Deep Robust Subjective Visual Property Prediction in Crowdsourcing

Qianqian Xu[1], Zhiyong Yang[2,3], Yangbangyan Jiang[2,3], Xiaochun Cao[2,3], Qingming Huang[1,3], Yuan Yao[4]

1. KLIIP, Inst. of Comput. Tech., Chinese Academy of Sciences. 2. SKLOIS, Inst. of Info. Engin., Chinese Academy of Sciences.
3. Univ. of Chinese Academy of Sciences. 4. Depart. of Math., Hong Kong Univ. of Sciences and Technology.

xuqianqian@ict.ac.cn, {yangzhiyong,jiangyangbangyan,caoxiaochun}@iie.ac.cn, qmhuang@ucas.ac.cn, yuany@ust.hk

CVPR LONG BEACH CALIFORNIA June 16-20, 2019

Introduction

- **Subjective Visual Property (SVP)**: measures a user's subjective perception and feeling, w.r.t. a certain property in images/videos.
  → *e.g.*, properties of consumer goods such as shininess of shoes, images/videos interestingness.

Fig. 1: Samples.

- **Pairwise comparison based methods**
  → Pairwise comparisons are more reliable than ratings.
  → Contamination in crowdsourced labels from **bad users** — *impatient users, malicious users, or robots.*
- **Our goal**
  (a) To learn a deep SVP prediction model from a set of sparse and noisy pairwise comparison labels, with the ranking patterns preserved.
  (b) To detect and remove all the noisy annotations along with the training process.
- **Why not majority voting?**
  → It has been found that when pairwise local rankings are integrated into a global ranking, it is possible to detect outliers that can cause global inconsistency and yet are locally consistent, *i.e.*, supported by majority votes.

Framework

Fig. 2: Overview of our approach.

(1) Constructing a comparison graph from the crowdsourcing annotations, which is contaminated with outlier labels.
(2) Our proposed generalized deep probabilistic framework, where an outlier indicator $\boldsymbol{\gamma}$ is learned along with the network parameters Θ.
(3) Our framework will output a clean graph on the training set, where contaminated annotations are eliminated. Furthermore, our model could predict a rank-preserved score for each unseen instance.

Problem definition

- **Input**: A pairwise comparison graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the vertex set (all image items), and $\mathcal{E}$ is the set of comparison edges.
  → An edge $(i_1, i_2, j)/(i_2, i_1, j)$ – for i-th pair (i_1, i_2), the user j believes that the item i_1 holds a *stronger/weaker* presence of the SVP than i_2
  → Label y_{ij} – for a given pair i and annotator j,

$$y_{ij} = \begin{cases} 1, & \text{if } (i_1, i_2, j) \in \mathcal{E}; \\ -1, & \text{if } (i_2, i_1, j) \in \mathcal{E}. \end{cases} \tag{1}$$

- **Output**:
  (a) the relative score s_{i_1} and s_{i_2} of images in the pair.
  (b) an outlier indicator $\gamma(y_{ij})$ which could automatically remove the abnormal directions on $\mathcal{G}$.

A deep robust SVP prediction model

- Parameters
  → Network parameters Θ
  → **Outlier indicator** $\gamma(y_{ij})$: to model the noise nature of the annotations
    ∗ During the training process, our prediction is an additive mixture of the reliable score and the outlier indicator.
    ∗ $\gamma(y_{ij}) \neq 0$ **only if** the corresponding direction is not reliable, *e.g.*, in Fig.2,
      · The correct direction – we have a $\gamma(1)$ for $(1, 2, 1)$ and $(1, 2, 2)$
      · The contaminated direction – we have a different gamma with value $\gamma(-1)$ for $(2, 1, 3)$.
      · For the pair $(1, 2)$, $\gamma(1) = 0$ and $\gamma(-1) \neq 0$ are pursued.
- With the predicted scores as $s(\cdot)$, we model the conditional distribution of the annotations along with the prior distribution of $\boldsymbol{\gamma}$ and Θ as:

$$y_{ij} \mid \boldsymbol{x}_{i_1}, \boldsymbol{x}_{i_2}, \Theta, \gamma(y_{ij}) \sim f(y_{ij}, s(\boldsymbol{x}_{i,1}, \boldsymbol{x}_{i,2}, \Theta) + \gamma(y_{ij})),$$
$$\gamma(y_{ij}) \mid \lambda_1 \sim h(\gamma(y_{ij}), \lambda_1), \quad \Theta \mid \lambda_2 \sim g(\Theta, \lambda_2).$$

  → $s(\boldsymbol{x}_{i_1}, \boldsymbol{x}_{i_2}, \Theta) = s(\boldsymbol{x}_{i_1}, \Theta) - s(\boldsymbol{x}_{i_2}, \Theta)$ – the relative score of the annotation, to model the reliable pattern in the annotations.
  → $\gamma(y_{ij})$ – to model the unreliability. Since only outliers have a nonzero indicator, we model the randomness of $\gamma(y_{ij})$ with an i.i.d sparsity-inducing prior distribution (*e.g.*, Laplacian distribution).
  → $f(y_{ij}, s(\boldsymbol{x}_{i,1}, \boldsymbol{x}_{i,2}, \Theta) + \gamma(y_{ij}))$ – the conditional p.d.f. of the labels.
- Let $\boldsymbol{\gamma} = \{\gamma(y_{ij})\}_{(i_1,i_2,j)\in\mathcal{E}}$, $\boldsymbol{y} = \{y_{ij}\}_{(i_1,i_2,j)\in\mathcal{E}}$. According to the Maximum A Posterior (MAP) rule, we *maximize* the posterior probability:

$$\min_{\Theta,\boldsymbol{\gamma}} -\log\left(P(\Theta, \boldsymbol{\gamma} \mid \boldsymbol{y}, \boldsymbol{X}, \lambda_1, \lambda_2)\right).$$

- Let $\mathcal{D} : \{(i, j) : (i_1, i_2, j) \in \mathcal{E} \text{ or } (i_2, i_1, j) \in \mathcal{E}\}$ and f_{ij}, h_{ij}, g be short for the p.d.f., respectively. If h_{ij} is the p.d.f. of a Laplacian distribution and g is the p.d.f. of an element-wise Gaussian distribution, our **loss function** could be::

$$\min_{\Theta,\boldsymbol{\gamma}} \sum_{(i,j)\in\mathcal{D}} \ell(y_{ij} - s(\boldsymbol{x}_{i,1}, \boldsymbol{x}_{i,2}, \Theta) - \gamma(y_{ij})) + \lambda_1 \left\|\boldsymbol{\gamma}\right\|_1 + \lambda_2 \sum_{\theta\in\Theta} \theta^2,$$

where $\left\|\boldsymbol{\gamma}\right\|_1 = \sum_{(i,j)\in\mathcal{D}} |\gamma(y_{ij})|$.
- Here we provide two specific models where ℓ depends on the assumptions on the distributions of y_{ij}:
  → **Model A**: y_{ij} conditionally subjects to a Gaussian distribution. Then ℓ is the least square loss.
  → **Model B**: y_{ij} conditionally subjects to a Logistic-like distribution. Then ℓ is the Logistic loss.

Optimization

- **Fix $\boldsymbol{\gamma}$, Learn Θ**: Θ can be updated with gradient based solver.
- **Fix Θ, Learn $\boldsymbol{\gamma}$**
  → **Model A**: the subproblem becomes

$$\min_{\boldsymbol{\gamma}} \sum_{(i,j)\in\mathcal{D}} \frac{1}{2}(y_{ij} - s(\boldsymbol{x}_{i,1}, \boldsymbol{x}_{i,2}, \Theta) - \gamma(y_{ij}))^2 + \lambda_1 \left\|\boldsymbol{\gamma}\right\|_1.$$

With $c_{ij} = y_{ij} - s(\boldsymbol{x}_{i,1}, \boldsymbol{x}_{i,2}, \Theta)$, the proximal operator of ℓ_1 norm gives a closed-form solution:

$$\gamma(y_{ij}) = \max(|c_{ij}| - \lambda_1, 0) \cdot \text{sign}(c_{ij}). \tag{2}$$

  → **Model B**: the subproblem becomes

$$\min_{\boldsymbol{\gamma}} \sum_{(i,j)\in\mathcal{D}} \log(1 + \Delta_{ij}) + \lambda_1 \left\|\boldsymbol{\gamma}\right\|_1.$$

where $\Delta_{ij} = \exp(-y_{ij}(s(\boldsymbol{x}_{i,1}, \boldsymbol{x}_{i,2}, \Theta) + \gamma(y_{ij})))$.
We adopt the proximal gradient method to find a numerical solution.

Experiments

- Proposed models
  → **LS-Deep(Model A)** – least square based deep model.
  → **Logit-Deep(Model B)** – logistic regression based deep model.
- Ablation
  → **LS-Deep-w/o γ** – the factor γ in **LS-Deep** is removed.
  → **Logit-Deep-w/o γ** – the factor γ in **Logit-Deep** is removed.

Algorithm	ACC	F1	Prec.	Rec.	AUC
Maj-LS	.5555	.4673	.4369	.5022	.5650
LS-with γ	.5594	.4729	.4414	.5093	.5759
Maj-Logistic	.5421	.4687	.4264	.5205	.5489
Logistic-with γ	.5585	.4743	.4410	.5131	.5735
Maj-RankNet	.5611	.4804	.4445	.5227	.5792
Maj-RankBoost	.5425	.5991	.6458	.5587	.4507
Maj-RankSVM	.5838	.3858	.4517	.3367	.5665
Maj-GBDT	.5827	.3880	.4504	.3408	.5619
Maj-DART	.5940	.3668	.4648	.3029	.5690
URLR	.5765	.4633	.5748	.5131	.5762
LS-Deep-w/o γ	.7313	.6694	.6407	.7008	.8060
Logit-Deep-w/o γ	.7439	.6818	.6584	.7070	.8168
LS-Deep-with γ	.7967	.7414	.7323	.7508	.8784
Logit-Deep-with γ	.7917	.7370	.7228	.7518	.8739

Fig. 3: Experiment results and outlier examples on Human age dataset.

Algorithm	Bald	D.Hai	B.Eye	G.Look	Masc.	Mouth	Smile	Teeth	Foreh.	Young	Aver.
Maj-LS	.4767	.5368	.4787	.4788	.5588	.4774	.5220	.5073	.4759	.5162	.5029
LS-with γ	.5805	.6400	.5506	.5932	.6009	.5097	.5178	.5198	.5680	.5911	.5672
Maj-Logistic	.6123	.6716	.5146	.5890	.6253	.5032	.5031	.5322	.5724	.6599	.5784
Logistic-with γ	.6059	.6400	.5640	.6038	.6275	.5269	.5073	.5405	.5724	.6437	.5832
Maj-RankNet	.6123	.6421	.5551	.6208	.6275	.5097	.5304	.5468	.5899	.6275	.5862
Maj-RankBoost	.5996	.7053	.5236	.5975	.6231	.5097	.5199	.5094	.6053	.6032	.5797
Maj-RankSVM	.4852	.6526	.4180	.5805	.5588	.4882	.5283	.5156	.5482	.6397	.5415
Maj-GBDT	.5551	.6253	.4899	.5466	.5721	.4903	.5094	.5198	.5965	.6235	.5528
Maj-DART	.5508	.6337	.4899	.5339	.5698	.4989	.5597	.5364	.5943	.6134	.5581
URLR	.5889	.6538	.6505	.5258	.5614	.6319	.5311	.4968	.5446	.5570	.5742
LS-Deep-w/o γ	.5932	.7095	.5551	.6081	.5543	.5742	.6436	.6133	.5746	.6741	.6100
Logit-Deep-w/o γ	.5551	.6758	.5124	.6335	.6253	.5806	.6038	.6175	.5724	.6235	.6000
LS-Deep-with γ	.6335	.7684	.5551	.6377	.6253	.7312	.7421	.7547	.6469	.7308	.6826
Logit-Deep-with γ	.6631	.7726	.5798	.6419	.5965	.7032	.7358	.7069	.6075	.6862	.6694

Fig. 4: Experiment results and outlier examples on LFW-10 dataset.

Algorithm	Comf.	Fash.	Form.	Pointy	Brown	Open	Ornate	Aver.
Maj-LS	.7300	.7825	.7325	.7897	.6950	.7331	.7300	.7418
LS-with γ	.8150	.8125	.7975	.7860	.7275	.7444	.7625	.7779
Maj-Logistic	.7600	.7850	.7475	.7970	.6900	.7068	.7175	.7434
Logistic-with γ	.8375	.8175	.7825	.7934	.7250	.7444	.7525	.7790
Maj-RankNet [?]	.7425	.7850	.7200	.7860	.6925	.7444	.7300	.7429
Maj-RankBoost [?]	.7525	.7300	.7275	.7675	.6975	.6955	.6725	.7204
Maj-RankSVM [?]	.7425	.7925	.7925	.8081	.6850	.7331	.7200	.7534
Maj-GBDT [?]	.7075	.7325	.7425	.8007	.6750	.7519	.7550	.7379
Maj-DART [?]	.6900	.7275	.7375	.8376	.6975	.7857	.7125	.7412
URLR [?]	.8200	.8150	.7900	.7860	.7325	.7444	.7550	.7775
LS-Deep-w/o γ	.7100	.8075	.7400	.7749	.7725	.7669	.7050	.7538
Logit-Deep-w/o γ	.7100	.8025	.7500	.8044	.7525	.7857	.6975	.7575
LS-Deep-with γ	.8500	.8550	.8125	.8044	.8250	.7782	.8300	.8222
Logit-Deep-with γ	.8550	.8500	.8200	.8339	.8125	.7481	.8325	.8217

Fig. 5: Experiment results and outlier examples on Shoes dataset.

Conclusion This work explores the challenging task of SVP prediction from noisy crowdsourced annotations from a deep perspective. We present a simple but effective general probabilistic model to simultaneously predict rank preserving scores and detect the outliers annotations, where an outlier indicator γ is learned along with the network parameters Θ. Practically, we present two specific models with different assumptions on the data distribution. Furthermore, we adopt an alternative optimization scheme to update γ and Θ iteratively. In our empirical studies, we perform a series of experiments on three real-world datasets: Human age dataset, LFW-10, and Shoes. The corresponding results consistently show the superiority of our proposed model.