

BAI

智源学者成果展示——自然语言处理

作者 李涓子（清华大学/计算机系知识工程研究室）

2020年6月

项目负责人：李涓子

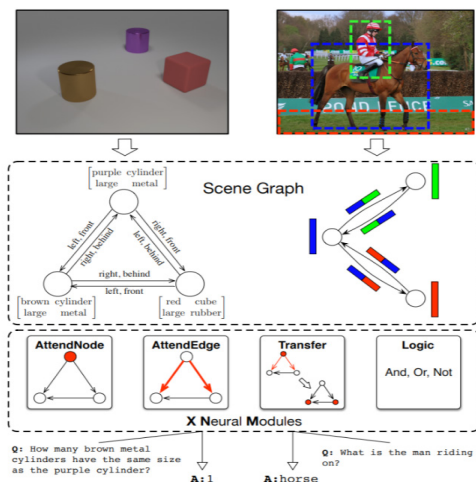
单位 / 团队：清华大学 / 计算机系知识工程研究室

成果名称：基于场景图的可解释显式视觉推理问答

成果简介：

视觉推理问答是融合文本和视觉的典型多模态任务，神经模块网络 (Neural Module Networks) 是目前的主流解决方法。针对已有方法直接处理图片的像素级特征且模块细节需仔细设计的问题，基于人脑推理建立在符号概念上的思想，团队首次提出使用场景图 (Scene Graph) 描述实体及其关系，定义了四种元模块对结点、关系及语义操作的建模，极大地简化了模块内部实现的设计。元模块设计完全基于注意力机制，保证了中间及最终结果具备可解释性。在标准数据集 CLEVR 上的结果表明，当场景图完全正确时，本方法可以达到 100% 的推理准确率，且参数量比其他模型小十倍。与现有的神经模块网络相比，本方法将“高级”推理与“低级”视觉感知区分开来，是可解释的推理方向非常有前景的发展方向。

Jiaxin Shi, Hanwang Zhang, **Juanzi Li**. (2019) Explainable and Explicit Visual Reasoning over Scene Graphs. CVPR2019. (CCF A 类)



The flowchart of using the proposed XNMs reasoning over scene graphs, which can be represented by detected one-hot class labels (left) or RoI feature vectors (colored bars on the right). Feature colors are consistent with the bounding box colors. XNMs have 4 meta-types. Red nodes or edges indicate attentive results. The final module assembly can be obtained by training an off-the-shelf sequence-to-sequence program generator.

图片来源：学者论文

Beijing Academy of Artificial Intelligence



微信关注
北京智源人工智能研究院