Social Relation Recognition in Egocentric Photostreams

E. Sánchez Aimar (1), P. Radeva (1), M. Dimiccoli (2)

(1) CVC, University of Barcelona
(2) IRI, CSIC-Technical University of Catalonia
IEEE ICIP 2019

Index

Problem definition

Can we conceptualize our social relations ?

Attachment

mother – child grandpa - grandchild

Reciprocity

friends siblings

Mating

lovers / spouses

teacher – student leader – subordinate

Hierarchical power

sport team members colleagues

Coalitional group

- **Bugental's social theory** [1]:

  Two-level hierarchical categorization (forming a partition):

  - 5 domains, undefined number K of relations per domain

[1] Bugental. "Acquisition of the algorithms of social life: A domain-based approach." Psychological bulletin, 2000.

Can we conceptualize our social relations ?

Attachment

mother – child grandpa - grandchild

Reciprocity

friends siblings

Mating

lovers / spouses

teacher – student leader – subordinate

Hierarchical power

sport team members colleagues

Coalitional group

- **Bugental's social theory** [1]:

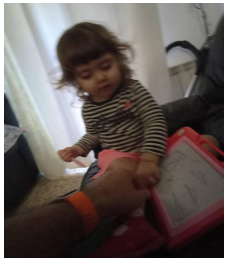
  Two-level hierarchical categorization (forming a partition):

  - 5 domains, undefined number K of relations per domain
- Sun et al. computational model of the Bugental's theory in *still images*[2]:
  - PIPA annotated dataset (K=16) + baseline (social cues + SVM)

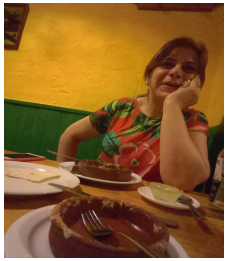
[1] Bugental. "Acquisition of the algorithms of social life: A domain-based approach." Psychological bulletin, 2000.
[2] Sun et al. "A Domain Based Approach to Social Relation Recognition". CVPR 2017.

Can we formalize the Bugental's theory in the domain of egocentric photostreams?

Attachment

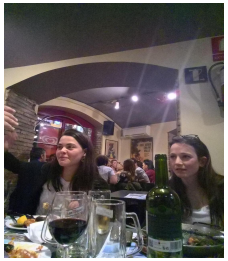


Father-Child

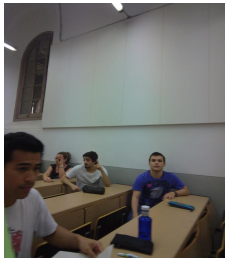


Mother-Child

Reciprocity



Friends

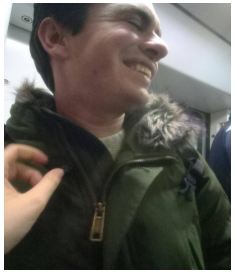


Classmates

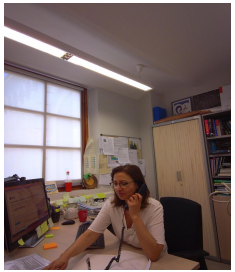
Coalitional group

Colleagues

- Wearable cameras allows to capture **unconstrained natural pictures** of social interactions.
- Useful to characterise a person's social patterns[3].



Lovers/Spouses



Leader-subordinate

Presenter-audience

Customer-staff

Mating

Hierarchical power

[3] Aghaei et al. "Towards social pattern characterization in egocentric photo-streams. CVIU 2018

Challenges of the egocentric photostream domain

Egocentric photostreams are sequences of images captured at *regular intervals* from egocentric point of view.

- 2 images per minute, during almost all day
- Camera-wearer is not visible
- Hard perspective/Camera shaking
- Large variability (poses, contexts, appearance, etc)

Proposed approach

Key contributions

1. Classification of a set of semantic attributes over time:
    i. leveraging the temporal evolution (**multidimensional time-series classification**)
    ii. exploiting hierarchical structure of the label space (**multi-task**)

2. **Egocentric dataset of social relations** in the form of photostreams (EgoSocialRelation)

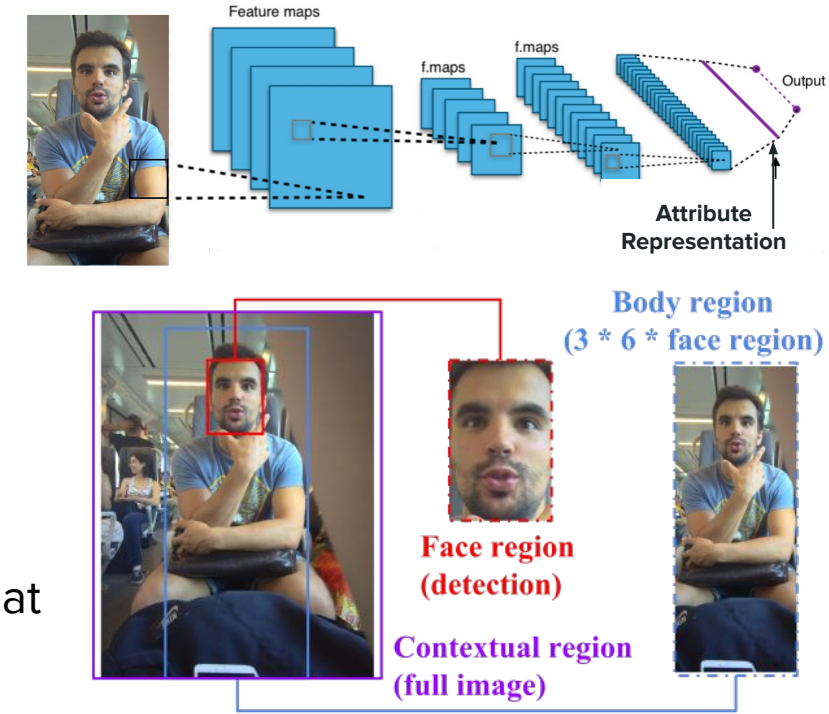
Semantic attributes

Ten attributes from multiple image regions:

- Age (face, body)
- Gender (face, body)
- Emotion (face)
- Head Appearance
- Clothing (body)
- Head Orientation (w.r.t. the camera)
- Activity (context)
- Proximity (w.r.t. the camera)

Use task-specific ***Convolutional Neural Nets*** at frame level.

Camera-user's ground truth age and gender (invariant to different situations)



Social relation classification

Input: set of 12 semantic attributes per each frame

Casted as a multidimensional time-series (d= 12) classification problem, solved by LSTM.

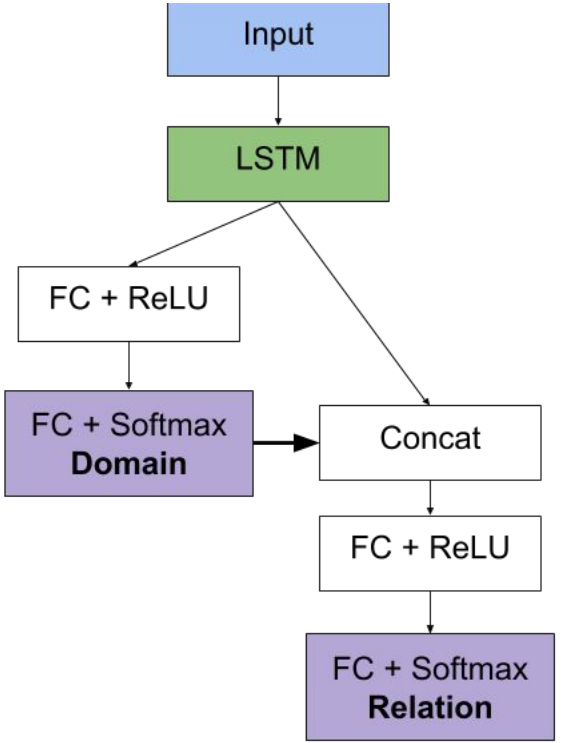
Can we exploit the hierarchical label space?

Multi-task learning:

- Highly correlated tasks improve generalization
- Model with shared parameters jointly optimized

Relation prediction is conditioned on domain output

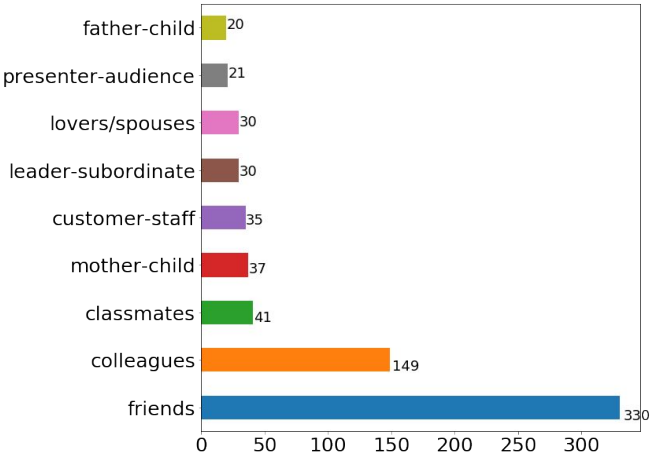
Multi-task Top-down model

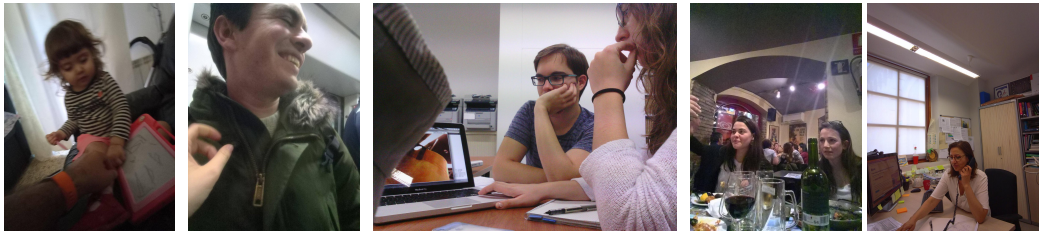


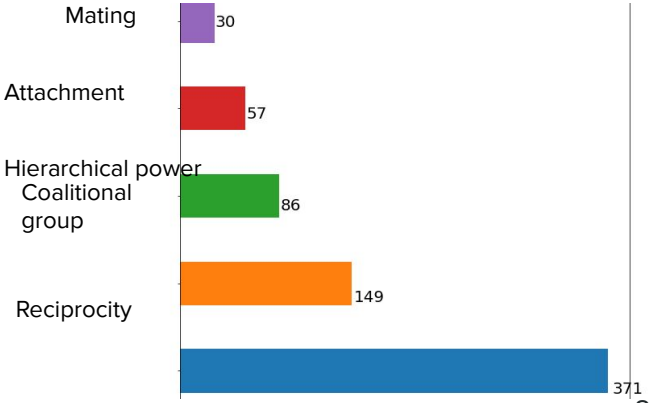
Experimental setting

EgoSocialRelation Dataset

- Extension of the ***EgoSocialStyle*** dataset [3]:
  - 213 new sequences (332 in total)
- Derived sub-segments for each visible person
  - # User-specific segments: 693
- Large class imbalance



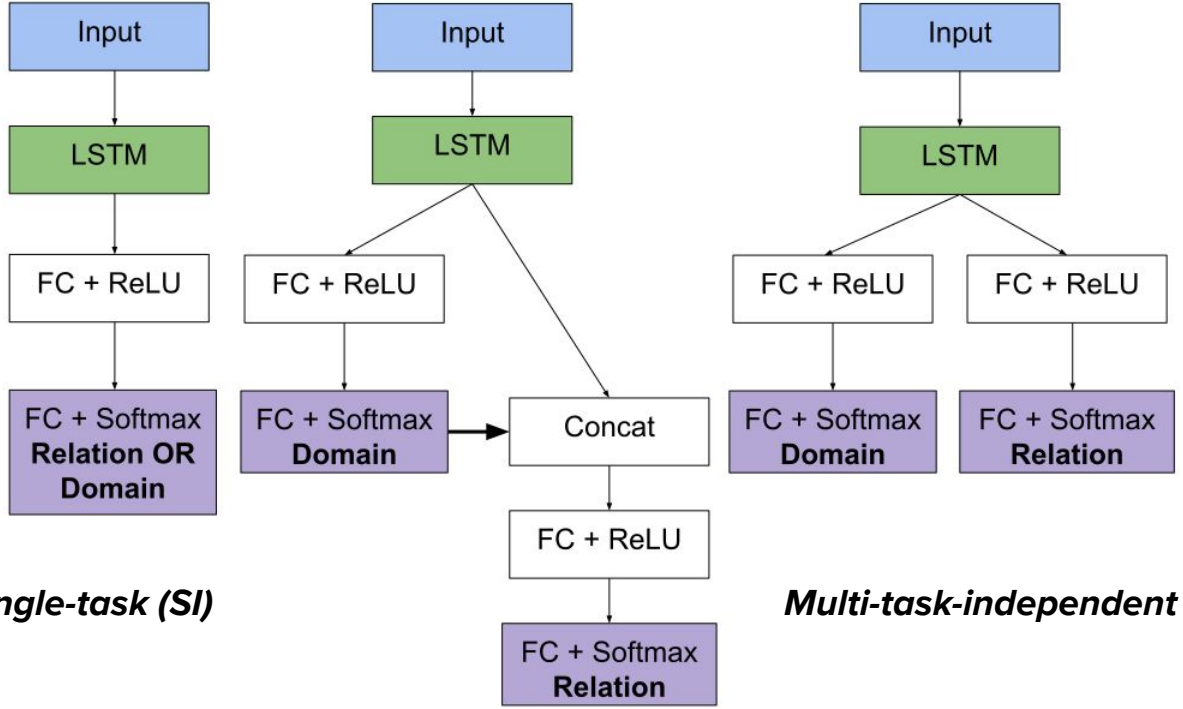




[3] Aghaei et al. "Towards social pattern characterization in egocentric photo-streams. CVIU 2018.

Ablation study and metrics

Multi-task Top-down (MT-TD)



Single-task (SI)

Multi-task-independent (MT-I)

Performance metrics :

1. Accuracy
2. F-1 score

Experimental results

Recognition results

	F1-score [%]	Acc [%]
REL-ONLY	32.19	57.10
REL-MT-I	31.06	54.90
REL-MT-TD	**33.26**	**58.60**

Relation recognition

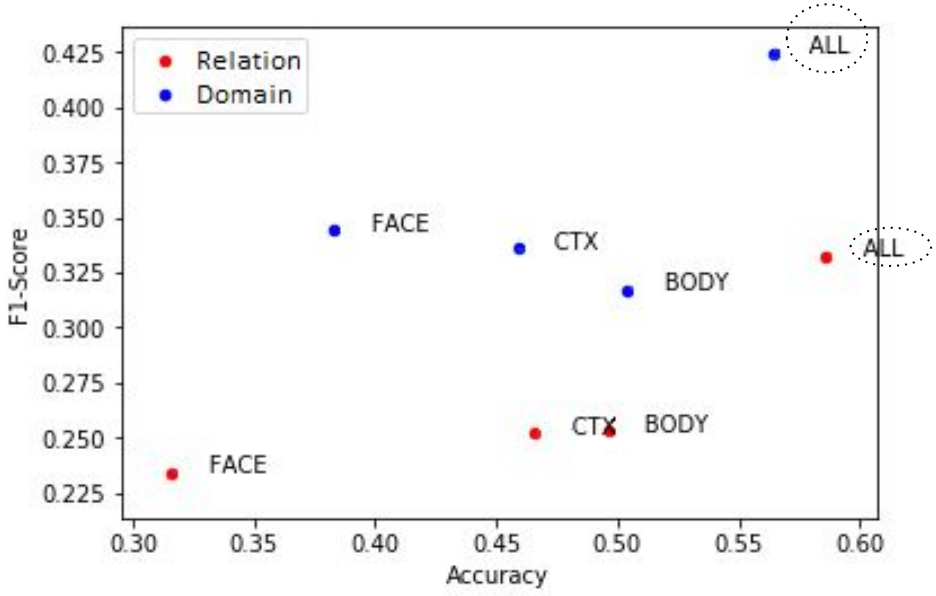
Multi-task without top-down conditioning is not effective

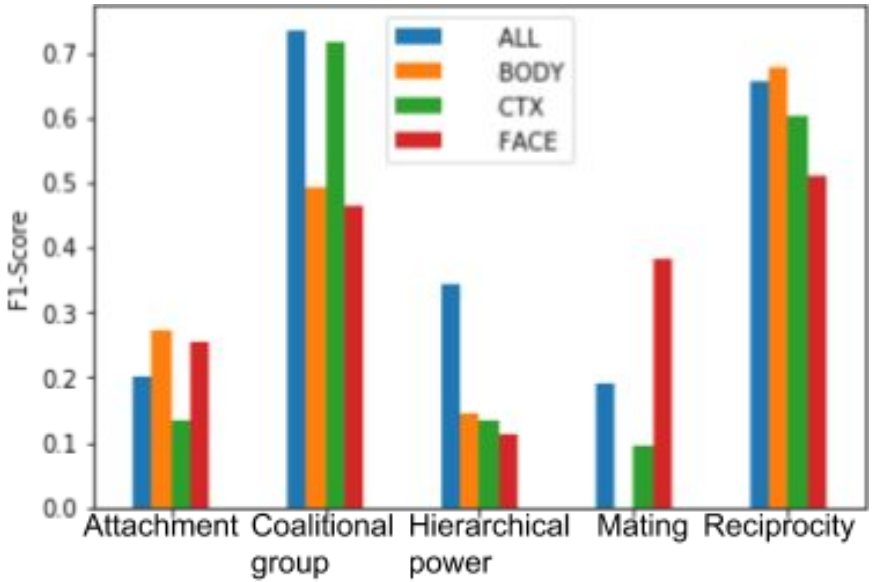
	F1-score [%]	Acc [%]
DOM-ONLY	**44.52**	**59.40**
DOM-MT-I	38.38	54.90
DOM-MT-TD	42.49	56.40

Domain recognition

Relation loss does not improve domain prediction

Analysis of Semantic Attributes Contribution





- Combining all attributes maximizes recognition performance
- Each domain *responds to* different social cues (as predicted by Bugental's theory)

Notation: **BODY**: "body age", "body gender", clothing + camera-user's info (age & gender)
FACE: "facial age", "facial gender", head pose, head appearance, emotion + camera-user's info
CTX: activity & proximity, **ALL**: all attributes.

Note: Multi-task top-down architecture is used in these experiments.

Conclusions

Conclusions

- Explore categorization of social relations following Bugental's conceptualization in the domain of egocentric photostreams
- Provide new dataset of rich social events captured under unconstrained conditions
- Propose a baseline which exploits semantic attributes, temporal redundancy and a hierarchical label space

Dataset: https://chest.iri.upc.edu/files/users/mdimiccoli/public html/DATASETS/EgoSocialRelation.zip

Code: https://github.com/emasa/social-relations-recognition-egocentric-photostreams

Bibliography

[1] D.B. Bugental. "*Acquisition of the algorithms of social life: A domain-based approach*." Psychological bulletin, 2000, 126(2), pp. 187.

[2] Q. Sun et al. "*A Domain Based Approach to Social Relation Recognition*". In 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 435-444.

[3] M. Aghaei et al. "*Towards social pattern characterization in egocentric photo-streams.*" Computer Vision and Image Understanding (CVIU), 2018, Vol. 171, pp. 104-117.