科研与英文学术论文写作

第十二讲：学术研究之创新性研究点
——如何找论文与进行文献调研？

于静 副研究员

中国科学院信息工程研究所

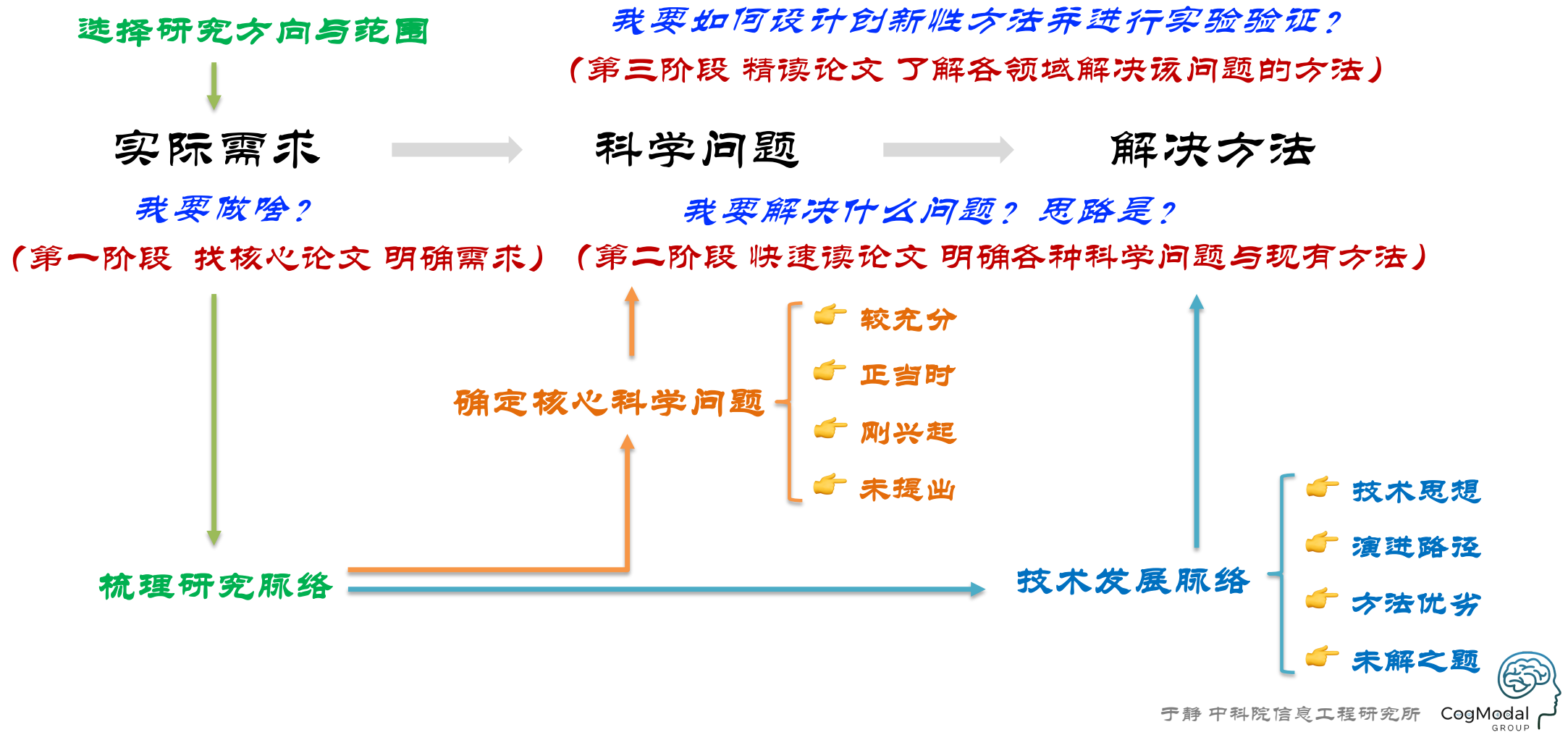
课程主页：https://mmlab-iie.github.io/course/

2022.08 @ Bilibili

中国科学院信息工程研究所
INSTITUTE OF INFORMATION ENGINEERING,CAS

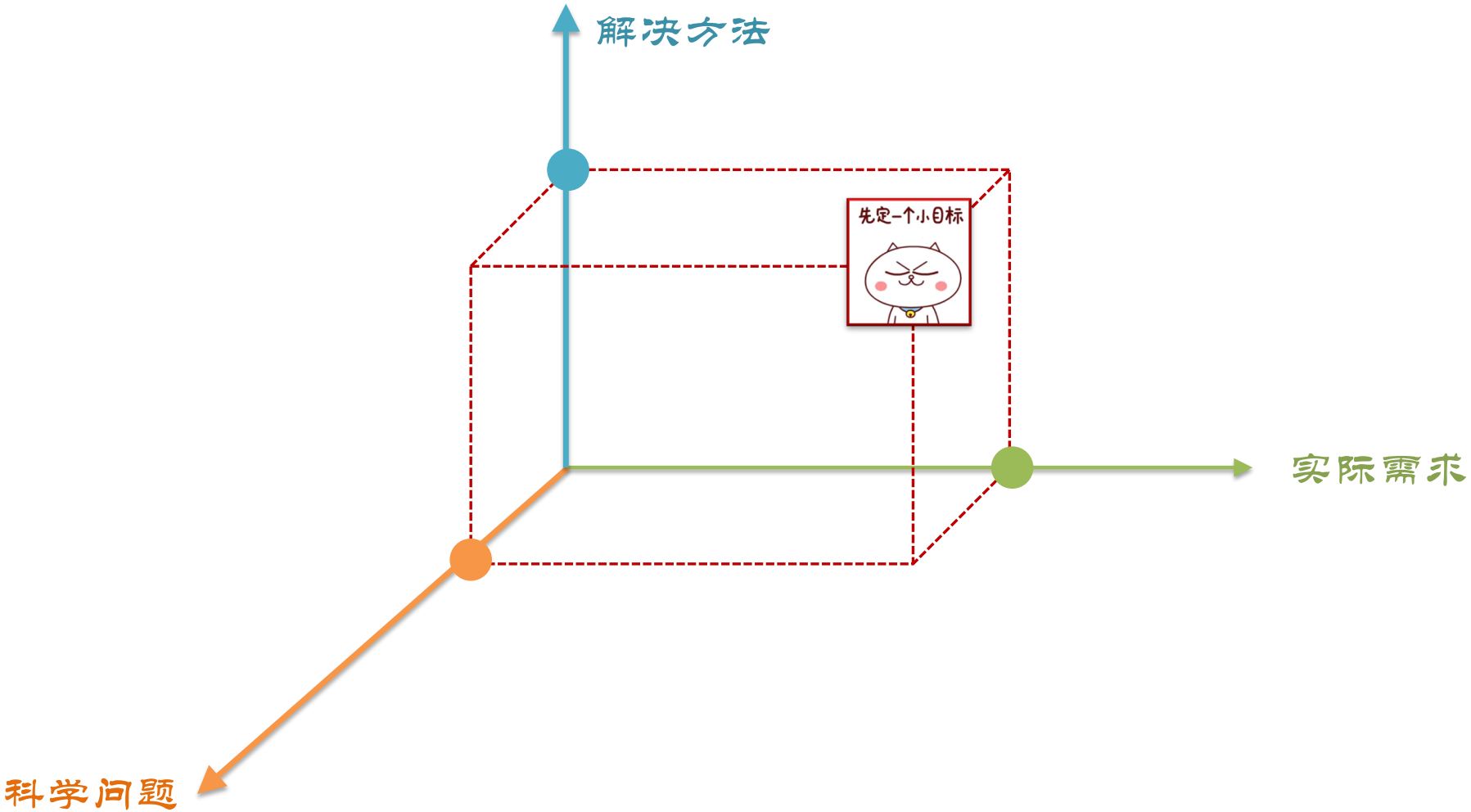
中国科学院大学
University of Chinese Academy of Sciences

文献调研与思考创新性研究点的关系？



于静 中科院信息工程研究所
CogModal GROUP

文献调研与思考创新性研究点的关系？

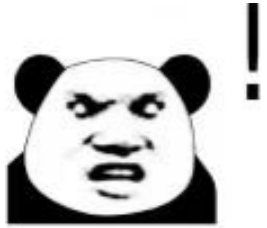


我们文献调研出了什么问题？

老师只给了一篇论文

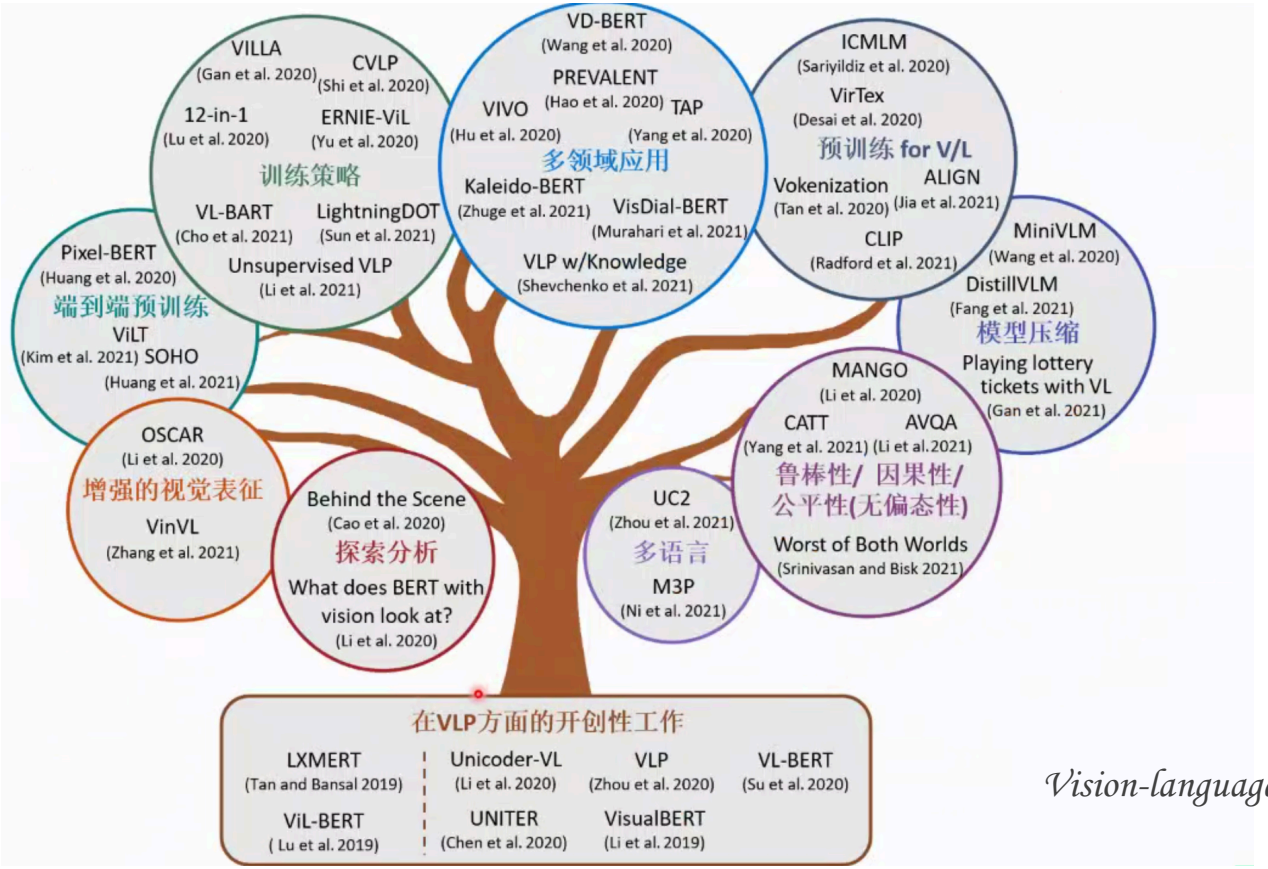
四个月后…

我去！这么多一样的想法！



从开始研究到写论文，只关注1～2篇论文，欠调研！

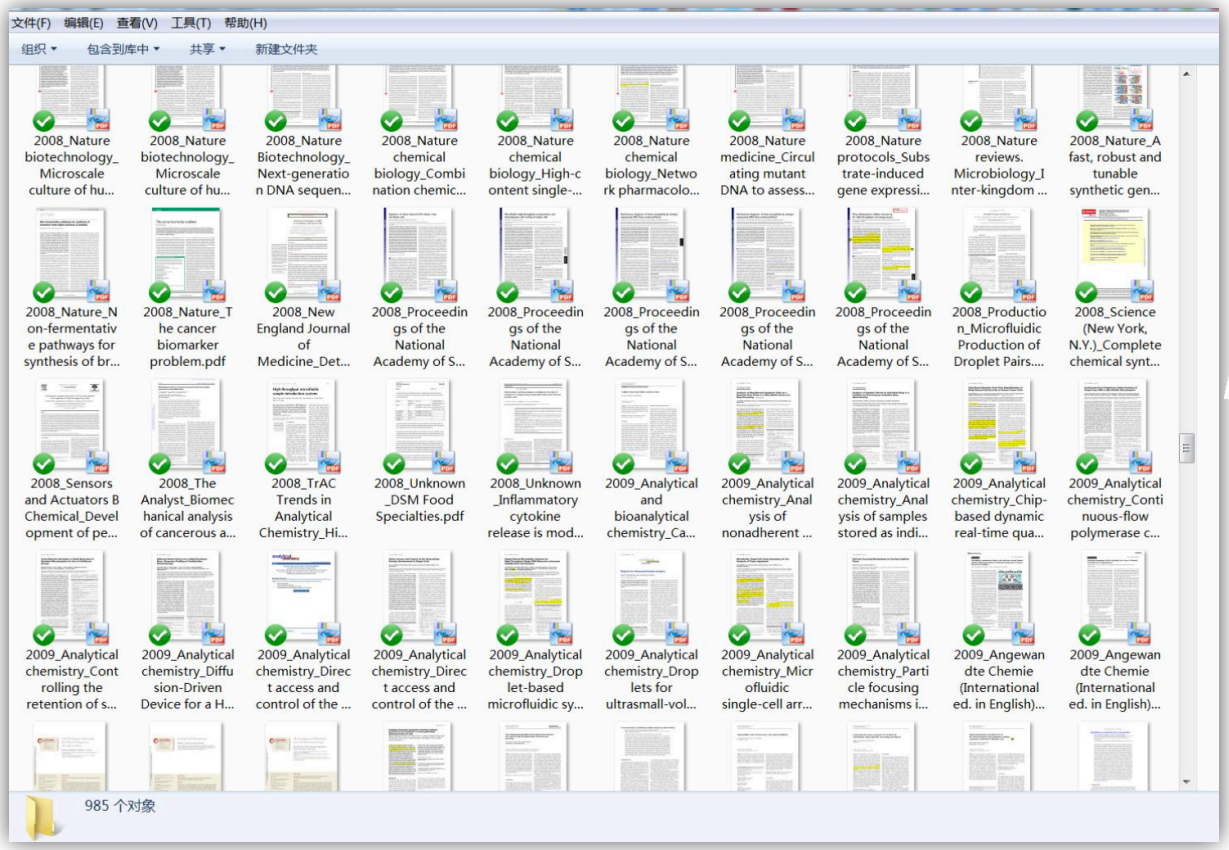
我们文献调研出了什么问题？



Vision-language pre-training

只知道 1～2 个关键词，检索片面，遗漏重要文献！

我们文献调研出了什么问题？

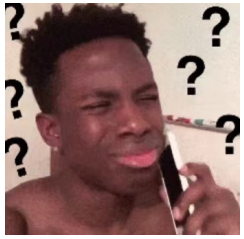


所有信息都新（立意、概念、方法），分不清重点，全看！

于静 中科院信息工程研究所

CogModal GROUP

我们文献调研出了什么问题？



肉菜怎么做？素菜怎么做？
不同调料什么作用？
怎么做一个喜欢口味的新菜？
…

鱼香肉丝 配料、做法、口味

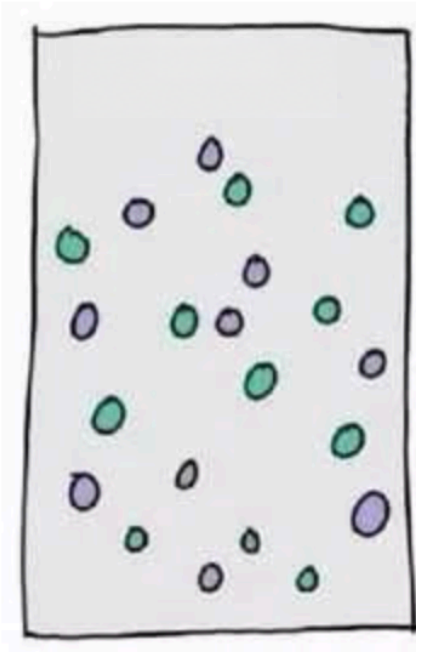
宫保鸡丁 配料、做法、口味

蓑衣黄瓜 配料、做法、口味

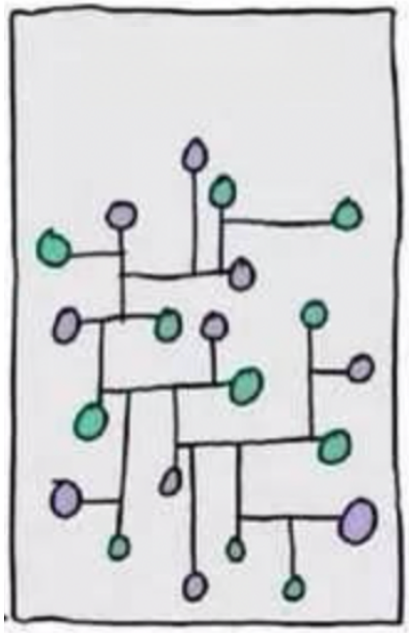
缺少深度思考与理解，看 1 篇到看 10 篇内功提升微小！

我们文献调研出了什么问题？

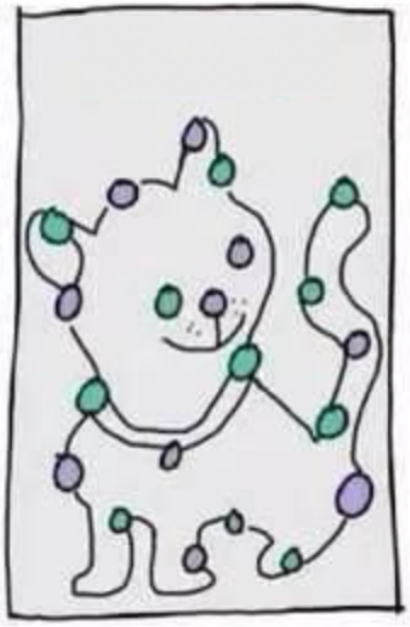
我看的论文



真实的研究体系

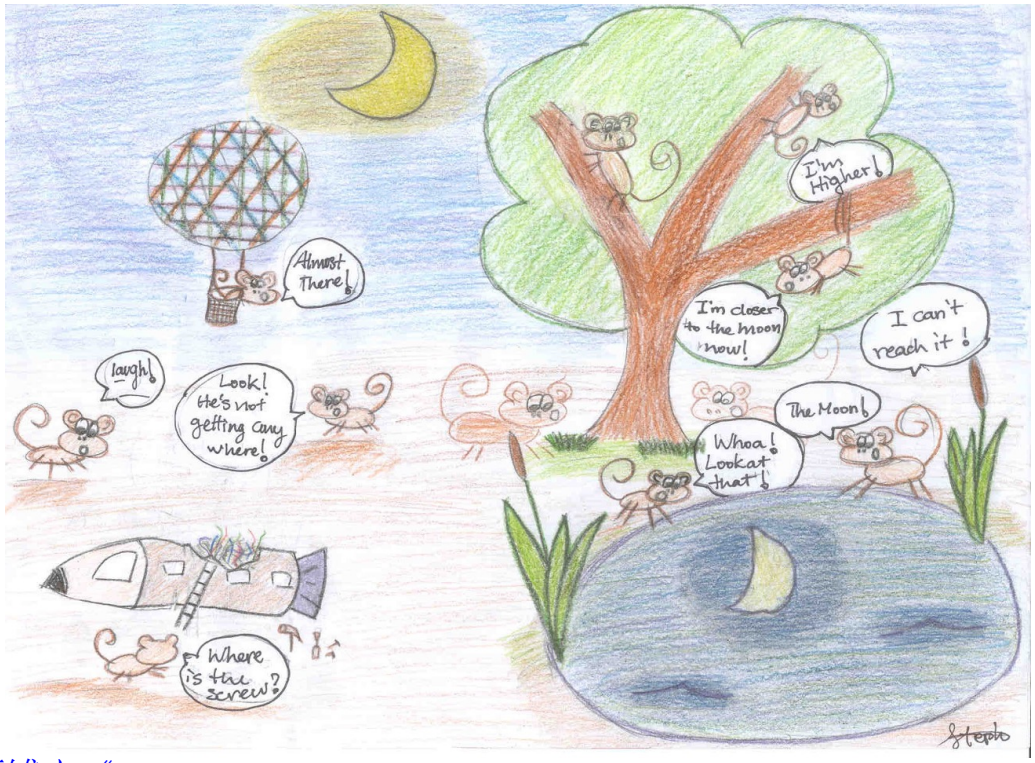


我理解的研究体系



零散看论文，不成体系，且存在认知偏差！

我们文献调研出了什么问题？



credit: 朱松纯老师博文《How to reach the moon: do we know that we are not doing research in the wrong way? (2010)》

难以辨别哪些文献在解决真问题！ 哪些在真解决问题！

我们文献调研出了什么问题？

只看1～2篇论文，没深度！

只搜1～2关键词，没广度！

海量盲目搜论文，没重点！

} 找论文的目标和方式

已有认知读论文，没提升！

零零散散读论文，没体系！

} 读论文的顺序和逻辑

跟风尽信牛论文，没思辨！

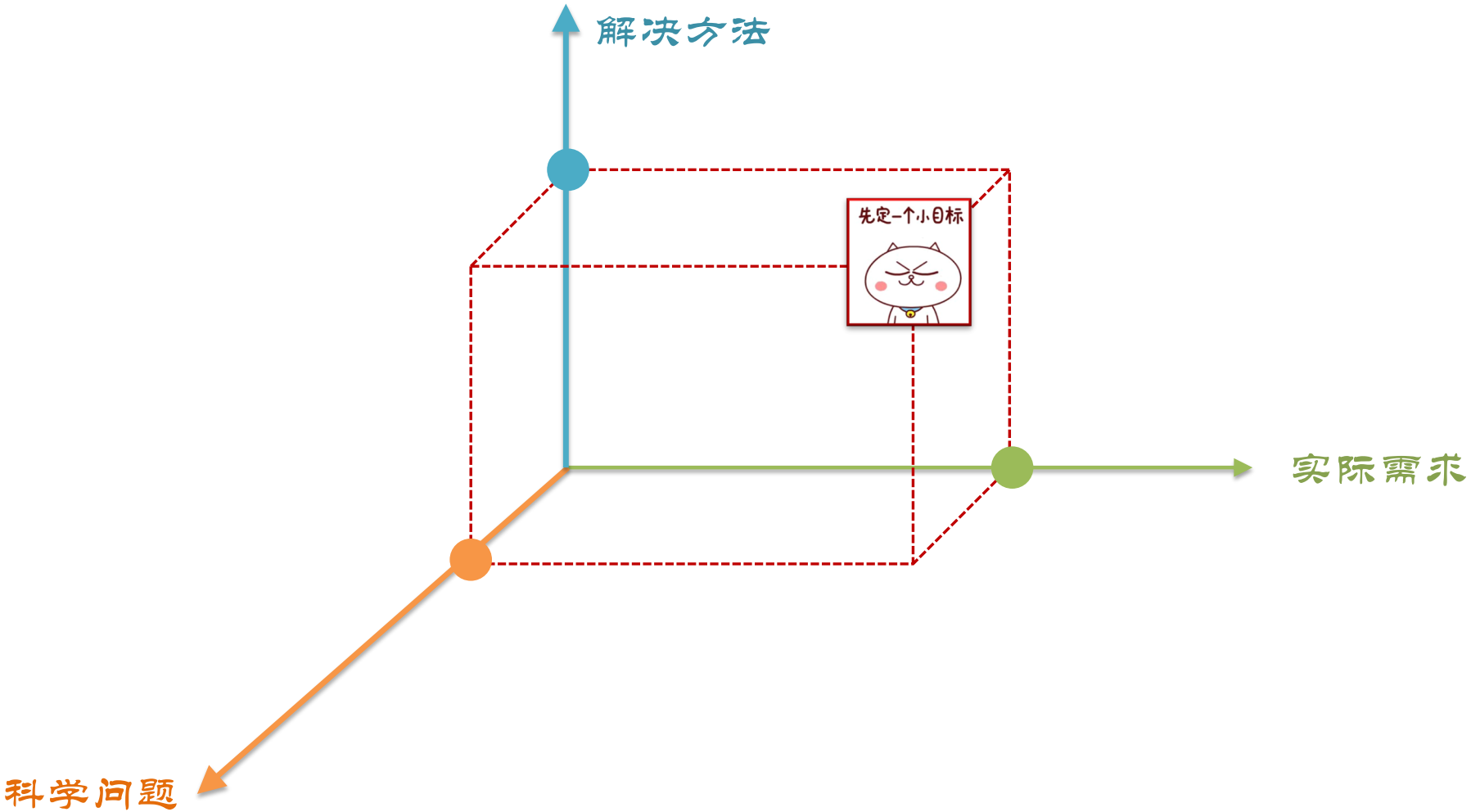
} 讲论文的深度和反馈

…

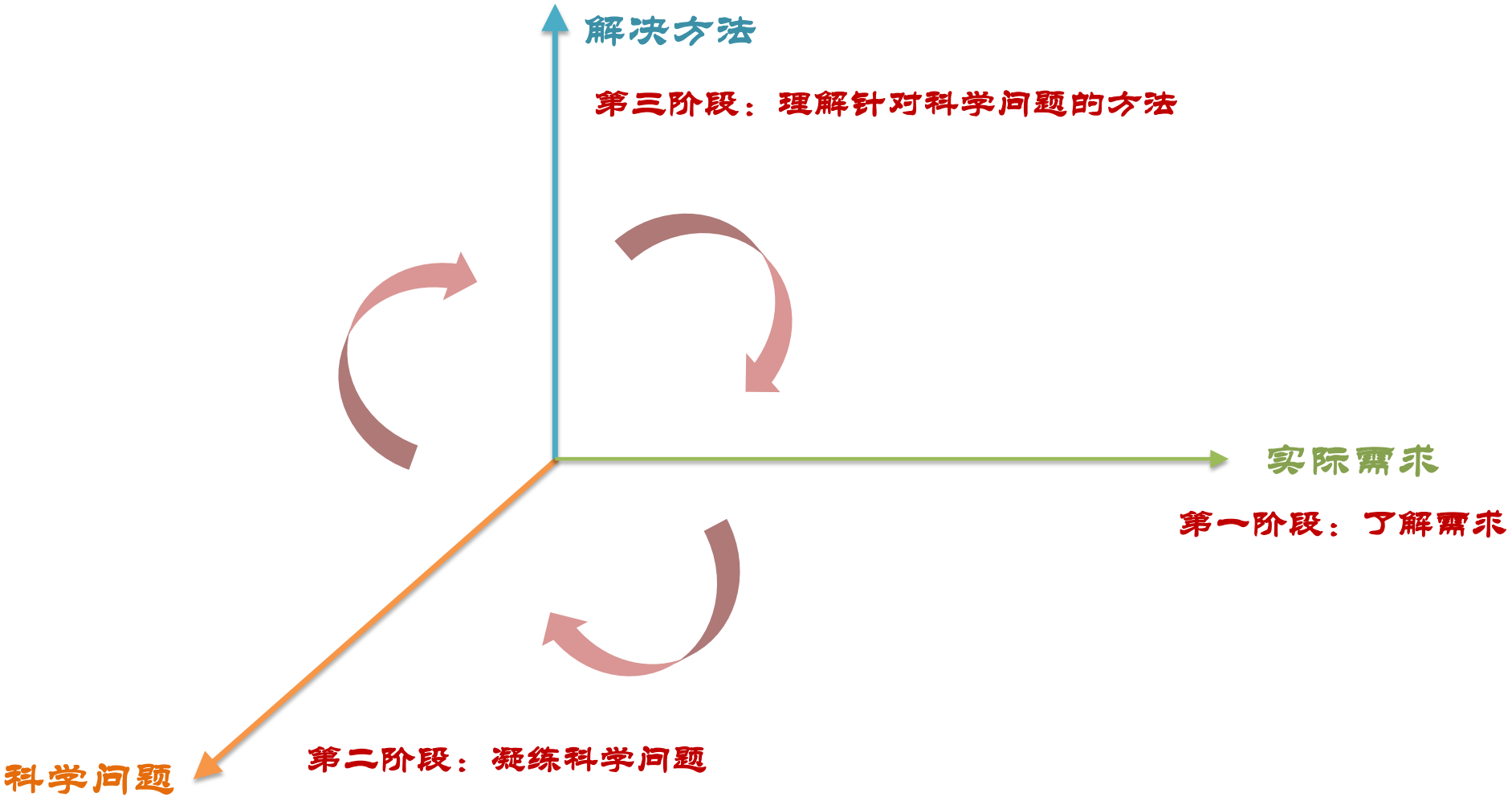
文献调研的

基本思路和方法

文献调研的目标——找到要解决的需求、问题和对应方法

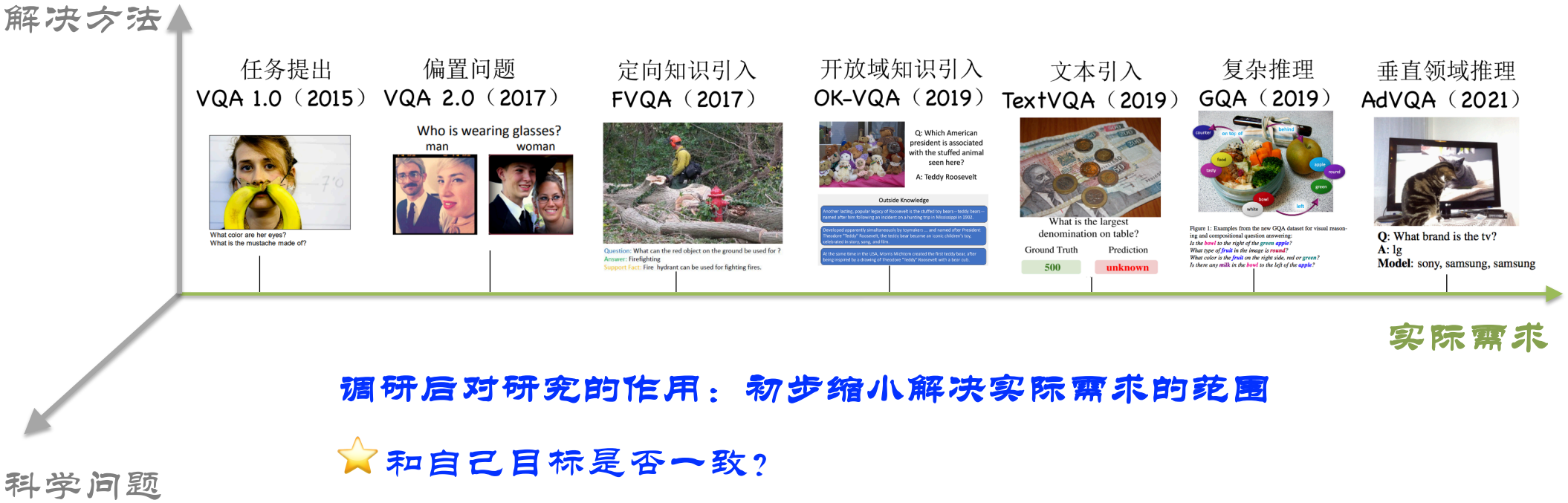


文献调研的基本思路



实际需求的调研目标——任务和技术能力的演进路径

从调研领域重要数据集开始



调研后对研究的作用：初步缩小解决实际需求的范围

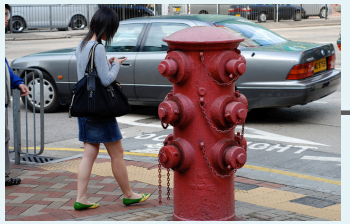
- ☆和自己目标是否一致？
- ☆数据上是否存在问题？
- ☆是否有延续研究空间？

实际需求包括哪些方面？

视觉问答

输入是什么？

图像



问题

图中地面上的红色圆柱体可以用来做什么？

知识

推理模型

希望具备的能力是什么？

如何根据视觉、语言、知识等多模态信息进行推理？

输出是什么？

答案

灭火

如何评价？

准确率 40%

如何调研需求——调研前五问

需求调研 *5W+1H* 思考：

- 任务的提出希望研究具备什么能力的AI算法？（Why）
- 任务对应的输入、输出、评价分别是什么？（What）
- 任务有哪些核心概念、相关概念？（What）
- 任务提出和比较的基准模型的技术思路是什么？（How）
- 达到上述能力，需要解决的科学问题是什么？（What，思考）
- 哪些技术挑战未被解决或未被验证解决？（Where，思考）

如何调研需求

入门阶段

盲目搜索 ×

- 刷顶会期刊
- 刷 arxiv
- 刷学术自媒体等

精准定位 ✓

- 讲座，报告
- 会议Tutorial
- 会议Workshop
- 综述论文

入门阶段从有影响力的经典工作开始阅读！

如何调研需求

近3～5年领域顶会相关 *Topic* 的 *Tutorial*

为什么是3～5年？

- 让思考和总结有“代际感”
  - 不是现在做的，不是过去1～2年内做的
  - 而是过去5年、10年、20年内做的
- 让思考和总结有“路径感”
  - 一生二，二生三，三生万物？
  - 三 = 一 + 二 ？

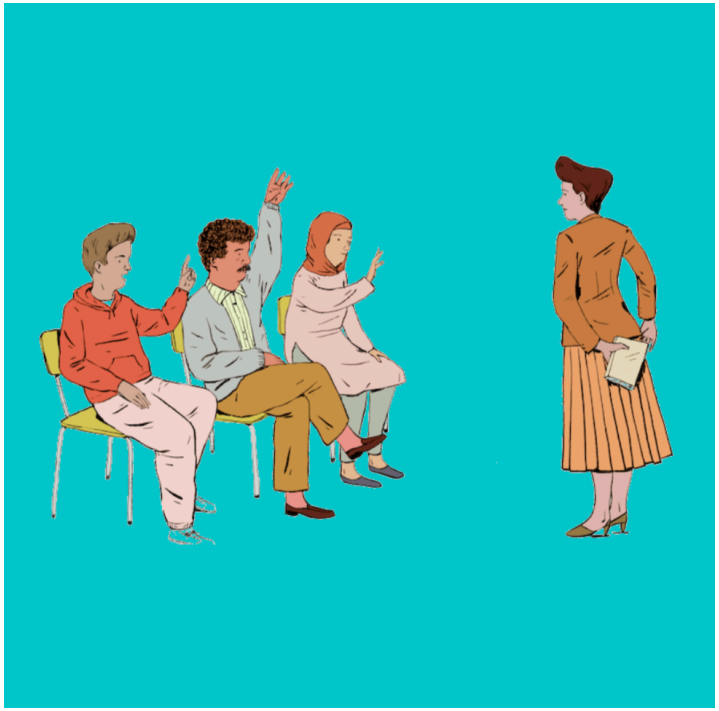
如何调研需求

近3～5年**领域顶会**相关 *Topic* 的 *Tutorial*

为什么是领域顶会？

- 前沿性强，领域认知度深
  - 聚焦过去1～2年热点研究问题
  - 领域近几年国际知名的学者/团队报告
  - *Tutorial* 系统梳理研究问题、发展脉络、技术挑战等
- 综合多个相关领域顶会的不同视角
  - 例如：从NLP、CV、IR等方向顶会搜集关于问答的tutorial

让我们一起来练习
——调研VQA



于静 中科院信息工程研究所

CogModal GROUP

如何调研需求——从调研顶会Tutorial 开始

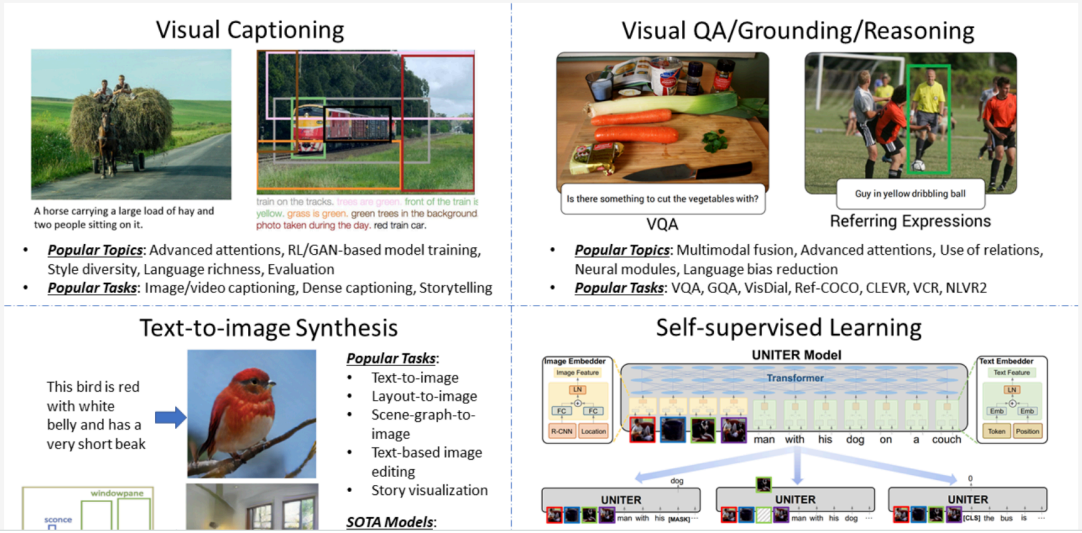
Recent Advances in Vision-and-Language Research

CVPR 2020 Tutorial

Click here to join our tutorial

Time: 06/15/2020, 1:15 - 5:00 PM PDT

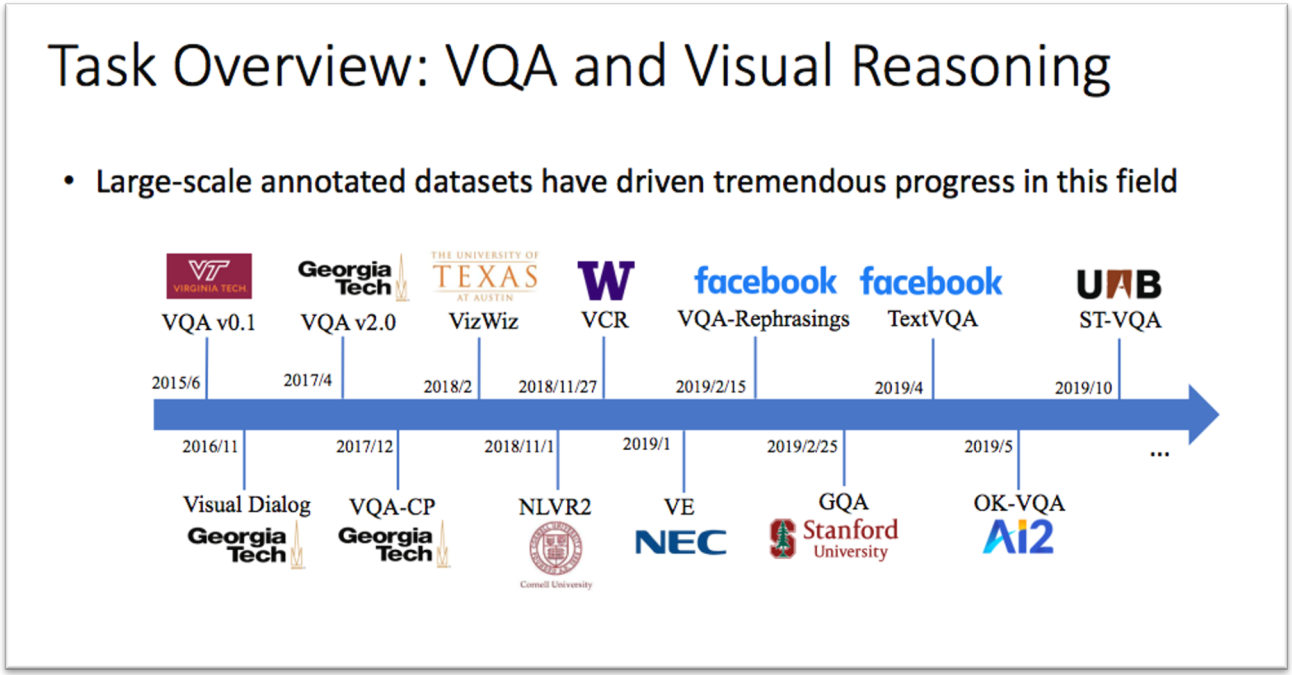
Location: Zoom



https://rohit497.github.io/Recent-Advances-in-Vision-and-Language-Research/

如何调研需求——从调研顶会Tutorial 开始

CVPR 2020 Tutorial： Recent Advances in Vision and Language—Visual Question Answering and Visual Reasoning



- 根据Tutorial了解技术发展脉络
- 下载对应论文，回答 5W+ 1H

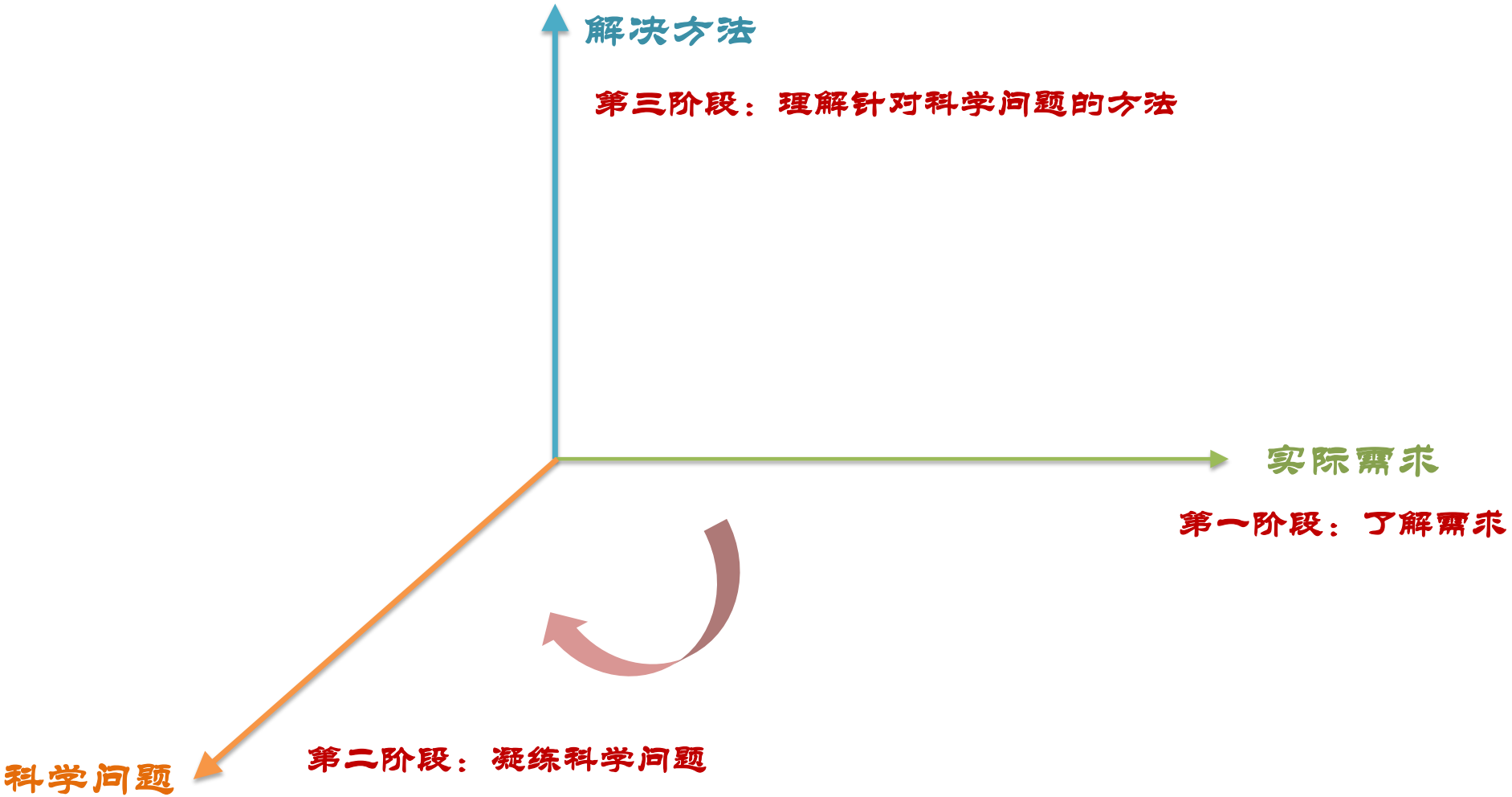
需求调研 *5W+1H* 思考：

- 任务的提出希望研究具备什么能力的AI算法？（Why）
- 任务对应的输入、输出、评价分别是什么？（What）
- 达到上述能力，需要解决的科学问题是什么？（What）
- 任务有哪些核心概念、相关概念？（What）
- 任务提出和比较的基准模型的技术思路是什么？（How）
- 哪些技术挑战未被解决或未被验证解决？（Where）

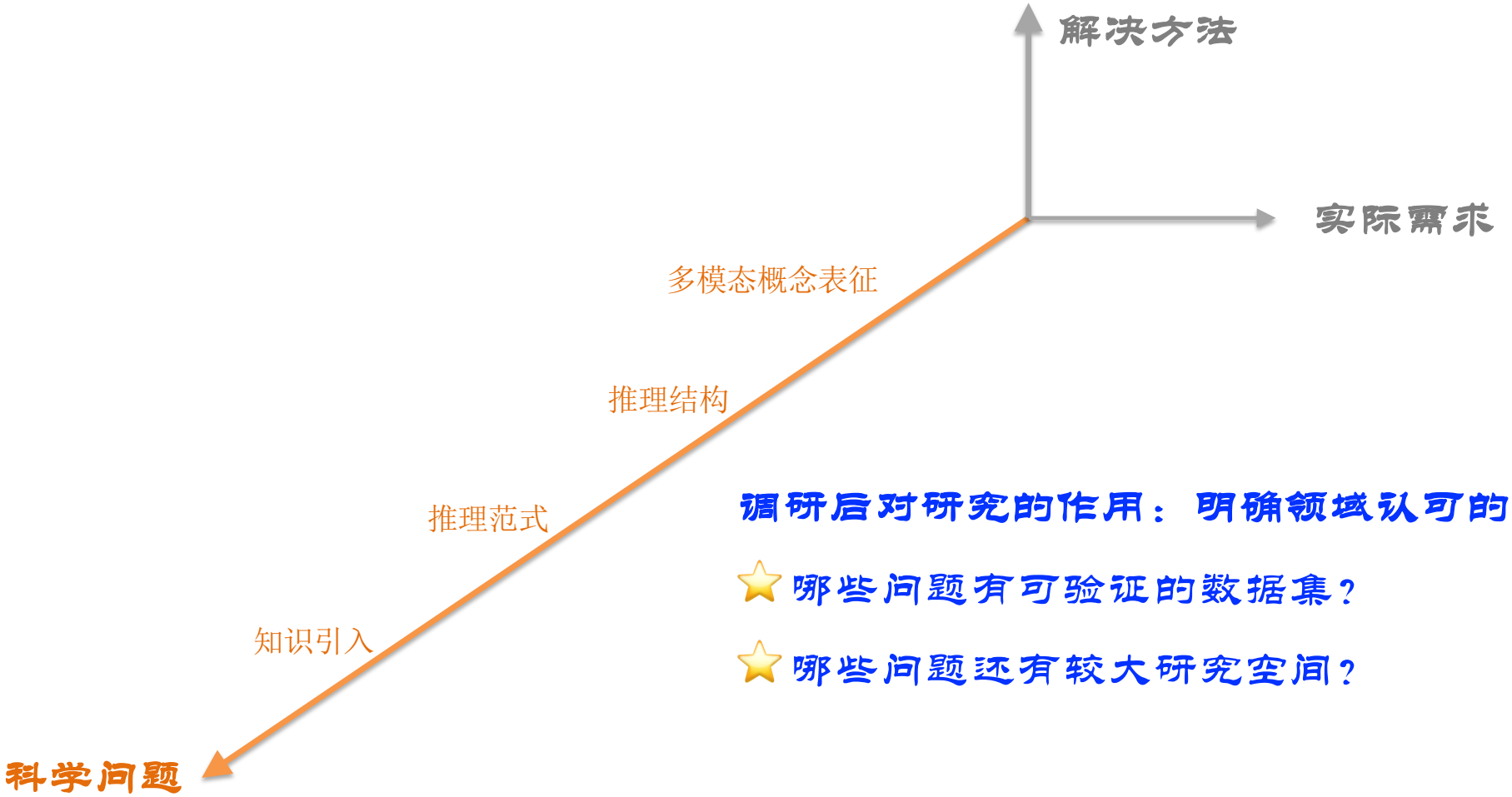
- 根据论文，零散梳理科学问题和方法

CVPR 2020 Tutorial Recent Advances in Vision and Language—Visual Question Answering and Visual Reasoning：
https://rohit497.github.io/Recent-Advances-in-Vision-and-Language-Research/slides/tutorial-part-2-vqa.pdf

文献调研的基本思路



科学问题的调研目标——凝练不同任务的本质问题



调研后对研究的作用：明确领域认可的本质问题

- ⭐ 哪些问题有可验证的数据集？
- ⭐ 哪些问题还有较大研究空间？

如何调研科学问题——调研前四问

科学问题调研 *3W+1H* 思考：

- 任务的科学问题是什么？（What）
- 任务分为哪几个子任务？每个子任务的科学问题是什么？（What）
- 每个科学问题解决的程度：是否有理论、实验或论据支撑？（Where）
- 每个科学问题，解决的技术思路有哪些？（How）

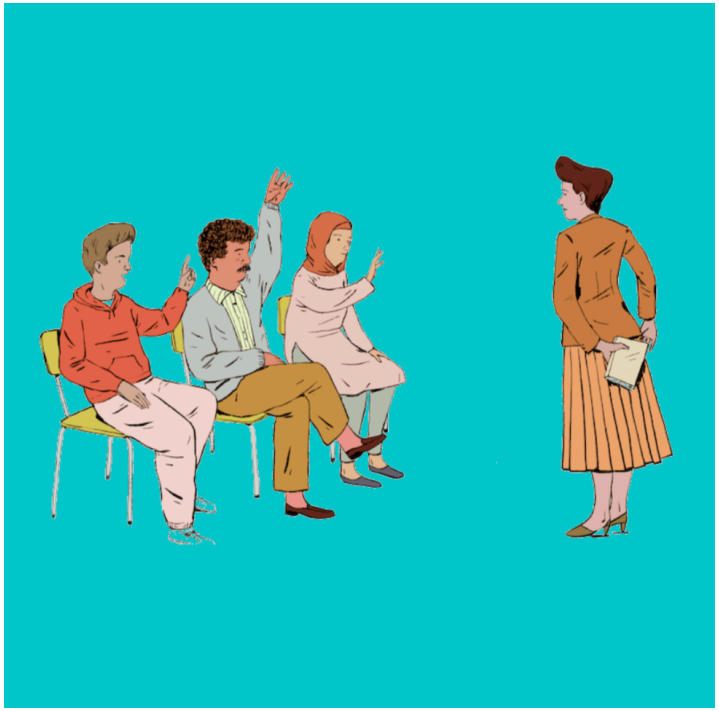
如何调研科学问题？

1. 近3～5年领域顶会相关 Topic 的Tutorial
2. 近3～5年领域顶会相关 Topic 的综述论文

不要依赖Tutorial和综述的总结！科学问题往往不直接阐述！

自己理解概念内涵，批判性接受，从论述间接总结科学问题！

让我们一起来练习
——调研VQA



于静 中科院信息工程研究所

CogModal GROUP

如何调研需求——从调研顶会Tutorial 开始

CVPR JUNE 19-24 2022 NEW ORLEANS LOUISIANA

HOME ORGANIZERS SPONSORS SUBMISSION ATTEND PROGRAM EXPO JOBS

Overview
Main Conference Schedule
Workshop Schedule
Tutorial List
Tutorial Selections
Doctoral Consortium
Reviewer Acknowledgements
Student Activities
Dataset Contributions
CVPR 2022 Paper Awards

TUTORIAL LIST

Tutorials	Primary Contacts	full/half day	Type	Date	Time (AM/PM)
A post-Marrian computational overview of how biological (human) vision works	Li Zhaoping	full	Contributed	6/19	Full day
Affine Correspondences and their Applications in Practice	Daniel Barath	full	Contributed	6/19	Full day
Beyond Convolutional Neural Networks	Neil Houlsby	half			
Building and Working in Environments for Embodied AI	Fanbo Xiang	half			
Contactless Health Monitoring using Cameras and Wireless Sensors	Wenjin Wang	half			
Deep AUC Maximization	Tianbao Yang	half			
Deep Visual Similarity and Metric Learning	Timo Milbich, Jenny Seidenschwarz, Ismail Elezi	half			
Denoising Diffusion-based Generative Modeling: Foundations and Applications	Karsten Kreis, Ruiqi Gao, Arash Vahdat	half			
Evaluating Models Beyond the Textbook: Out-of-distribution and Without Labels	Liang Zheng, Ludwig Schmidt	half			
Graph Machine Learning for Visual Computing	Guohao Li, Guocheng Qian, Jesus Zarzar	half			
High-degree polynomial networks for image generation and recognition	Grigorios Chrysos				
Human-centered AI for Computer Vision	Bolei Zhou				
Imaging Through Atmospheric Turbulence: Theory, Simulation, and R					Contributed
Inside Plato's do					Contributed
Medical Diagnosis Using Computer Vision: Foundations, Advances, Challenges	Soheila Borhani, Reza Borhani	half	Contributed		
Multimodal Machine Learning	Louis-Philippe Morency, Paul Liang, Amir Zadeh	half	Contributed		
Neural Fields in Computer Vision	Yiheng Xie, James Tompkin, Vincent Sitzmann, Srinath Sridhar	full	Contributed		
Performance Measures in Visual Detection and Their Optimization	Emre Akbas, Sinan Kalkan, Kemal Oksuz	half	Contributed		
Recent Advances in Vision-and-Language Pre-training	Zhe Gan	full	Contributed		
Sparsity Learning in Neural Networks and Robust Statistical Analysis	Yanwei Fu, Yuan Yao	half	Contributed		
Applications - Rendered.ai			track		
How to get quick and performant model for your edge application. From data to application	Paula Ramos, Raymond Lo	half	Industry-track		
OpenMMLab: A Foundational Platform for Computer Vision Research and Production	Kai Chen	half	Industry-track		
Towards always-on egocentric vision research using Meta's Aria glasses	Zhaoyang Lv, Edward Miller, Jeff Meissner	half	Industry-track		
Computational Imaging	Laura Waller, Aviad Levis	half	Invited		
Ego4D tutorial (joint with workshop)	Kristen Grauman et al.	half	Invited		

Tutorial 视角：
技术挑战 & 核心问题

Multimodal Machine Learning

Recent Advances in Vision-and-Language Pre-training

Tutorial 视角：
技术路线

CVPR 2022 tutorial list: https://cvpr2022.thecvf.com/tutorial-list

于静 中科院信息工程研究所 CogModal GROUP

如何调研需求——从调研顶会Tutorial 开始

希望具备的能力！

Multimodal Machine Learning ——Reasoning

Schedule

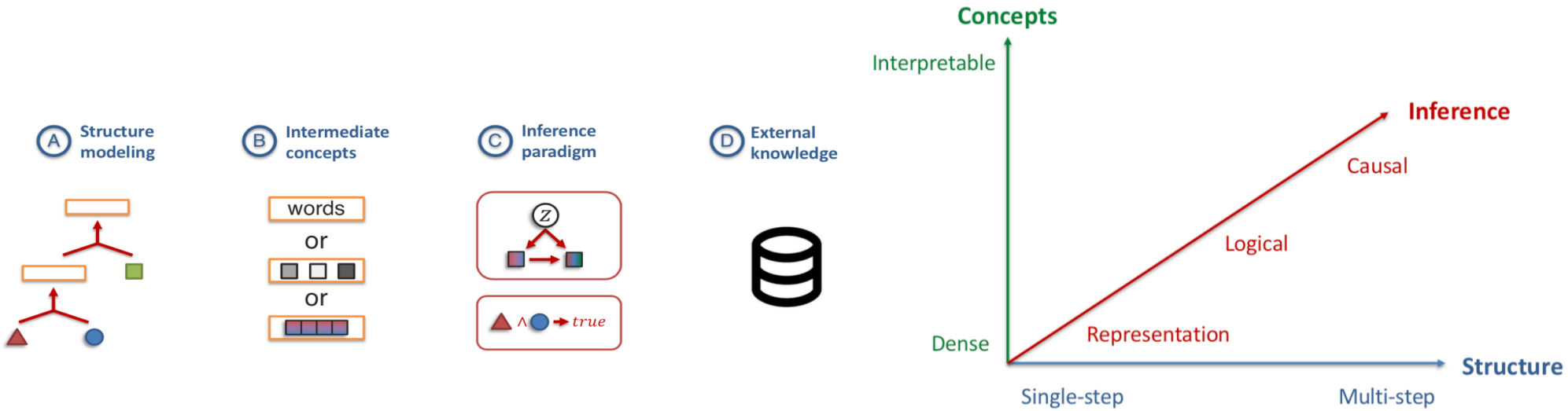
** Exact topics and schedule subject to c

Part	Topics	Readings
1	**Introduction** [slides] [video] • What is Multimodal? Definitions, dimensions of heterogeneity and cross-modal interactions. • Historical view and multimodal research tasks. • Core technical challenges: representation, alignment, transference, reasoning, generation, and quantification.	• Multimodal Machine Learning: A Survey and Taxonomy • Representation Learning: A Review and New Perspectives
2	**Representation** [slides] [video] • Representation fusion: additive, multiplicative, non-linear, complex fusion strategies. • Representation coordination: contrastive learning, vector-space models, canonical correlation analysis. • Representation fission: factorization, component analysis, clustering.	• Multiplicative Interactions and Where to Find Them • Multimodal Learning with Deep Boltzmann Machines • Tensor Fusion Network for Multimodal Sentiment Analysis • Gated Multimodal Units for Information Fusion • Does My Multimodal Model Learn Cross-modal Interactions? It's Harder to Tell Than You Might Think! • On Deep Multi-View Representation Learning: Objectives and Optimization • Unifying Visual-Semantic Embeddings with Multimodal Neural Language Models • Learning Transferable Visual Models From Natural Language Supervision • Learning Factorized Multimodal Representations • Multi-view Clustering: A Survey
3	**Alignment** [slides] [video] • Connections: grounding, optimal transport, distribution matching. • Aligned representations: attention models, multimodal transformers, graph neural networks. • Segmentation: time warping, CTC, temporal alignment, clustering	• Deep Canonical Correlation Analysis • Graph Optimal Transport for Cross-domain Alignment • Deep Canonical Time Warping for Simultaneous Alignment and Representation Learning of Sequences • Deep Visual-semantic Alignments for Generating Image Descriptions • Cross-Modal Generalization: Learning in Low Resource Modalities via Meta-Alignment • ViLBERT: Pretraining Task-Agnostic Visiolinguistic Representations for Vision-and-Language Tasks • Multimodal Transformer for Unaligned Multimodal Language Sequences • Decoupling the Role of Data, Attention, and Losses in Multimodal Transformers
4	**Reasoning** [slides] [video] • Structure: hierarchical, graphical, temporal, and interactive structure, structure discovery. • Concepts: dense and neuro-symbolic. • Inference: logical and causal inference. • Knowledge: external knowledge bases, commonsense reasoning.	• Learning to Compose and Reason with Language Tree Structures for Visual Grounding • The Neuro-Symbolic Concept Learner: Interpreting Scenes, Words, and Sentences From Natural Supervision • A Survey of Reinforcement Learning Informed by Natural Language • Dynamic Memory Networks for Visual and Textual Question Answering • Multimodal Memory Modelling for Video Captioning • ICON: Interactive Conversational Memory Network for Multimodal Emotion Detection • VQA-LOL: Visual Question Answering Under the Lens of Logic • Towards Causal VQA: Revealing and Reducing Spurious Correlations by Invariant and Covariant Semantic Editing • Building a Large-scale Multimodal Knowledge Base System for Answering Visual Queries • KAT: A Knowledge Augmented Transformer for Vision-and-Language
5	**Generation** [slides] [video] • Summarization, translation, and creation. • Model evaluation and ethical concerns.	• VisualGPT: Data-efficient Adaptation of Pretrained Language Models for Image Captioning • DALL·E: Creating Images from Text and DALL·E 2 • On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? • The Social Impact of Deepfakes • What a machine learning tool that turns Obama white can (and can't) tell us about AI bias • What comprises a good talking-head video generation?: A Survey and Benchmark • Defending Against Neural Fake News • Lessons from the PULSE Model and Discussion • Cross-modal Coherence Modeling for Caption Generation • Multimodal Abstractive Summarization for How2 Videos • Extracting Training Data from Large Language Models
6	**Transference** [slides] [video] • Transfer via pre-trained models: pre-trained models, prefix tuning, representation tuning, multitask models. • Co-learning: co-learning via representation and generation.	• Vokenization: Improving Language Understanding via Contextualized, Visually-Grounded Supervision • Foundations of Multimodal Co-learning • Multimodal Prototypical Networks for Few-shot Learning • SMIL: Multimodal Learning with Severely Missing Modality • Multimodal Co-learning: Challenges, Applications with Datasets, Recent Advances and Future Directions • What Makes Multi-modal Learning Better than Single (Provably) • Found in Translation: Learning Robust Joint Representations by Cyclic Translations Between Modalities • Zero-Shot Learning Through Cross-Modal Transfer • 12-in-1: Multi-Task Vision and Language Representation Learning • A Survey of Reinforcement Learning Informed by Natural Language
7	**Quantification** [slides] [video] • Dimensions of heterogenity: modality importance, dataset biases, social biases, noise topologies and robustness. • Cross-modal interactions: interpreting cross-model connections and interactions. • Learning: learning and optimization challenges.	• MultiBench: Multiscale Benchmarks for Multimodal Representation Learning • HighMMT: Towards Modality and Task Generalization for High-Modality Representation Learning • Missing Modalities Imputation via Cascaded Residual Autoencoder • M2Lens: Visualizing and Explaining Multimodal Models for Sentimen Analysis • VL-InterpreT: An Interactive Visualization Tool for Interpreting Vision-Language Transformers • The Mythos of Model Interpretability • Interpretable Machine Learning: Moving From Mythos to Diagnostics • The Disagreement Problem in Explainable Machine Learning: A Practitioner's Perspective • Do explanations make VQA models more predictable to a human? • Women also Snowboard: Overcoming Bias in Captioning Models • Measuring Social Biases in Grounded Vision and Language Embeddings • Multimodal datasets: misogyny, pornography, and malignant

CVPR 2022 "Multimodal Machine Learning" tutorial: https://cmu-multicomp-lab.github.io/mmml-tutorial/schedule/

lodal GROUP

如何调研需求——从调研顶会Tutorial 开始



如何调研科学问题？

1. 近3～5年领域顶会相关 Topic 的Tutorial
2. 近3～5年领域顶会相关 Topic 的综述论文
3. 任务数据集上**经典论文**调研与科学问题归纳

什么是经典论文（什么不是）？

- 是否和你所研究的任务（需求）相关？
- 论文是否在回答本质问题（*why*）？而不只是*incremental*的解决方案（*how*）？
- 论文作者/团队是否在该领域持续深耕？有哪些代表作？
- 论文是否是顶会顶刊？且被广泛认可（*best paper*/被高引/*tutorial*介绍/…）？

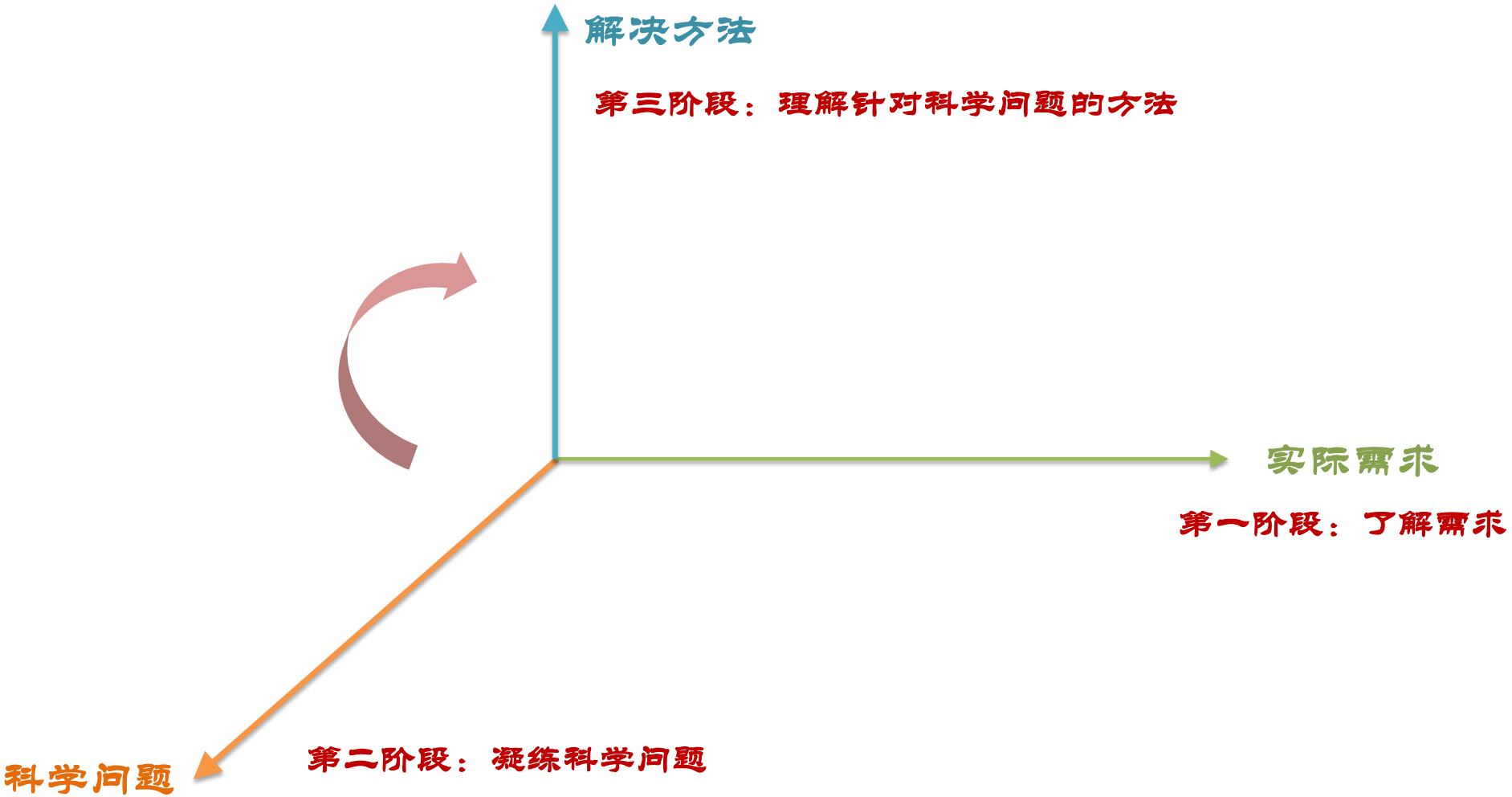
如何调研科学问题？

1. 近3～5年领域顶会相关 Topic 的Tutorial
2. 近3～5年领域顶会相关 Topic 的综述论文
3. 任务数据集上经典论文调研与**科学问题归纳**

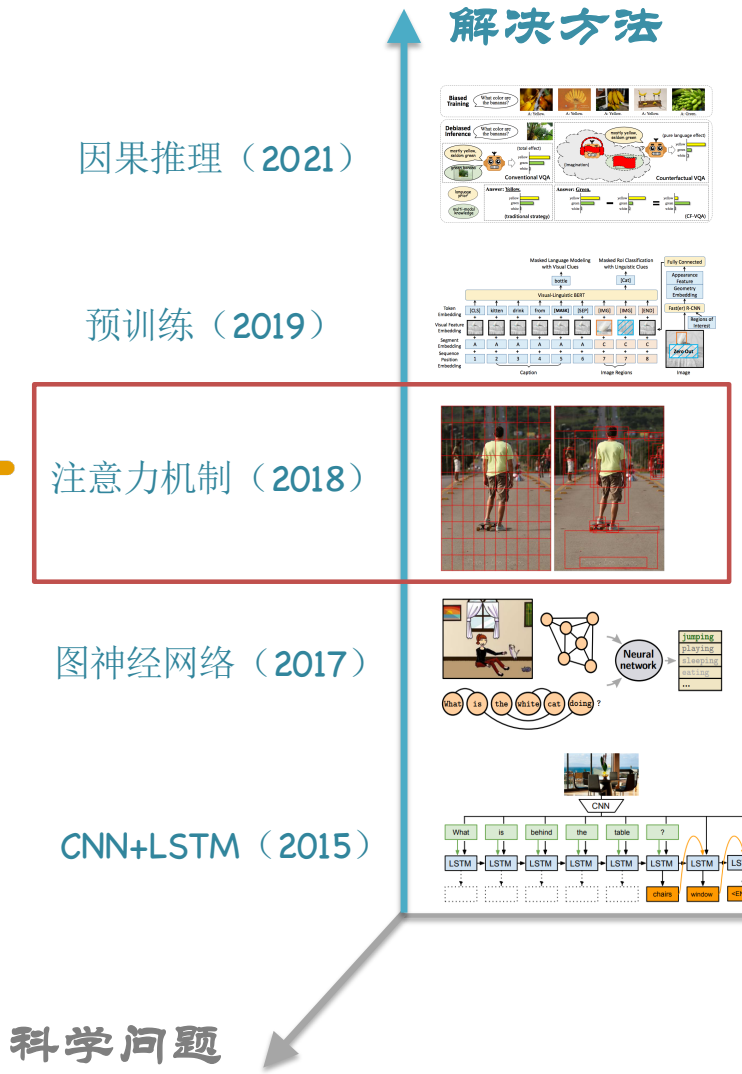
如何归纳科学问题？

- 列出（解读）每篇论文的科学问题（以及子任务的科学问题）
- 绘制思维导图，将论文列在所解决的科学问题后

文献调研的基本思路



解决方法的调研目标——把握不同方法的优势和局限



调研后对研究的作用：明确问题如何创新性解决？

- ⭐现有方法为何不能完善解决问题？
- ⭐哪些未考虑的视角可以更好解决？
- ⭐注意：并非方法有哪些增量改进空间！

如何调研解决方法——调研前四问

解决方法调研 *3W* 思考：

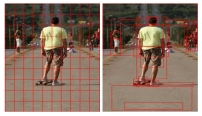
- 方法为什么能解决科学问题？（Why）
  - 三段论：前提、论据、结论
  - 提出新概念的内涵和外延
  - 实验结果如何支撑作者观点和发现
- 方法的基本技术思路是什么？（What）
- 方法对应具有“代际感”的技术是什么？（What）

如何调研解决方法

Follow 相关工作&被引工作

2. Related Work

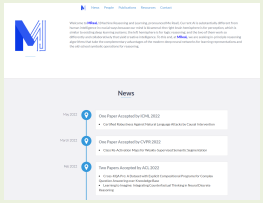
Single-purpose vision-language models. Over the last decade, specialized and effective approaches have been developed for vision-language tasks, including image captioning [10,21,25,33,47,54], phrase grounding [37,38,43], referring expression comprehension [22, 34], visual question answering (VQA) [2,11,16,48,51,55], visual dialog [7], and text-to-image generation [6]. Advances that have pushed the performance envelope include cross-model transformer architectures [45], powerful self-supervised [3, 8, 28] and multitask [41] language models, pretrained visual representations from object and attribute detectors [1, 57] or text conditioned detectors [20], and large-scale image/video-text [19, 27, 39, 56] pretraining.



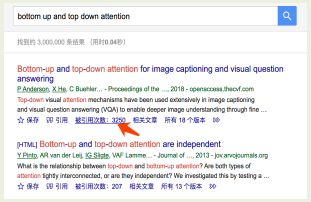
注意力机制（2018）

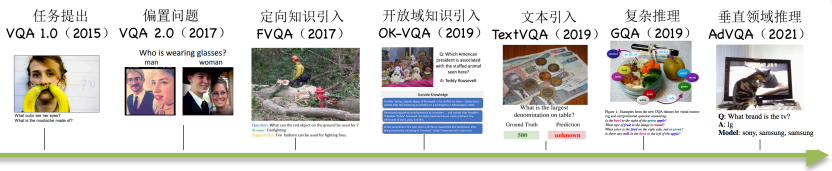
解决方法

Follow研究团队



Follow 被引用论文





实际需求

多模态概念表征

推理结构

推理范式

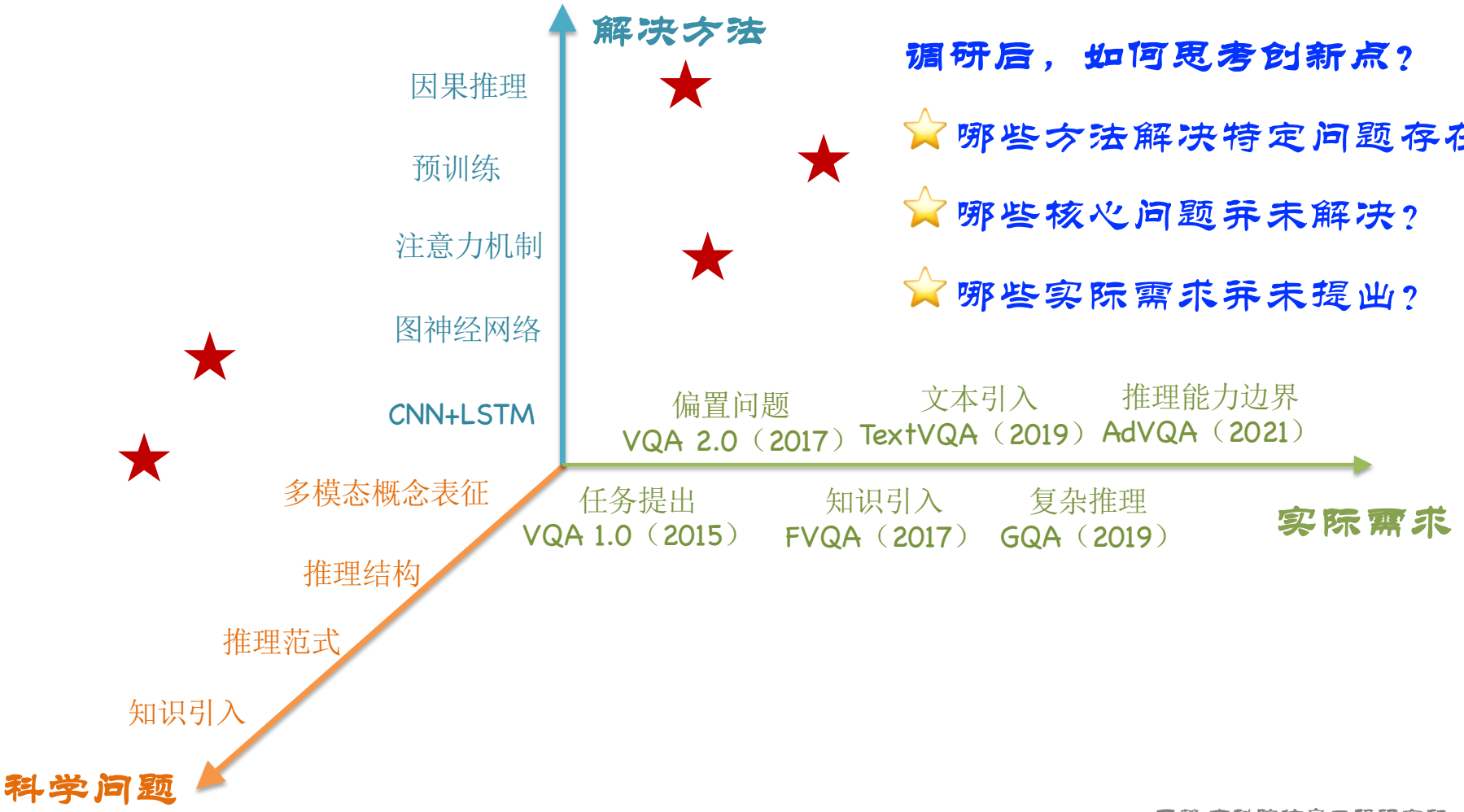
知识引入

科学问题

Follow 顶会顶刊不同Track

Session Title
Machine Learning
Statistical Methods
Optimization Methods

文献调研与思考创新性研究点的关系？



下一讲你将学到什么？


解决方法
因果推理
预训练
注意力机制
图神经网络
CNN+LSTM
偏置问题
VQA 2.0（2017）
文本引入
TextVQA（2019）
推理能力边界
AdVQA（2021）
如何根据科学问题梳理研究脉络？
而不是方法！
多模态概念表征
推理结构
推理范式
知识引入
任务提出
VQA 1.0（2015）
知识引入
FVQA（2017）
复杂推理
GQA（2019）
实际需求
如何归纳科学问题？
科学问题

欢迎大家在B站留言交流！

于静

邮箱：yujing02@iie.ac.cn

课程主页： https://mmlab-iie.github.io/course/

研究组主页：https://mmlab-iie.github.io/

知乎专栏：https://www.zhihu.com/column/c_1284803871596797952

课程主页

研究组主页

知乎专栏

中国科学院信息工程研究所
INSTITUTE OF INFORMATION ENGINEERING,CAS

中国科学院大学
University of Chinese Academy of Sciences