软件所在因果图表示学习方面获进展

作者：writer 来源：中国科学院

本文原地址：https://www.iikx.com/news/progress/25587.html

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

近日，中国科学院软件研究所天基综合信息系统重点实验室研究团队撰写的题为*Rethinking Causal Relationships Learning in Graph Neural Networks*的研究成果，被人工智能领域顶级学术会议Association for the Advancement of Artificial Intelligence（AAAI）接收。该研究首次提出了新颖的合成数据集CRCG来分析GNN的因果建模能力，并进一步构建了轻量级GNN因果增强模块R-CAM。

图表示学习的挑战之一在于将复杂的图映射为相对简单、紧凑的向量表示，且可保留重要的结构和语义洞察力。GNN结合神经网络成为应对这一挑战的有力工具。而GNN建模数据和标签之间通常是统计关系，相较于因果关系，无法保证完全反映真实世界，在复杂的图数据中可能会降低可靠性。为增强GNN的因果建模能力，使GNN能够把握数据和标签之间的因果关系，该团队希望通过构造数据集来观察数据中的混杂因素如何影响GNN训练。

基于上述分析，该研究遵循统计因果学的规律，从因果学习的角度全面分析各种GNN模型，构建了人工合成的数据集CRCG（其中，数据和标签之间的因果关系已知且可控）。同时，研究通过理论基础分析进一步保证生成数据的合理性，引入轻量级且高度适应性的GNN模块，增强GNN在各种任务中的因果学习能力。

该团队利用CRCG进行了一系列实验，比较了五种因果增强GNN和传统GNN在不同场景下的性能差异。在图数据中存在混杂因素的情况下，该研究采用因果增强方法的GNN呈现有效性。实验从概率、数据量、显著度三个层面来实证分析混杂因素的影响。结果证明，混杂因素与因果因素之间的概率关系是影响因果增强GNN算法有效性的首要因素。基于这一结果，该团队提出了基于表示的因果关系增强模块R-CAM。研究通过人工合成数据集和真实数据集的多次对比实验发现，集成R-CAM后，多数算法在跨数据集上有平均超过5%的精度提升，这验证了R-CAM在强调数据中的因果信息方面的有效性。

论文链接

Method	P=5%	P=20%	P=40%	P=60%	P=80%	P=100%
ERM	34.21±1.56	28.86±1.17	25.94±1.63	24.43±1.40	23.15±1.10	22.62±1.79
ASAP	31.54±1.67 (-2.67)	26.05±1.40 (-2.81)	23.62±1.23 (-2.32)	23.24±1.08(-1.19)	22.71±1.48 (-0.44)	22.35±1.04 (-0.27)
DIR	38.54±0.99(+4.33)	32.62±1.29(+3.76)	27.15±1.38(+1.21)	26.86±0.87(+2.43)	24.68±0.94(+1.53)	20.71±1.17(-1.91)
CIGA	45.16±1.29(+10.95)	40.48±1.08(+11.62)	26.06±0.86(+0.12)	24.74±1.04(+0.31)	23.05±1.28(-0.10)	19.76±0.95(-2.86)
RCGRL	34.94±0.96(+0.73)	31.96±1.17(+3.10)	24.83±0.69(-1.11)	23.72±0.75(-0.71)	23.51±0.62(+0.36)	21.26±0.53(-1.36)
DISC	41.25±0.83(+7.04)	40.00±0.98(+11.14)	37.00±0.92(+11.06)	35.15±1.35(+10.72)	33.50±1.08(+10.35)	23.60±0.64(+0.98)

Method	Size=1	Size=3	Size=8	Size=15	Size=20	Size=30
ERM	35.40±0.98	32.70±1.12	30.30±0.69	28.80±0.73	27.70±0.57	27.30±1.19
ASAP	26.10±0.73(-9.30)	25.40±1.49(-7.30)	25.00±1.26(-5.30)	24.80±1.17(-4.00)	24.70±1.08(-3.00)	24.20±0.59(-3.10)
DIR	35.80±0.86(+0.40)	33.70±1.13(+1.00)	30.50±0.96(+0.20)	29.40±0.73(+0.60)	28.30±0.82(+0.60)	25.50±0.79(-1.80)
CIGA	28.25±1.31(-7.15)	26.40±0.76(-4.60)	24.90±0.94(-5.40)	24.30±1.24(-4.50)	24.20±0.71(-3.50)	23.00±0.84(-4.30)
RCGRL	32.20±0.93(-3.40)	28.10±0.65(-4.60)	27.90±1.17 (-2.40)	27.80±0.71(-1.00)	25.50±1.06(-2.20)	24.50±1.20(-2.80)
DISC	41.70±0.85(+6.30)	40.10±1.06(+7.40)	39.10±0.62(+8.80)	38.40±1.24(+9.60)	37.10±1.18(+9.40)	36.30±0.95(+9.00)

Method	Very low	Low	Medium	High	Very high	Extremely high
ERM	33.10±0.78	32.90±1.11	31.60±0.89	31.20±0.76	29.50±1.13	27.80±0.97
ASAP	40.50±1.22(+7.40)	38.10±0.87(+5.20)	37.70±0.59(+6.10)	36.40±0.71(+5.20)	36.00±1.04(+6.50)	34.20±0.96(+6.40)
DIR	36.00±1.12(+2.90)	35.70±0.93(+2.80)	34.50±0.74(+2.90)	34.30±1.16(+3.10)	33.10±0.83(+3.60)	33.00±1.18(+5.20)
CIGA	32.50±0.94(-0.60)	31.10±1.07(-1.80)	29.90±1.18(-1.70)	29.80±0.86(-1.40)	25.50±1.23(-4.00)	25.10±0.92(-2.70)
RCGRL	30.10±1.14(-2.00)	29.00±0.98(-3.90)	27.40±1.28(-4.20)	27.20±1.36(-4.00)	25.80±0.76(-3.70)	25.30±1.07(-2.50)
DISC	43.65±0.96(+10.55)	41.00±1.03(+8.10)	39.55±1.42(+7.95)	39.10±0.75(+7.90)	38.95±0.83(+9.45)	37.25±1.25(+9.45)

CRCG实验结果图

研究团队单位：软件研究所

更多 科学进展 请访问 https://www.iikx.com/news/progress/

本文版权归原作者所有，请勿用于商业用途，爱科学iikx.com转发