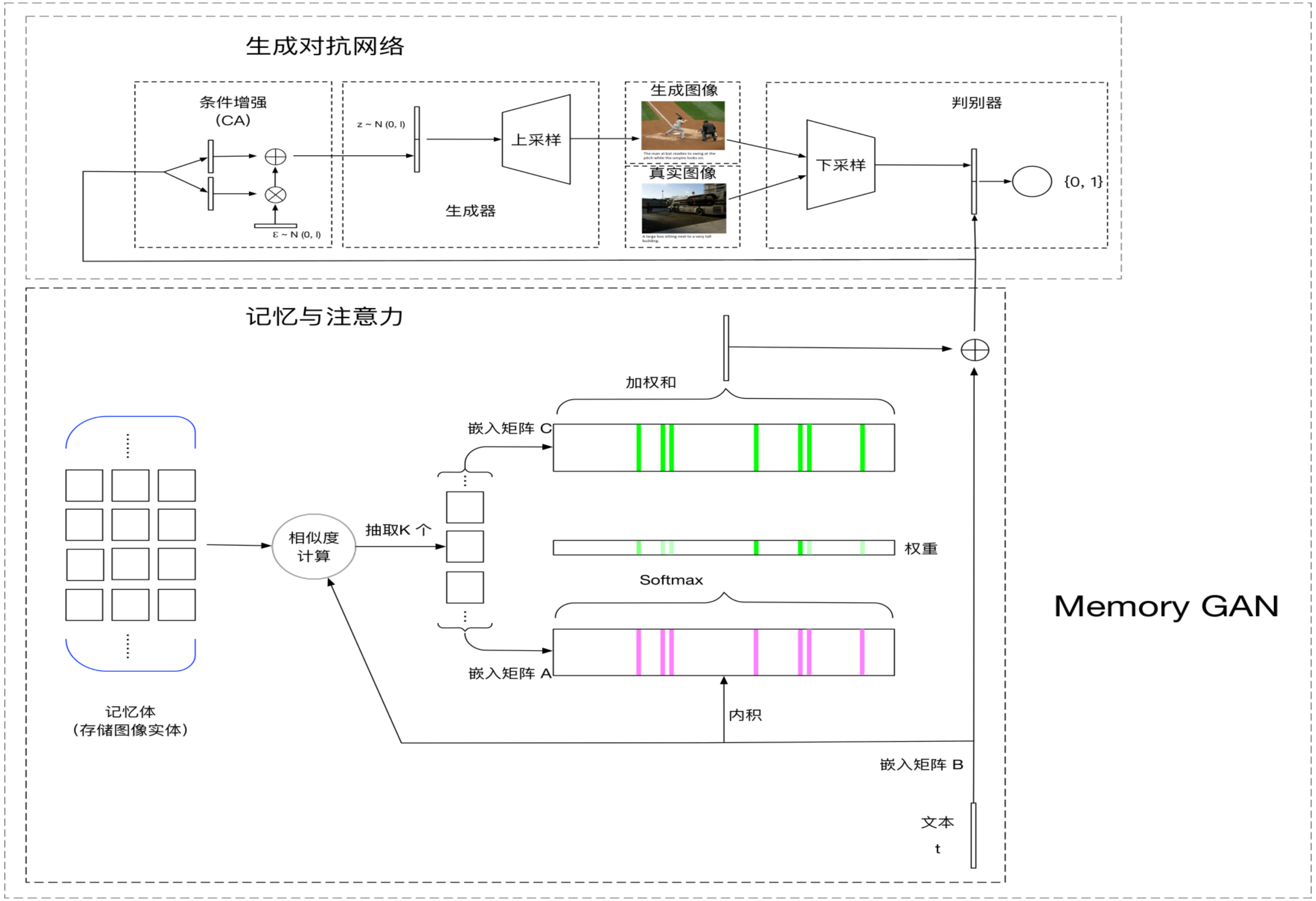


Table 5 Accuracy of obtaining the right answer when using different knowledge memories

Number of first sentences given	Accuracy (%)			
	TE-KMN^{-K}	TE-KMN^{wiki}	TE-KMN^{train}	TE-KMN
1	41.33	42.08	64.58	**66.07**
2	52.50	53.28	68.71	**70.77**
3	54.74	55.66	70.58	**72.91**
4	56.32	58.60	71.45	**73.94**
5	57.70	61.31	72.00	**74.46**

进一步引入知识记忆网络(knowledge memory network)，利用外部知识来加强问答学习性能

你说我画：从认知到创意



生成对抗网络
条件增强
(CA)
z ~ N (0, I)
上采样
生成器
ε ~ N (0, I)
生成图像
真实图像
判别器
下采样
{0, 1}
记忆与注意力
加权和
嵌入矩阵 C
相似度
计算
抽取K 个
权重
Softmax
嵌入矩阵 A
内积
记忆体
(存储图像实体)
嵌入矩阵 B
文本
t
Memory GAN

提纲

1、记忆是认知的基石

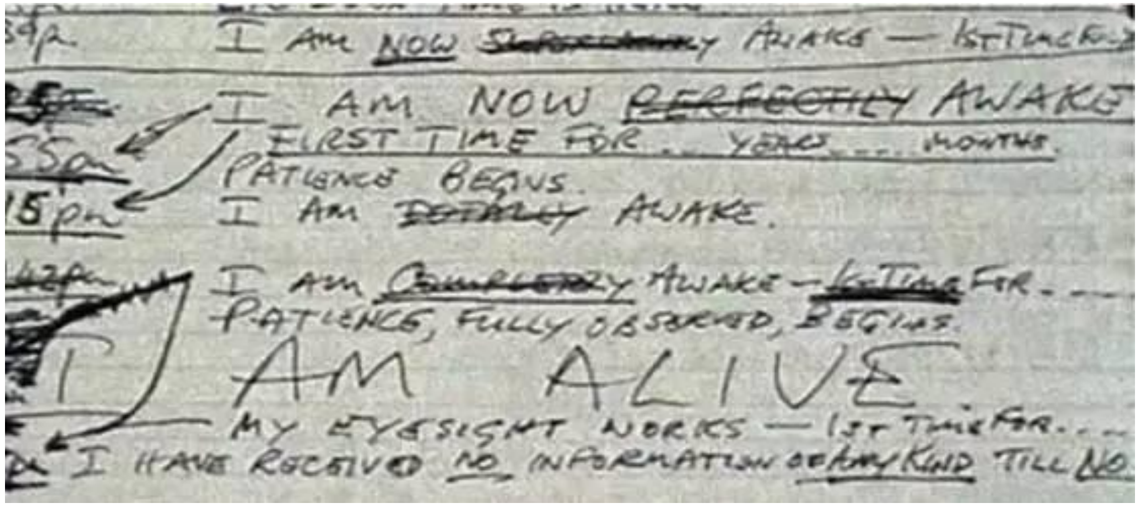
2、若干工作

3、总结

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



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。

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