

# 基于消息传递机制的社会关系理解方法研究

## Social Relationship Understanding via Message Passing Mechanism

李雷来 (17214642)

数据科学与计算机学院  
工程(软件工程)

2019 年 5 月 19 日



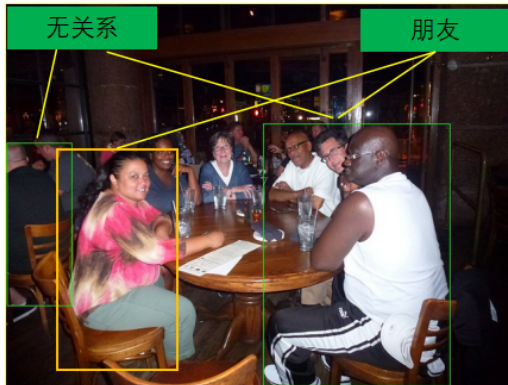
# Table of Contents

- 1 研究背景与相关工作
- 2 PPRN模型
- 3 实验设计与分析
- 4 总结与展望
- 5 附录



# 研究背景

- 社会关系理解的目的是推断出图片中人之间的社会关系，输入是图片和人的包围盒坐标
- 不同的社会关系是相互独立的，由关系域再到具体的社会关系。



# 相关工作

## 社会关系理解模型

- Dual-glance[2]利用人周边的物体区域来帮助判断人之间的社会关系
- GRM[3]通过引入物体类别和关系类别共现频率的先验知识，进而提高预测效果



# 相关工作

## 社会关系理解模型

- Dual-glance[2]利用人周边的物体区域来帮助判断人之间的社会关系
- GRM[3]通过引入物体类别和关系类别共现频率的先验知识，进而提高预测效果



# 相关工作

## 社会关系理解模型

- Dual-glance[2]利用人周边的物体区域来帮助判断人之间的社会关系
- GRM[3]通过引入物体类别和关系类别共现频率的先验知识，进而提高预测效果

## 模型缺点

- 场景中的关系往往存在关联，但现有工作忽略人对关系之间的信息交互。
- 引入的周边物体特征会把图中所有的关系预测为一样，但存在一张图片包含多种关系类别。
- 额外的检测模型检测出物体区域与类别并不一定正确，引入人周边物体的信息存在噪声。

# 社会关系图谱

- 对于图片中的所有人对的社会关系的检测任务，抽象为图片的社会关系图谱生成。

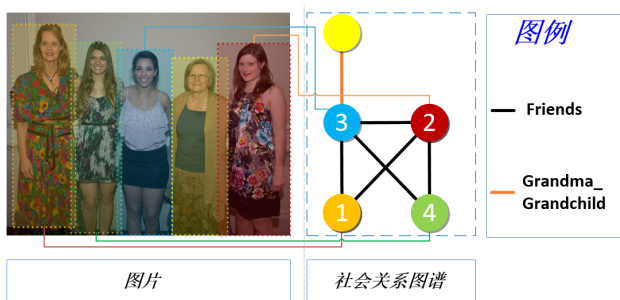


Figure 2: PIPA-relation[1]数据集中的一个图例



# 挑战

## 挑战

- 为利用上人对关系间交互的特征，如何设计能捕获人对间社会关系交互的机制。
- 如何避免引入不确定的信息，并有效的利用上数据集中已有的标注信息来提升检测效果。

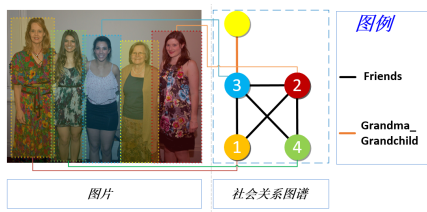


Figure 3: PIPA-relation[1]数据集中的一个图例





# 本文工作

- 对于社会关系理解任务，提出了新的概念：社会关系图谱
- 为捕获人对间交互的信息，本文提出了一个迭代的人对关系网络。
- 提出了一个多任务的损失函数，包括关系损失和关系域损失，并在PISC和PIPA-Relation数据集上验证了PPRN模型的有效性。



# Table of Contents

- 1 研究背景与相关工作
- 2 PPRN模型**
- 3 实验设计与分析
- 4 总结与展望
- 5 附录



# 问题定义

- 图片中所有人与人之间的社会关系定义为社会关系图谱，图的每个节点表示一个人对的关系。
- 图谱中关系节点的类别预测可以认为一个图推理问题，并且消息传递是图推理的一种方式。

社会关系理解问题可以定义为公式??:

$$\begin{aligned} x^* &= \operatorname{argmax}_x Pr(x \mid I, B_i) \\ Pr(x \mid I, B_i) &= \prod_{i=1}^N \prod_{j \neq i}^N Pr(x_{i \rightarrow j}^{relation} \mid I, B_i) \end{aligned} \quad (1)$$



# PPRN模型框架图

PPRN包括三个模块: 特征提取模块、消息传递模块、多任务损失模块。

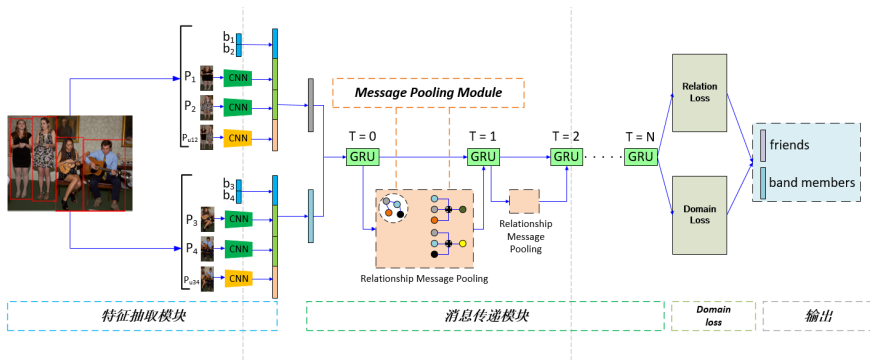
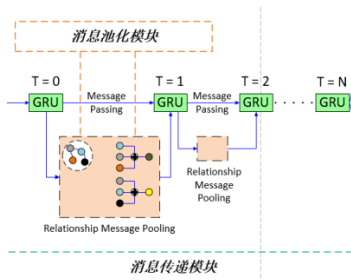


Figure 4: PPRN的模型框架图



# 消息传递模块

- PPRN采用RNN来解决推理问题，由于GRU的简单和有效，RNNs的神经元采用GRU Unit。
- 第 $t$ 步的隐藏层状态表示当前社会关系图谱的关系节点，每一步GRU的输入来自消息池化模块的输出。
- 消息池化模块完成不同人对关系节点的信息传递。



# 消息池化函数

- 由于每个GRU会接收多个其它节点的消息，需要一个聚集函数来结合所有的消息得到一个有意义的表征，消息池化函数为每个收到的消息计算一个权重并利用权重来结合这些收到的消息。
- 对于在第t步的节点i，其前一步的GRU Unit的隐藏层状态为 $h_{i,t-1}$ ，在第t步来自其他节点的消息表示为 $m_{i,t}$ ，并且 $m_{i,t}$ 为下一步GRU的输入 $x_t$ 。

$$m_{i,t} = \sum_{j \neq i} \sigma(\mathbf{w}^T [\mathbf{h}_{i,t-1}, \mathbf{h}_{j \rightarrow i, t-1}]) \mathbf{h}_{j \rightarrow i, t-1} \quad (2)$$



# Table of Contents

- 1 研究背景与相关工作
- 2 PPRN模型
- 3 实验设计与分析
- 4 总结与展望
- 5 附录



# 数据集

Table 1: PISC、PIPA-relation数据集的统计表

数据集	Train	Valid	Test	#train	#valid	#test
PISC-coarse	13142	4000	4000	14536	2 36	15497
PISC-fine	16828	500	1250	55400	1505	3691
PIPA-relation	5857	261	2452	13729	709	5106

大多数图片只包含一种关系类型，少数包含2中或多种。

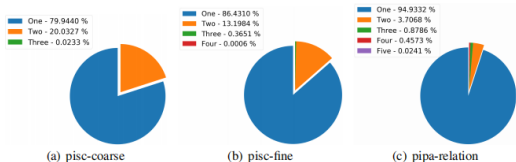


Figure 6: PISC,PIPA-Relation的关系类别





# 实验结果

Table 2: PISC-coarse数据集实验结果, 单位为百分比(%)

模型	Intimate	Non-Intimate	No Relation	mAP
Union CNN [4]	72.1	81.8	19.2	58.4
Pair CNN [2]	70.3	80.5	38.8	65.1
Pair CNN + BBox + Union [2]	71.1	81.2	57.9	72.2
Pair CNN + BBox + Global [2]	70.5	80.0	53.7	70.5
Dual-glance [2]	73.1	<b>84.2</b>	59.6	79.7
GRM [3]	81.7	73.4	65.5	<b>82.8</b>
PPRN	<b>81.9</b>	67.3	<b>74.7</b>	81.8



# 实验结果

Table 3: PISC-fine数据集实验结果，单位为百分比(%)

模型	Friends	Family	Couple	Professional	Commercial	No Relation	mAP
Union CNN [4]	29.9	58.5	70.7	55.4	43.0	19.6	43.5
Pair CNN [2]	30.2	59.1	69.4	57.5	41.9	34.2	48.2
Pair CNN + BBox + Union [2]	32.5	62.1	73.9	61.4	46.0	52.1	56.9
Pair CNN + BBox + Global [2]	32.2	61.7	72.6	60.8	44.3	51.0	54.6
Dual-glance[2]	35.4	68.1	76.3	70.3	57.6	60.9	63.2
GRM [3]	59.6	64.4	<b>58.6</b>	76.6	39.5	67.7	68.7
PPRN	<b>61.0</b>	67.1	56.2	<b>76.9</b>	46.0	68.1	69.7
PPRN-Ld(PPRN+domain loss)	58.2	<b>68.0</b>	<b>74.6</b>	63.3	<b>67.6</b>	<b>70.3</b>	<b>72.0</b>



# PIPA-relation实验结果与分析

Table 4: PIPA-relation实验结果，单位为百分比(%)

模型	accuracy
Two stream CNN [5]	57.2
Dual-Glance [2]	59.6
GRM [3]	62.3
PPRN	<b>64.7</b>
PPRN+d(PPRN+domain loss)	63.6

## 结论与分析

- PPRN在fine-level的识别任务上超过了之前所有的模型，但是在coarse-level轻微的低于之前最优模型。
- 通过引入关系域损失，模型在PISC-fine数据集上有显著的提高。模

# 实验结果分析

Table 5: 不同池化策略和引入周边物体区域的实验结果

模型	PISC coarse		PISC fine	
	accuracy	mAP	accuracy	mAP
RCNN	-	63.5	-	48.4
PPRN(max pooling)	74.3	80.8	64.1	68.1
PPRN(avg pooling)	74.6	80.1	63.8	68.3
PPRN(attention)	<b>75.1</b>	<b>81.8</b>	<b>65.7</b>	<b>69.7</b>
PPRN+d	-	-	<b>66.2</b>	<b>72.0</b>
PPRN+d+obj	74.9	81.2	65.3	69.1

## 结论与分析

- RCNN, Dual-glance和GRM模型均引入额外的Faster-RCNN模型来提取图片人周围的问题特征,但是在PPRN中引入后效果反而出现轻微下降。

# 样例分析

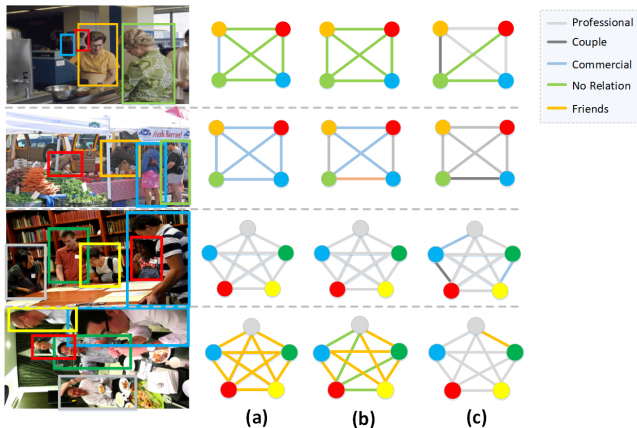


Figure 7: (a)标注(b) PPRN (c) GRM



# Table of Contents

- 1 研究背景与相关工作
- 2 PPRN模型
- 3 实验设计与分析
- 4 总结与展望**
- 5 附录



# 总结与展望

## 总结

- 提出一个生成图片社会关系图谱的人对关系网络，PPRN利用RNNs在一个在不同人对间进行推理，在特征提取的基础上进一步提升了识别的效果。
- 提出引入关系域损失，提升关系识别的效果
- 在现有的大型社会关系理解数据集验证了PPRN模型的有效性。

## 展望

- 引入更多的未标注人的标签信息，例如人的年龄和性别。
- 将当前方法应用到视频理解的任务上，例如视频的人物关系检测。



# Table of Contents

- 1 研究背景与相关工作
- 2 PPRN模型
- 3 实验设计与分析
- 4 总结与展望
- 5 附录





# 攻读硕士学位期间科研成果

- ① 中山大学.一种基于二次主题空间投影的场景图谱低维空间嵌入方法; (专利, 学生第一作者, 申请未公开)(与本文第1章关于场景图谱的内容相关)
- ② Understanding Social Relationship with Person-pair Relations. 投稿于IJCAI-2019(CCF-A类会议, 学生第一作者, 录用边缘)(与本文第3、4章相关)



# 盲审意见修改

- ① 此外，文章的图是否部分引用了现有文献的？这个需要作者检查，如果用了别人的图，需要加标注  
论文17页，图2-5，已添加参考文献原图的引用。论文38页，图4-1，已添加参考文献原图的引用
- ② 在进行与已有工作对比时，除了展示本文方法的优势之外，建议也分析一下所存在的劣势  
论文45页，实验分析的第五点添加了关于本方法的一些劣势，例如在粗粒度的任务上未取得最优的效果
- ③ 绪论，实验设计等章节部分内容表述不明确，需要仔细检查，修改已修改论文绪论、实验部分表述



# 初审意见修改

- ① 答辩ppt英文题目与论文题目不一致  
已修正



感谢各位评审老师的聆听与指导!



# 参考文献 I



Qianru Sun, Bernt Schiele, and Mario Fritz. “A Domain Based Approach to Social Relation Recognition”. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, July. 2017*, pp. 435–444.



Junnan Li et al. “Dual-Glance Model for Deciphering Social Relationships”. In: *Proceedings of IEEE International Conference on Computer Vision (ICCV), Venice, Italy, October. 2017*, pp. 2669–2678.



## 参考文献 II



Zhouxia Wang et al. “Deep Reasoning with Knowledge Graph for Social Relationship Understanding”. In: *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI)*, Stockholm, Sweden, July. 2018, pp. 1021–1028.



Cewu Lu et al. “Visual Relationship Detection with Language Priors”. In: *Proceedings of 14th European Conference On Computer Vision (ECCV)*, Amsterdam, Netherlands, October. 2016, pp. 852–869.



## 参考文献 III



Ning Zhang et al. “Beyond frontal faces: Improving Person Recognition using multiple cues”. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, June. 2015*, pp. 4804–4813.

