Understanding social relationship with person-pair relations

Sarit Kraus
Department of Computer Science, Bar-Ilan University, Israel
pcchair@ijcai19.org

Abstract

Social relationships understanding is to infer the relations among people from images and videos, which has attracted increasing attention in computer vison recently. A great progress has been made since the rise of deep learning. However, they mostly focuses on the facial attributes or contextual objects cues without taking into account the interaction among person pairs. Motivated by scene graph generation, we carefully studied the datasets and found the social relations in a still image always have high semantic relevance. For instance, if two person pair in an image are *Friends*, then the third one is always friends or at least other *Intimate* relations but not *No Relation*. Therefore, to capture this interaction cues, we propose a novel end-to-end trainable Person-Pair Relation Network (PRN) using standard RNNs, a graph inference network that learns iteratively to improve its predictions via message passing among person pair nodes. Extensive experiments on PISC and PIPA-Relation show the superiority of our method over previous methods.

1 Introduction

Social relationships are closely related to our daily life [Barr *et al.*, 2014]. After understanding the social relationship between the person pair, we can easily explain their behavior. For machines, only when they fully understand the social relationships, can they further understand and infer the human behavior in our social life, so as to make a better response.

This part will be written by **liangjinrui** . This part will consist of the following sections:

- The significance of social relation understanding
- **[Example Figure]**
- Analyze the shortcomings of the current model and introduce our model(based ILP)
- Challenges
- Contributions

2 Related Work

This part will be written by **liangjinrui** and it will be surveyed by **liangjinrui** and **chenhaicheng** by January 31.

2.1 Social Relationship Understanding

The foundation of social network is the social relationships understanding, an important multidisciplinary problem that has attracted increasing attention in computer vision recently. A much number of studies that aim to infer social relationships from images [Wang *et al.*, 2015; Li *et al.*, 2017; Wang *et al.*, 2018; Wang *et al.*, 2010; Zhang *et al.*, 2015b] and videos [Ding and Yilmaz, 2010; Ramanathan *et al.*, 2013; Vinciarelli *et al.*, 2009] have been made since the rise of deep learning. For instance, motivated by psychological sudies, [Zhang *et al.*, 2015b] and [Dibeklioglu *et al.*, 2013] exploit social relationships based on facial attributs such as expression and head pose, and affective behaviour analysis. Besides, [Li *et al.*, 2017] and [Wang *et al.*, 2018] discover that contextual cues around people play a significant role in social realtionship infering. Concretely, [Li *et al.*, 2017] proposed a dual-glance model for social relationship, where the first glance makes a coarse relationship prediction for a given person pair and then the second one refines the prediction by using the objects around the pair. [Wang *et al.*, 2018] build a knowledge graph and employed Gated Graph Neural Network (GGNN) [Li *et al.*, 2015] to integrate the graph into the Graph Reasoning Model (GRM), a deep neural network where a proper message propagation and graph attention mechanism are introduced to explore the interaction between person pair and the contextual objects.

Unlike the aforementioned works which mainly focus on facial attributs or contextual object cues, we detaily studied the two classic datasets PISC [Li *et al.*, 2017] and PIPA-relation [Sun *et al.*, 2017] and found the social relations in a still image have high semantic relevance. Based on this discovery, we designed a novel end-to-end trainable Person-Pair Relation Network (PRN), a graph inference network to capture this semantic relevance cues via message passing among person pair nodes.

2.2 Message Passing

Introduction to Message Passing, written by **liangjinrui**.

2.3 Rules and ILP

Introduce rules and ILP referring to [Wang *et al.*, 2015]

3 SRDR model

This part will be written by **liangjinrui** and **chenhaicheng**.
[model figure]

[Introduce the total model]

3.1 Social Relationship Detection Model

3.2 Imposing Rules

3.3 Integrating by Integer Linear Programming

4 Experiments

This part will be written by **chenhaicheng**.

4.1 Experiment Setting

Datasets. In this work, two datasets were used to evaluate our proposed method and other existing ones. The first one is the large-scale People in Social Context (PISC) [Li *et al.*, 2017] with 22,670 images and contains two-level recognition tasks: **3 Coarse-level relationship**, namely *No Relation, Intimate Relation, None-Intimate Relation* and **6 Fine-level relationship**, i.e., *Friend, Family, Couple, Professional, Commerical, No Relation*. The second one is the People in Photo Album Relation (PIPA-Relation) [Sun *et al.*, 2017], an extension verson of People in Photo Album (PIPA) [Zhang *et al.*, 2015a] with 37107 images. It also annotates 26,915 person pairs on two-level recognition tasks: **5 Social Domains** and **16 Social Relations** based on these domains. The train/val/test in PISC are 13,142/4,000/4,000 images with 14,536/25,636/15,497 person pairs on coarse level relationship, and 16,828/500/1,250 images with 55,400/1,505/3,691 person pairs on fine level relationshp, respectively. In PIPA-relation, we follow [Wang *et al.*, 2018] and focus on recognizing its 16 relationships in the experiment. The train/val/test in it are 13,729/709/5,106 person pairs.

Implementation Details. During our work, We adopt the same strategy as previous works including [Li *et al.*, 2017] and [Wang *et al.*, 2018]. First, we fine-tune the ResNet-101 model [He *et al.*, 2016] , and we set the a lower learning rate as 0.0001. For the message passing progation model, the dimension of hidden size is set as 512. The iteration time T is set as 4 and learning rate as 0.0001. Similer to [Wang *et al.*, 2018], we the fine-tuning model utilized SGD, and the message passing module is trained with ADAM.

4.2 Comparisons with State-of-the-Art Methods

we compare our proposed model with existing state-of-the-art methods on both PISC and PIPA-Relation datasets.Formally, the compared methods are as followed:

Performance on the PISC dataset

UnionCNN Following [Lu *et al.*, 2016], it generates a single CNN model to predicate relations. In this task, we also feeds the union region of person pair to s single CNN for classification.

Pair CNN[Li *et al.*, 2017] consists of two equivalent CNNs with shared weights to extracted features for image for two individuals.

Pair CNN + BBox + Union[Li *et al.*, 2017] incorporates spatial location information of two bounding box that based the previous pair CNN and Union CNN.

Dual-glance[Li *et al.*, 2017] implements coarse and fine prediction which includes three and six relationships. Dual-glance employing pair CNN + BBox + BBox + Union and utilized surrounding region proposal to refine the prediction.

GRM[Wang *et al.*, 2018] propse a graph reasoning model that unifies the frequence of co-concurrences of each relationship-object pair to facilitate social relation.

Similar to the model of GRM, we also adopt the per-class recall and mean average precision(mMP) to evaluate our model. The experiments data are reported in Table 1. First, both Pair CNN + BBox + Union, Pair CNN + BBox + Global, Dual-glance are incur extra Faster-RCNN[Ren *et al.*, 2015] to extract the local contextual cues(object proposal). GRM utilized the object proposal to construct a semantic-aware knowledge graph to reason about the social relationship. It is notable that both of them incur extra detection annotations that contains noisies. dasdasdasd

Performance on the PIPA-Relation dataset

On this dataset, we also compare our proposed model with the existing Two stream CNN[Sun *et al.*, 2017],Dual-glance[Li *et al.*, 2017] and GRM[Wang *et al.*, 2018].Two stream CNN takes

4.3 Experiment Results

4.4 Experiment Analysis

4.5 Ablation Study

4.6 Case Study

5 Conclusion

This part will be written by **liangjinrui**.

Table 1: Recall-per-class and mean average precision (mAP) evaluating our PRN model and previous methods on PISC

Methods	Coarse relationships				Fine relationships						
	Intimate	Non-Intimate	No Relation	mAP	Friends	Family	Couple	Professional	Commerical	No Relation	mAP
Union CNN [Lu *et al.*, 2016]	72.1	81.3	19.2	58.4	29.9	58.5	70.7	55.4	43.0	19.6	43.5
Pair CNN [Li *et al.*, 2017]	70.3	80.5	38.8	65.1	30.2	59.1	69.4	57.5	41.9	34.2	48.2
Pair CNN + BBox + Union [Li *et al.*, 2017]	71.1	81.2	57.9	72.2	32.5	62.1	73.9	61.4	46.0	52.1	56.9
Pair CNN + BBox + Global [Li *et al.*, 2017]	70.5	80.0	53.7	70.5	32.2	61.7	72.6	60.8	44.3	51.0	54.6
Dual-glance [Li *et al.*, 2017]	73.1	84.2	59.6	79.7	35.4	68.1	76.3	70.3	57.6	60.9	63.2
Ours											

References

[Barr *et al.*, 2014] Jeremiah R. Barr, Leonardo A. Cament, Kevin W. Bowyer, and Patrick J. Flynn. Active clustering with ensembles for social structure extraction. In *IEEE Winter Conference on Applications of Computer Vision, Steamboat Springs, CO, USA, March 24-26, 2014*, pages 969–976, 2014.

[Dibeklioglu *et al.*, 2013] Hamdi Dibeklioglu, Albert Ali Salah, and Theo Gevers. Like father, like son: Facial expression dynamics for kinship verification. In *IEEE International Conference on Computer Vision, ICCV 2013, Sydney, Australia, December 1-8, 2013*, pages 1497–1504, 2013.

[Ding and Yilmaz, 2010] Lei Ding and Alper Yilmaz. Learning relations among movie characters: A social network perspective. In *Computer Vision - ECCV 2010, 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part IV*, pages 410–423, 2010.

[He *et al.*, 2016] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 770–778, 2016.

[Li *et al.*, 2015] Li-Jia Li, David A. Shamma, Xiangnan Kong, Sina Jafarpour, Roelof van Zwol, and Xuanhui Wang. Celebritynet: A social network constructed from large-scale online celebrity images. *TOMCCAP*, 12(1):3:1–3:22, 2015.

[Li *et al.*, 2017] Junnan Li, Yongkang Wong, Qi Zhao, and Mohan S. Kankanhalli. Dual-glance model for deciphering social relationships. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*, pages 2669–2678, 2017.

[Lu *et al.*, 2016] Cewu Lu, Ranjay Krishna, Michael S. Bernstein, and Fei-Fei Li. Visual relationship detection with language priors. In *Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part I*, pages 852–869, 2016.

[Ramanathan *et al.*, 2013] Vignesh Ramanathan, Bangpeng Yao, and Fei-Fei Li. Social role discovery in human events. In *2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, June 23-28, 2013*, pages 2475–2482, 2013.

[Ren *et al.*, 2015] Shaoqing Ren, Kaiming He, Ross B. Girshick, and Jian Sun. Faster R-CNN: towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, pages 91–99, 2015.

[Sun *et al.*, 2017] Qianru Sun, Bernt Schiele, and Mario Fritz. A domain based approach to social relation recognition. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 435–444, 2017.

[Vinciarelli *et al.*, 2009] Alessandro Vinciarelli, Maja Pantic, and Hervé Bourlard. Social signal processing: Survey of an emerging domain. *Image Vision Comput.*, 27(12):1743–1759, 2009.

[Wang *et al.*, 2010] Gang Wang, Andrew C. Gallagher, Jiebo Luo, and David A. Forsyth. Seeing people in social context: Recognizing people and social relationships. In *Computer Vision - ECCV 2010 - 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part V*, pages 169–182, 2010.

[Wang *et al.*, 2015] Quan Wang, Bin Wang, and Li Guo. Knowledge base completion using embeddings and rules. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015*, pages 1859–1866, 2015.

[Wang *et al.*, 2018] Zhouxia Wang, Tianshui Chen, Jimmy S. J. Ren, Weihao Yu, Hui Cheng, and Liang Lin. Deep reasoning with knowledge graph for social relationship understanding. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden.*, pages 1021–1028, 2018.

[Zhang *et al.*, 2015a] Ning Zhang, Manohar Paluri, Yaniv Taigman, Rob Fergus, and Lubomir D. Bourdev. Beyond frontal faces: Improving person recognition using multiple cues. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015*, pages 4804–4813, 2015.

[Zhang *et al.*, 2015b] Zhanpeng Zhang, Ping Luo, Chen Change Loy, and Xiaoou Tang. Learning social relation traits from face images. In *2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015*, pages 3631–3639, 2015.