

跨媒体群智计算

浙江大学计算机学院

吴飞

2016年8月29日

提纲

一、三元空间涌现群智

二、群智可计算思路

三、跨媒体异构序列深度学习

四、若干思考

973项目：面向城市管理的三元空间大数据计算理论 (2015.1-2019.11，首席科学家：朱文武教授)

三元空间城市大数据



事件和主题

互联网信息

百度每天搜索100亿；
处理数据量100PB

空间位置信息

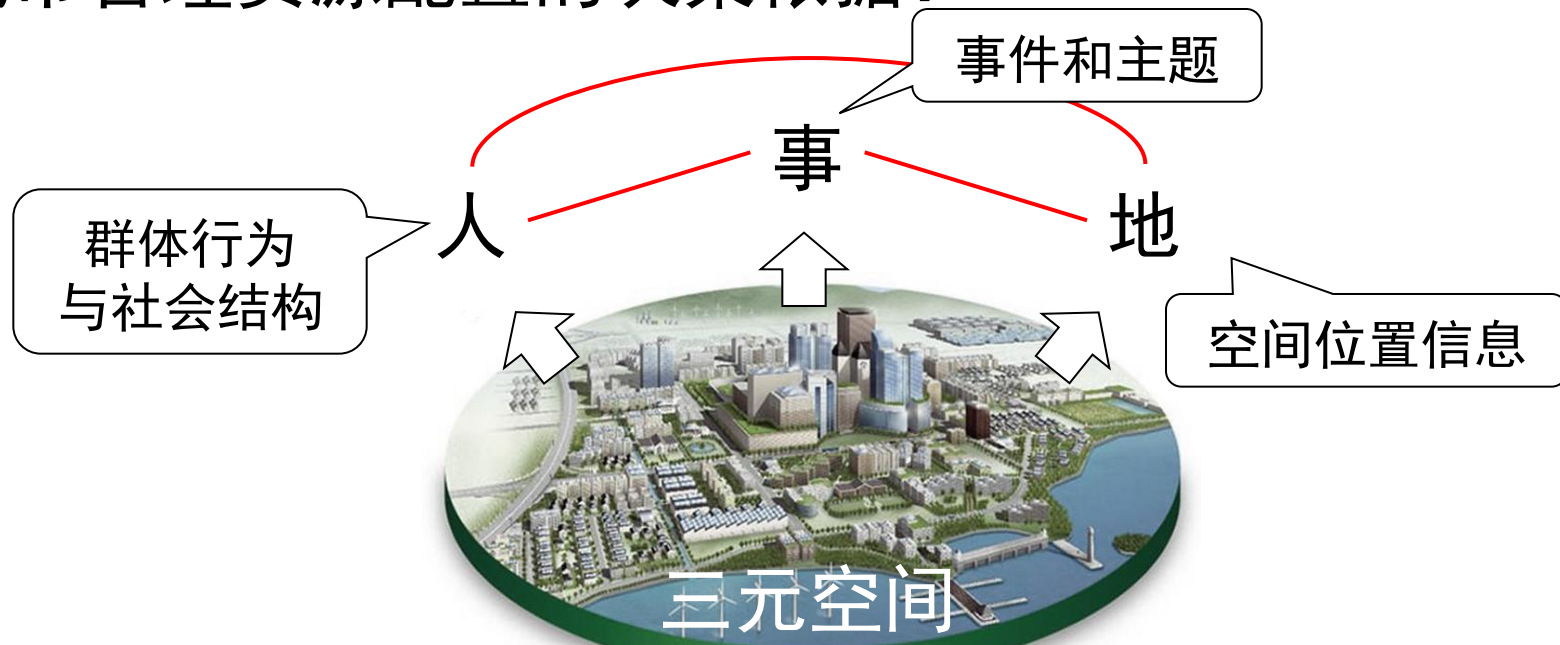
空间位置信息

增加1亿GPS设备

融合的三元空间城市大数据是
城市运行体征的全面量化表现

三元空间大数据将带来城市管理的巨大变革

城市管理资源配置的决策依据：



打通三元空间的数据孤岛，
实现三元空间城市大数据“人-事-地”的关联分析

传统方法



数据驱动的机器学习方法

Machine learning: trends, perspectives, and prospects, *Science*, 2015

优点	问题与不足
拓展性强、若干领域取得显著进展(以深度学习为代表)	依赖(高质量)数据及可解释性差(黑盒子, 与贝叶斯比较)

以涌现数据为核心



群智感知与认知方法

The power of crowds, *Science*, 2016

优点	问题与不足
解释性强(以社会学模型为代表, 如弱链接)	可计算性差

以用户交互为核心

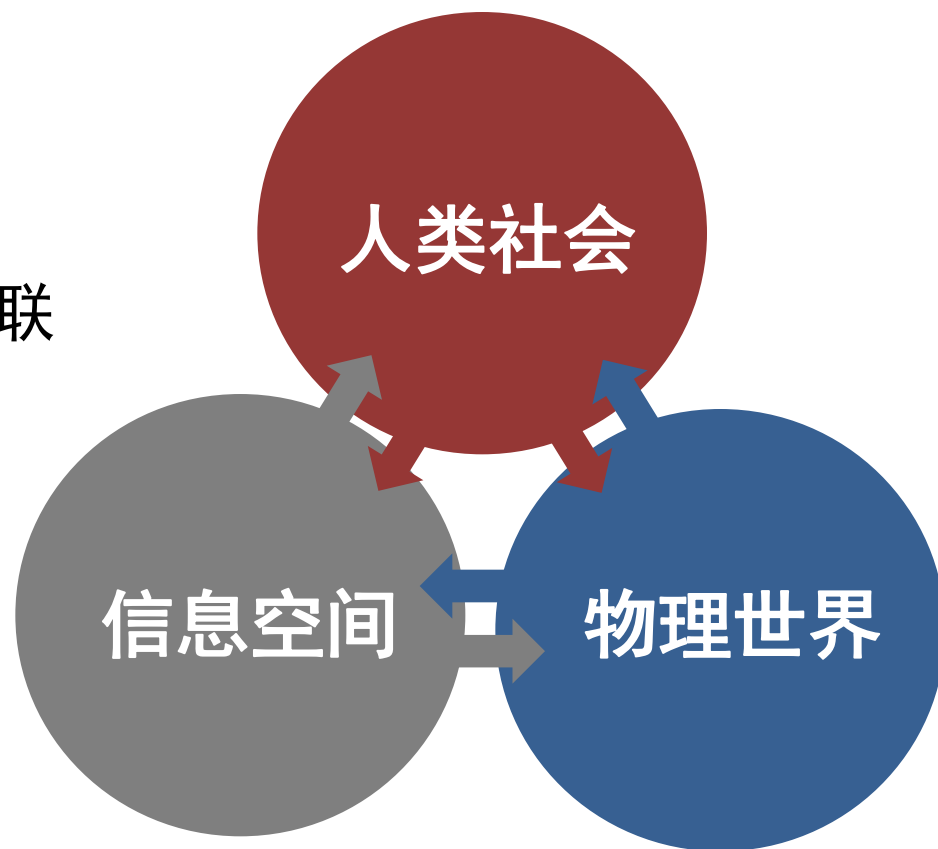
核心问题

异构
关联

建立三元空间人-事-地关联

知识
发现

理解三元空间知识规律

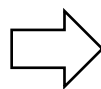


揭示三元空间关联映射与知识生成机理

科学问题分析

挑战一：关联复杂性

多元异构关联复杂



科学问题1： 三元空间
异构数据的关联表征

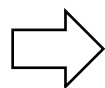


如何对缺乏数据驱动的信息表征理论进行表征？

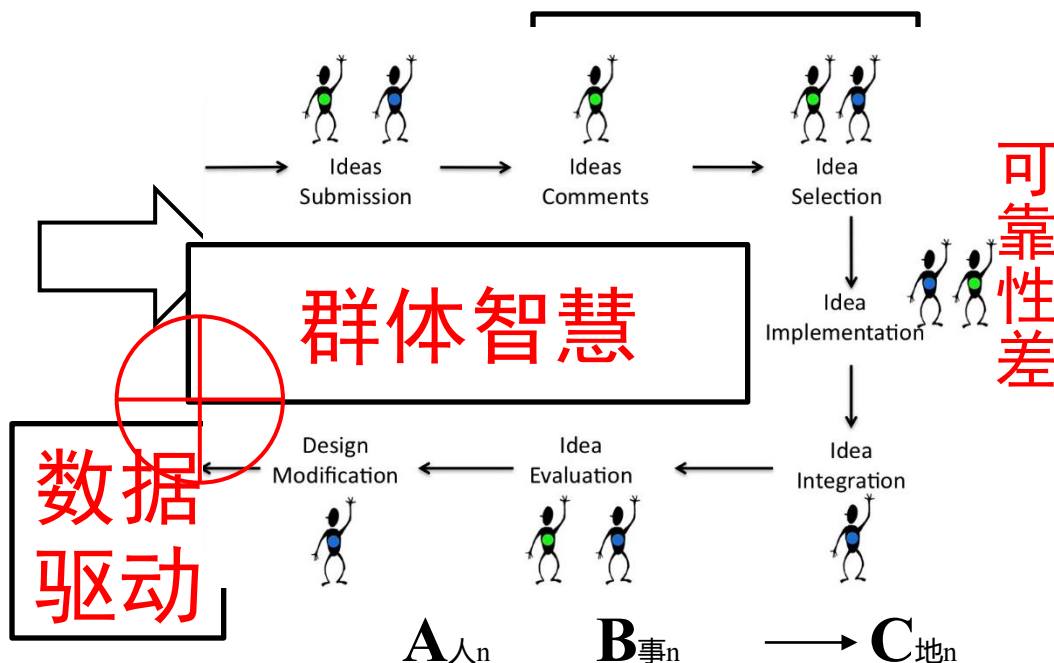
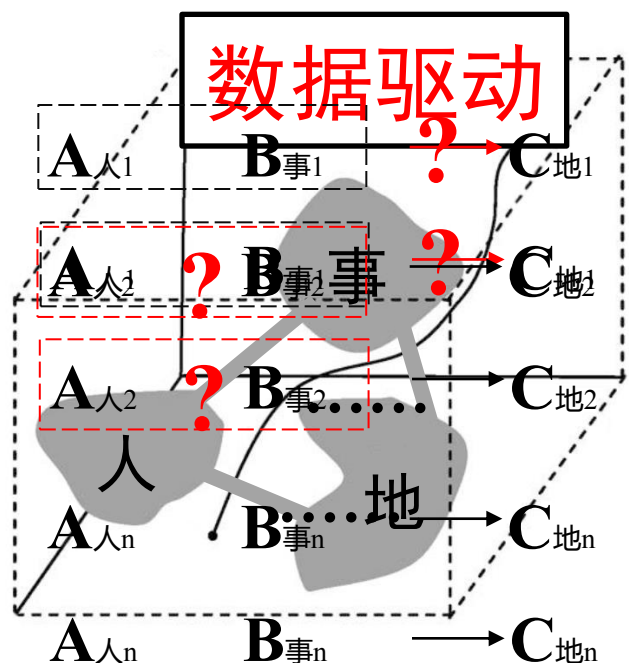
科学问题分析

挑战二：认知复杂性

三元空间数据不确定



科学问题2：数据驱动与群体智慧协同的知识发现机理

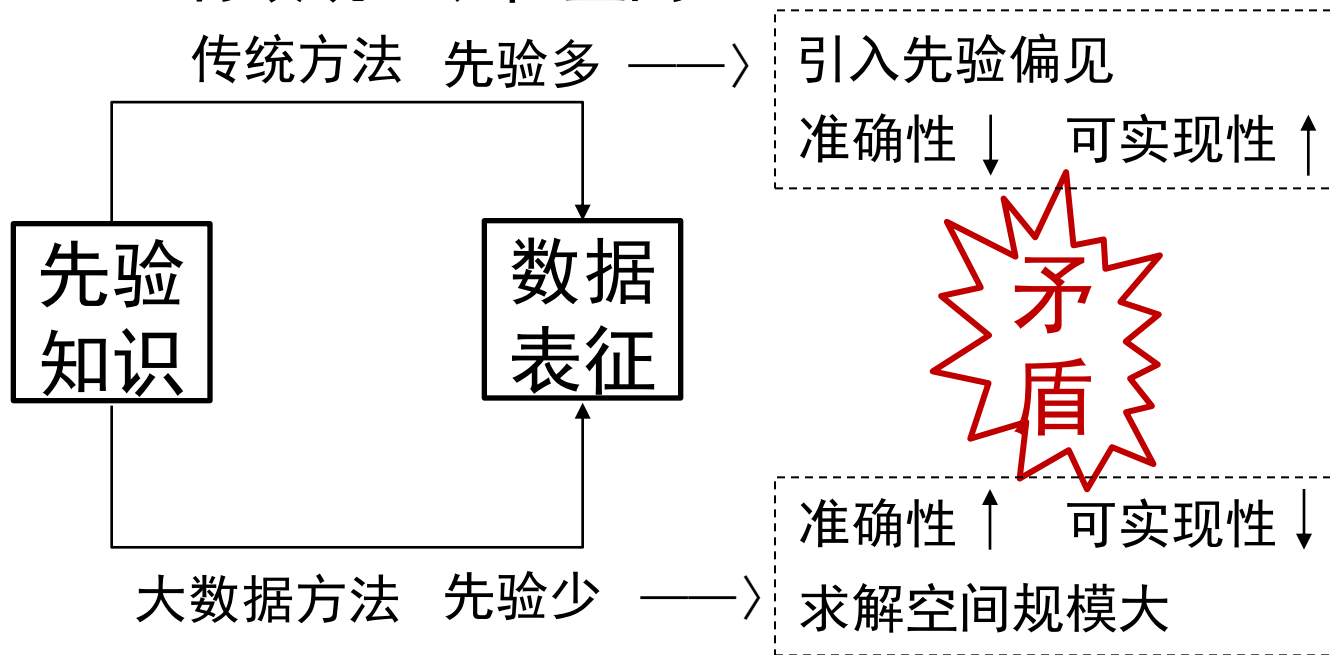


缺乏数据驱动与群体智慧相结合的知识发现理论

科学问题1: 三元空间异构数据的关联表征

难题

如何突破先验局限，从数据中学习出三元空间的有效统一表征空间？

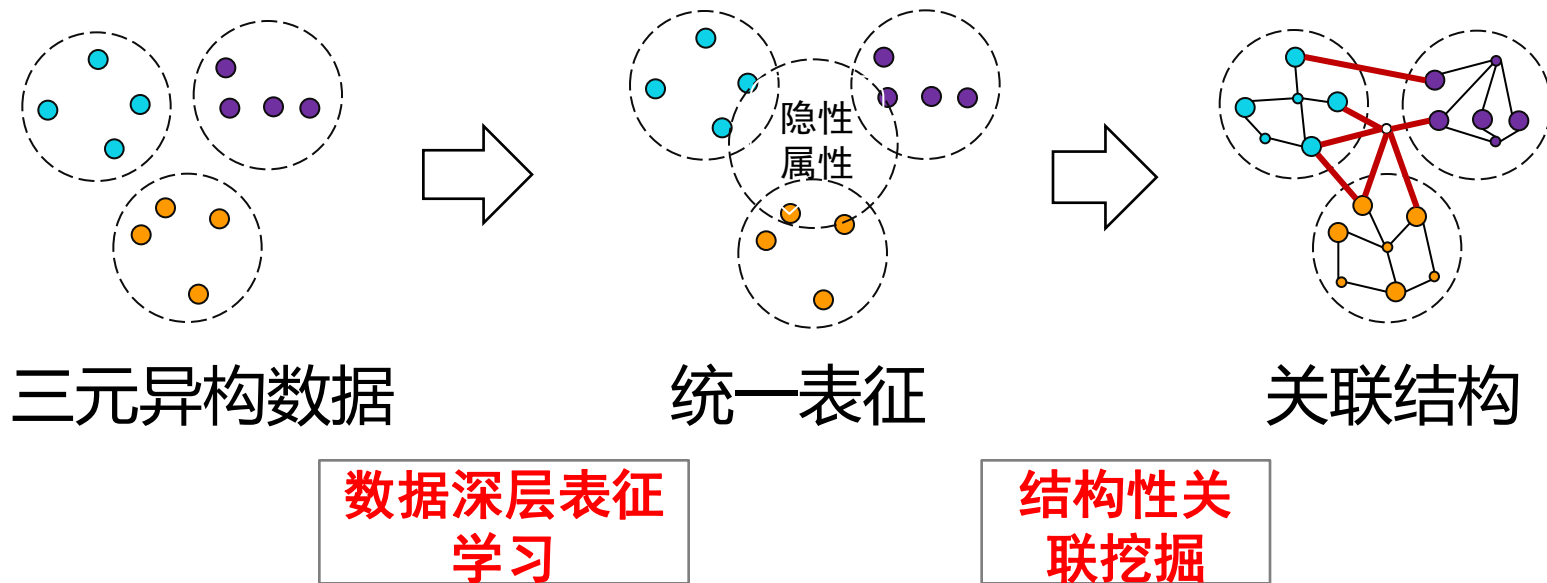


如何在数据表征中减少先验知识，
并保证超大规模求解空间的可实现性？

科学问题1: 三元空间异构数据的关联表征

研究思路

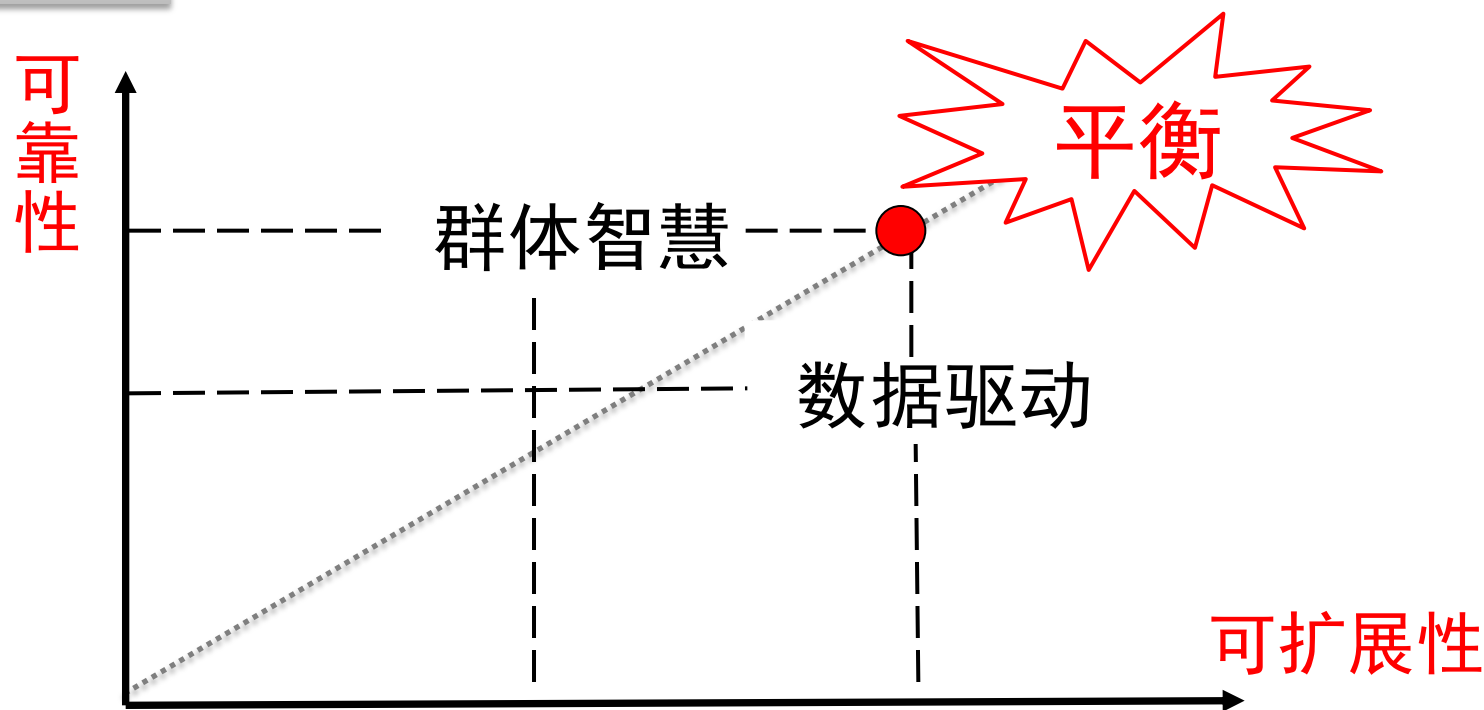
数据驱动的三元异构数据深层表征学习与关联计算



科学问题2：数据驱动与群体智慧协同的知识发现机理

难题

如何在不确定数据中发现可靠的知识？

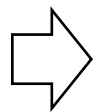
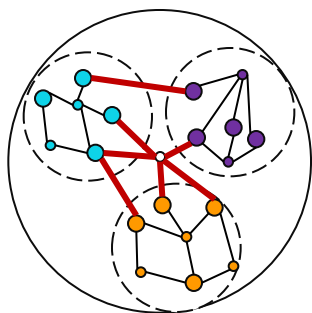


如何实现群体智慧与数据驱动方法的有机结合？

科学问题2：数据驱动与群体智慧协同的发现知识机理

方案 1. 群智协同模型

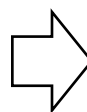
关联结构



子任务



子任务
分解设计



群智模型



$$f(x, y, z)$$

博弈与
竞争机制设计

课题1：三元空间异构数据的深层表征与关联计算

- 主要研究内容

- 三元空间数据的深层表征建模：基于深度计算理论，对多元数据统一特征空间进行建模

- 异构关联表达的深度计算：研究面向深度计算的大规模、非线性、非凸问题的最优化理论和方法以及并行计算框架

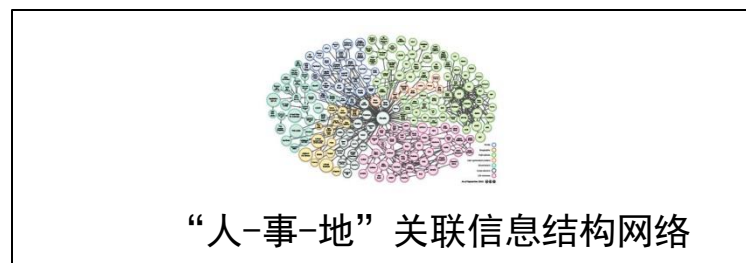
研究三元异构数据统一表征，建立“人-事-地”全局结构关联

- 三元空间结构关联构建：通过对数据关联性和全局一致性进行建模，构建多元数据的关联信息结构网络

- 负责人：朱文武

- 单位：清华大学 百度公司

研究方案



课题2：三元空间的群智计算

- 主要研究内容

- 群智协同模型：建立可计算社会模型来刻画个体/群体竞争和激励等机制。

研究群智协同模型，通过群体智慧与数据驱动相结合方法发现知识

- 群智与数据相结合的知识推理：建立数据驱动、人类知识引导和个体/群体交互约束的知识图谱构建新方法。

- 负责人：吴飞

- 单位：浙江大学 北京大学

研究方案



三元

空间

知识

图谱



课题3：面向城市管理的示范应用

- 主要研究内容

- 多尺度融合的态势分析：发掘三元空间的时空演化规律，实现时空态势的精准预测
- 人事地三元融合搜索：理解用户查询意图 合成答案

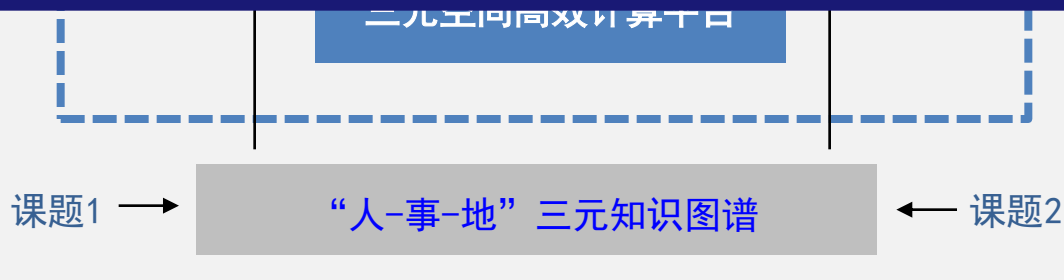
研究方案

示范验证结果



构建面向城市管理的应用平台，针对交通出行预测和流感预测开展技术验证和应用示范。

- 负责人：余凯(现：张潼)
- 单位： 百度公司 清华大学



提纲

一、三元空间涌现群智

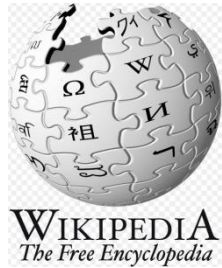
二、群智可计算思路

三、跨媒体异构序列深度学习

四、若干思考

不断涌现的众包数据感知三元空间

□ Voluntary (explicit)

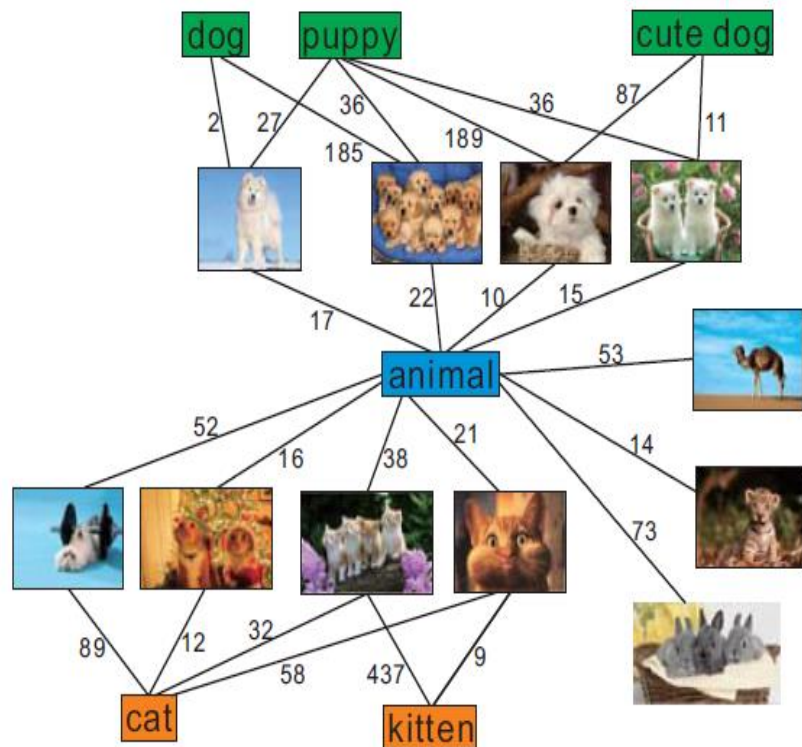


□ Incentive-based (explicit)



不断涌现的众包数据感知三元空间

□ Implicit crowded data



搜索引擎数据

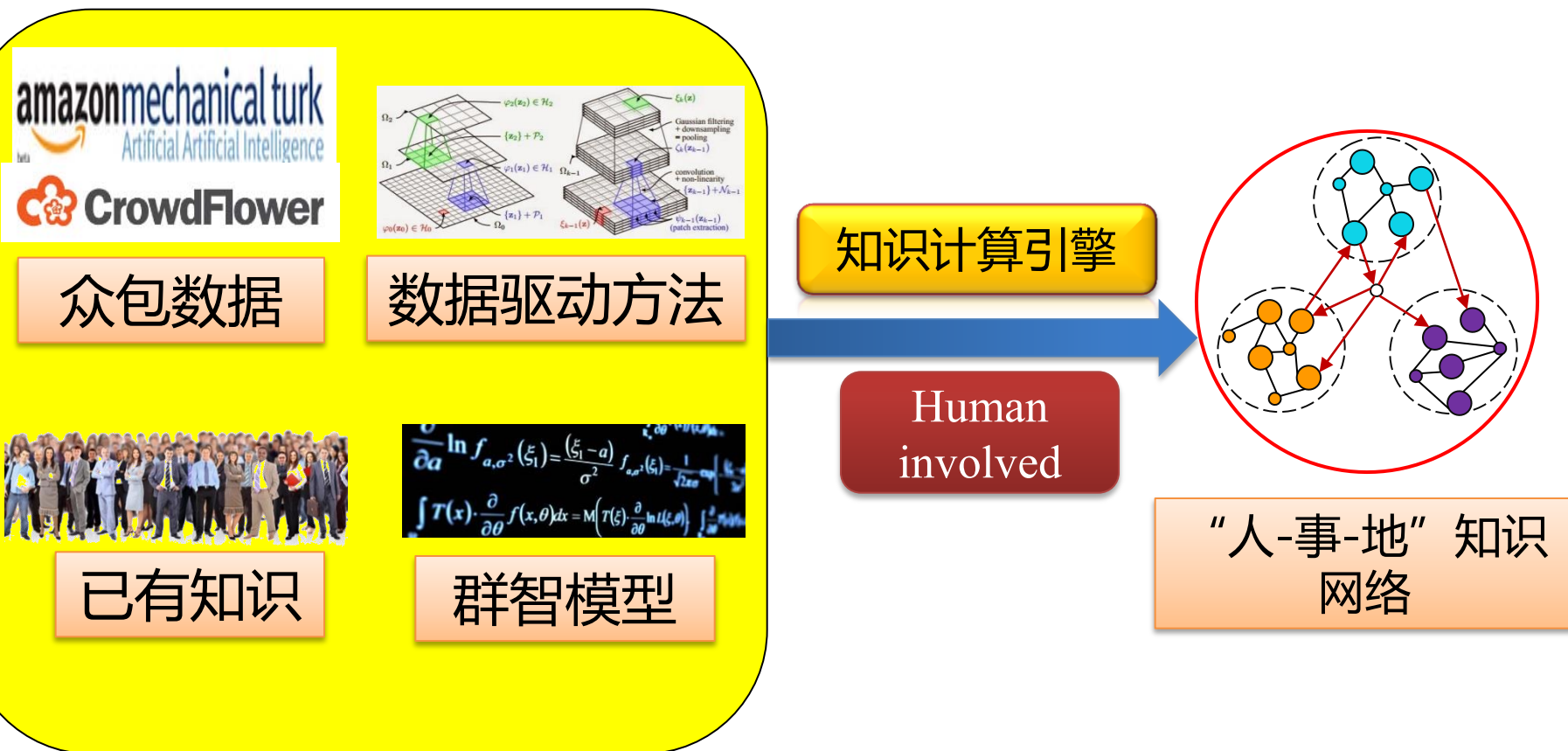
Microsoft Clickture Dataset



地图搜索数据(Baidu)

群智可计算

研究 “智能算法和群体智慧” 相结合的人机协同的知识挖掘方法。



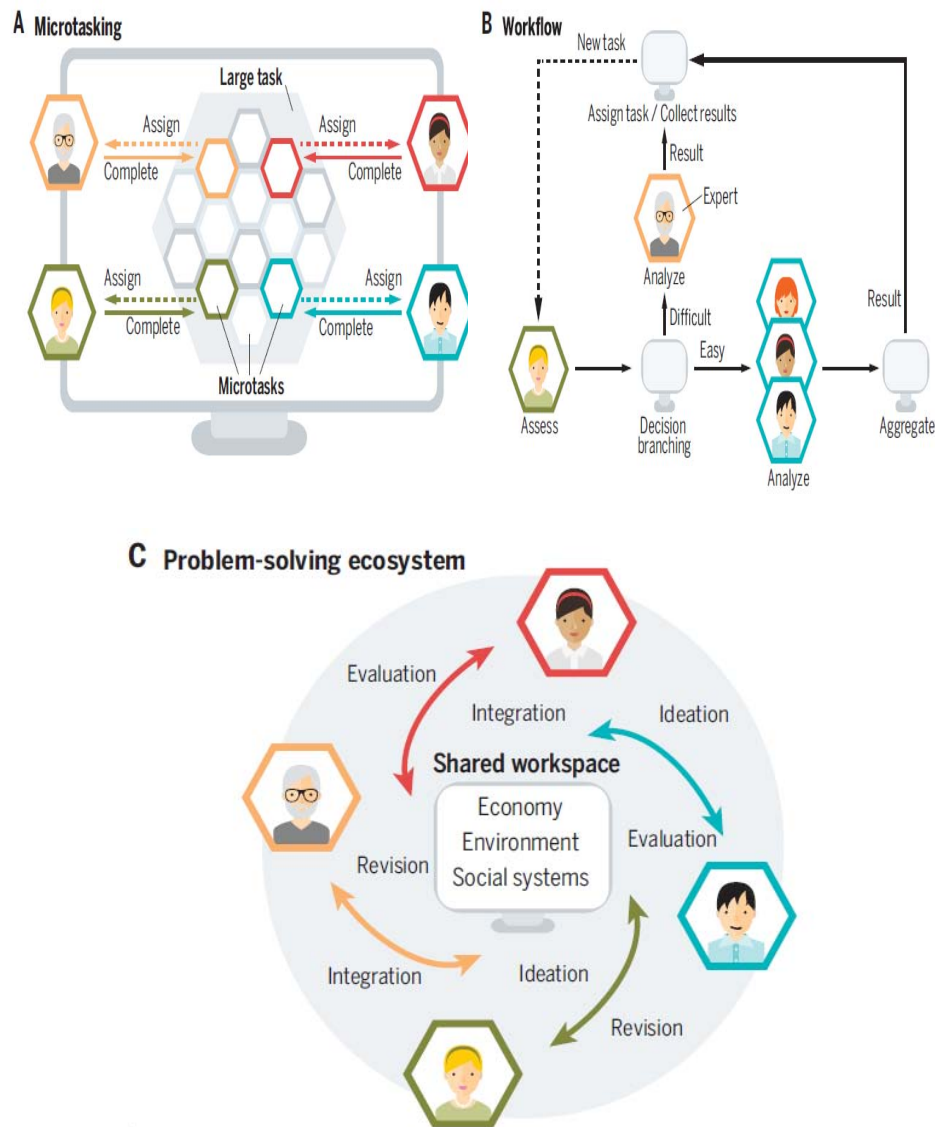
主要研究内容

- **群智协同模型：**建立可计算社会模型来刻画个体/群体竞争和激励等机制。
- **基于群智的知识生成：**研究高效众包策略，综合利用群体智力来挖掘知识。
- **群智与数据相结合的知识推理：**建立数据驱动、人类知识引导和个体/群体交互约束的知识图谱构建新方法。



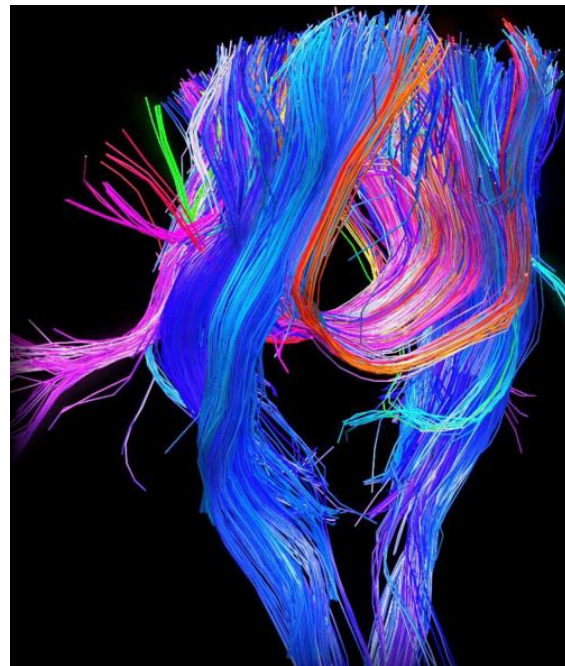
从个体智慧到群智使能

- 《Science》2016年1月1日发表“群智之力量(The Power of Crowds, Vol.351, Issues 6268)”的论文认为：结合群体智慧与机器性能来解决快速增长难题。
- 其将群智计算按难易程度分为三种类型：实现任务分配的**众包模式(Crowdsourcing)**、较复杂支持**工作流模式的群智(Complex workflows)**以及最复杂的**协同求解问题的生态系统模式(Problem solving ecosystem)**。
- 能用众力，则无敌于天下矣；能用众智，则无畏于圣人矣



从个体智慧到群智使能

- ❑ 大规模松散个体通过互联网进行显式或隐式的交互，在群体层次上表现出超越任一个体或传统组织的智能行为。
- ❑ 美国普林斯顿大学Connectome项目开发了EyeWire游戏，玩家对三维显微图像中单个细胞及其神经元连接进行涂色，完成复杂脑神经图的三维图像。第一次提供了哺乳动物视网膜的神经元结构和组织如何产生检测运动的功能。145个国家的165000多名公民科学家（玩家）的参与。
- ❑ 类似还有：Wiki百科、百度问答、知乎问答等等



美国普林斯顿大学Connectome项目
通过群体参与来探讨大脑结构和功
能之间映射的神经通路

<http://www.humanconnectomeproject.org/>

群智可计算

□ 群体之间的强弱链接产生群智

■ 异构序列数据的学习

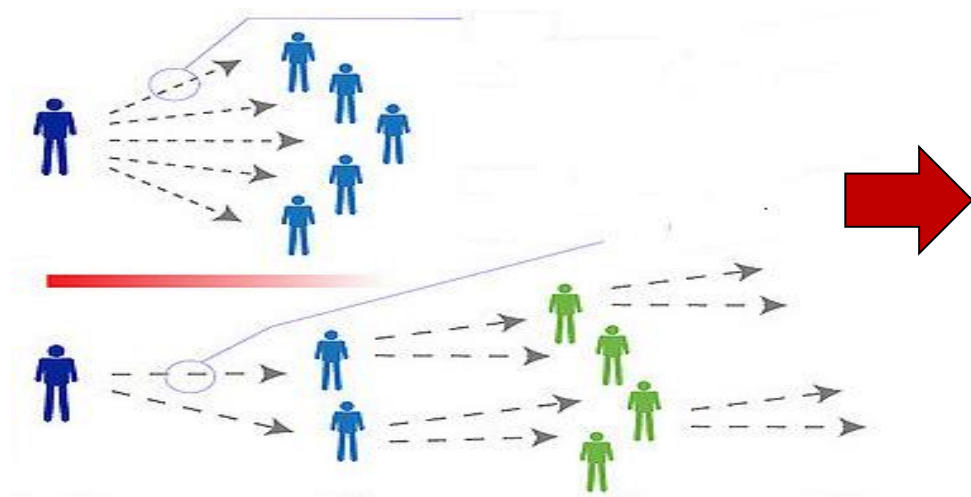
- 数据耦合：图像-文本-视频
- 用户链：用户-用户-用户



群智可计算

□ 群体之间的相互影响产生群智

■ 个体的激励、竞争和协作等



Crowd Workers



Q-A问答数据

群智可计算

□ 群体的差异性产生群智

- 个体感知差异产生群智：感知偏好和习惯等



个体感知差异性

群智可计算

- 三元空间中个体通过交互形式产生的众包数据反映了其对客观世界的理解，需要有效利用众包数据所蕴含个体直觉和经验来提升知识推理性能。

不同用户给出的图
文描述标注数据



During a gay pride parade in an Asian city, some people hold up rainbow flags to show their support.

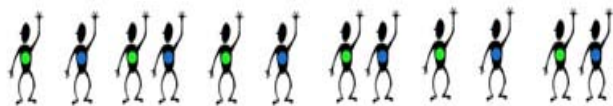
A group of youths march down a street waving flags showing a color spectrum.

Oriental people with rainbow flags walking down a city street.

A group of people walk down a street waving rainbow flags.

People are outside waving flags .

用户与
数据的
交互



个体在自然语言和视觉
感知过程中的差异

Baidu 知道

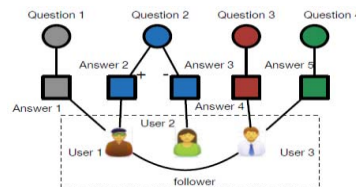
Quora

The best answer to any question

知乎

与世界分享你的知识、经验和见解

信息社区中不同
用户的问答数据



个体在答案凝练过程中的思想交互

群智可计算

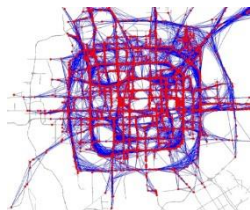
信息
空间



社交
空间



物理
空间



三元空间



三元空间异构序列数据学习

强弱链接产生群智

三元空间“数据-个人”交互建模

相互影响产生群智

三元空间群体差异性感知计算

差异累加产生群智

挑战难点

提纲

一、三元空间涌现群智

二、群智可计算思路

三、跨媒体异构序列深度学习

四、若干思考

跨媒体计算

- 互联网时代涌现的类型复杂多样、来源广泛的海量数据，呈现跨媒体特点：
 - 不同类型和不同来源的媒体数据之间相互耦合



跨媒体计算

□ **跨媒体计算**：本质是通过学习和推理，来实现从一种媒体类型到另外一种媒体类型的**跨越**。

- A woman wearing a black coat and sunglasses is standing in front of a man walk away with a backpack.



- A dog jumps up towards a woman in a car, while another dog is outside a car.



A white car is drifting. Cars racing on a road surrounded by lots of people. Cars are racing down a narrow road. A race car races along a rack. A car is drifting in a fast speed.

跨媒体检索：文本句子到图像

图文生成：视频到文本描述句子

跨媒体计算面临的新挑战

- *Nature* (463:26-32,2010)、*Science* (334:1502-1503,2011): 对存在强弱耦合的不同类型媒体数据进行关联学习是智能计算面临巨大科学挑战。



跨媒体计算面临三个难点问题

□ 需突破三个难点问题，进行源头创新

异构鸿沟：

打通不同类型媒体之间相似关联

语义鸿沟

推理手段

难点问题



难点1：如何度量视觉、听觉和自然语言等异构信息之间的相似度

跨媒体计算面临三个难点问题

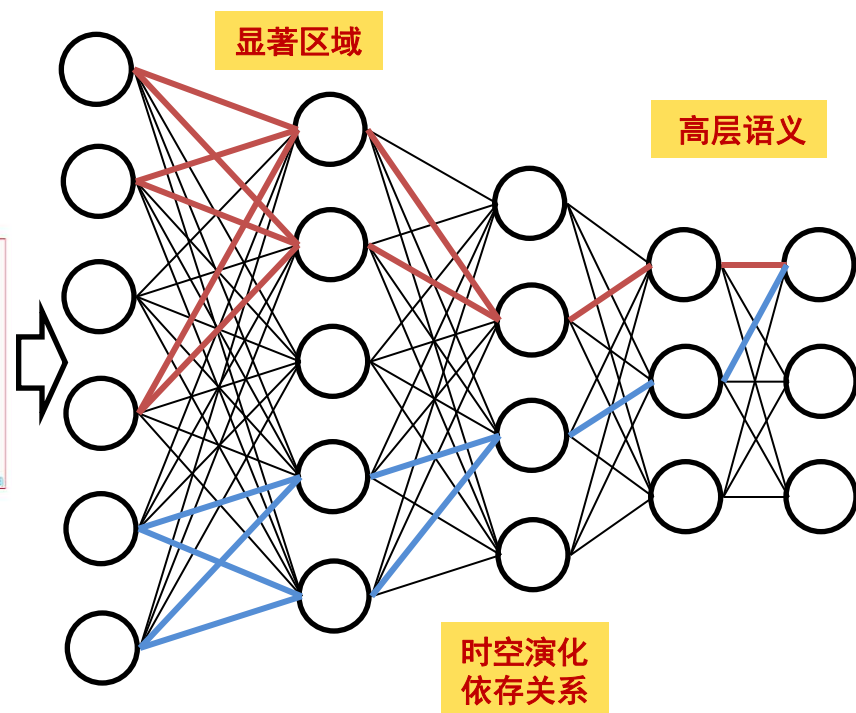
□ 需突破三个难点问题，进行源头创新

异构鸿沟

语义鸿沟：
从底层特征逐层抽象高层语义

推理手段

难点问题



难点2：如何在从底层特征到高层语义学习过程中协同多模态隐性结构

跨媒体计算面临三个难点问题

□ 需突破三个难点问题，进行源头创新

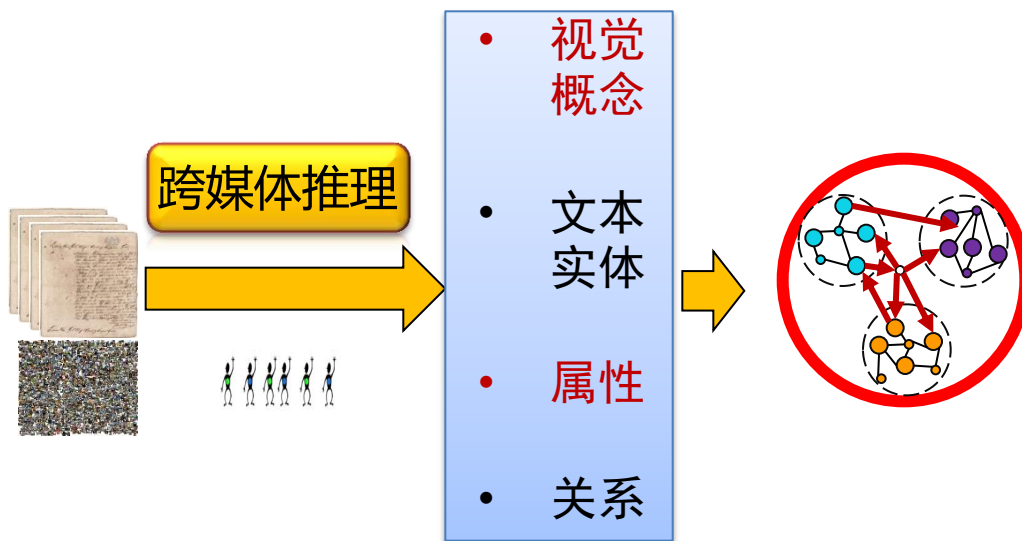
异构鸿沟

语义鸿沟

推理手段：

对跨媒体数据深度推理

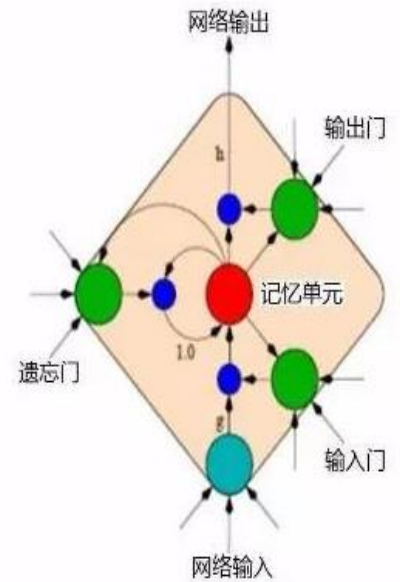
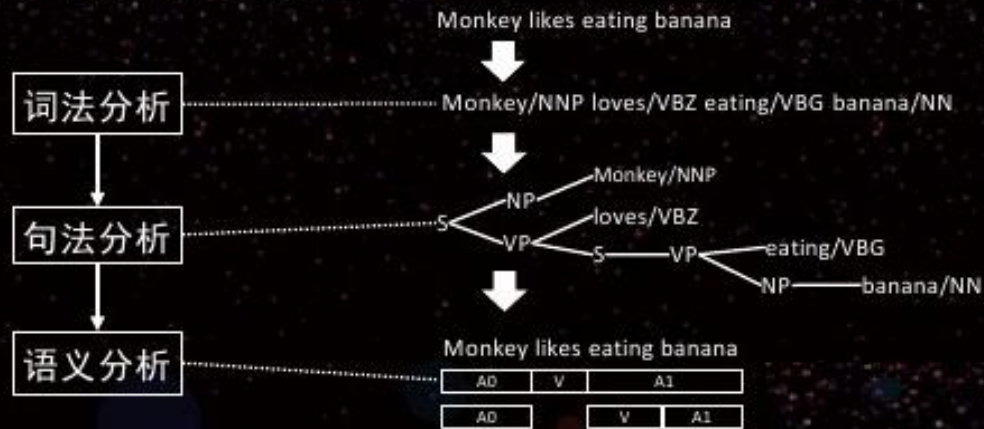
难点问题



难点3：如何将个体直觉和经验与机器计算优势相结合，拓展推理手段

(同构)端到端的序列学习

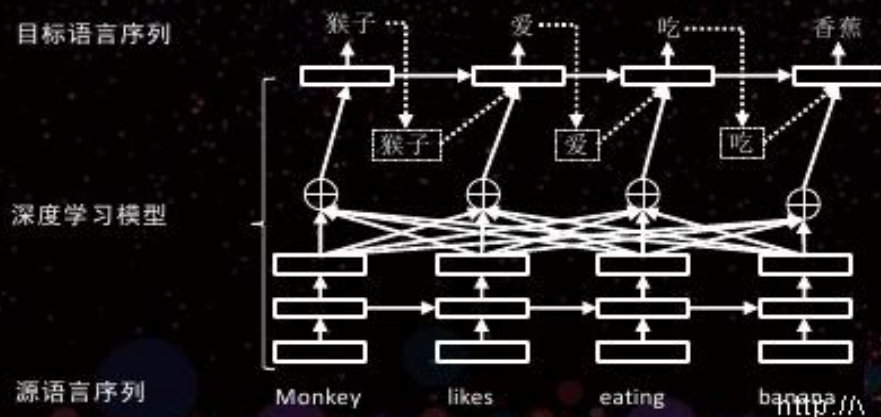
传统的语言理解系统



长短时记忆单元

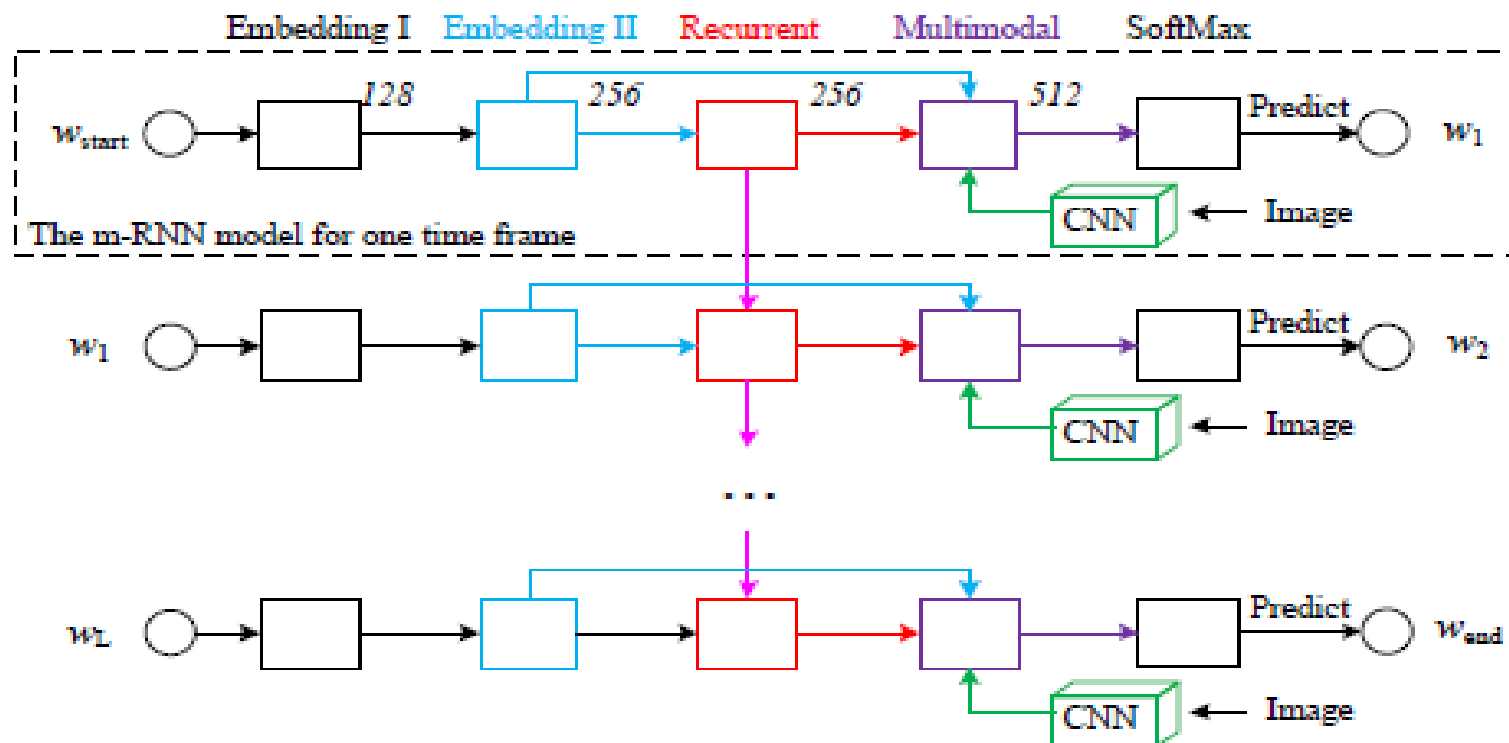
LSTM

端到端的机器翻译



(异构)端到端的序列学习

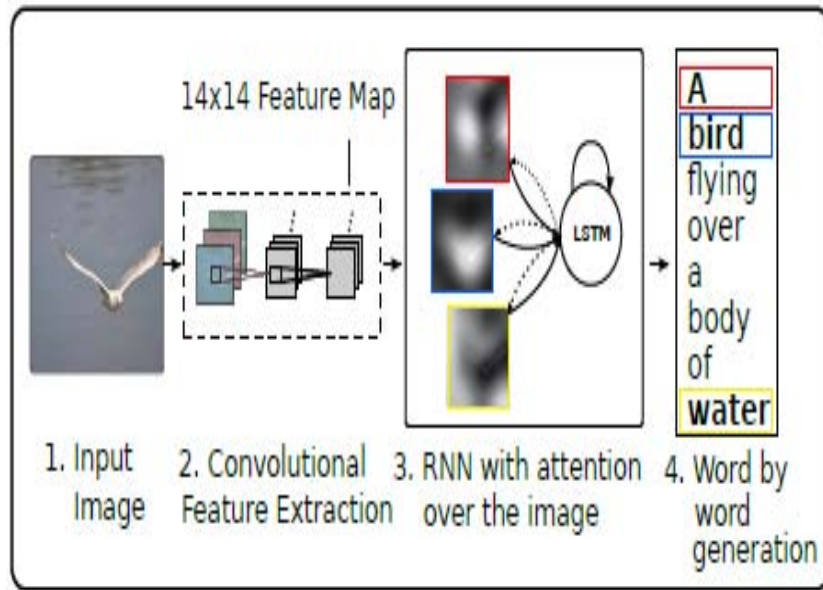
- LSTM: 在每次产生下一个单词时，既要利用当前单词信息，也要利用图像全部信息
- 学习 sentence descriptions 和 images 之间的 alignment



Mao, J.; Xu, W.; Yang, Y.; Wang, J.; Huang, Z.; and Yuille, A., Deep captioning with multimodal recurrent neural networks (m-rnn), ICLR 2015

(异构)端到端的序列学习

- LSTM with attentional model: 在每次产生下一个单词时, 既要利用当前单词信息, 也要利用图像中salient regions信息
- 学习sentence descriptions和salient regions之间的alignment



- Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel and Yoshua Bengio, Show, Attend and Tell: Neural Image Caption Generation with Visual Attention, ICML-15, 2048-2057
- Junqi Jin, Kun Fu, Runpeng Cui, Fei Sha, Changshui Zhang, Aligning where to see and what to tell: image caption with region-based attention and scene factorization

近期研究工作

三元空间

感知理解

- 深度随机游走
- 层次化上下文attention-LSTM
- 三元组深度学习模型

三元空间

群智建模

- 因果依赖与时序影响LSTM
- 群体差异GroupTalk模型

三元空间

知识图谱

- 众包深度特征命名体消歧
- 知识图谱嵌入学习与关系预测

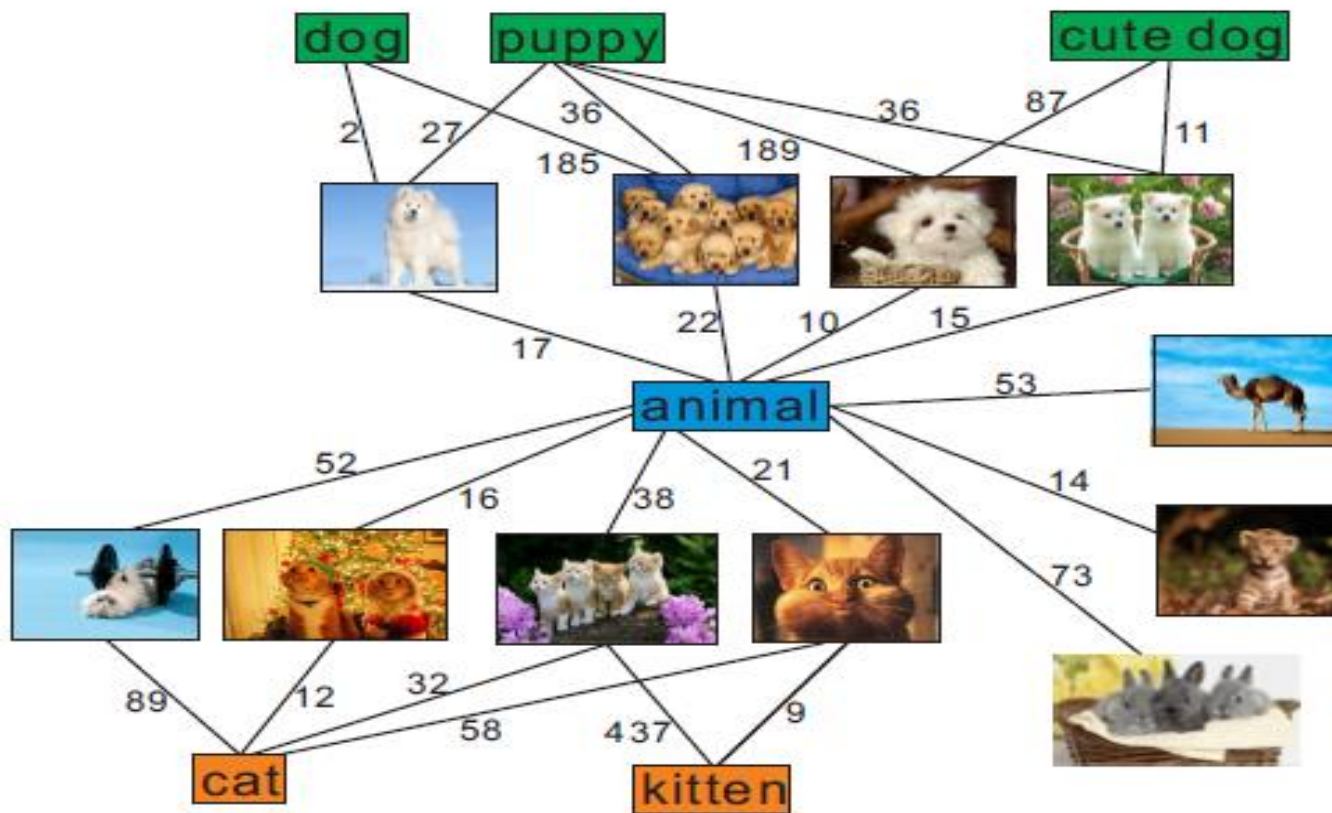
异构序列学习

个体/群体交互

知识图谱构建

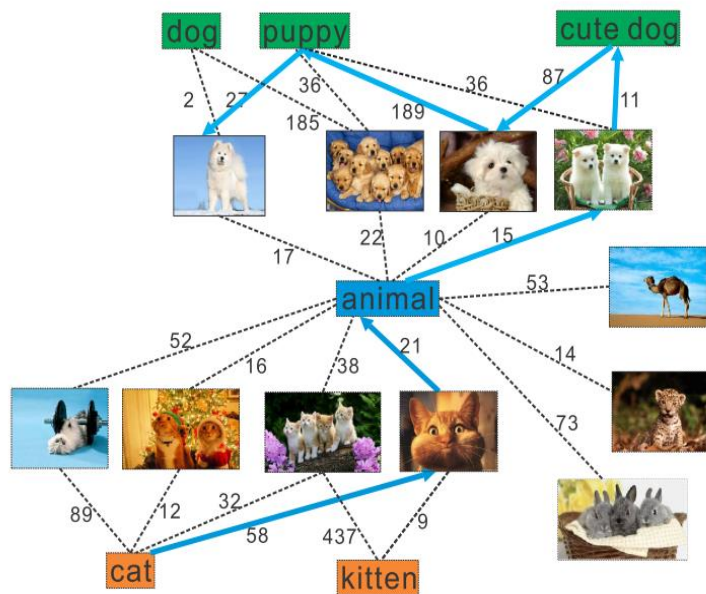
工作(1):基于深度随机游走的众包数据学习

- 互联网搜索引擎记录了用户海量点击数据，如何从这些数据中学习和理解用户查询意图、偏好以及语义？

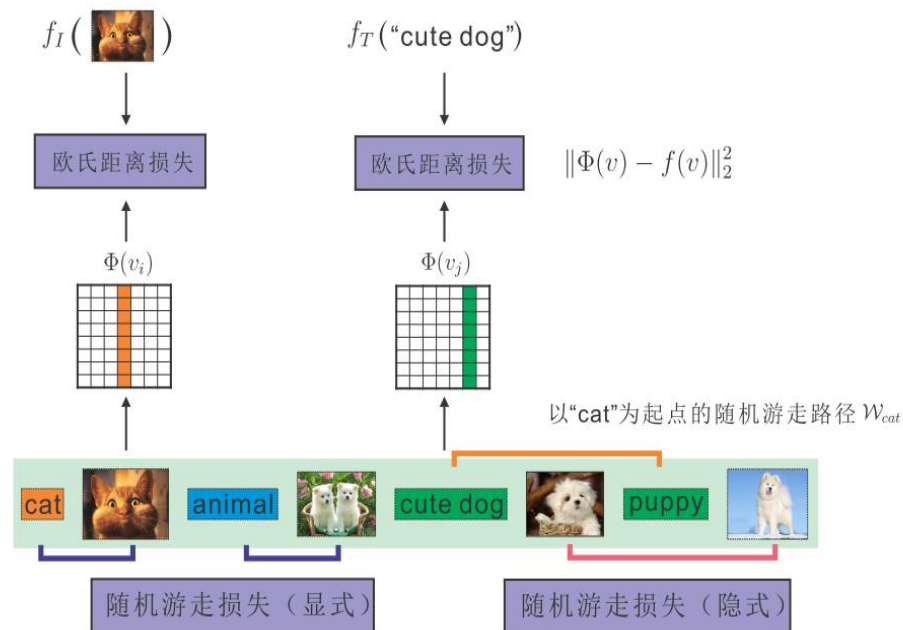


工作(1):基于深度随机游走的众包数据学习

思路：深度随机游走神经网络(Deep Random Walk Neural Network)



(A) 随机游走生成

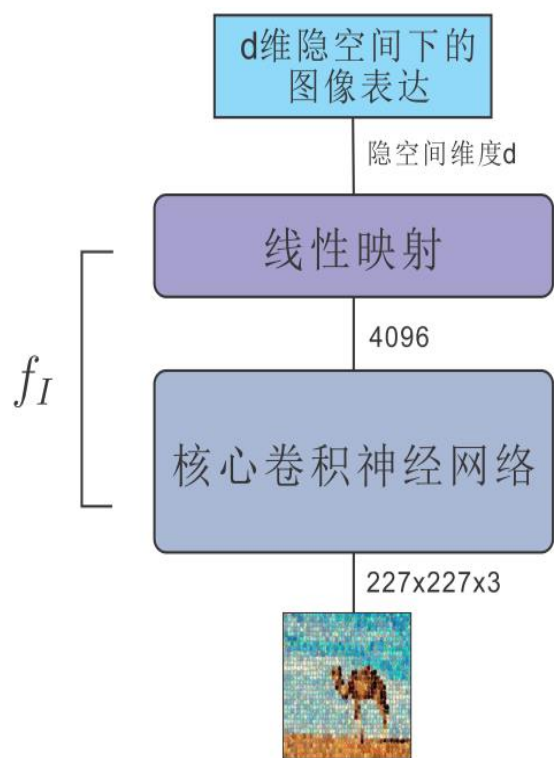


(B) 学习表达以及映射函数

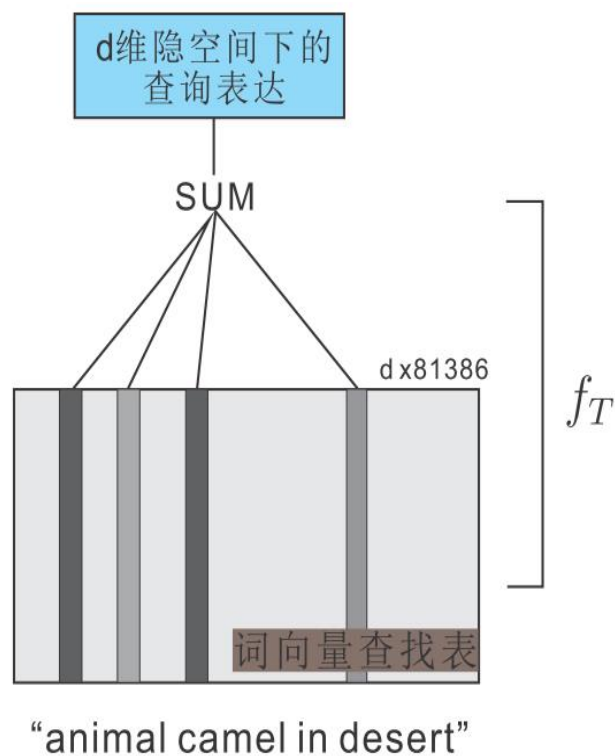
从群体产生的点击数据来构造异构序列，通过深度学习来刻画用户的点击意图和偏好

工作(1):基于深度随机游走的众包数据学习

- 通过深度学习(卷积神经网络和Skip-gram)对序列中异构节点分别编码，学习其特征表达。



(A) 视觉映射函数



(B) 文本映射函数

工作(1):基于深度随机游走的众包数据学习

- 最小化截断随机游走的损失，同时考虑顶点在点击图中的表达应该和其神经网络的输出应该相近。

$$\begin{aligned} \min_{\Phi, f} \quad & \sum_{v \in V} \sum_{v_i \in W_v} \sum_{1 \leq j \leq w} -\log \Pr(v_{i+j} | v_i) \\ \text{s.t.} \quad & \|\Phi(v) - f(v)\|_2^2 \leq C, v \in V \end{aligned}$$

- C为无穷大时，上述优化问题等价于DeepWalk模型，其仅仅学习得到顶点的隐表达（点击表示）而不考虑顶点的内容。

工作(1):基于深度随机游走的众包数据学习

□ 实验及分析

■ Clickture数据集

Clickture 数据集的统计信息

	训练集	测试集
图像数目	1M	79665
查询文本数目	11.7M	1000
查询文本 - 图像对数目	23.1M	79926
词汇数目	915K	/
词汇数目(经过预处理后 ^a)	81K	/
总点击数	82.3M	/

^a 我们提取了单词词干,并去除了停用单词(例如“the”、“photo”和“pic”等)以及出现次数不超过 10 次的稀有词。

工作(1):基于深度随机游走的众包数据学习

□ 实验及分析

■ 用户检索意图

france		shanghai		huawei	
germany	0.6391	seattle	0.5422	smartphone	0.6451
spain	0.5658	singapore	0.5331	cellphone	0.6184
ukrain	0.5575	seatle	0.5215	phone	0.6065
city	0.5571	skyline	0.5204	iphone	0.5989
espana	0.5545	hongkong	0.5197	samsung	0.5893
parisfrance	0.5523	nyc	0.5112	lg	0.5834
prague	0.5498	chicago	0.5019	nokia	0.5802

ps4		nba	
playstation	0.7656	basketball	0.6393
ps	0.6564	drose	0.5988
playstation4	0.6468	derrick	0.5963
plastation	0.6443	lebronjames	0.5800
xbox720	0.6412	bakettball	0.5589
ps3	0.6039	dunk	0.5120
xbox	0.5960	miamiheat	0.5051

工作(1):基于深度随机游走的众包数据学习

□ 实验及分析

■ 用户检索意图

错误拼写	正确的拼写
wrld worl wourld wrold wolrd	world
aple appple aplle appele applwe aapple	apple
anmal animla animal aniaml anime anmiale	animal
gerl gril gile gilr grls	girl
goole googele gogle gooogle googal goolge	google
gutar gitar guiter gutair giutar gituar quitar	guitar

错误拼写的单词及其正确的拼写方式

工作(1):基于深度随机游走的众包数据学习

□ 实验及分析

■ 用户检索意图

输入的查询文本	搜索建议
barack obama	2012bracket barack obama barack childhood obama
bathroom shelf	bathroom shelf small bathroom shelf wall
hairstyle short	hairstyle olderwomen short hairstyle kinkey short
cuban flag	cuban flag large cuban flag small
cupcake pail	collect cupcake cost cupcake cupcake price

搜索建议

工作(1):基于深度随机游走的众包数据学习

□ 实验及分析

■ 与点击次数相比，可得到更好的排序

查询: “*animal best*”



查询: “*beer*”

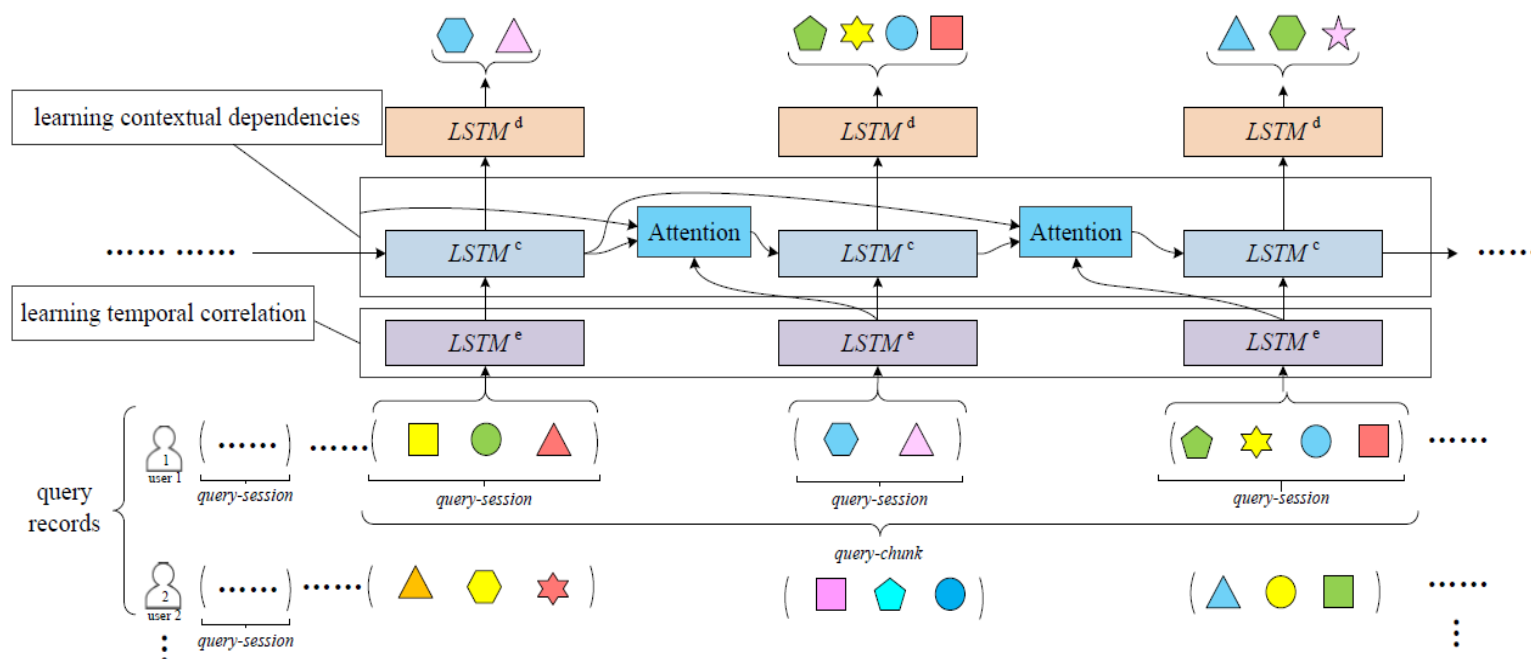


查询: “*dell computer*”



工作(2): 层次化上下文attention-LSTM

- 目的: 从地图搜索数据中感知城市中群体行为
- 在层次化LSTM中引入注意力(attention)机制, 提出了一种层次化上下文attention-LSTM模型(hierarchical contextual attention LSTM)



- Pingbo Pan, Zhongwen Xu, Yi Yang, Fei Wu, Yueting Zhuang, Hierarchical Recurrent Neural Encoder for Video Representation with Application to Captioning, CVPR 2016
- Fei Wu, Jun Song, Haishan Wu, Tong Zhang, Zhongfei Zhang, Wenwu Zhu, Hierarchical Contextual Attention Recurrent Neural Network for Map Query Suggestion, TKDE (under review)

工作(2): 层次化上下文attention-LSTM

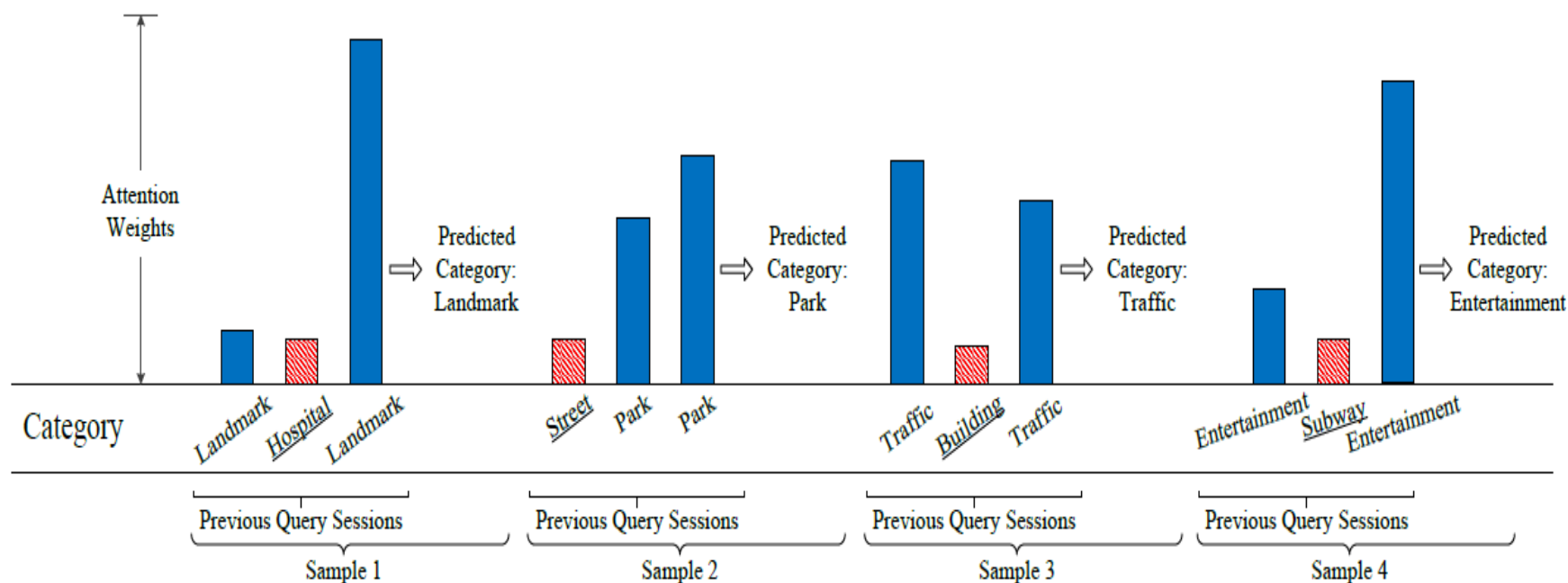
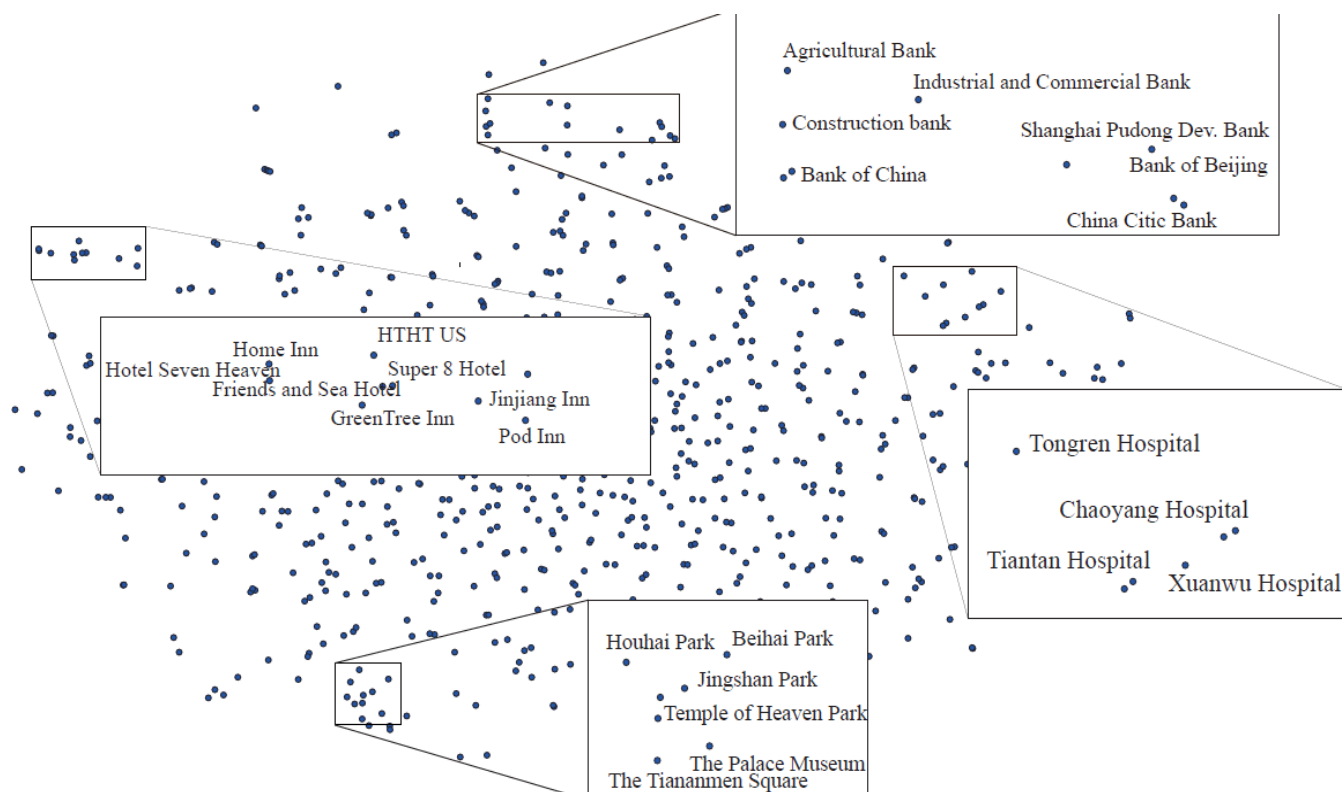


Fig. 6: The attention weights learned by our model while predicting the next query. The height of each bar equals to the weight learned for that query session and the category each query belongs to is also given. The bar of each noised query (i.e., whose belonging category is not consistent with that of its neighbouring queries) is denoted with an oblique line and the category name of each noised query is denoted with underline. The queries in each example are arranged in a temporal order and the category names at the right side of each arrow are the category names of the next ground truth query.

群体用户地图检索模式发现

工作(2): 层次化上下文attention-LSTM

- 与百度公司合作，收集了2015年12月北京地区一个月手机地图查询数据，包含了118,468用户提交的6,590,157次地图检索请求，这些地图检索请求涵盖了21,358地点查询（如颐和园等）。



工作(3): 刻画因果依赖和时序影响机制

□ 个体和个体之间是由大量元素及其复杂关系组成的系统，传统上采用复杂网络理论、进化博弈论和基于Agent模式等方法来研究个体之间存在的相互影响。

□ 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》)



Is it illegal to decorate vehicle with flag or sticker on National Day?	Question
What do you think? Are they going to arrest every driver who ventures out this National Day?	Potential
Don't worry about the safety of yourself or other road users. It's National day. Windows are for decoration and not actually looking out of.	Bad
Not on National Day.	Good
My car has failed inspection twice now because it has stickers on the back window. They just overlook it every time when I renew the registration. As long as you pass the technical part...	Good
Remember not to place them besides or near your plate.	Good
Don't forget it's National Day.	Good

Q-A问答中观点逐步凝练形成

工作(3): 刻画因果依赖和时序影响机制

□ 在Q-A问答中考虑question-answer之间因果依赖以及answers-answer之间时序影响

□ question-answer causation influence (how appropriate a posted answer)

□ answers-answer temporal interaction (how evolved answer formation)

□ 图中 $h(1)$ 为 question-answer 的交互信息, $h(2)$, $h(3)$... $h(t-1)$ 为 answers-answer 之间的相互影响

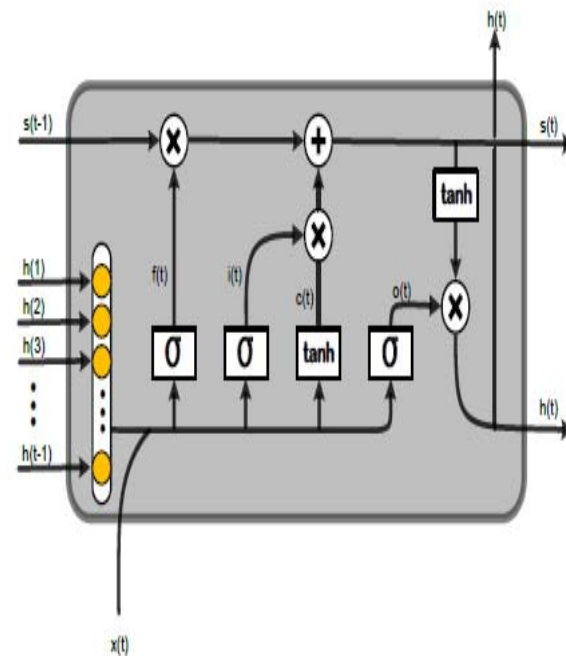
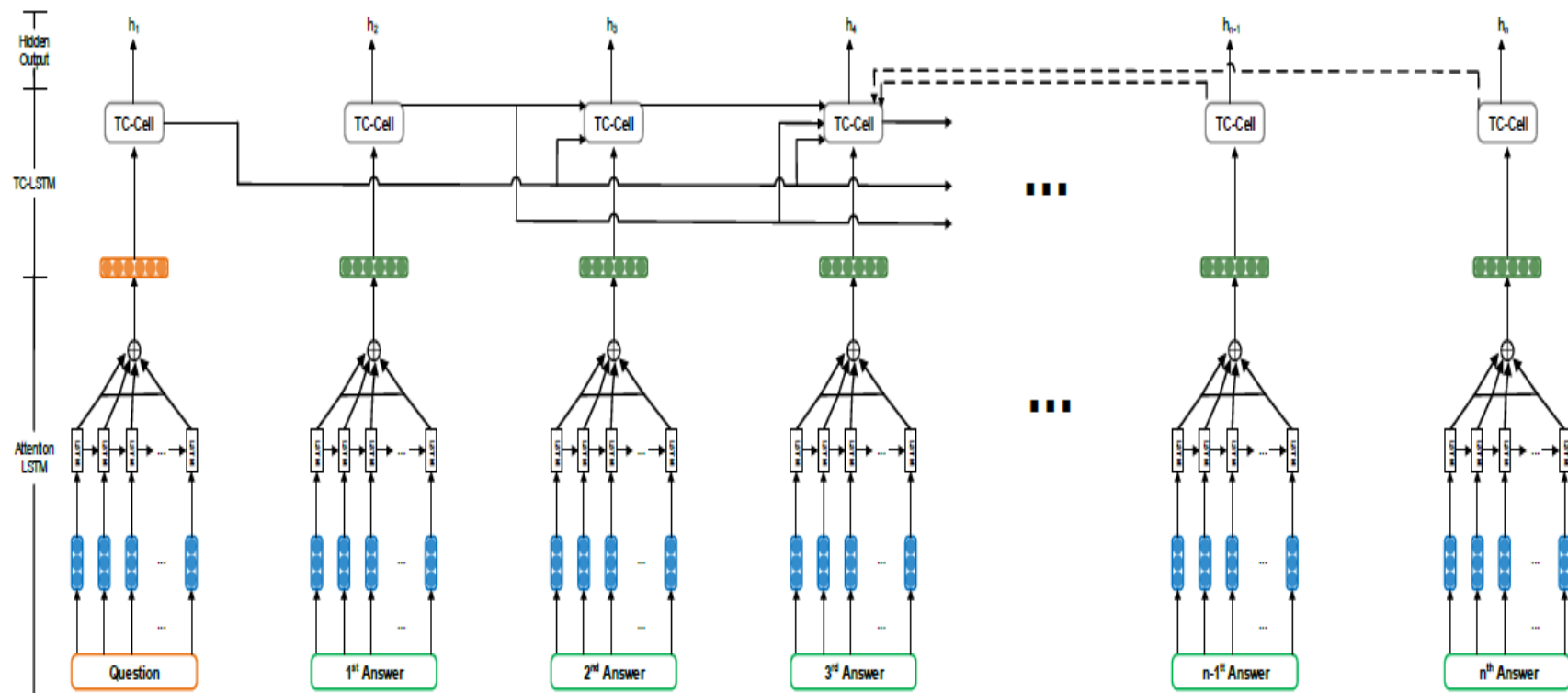


Fig. 6: The detail of TC-cell. $i(t)$, $o(t)$, $f(t)$, $c(t)$, $s(t)$, $h(t)$ are inner gates. $x(t)$ is the answer input, $s(t-1)$ is the output of previous state gate and $h(1)$, $h(2)$, $h(3)$, $\dots$, $h(t-1)$ are the hidden output from all the previous cells.

工作(3): 刻画因果依赖和时序影响机制



- 综合考虑question-answer之间影响与answers-answer之间交互后所得到的TC-LSTM网络
- 此处以第四个节点为例，第四个节点在接收第三个答案的同时，还会接收来自question和前两个answer的隐藏层输出，并且会影响到其之后的每一个节点。

工作(3): 刻画因果依赖和时序影响机制

Methods	Good	Bad	Potential
SVM	80.24	76.22	0.00
CRF	79.50	75.25	5.63
DBN	76.72	70.98	12.97
mDBN	78.33	69.89	14.23
CNN	76.70	75.04	12.64
R-CNN	78.12	75.73	15.82
QA-LSTM	77.49	74.27	12.25
C-LSTM	76.14	73.97	11.02
T-LSTM	75.63	72.36	10.19
TC-LSTM ⁻	81.27	76.12	21.48
TC-LSTM(Random)	77.04	73.91	18.25
TC-LSTM($\beta_1 = 0.37$)	83.69	77.13	25.56

TABLE 2: F1 scores for each answer class (%) (best scores are boldfaced)

Methods	Precision	Recall	F1
SVM	50.23	54.28	52.15
CRF	53.62	53.30	53.46
DBN	54.98	52.21	53.56
mDBN	55.47	52.89	54.15
CNN	55.60	54.00	54.79
R-CNN	56.84	56.18	56.51
QA-LSTM	53.57	55.82	54.67
C-LSTM	54.00	53.42	53.71
T-LSTM	53.87	51.64	52.73
TC-LSTM ⁻	60.45	58.81	59.62
TC-LSTM(Random)	55.79	57.02	56.40
TC-LSTM($\beta_1 = 0.37$)	63.76	60.58	62.13

TABLE 3: Macro-averaged results (%) (best scores are boldfaced)

Method \ i^{th}	1	2	3	4	5	6
SVM	79.04	41.04	21.13	11.16	5.71	2.98
CRF	78.80	42.42	22.78	12.47	6.71	3.60
DBN	77.03	42.76	23.41	13.05	7.18	3.96
mDBN	78.53	43.30	23.96	13.33	7.39	4.10
CNN	76.98	43.16	24.27	13.83	7.65	4.29
R-CNN	78.25	44.80	25.45	14.80	8.41	4.81
QA-LSTM	76.81	43.98	25.08	14.36	8.22	4.71
C-LSTM	75.53	42.30	23.67	13.26	7.43	4.16
T-LSTM	74.27	40.80	22.57	12.36	6.79	3.74
TC-LSTM ⁻	78.59	47.10	28.26	16.80	10.17	6.19
TC-LSTM($\beta_1 = 0.37$)	81.68	50.07	31.04	19.25	11.93	7.40

TABLE 4: The precision performance when the first good answer occurring at the i^{th} ($i = 1, 2, 3, 4, 5$ or 6) position is identified as the *Good* answer. (%) (best scores are boldfaced)

工作(4): 群智感知差异性学习

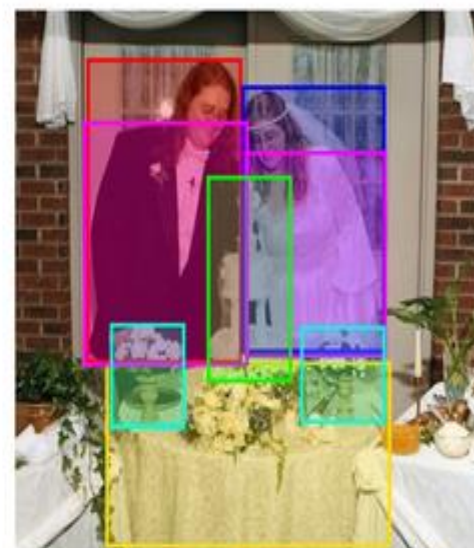
- 对群体之间的注意力感知差异性进行建模，即如何来刻画群体注意力感知差异性。



A man with pierced ears is wearing glasses and an orange hat.
A man with glasses is wearing a beer can crotched hat.
A man with gauges and glasses is wearing a Blitz hat.
A man in an orange hat starring at something.
A man wears an orange hat and glasses.



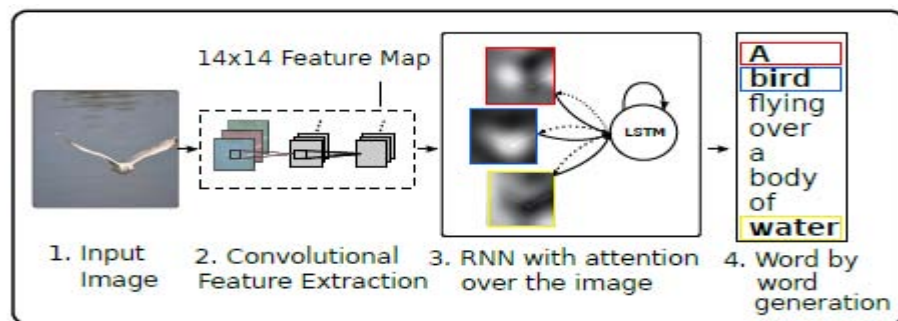
During a gay pride parade in an Asian city, some people hold up rainbow flags to show their support.
A group of youths march down a street waving flags showing a color spectrum.
Oriental people with rainbow flags walking down a city street.
A group of people walk down a street waving rainbow flags.
People are outside waving flags.



A couple in their wedding attire stand behind a table with a wedding cake and flowers.
A bride and groom are standing in front of their wedding cake at their reception.
A bride and groom smile as they view their wedding cake at a reception.
A couple stands behind their wedding cake.
Man and woman cutting wedding cake.

工作(4): 群智感知差异性学习

- 深度学习方法与注意力模型结合取得了较好成功, 如Neural Talk



A woman is throwing a frisbee in a park.



A dog is standing on a hardwood floor.



A stop sign is on a road with a mountain in the background.



A little girl sitting on a bed with a teddy bear.



A group of people sitting on a boat in the water.

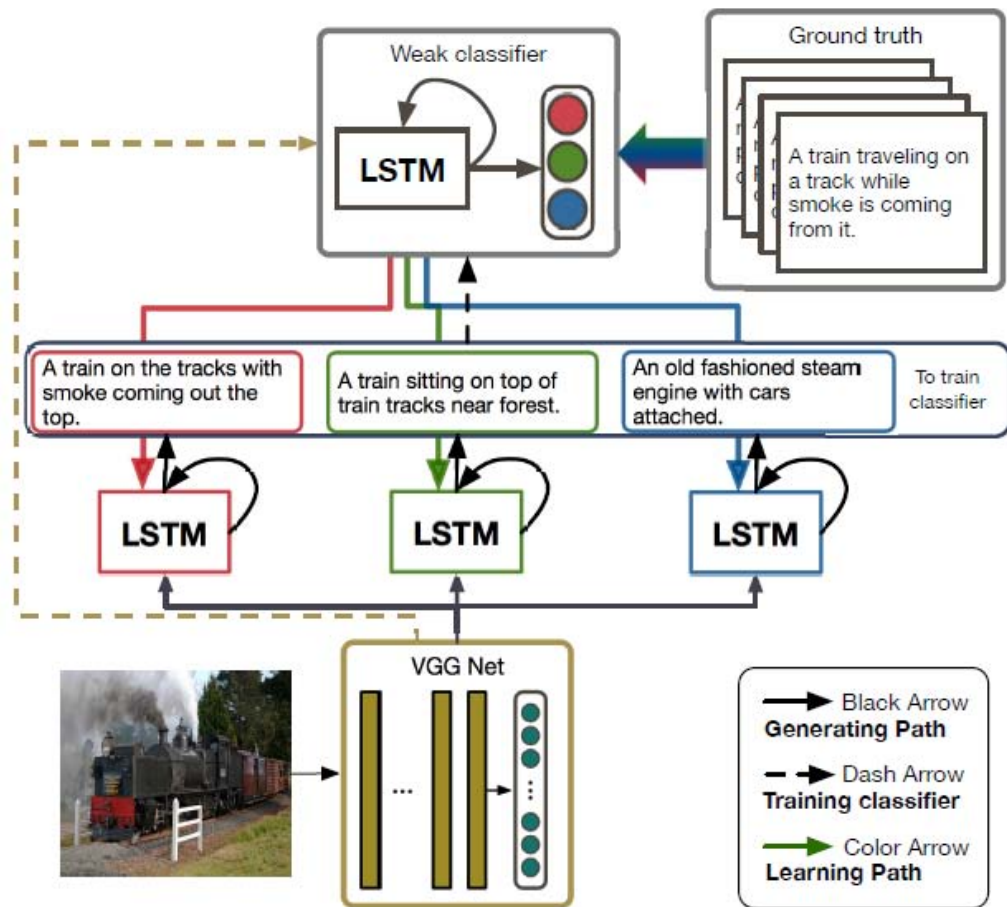


A giraffe standing in a forest with trees in the background.

Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel and Yoshua Bengio, Show, Attend and Tell: Neural Image Caption Generation with Visual Attention, [ICML 2015](#)

工作(4): 群智感知差异性学习

- Crowd Attentional GroupTalk机制的注意力感知差异学习方法:
- 引入反馈神经网络(LSTM)对语言差异偏好性(非线性)进行识别
- 再进一步将识别后数据用于训练特定语言模型以产生差异化视觉描述



刻画群智差异性的GroupTalk

工作(4): 群智感知差异性学习

Table 2: The demonstration of generated sentences by base model and GroupTalk on MS COCO.



Base Model with beam search size 3:

A large jetliner sitting on top of an airport tarmac.

A large jetliner sitting on top of an airport runway.

A large passenger jet sitting on top of an airport runway.

GroupTalk:

Model 1: A large jetliner sitting on top of an airport runway.

Model 2: An airplane sitting on the tarmac at an airport.

Model 3: A large air plane on a run way.

Table 3: Diversity evaluation. $\downarrow / \uparrow$ depicts that a larger value indicates a less/more diverse result.

Dataset	Model	mB-1 $\downarrow$	mB-2 $\downarrow$	mB-3 $\downarrow$	mB-4 $\downarrow$	DIV-1 $\uparrow$	DIV-2 $\uparrow$	DIV-3 $\uparrow$	DIV-4 $\uparrow$
Flickr 8K	Base Model (B)	0.83	0.77	0.72	0.69	0.49	0.63	0.68	0.72
	GroupTalk	0.53	0.37	0.27	0.20	0.62	0.85	0.91	0.94
Flickr 30K	Base Model (B)	0.85	0.79	0.75	0.71	0.48	0.62	0.66	0.70
	GroupTalk	0.60	0.44	0.35	0.28	0.58	0.81	0.87	0.92
MS COCO	Base Model (B)	0.87	0.80	0.74	0.70	0.38	0.54	0.61	0.67
	GroupTalk	0.73	0.59	0.49	0.41	0.47	0.71	0.81	0.86

Table 4: Statistics of word usage preferences.

Model	#plane	#airplane	#jetliner	#jet
Model 1	0	14	64	10
Model 2	13	84	0	0
Model 3	72	25	0	0

工作(5): 群智感知与数据驱动的学习模型

- 重点: 如何在知识图谱构建的实体链接过程中实现更好的命名体消歧。

or postnatal infection⁵³ and to
jelae. The study population in
genital HCMV infection⁵⁴. T
al load was measured by quan
and gN4c were prevalent in the

Description

Synonyms

Tree

Infection with CYTOMEGALOVIRUS, characterized by enlarged cells bearing intranuclear inclusions. Infection may be in almost any organ, but the salivary glands are the most common site in children, as are the lungs in adults.[*Cytomegalovirus Infections* (D003586), **MESH**]. Human cytomegalovirus a species of the Cytomegalovirus genus of viruses, which in turn is a member of the viral family known as "Herpesviridae" or herpesviruses. It is typically abbreviated as HCMV or, commonly but more ambiguously, CMV. It is also known as human herpesvirus-(HHV-5). Within "Herpesviridae", HCMV belongs to the "Betaherpesvirinae" subfamily which also includes cytomegaloviruses from other mammals.[*Human cytomegalovirus* (8499164), **Wikipedia**].

n per sample) was used to mea
roeter) and PAO1 biofilms. This
sal.
5b gene for proline⁴-analog res
detoxifying the proline⁴ analog
lative stresses. Mpr1 mediates
-pyrroline-5-oxo⁷boxylate⁷, lea

Description

Synonyms

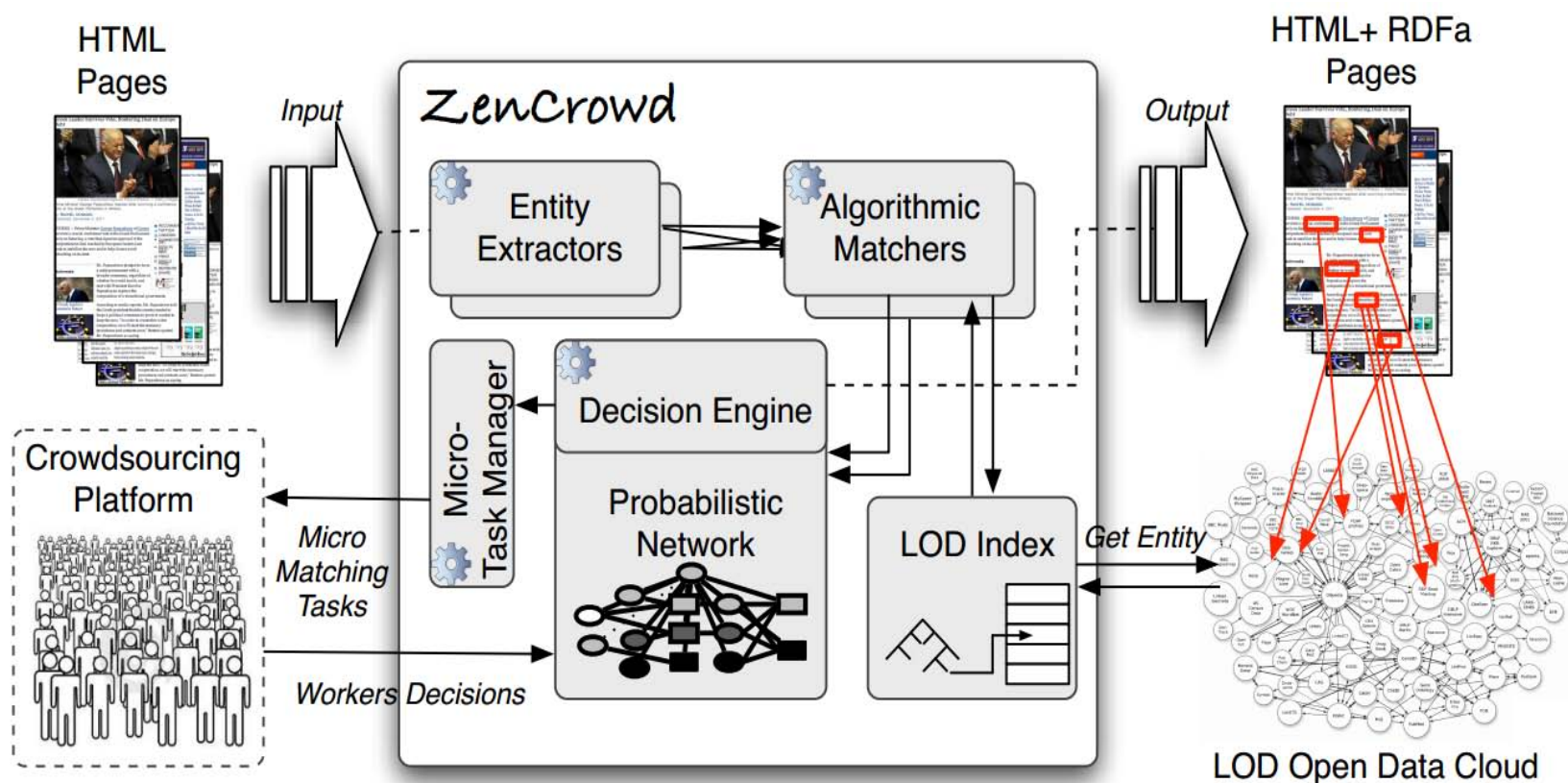
Tree

An alpha-amino acid that is pyrrolidine bearing a carboxy substituent at position 2.[*proline* (26271), **CHEBI**]. Proline (abbreviated as Pro or P) is an α -amino acid, one of the twenty DNA-encoded amino acids. Its codons are CCU, CCC, CCA, and CCG. It is not an essential amino acid, which means that the human body can synthesize it. It is unique among the 20 protein-forming amino acids in that the amine nitrogen is bound to not one but two alkyl groups, thus making it a secondary amine. The more common form has "S" stereochemistry.[*Proline* (38811), **Wikipedia**].

知识图谱中命名体链接

工作(5): 群智感知与数据驱动的学习模型

- 命名实体消歧问题上，众包数据引入取得了良好的效果(如 Zencrowd方法)



工作(5): 群智感知与数据驱动的学习模型

□ 众包用户常常能在使算法迷惑的问题上做出正确的判断

...trains will share a track between New York Avenue and Rhode Island Avenue stations...	用户	33	RIA (Station) ¹
		26	RIA (Station)
		32	
		47	RIA (Avenue) ²
	算法	TFIDF + 余弦距离	RIA (Avenue)

1. http://dbpedia.org/page/Rhode_Island_Avenue_%E2%80%93_Brentwood_%28WMATA_station%29

2. http://dbpedia.org/page/Rhode_Island_Avenue_%28Washington,_D.C.%29

□ 众包用户也因各种原因在显而易见的问题上产生噪声

- 如何学习用户对问题的判断
- 如何减少因用户能力不同而产生的噪声

. ...they planned to work closely with the United States , Europe and India to plan ...	用户	7	These United States (rock band) ¹
		59	Electoral College (institution) ²
		39	United States (state) ³
	算法	TFIDF + 余弦距离	United States (state)

1. http://dbpedia.org/page/These_United_States

2. http://dbpedia.org/page/Electoral_College_%28United_States%29

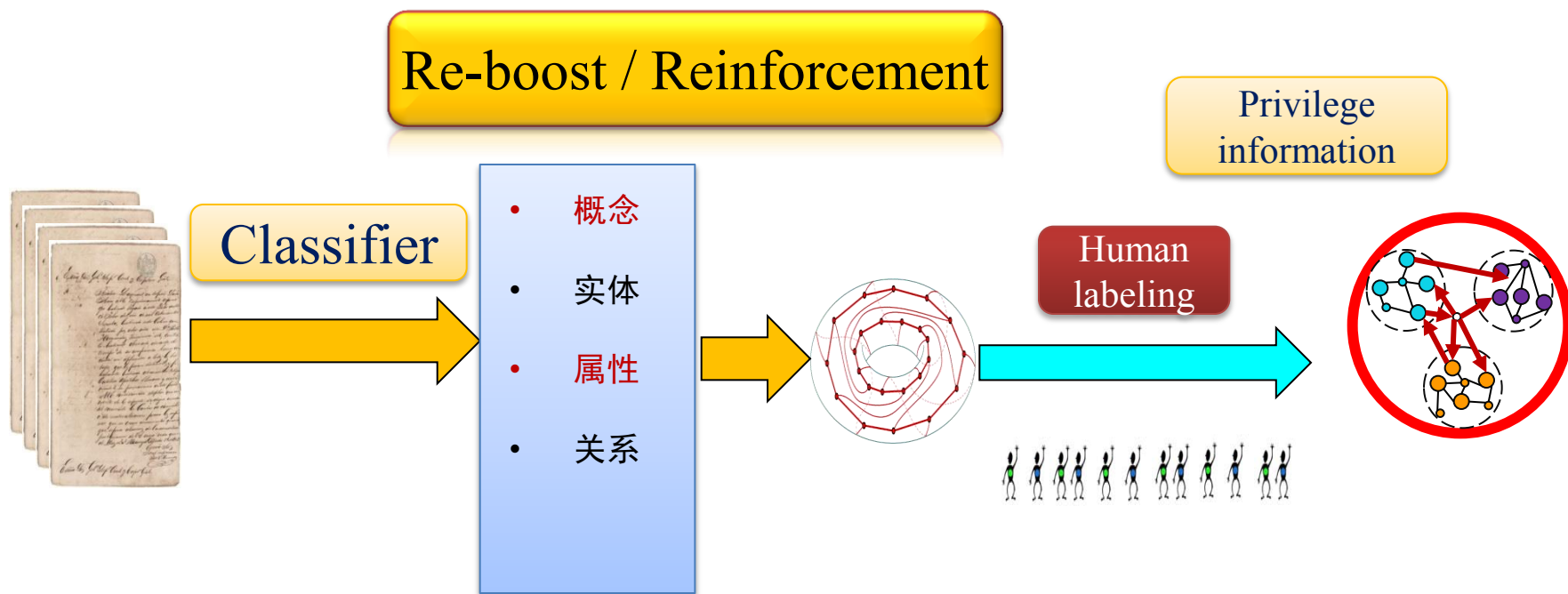
3. http://dbpedia.org/page/United_States

实体消歧问题的用户标注实例

注：以上数据摘自ZenCrowd提供的数据集

工作(5): 群智感知与数据驱动的学习模型

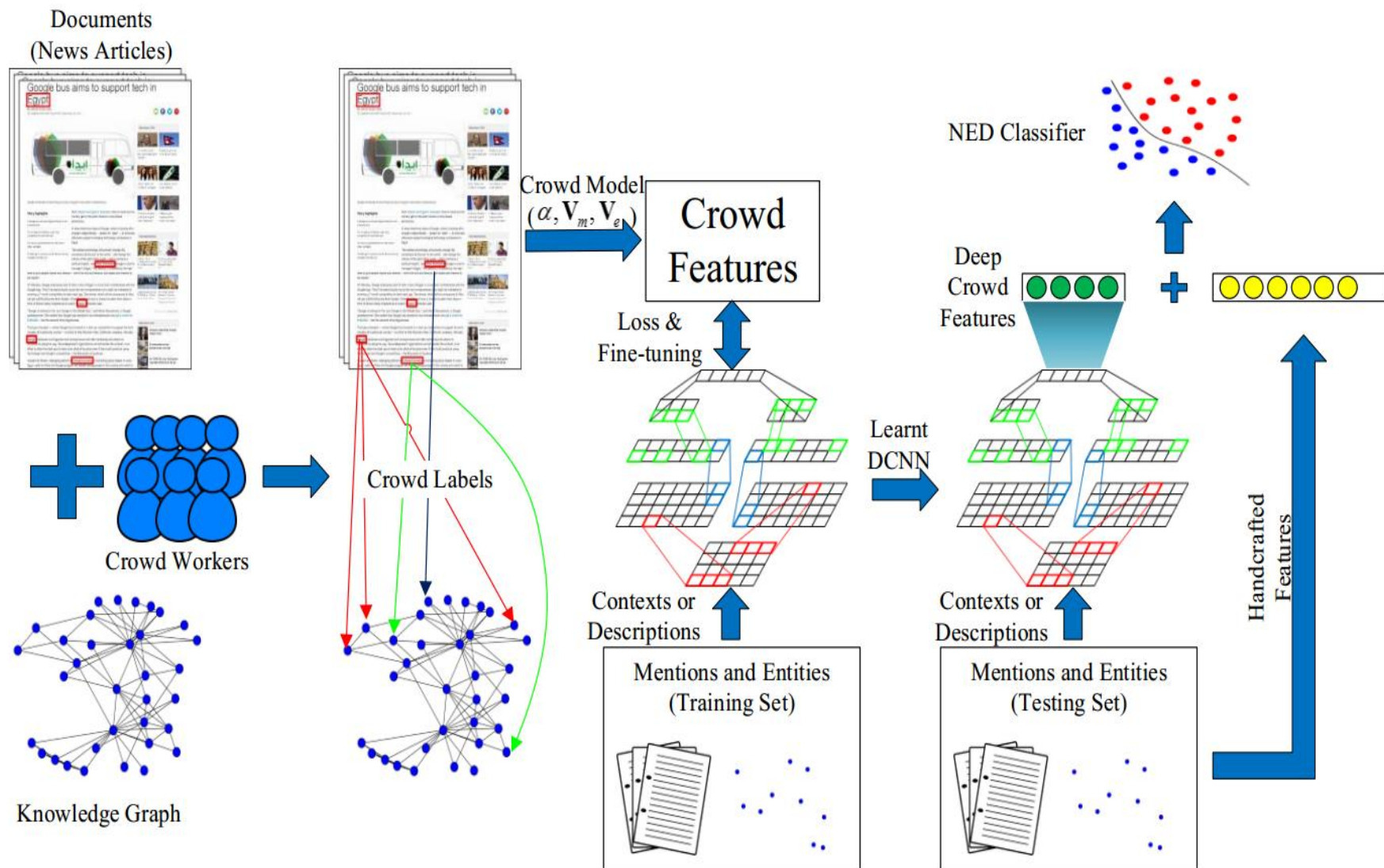
□思路：将众包标注结果及标注用户作为一种“privilege information”，提升机器学习性能。



V. Vapnik and A. Vashist, A new learning paradigm: Learning using privileged information, *Neural Networks*, 22(5-6):544–557, 2009

工作(5): 群智感知与数据驱动的学习模型

□ 基于众包深度特征学习的命名实体消歧



工作(5): 群智感知与数据驱动的学习模型

- 为降低众包数据中的噪声，对用户个人能力进行建模
- 根据已有的众包标签，通过优化表达式(1)，估计出各位用户的能力，同时也将表面形式(surface form)和实体(entity)向量化，得到它们对应的“众包特征”

Using notation $v(\cdot)$ as the corresponding d -dimensional crowd feature vector of a mention or an entity, then the sets of crowd feature vectors are denoted as $\mathbf{V}_m = \{v_m(i) := v(M_i)\}_{i=1}^I$ and $\mathbf{V}_e = \{v_e(j) := v(E_j)\}_{j=1}^J$.

We introduce parameter $\alpha = \{\alpha_k\}_{k=1}^K$ constrained by $\|\alpha\|_2 = 1$ in order to resolve conflicts by modelling the ability of workers. Worker U_k always gives the answer in “common sense” when $\alpha_k = 1$ while the worker always gives the opposite answers when $\alpha_k = -1$, and the worker gives random answers when $\alpha_k = 0$.

In our crowd model, α , $\mathbf{V}_m$ and $\mathbf{V}_e$ are learned to minimize the loss function as follows:

$$L_1(\alpha, \mathbf{V}_m, \mathbf{V}_e) = \sum_{q=1}^Q \sum_{j \in \mathbf{e}_q} -\alpha_{u_q} \delta^*(j, t_q) \langle v_e(j), v_m(m_q) \rangle + \frac{\eta}{2} (\|v_e(j)\|^2 + \|v_m(m_q)\|^2) \quad s.t. \|\alpha\|_2 = 1 \quad (1)$$

where $\langle \cdot, \cdot \rangle$ is the inner product of two vectors, and $\delta^*(\cdot, \cdot)$ is the modified version of Kronecker delta function given by

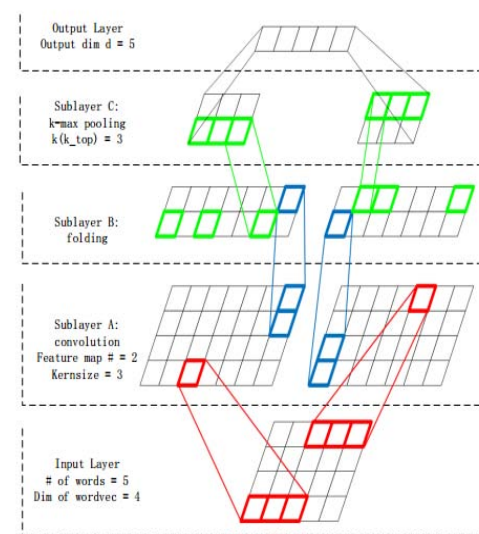
$$\delta^*(x, y) = \begin{cases} 1, & \text{if } x = y \\ -1, & \text{if } x \neq y \end{cases} \quad (2)$$

众包模型的损失函数

工作(5): 群智感知与数据驱动的学习模型

□ 动态卷积神经网络(Dynamic Convolutional Neural Network, DCNN)学习句子的特征表达

□ 将DCNN学习得到的特征与众包特征之间的欧氏距离作为损失函数，对DCNN进行微调。



DCNN结构示意图

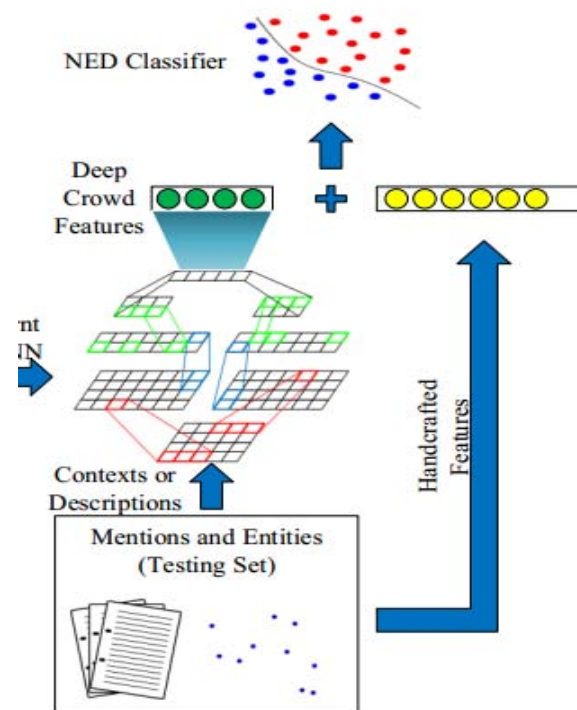
We finetune the DCNN with a loss function based on euclidean distances as follows:

$$L_2(\theta) = \sum_{i \in \mathbf{M}_c} \|\gamma(s_m(i); \theta) - v_m(i)\|^2 + \sum_{j \in \mathbf{E}_c} \|\gamma(s_e(j); \theta) - v_e(j)\|^2 + \xi \|\theta\|^2 \quad (3)$$

细粒度调整的损失函数

工作(5): 群智感知与数据驱动的学习模型

- 在没有被众包标注的数据上，通过学好的DCNN将基础特征映射为众包特征，通过扩展特征空间的方式提高算法的效率
- 使用众包特征前后的效果对比如下



(a) The accuracy when deep crowd feature dimension is 5(the TF-IDF dimension is 18478)

ratio of data in training set to testing set	9:1	8:2	7:3
TFIDF+DCF	78%	71%	69%
TFIDF	74%	70%	69%

(b) The accuracy when deep crowd feature dimension is 300(the TF-IDF dimension is 18478)

ratio of data in training set to testing set	9:1	8:2	7:3	5:5
TFIDF+DCF	77%	71%	69%	67%
TFIDF	74%	70%	69%	69%

工作(5): 群智感知与数据驱动的学习模型

- 用Crowd features来提升deep features性能，进行命名体去歧处理，以实现知识图谱中Entity Linking目的。
- Disambiguate Named Entities with Deep Supervised Learning via Crowd Labels：建立数据驱动和群智引导的知识图谱构建
- 包含：
 - *Crowd Model*：判断crowd workers的能力大小
 - *Joint Optimization*：对crowd features和 learning features进行联合优化
 - *Entity Disambiguation*：对知识图谱中实体进行歧义消除

提纲

一、三元空间涌现群智

二、群智可计算思路

三、跨媒体异构序列深度学习

四、若干思考

人工智能的诞生

□ 1956 年，斯坦福大学 John McCarthy 教授、麻省理工学院 Marvin Lee Minsky 教授、卡内基梅隆大学的 Herbert Simon 和 Allen Newell 教授（以上四位皆为图灵奖获得者）、信息理论之父贝尔实验室的 Claude Elwood Shannon、IBM 公司 Nathaniel Rochester 等学者在美国达特茅斯（Dartmouth）学院首次确立了“人工智能”概念。

A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence

August 31, 1955

*John McCarthy, Marvin L. Minsky,
Nathaniel Rochester,
and Claude E. Shannon*

■ The 1956 Dartmouth summer research project on artificial intelligence was initiated by this August 31, 1955 proposal, authored by John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon. The original typescript consisted of 17 pages plus a title page. Copies of the typescript are housed in the archives at Dartmouth College and Stanford University. The first 5 papers state the proposal, and the remaining pages give qualifications and interests of the four who proposed the study. In the interest of brevity, this article reproduces only the proposal itself, along with the short autobiographical statements of the proposers.

guage, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.

The following are some aspects of the artificial intelligence problem:

1. Automatic Computers

If a machine can do a job, then an automatic calculator can be programmed to simulate the machine. The speeds and memory capacities of present computers may be insufficient to simulate many of the higher functions of the human brain, but the major obstacle is not lack

Research Project on Artificial Intelligence, August 31, 1955, Dartmouth

人工智能的诞生

报告列举了 Artificial Intelligence 所面临的 7 类问题

- Automatic Computers
- How Can a Computer be Programmed to Use a Language
- Neuron Nets
- Theory of the Size of a Calculation
- Self-improvement
- Abstractions (intuition)
- Randomness and Creativity

智能计算三个主要的轨迹方向

□ 逻辑 (Logic):

- 解决人工智能推理复杂这一困难，如 search, constraint satisfaction problems, first order logic, planning and Inductive logic programming等方法。

□ 概率模型(Probabilistic Graphical Models)

- 解决人工智能推理非确定性这一困难。

□ 深度学习(Deep Learning)

- 含有大量非线性隐含层，使得模型对原始数据的表达能力非常强(逐层抽象)。

人工智能理论研究近年来的三条主线

expressive clarity of logic

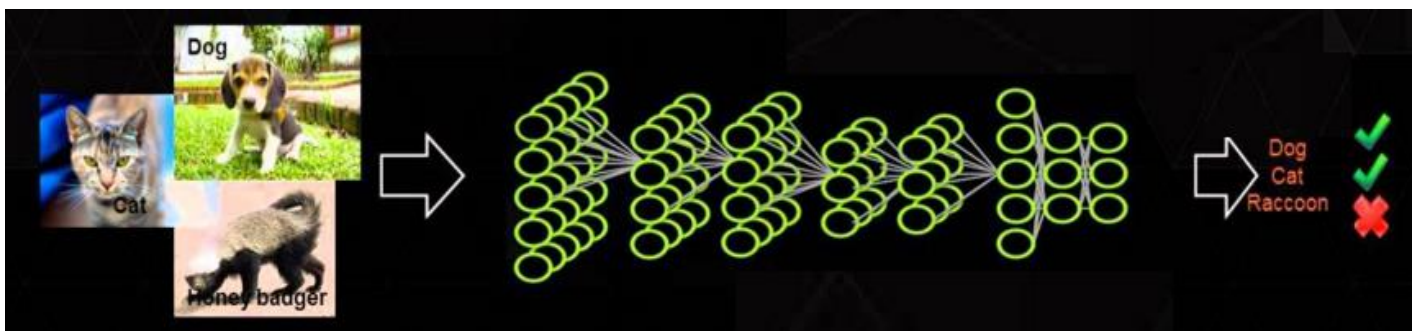
证明和推理为目标

(rich) expressive completeness of
probability

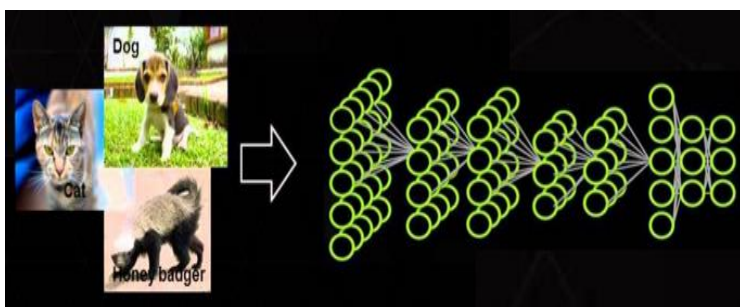
对数据的探索(exploratory maths)为目标, 如聚
类和分类等

难点1：从深度学习到深度推理

- 深度学习本质上是一种“端到端特征学习”机制，再将学习得到的特征(一般是最后一层全连接层)用于后续分类与识别。



深度学习

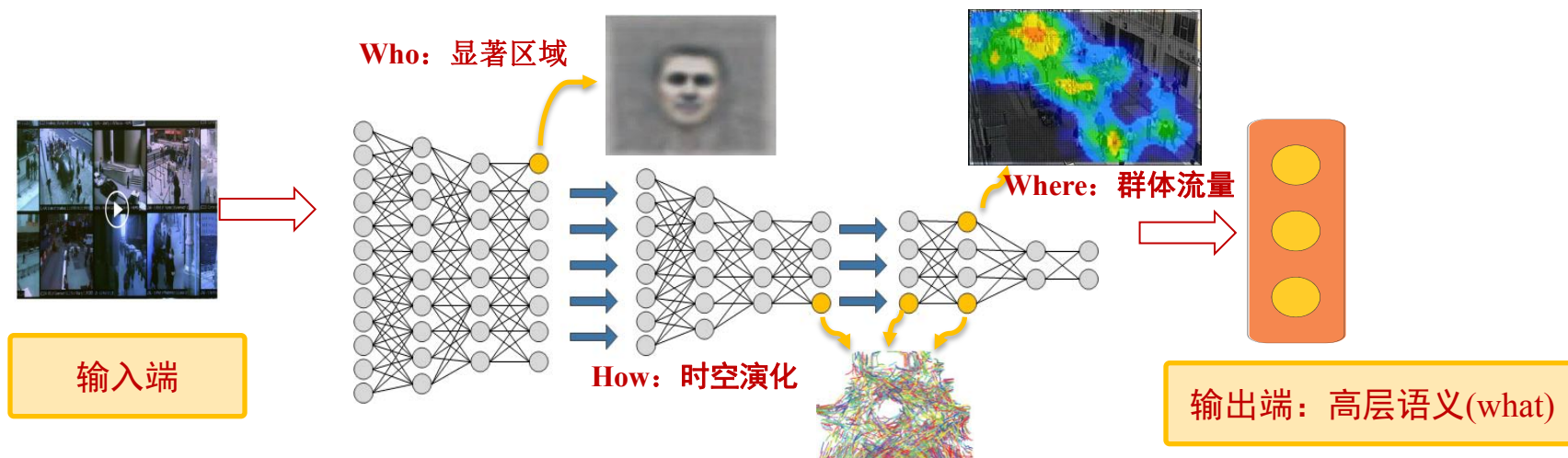
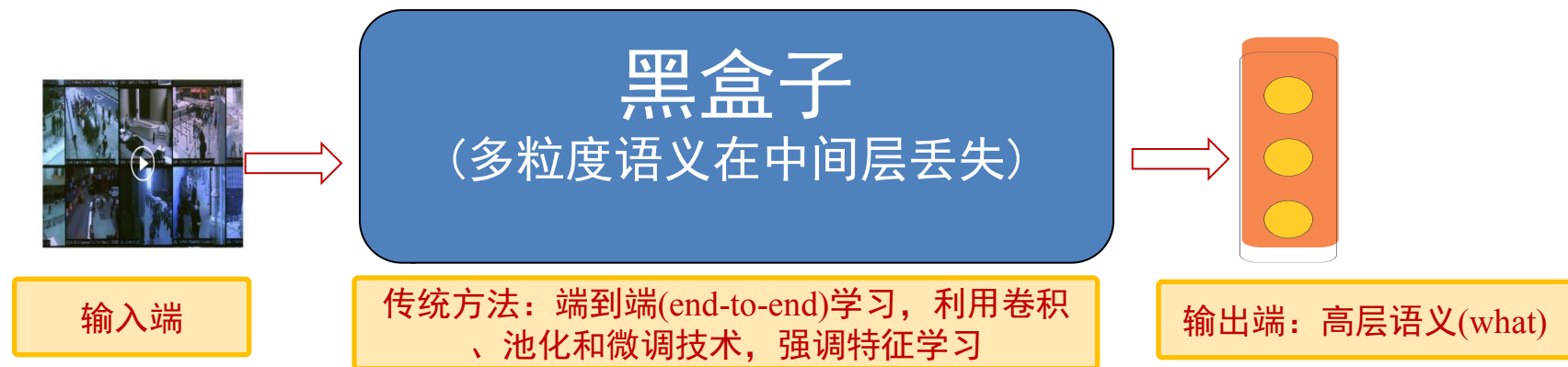


- 条件随机场
- 长短时记忆模型
- 随机森林

深度学习 + 推理模型

难点1：从深度学习到深度推理

- 如何有效利用深度学习所得到的中间层(feature maps) 甄别不同粒度线索而进行推理：



如何从中间层学习和析取多粒度语义，并且与输入图像和输出语义之间建立联系

难点1：从深度学习到深度推理

□ 结合深度学习与逻辑规则，建立数据驱动和知识引导相结合方法

:

□ 深层模型解释性差（黑盒子）

□ 逻辑规则可较好建模人类的认知思维和先验知识

□ 将一阶逻辑(*teacher network*)与深度模型(*student network*)结合起来。

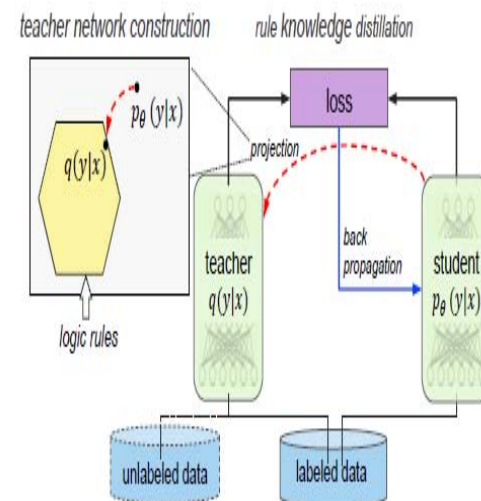


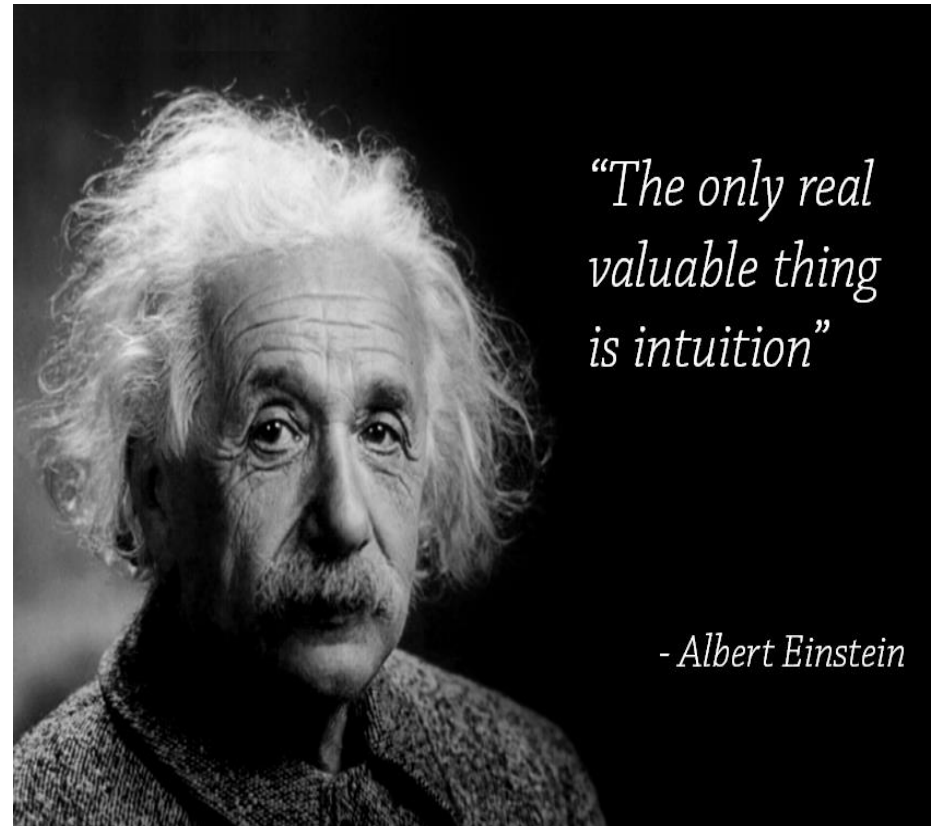
Figure 1: Framework Overview. At each iteration, the teacher network is obtained by projecting the student network to a rule-regularized subspace (red dashed arrow); and the student network is updated to balance between emulating the teacher's output and predicting the true labels (black/blue solid arrows).

难点1：从深度学习到深度推理

- 建立更好的推理数据结构：张量积
- 已有推理方法：谓词逻辑推理或统计推理（Path Ranking and Random Walk）基于树或图。
- 基于张量积表示(Tensor Product Representations)的知识推理：
 - 任意一个纯符号树状或图状结构（可以高效率且具有强解释性用在逻辑推理上）可跟一个高维度张量建立起同构。
 - 这种同构就可有效实现结构到结构的符号映射（比如自然语言或计算机程序的输入输出），但同时又能直接用深度神经网络的方法去学习和优化这种结构映射（这包括复杂多步的逻辑推理）。

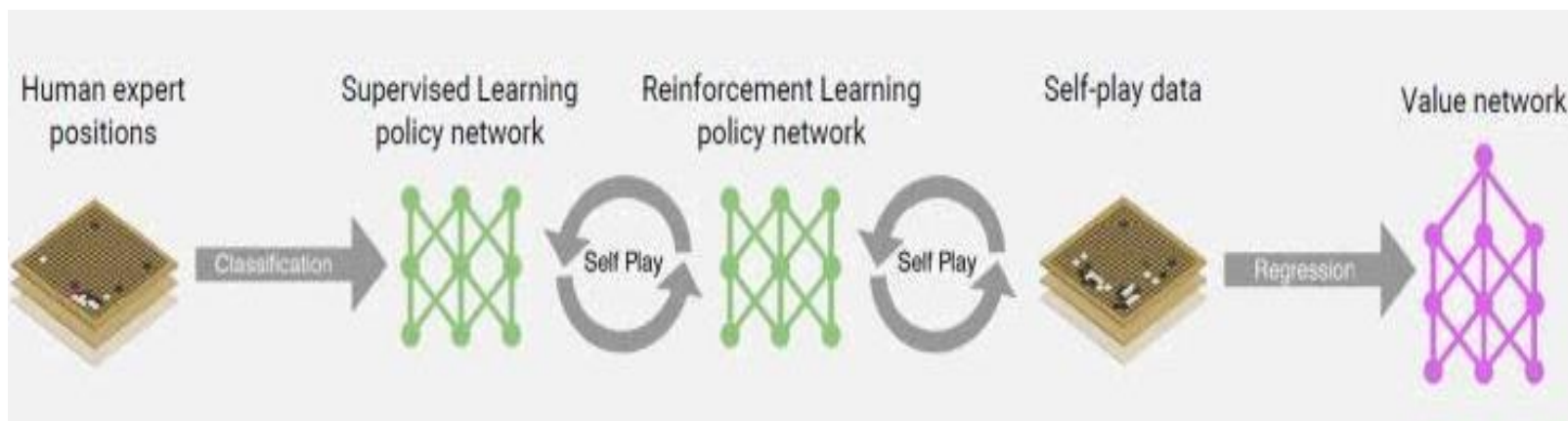
难点2：智能计算中隐性知识和显性知识的协同

- Humans are the only species that combines **intuitive** (*implicit*) and **symbolic** (*explicit*) knowledge, with the dual capacity to transform the former into the latter and in reverse to improve the former with the latter's feedback.
- **Intuition:** The ability to understand something immediately, without the need for conscious reasoning



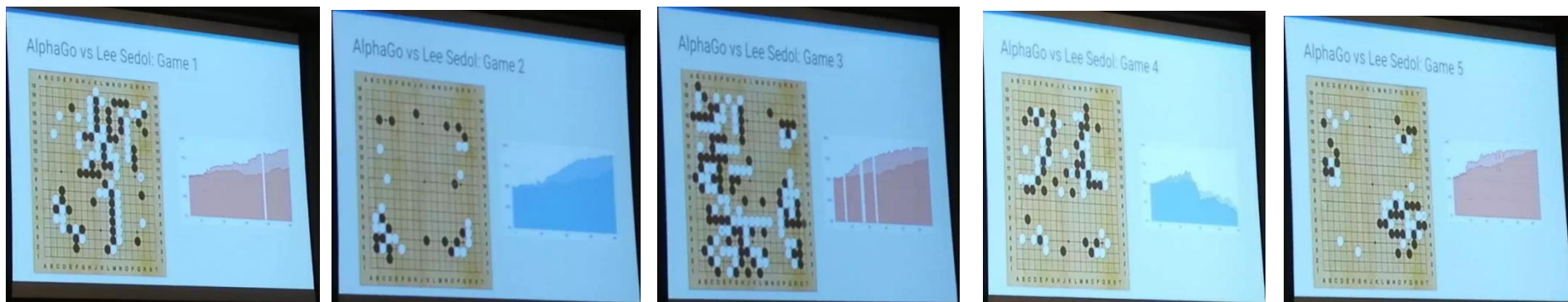
难点2：智能计算中隐性知识和显性知识的协同

- 通过深度学习的非线性映射以及蒙特卡洛搜索的跳跃性将隐性直觉和显性知识结合起来。



AlphaGo训练过程

难点2：智能计算中隐性知识和显性知识的协同



从左到右给出了AlphaGo价值网络计算的每局获胜概率(第四局失利)

- 第三局中AlphaGo在第37手走出了人类选手只有万分之一机会下的一步棋(创新之举)。
- 李世石在第四局78步下出“神之一手”后，AlphaGo仍“感觉”良好，直到87步才反应被骗，随后计算所得获胜概率一直很低。
- 据说目前AlphaGo仍然无法完胜第四局。

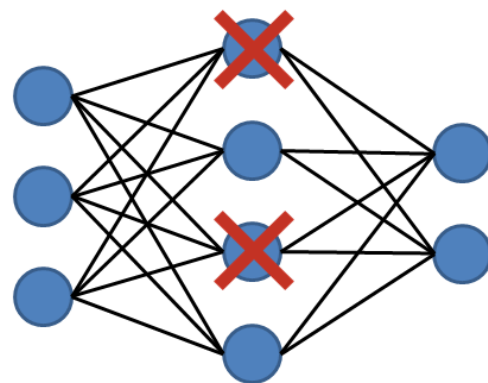
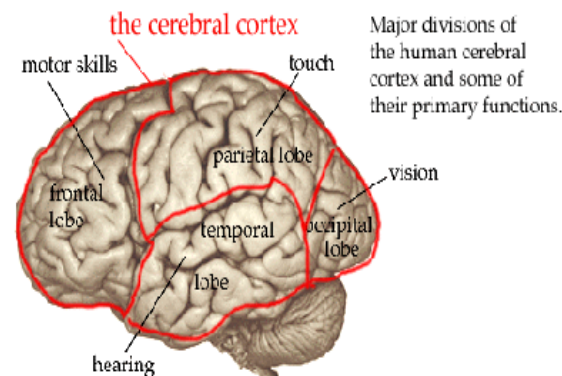
难点2：智能计算中隐性知识和显性知识的协同

随机性对深度学习中极为重要，如dropout、变分编码和异步随机下降

□从 Bayesian 角度解释深度学习中“dropout”机制。

□dropout是一种在训练过程中随机将50%左右神经元节点及其链接丢掉而防止模型过学习的一种策略，从而可对不确定建模。

□注：大脑中只有5%神经元同时被激活和工作。

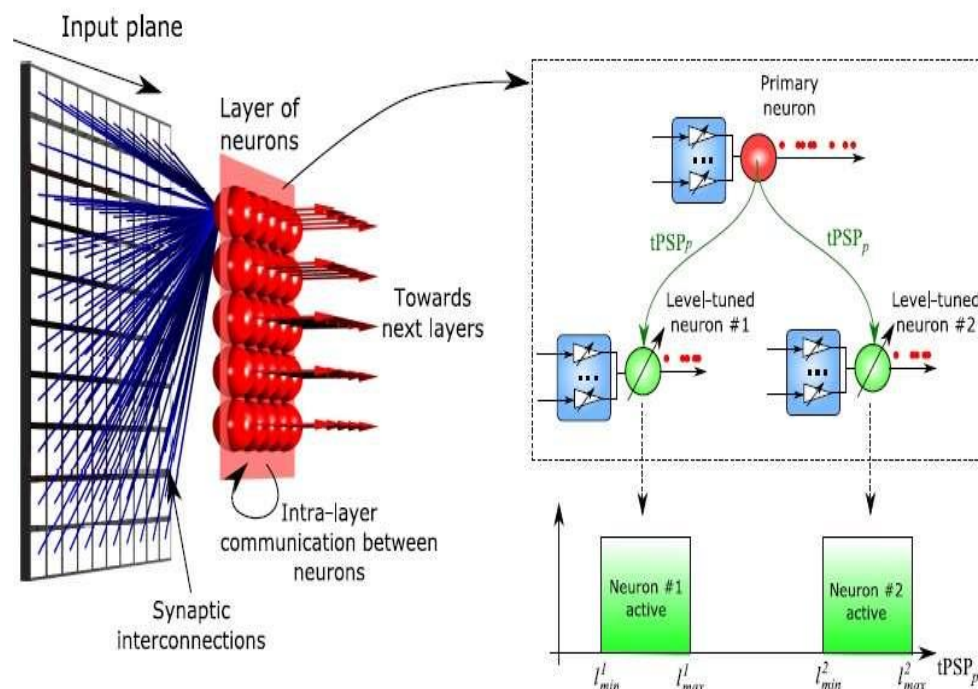


Dropout

难点2：智能计算中隐性知识和显性知识的协同

随机性在智能学习中很重要

- IBM苏黎世研究中心8月初宣布制造出世界上首个人造纳米尺度随机相变神经元 (stochastic phase-change neurons), 可实现高速无监督学习。目前, IBM已经构建了由500个该神经元组成的阵列, 并让该阵列以模拟人类大脑的工作方式进行信号处理。
- 随机性意味着在相同的输入信号下, 多个相变神经元的输出会有所不同。



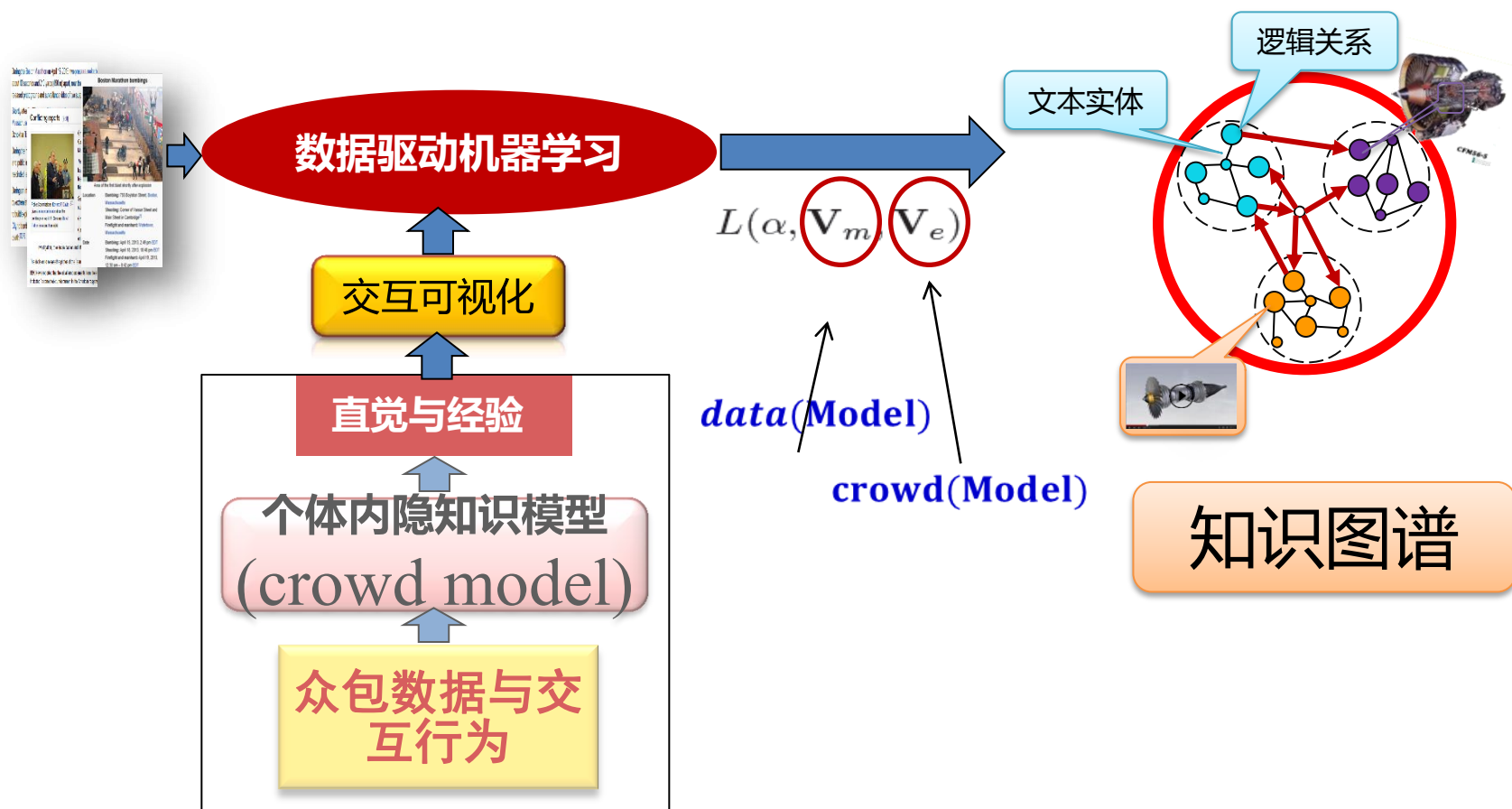
Neuromorphic architecture with level-tuned neurons. The internal state of a primary neuron is used to enable a set of level-tuned neurons.

Tomas Tuma, Angeliki Pantazi, Manuel Le Gallo, Abu Sebastian, Evangelos Eleftheriou, Stochastic phase-change neurons, Nature Nanotechnology, 2016, May

难点2：智能计算中隐性知识和显性知识的协同

□ 有效的协同方式：

□ 可视交互呈现隐性知识，建立人在回路的计算模式。



总结

- 不断出现的数据
- 不断演化的先验
- 不断涌现的群智
- 数据、模型和算法的综合考虑，形成从数据到知识的能力。

谢谢大家
敬请批评指正

Email: wufei@cs.zju.edu.cn