A Probabilistic Graphical Model for Analyzing the Subjective Visual Quality Assessment Data from Crowdsourcing

Jing Li, Alibaba Group, China

Suiyi Ling, Capacites, France

Zhi Li, Netflix, USA

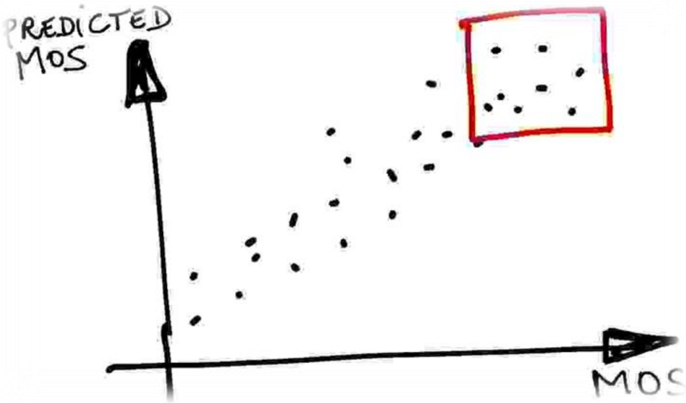
Junle Wang, Tencent, China

Patrick Le Callet, University of Nantes, France

Visual Quality Assessment

Subjective Quality: ratings from observers using scale (ACR, DSIS, SSQE, DSCQS, SAMVIQ …)
Averaged across observers => A.K.A MOS (Mean Opinion Score)

Objective Quality: predict a quality score

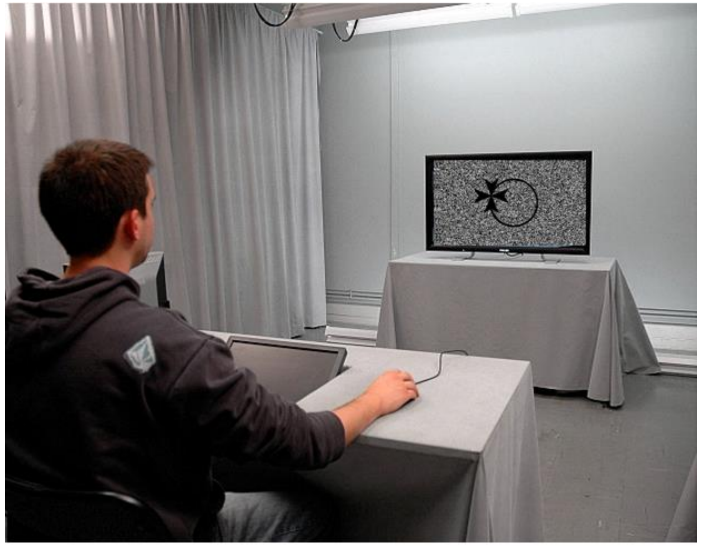


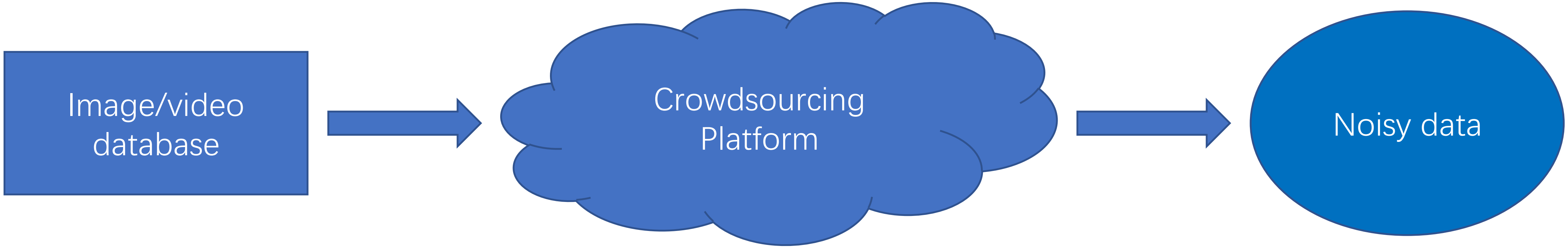
Deep Learning needs big data

- Traditional way to get quality data:
  Everything is in ITU standard
  - well control
  - time consuming

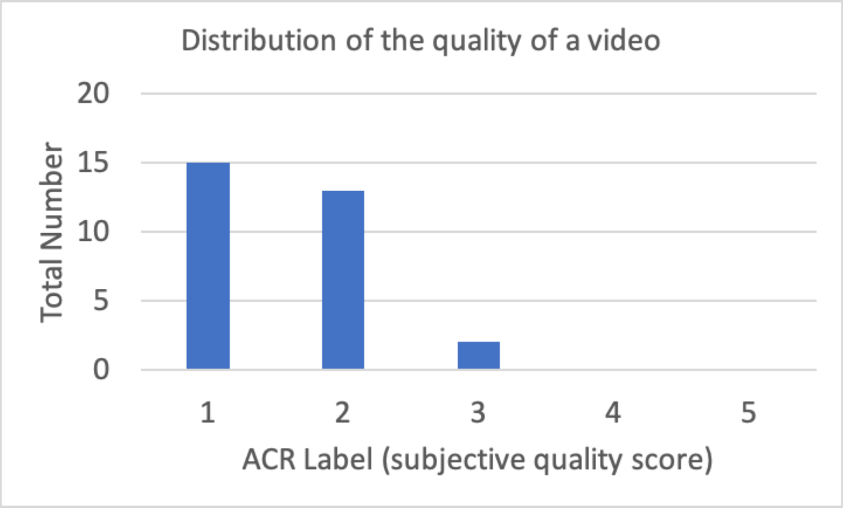
- Solution: Crowdsourcing

ITU standard test room





A model to recover ground truth



- Regarding subjective quality data
  - The distribution is not gaussian

But an ordinal categorical distribution

- Regarding annotator's behavior
  - he/she does not always give wrong/random answer

Should count on probability of abnormal behavior

The Proposed Annotation Model: GPM

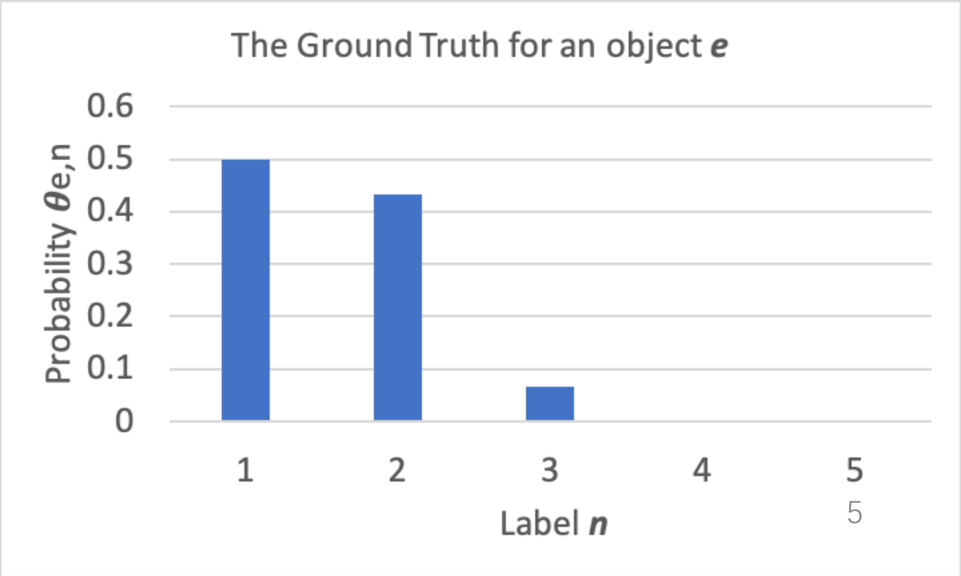
- Ground truth is an ordinal categorical distribution
    - For example, in ACR test, N = 5

$$Cat(y|\theta_{\mathbf{e}}) = \prod_{n=1}^{N} \theta_{e,n}^{[y=n]}$$

$\theta_{e,n}$ the probability of obtaining label *n* in one trial for object *e*

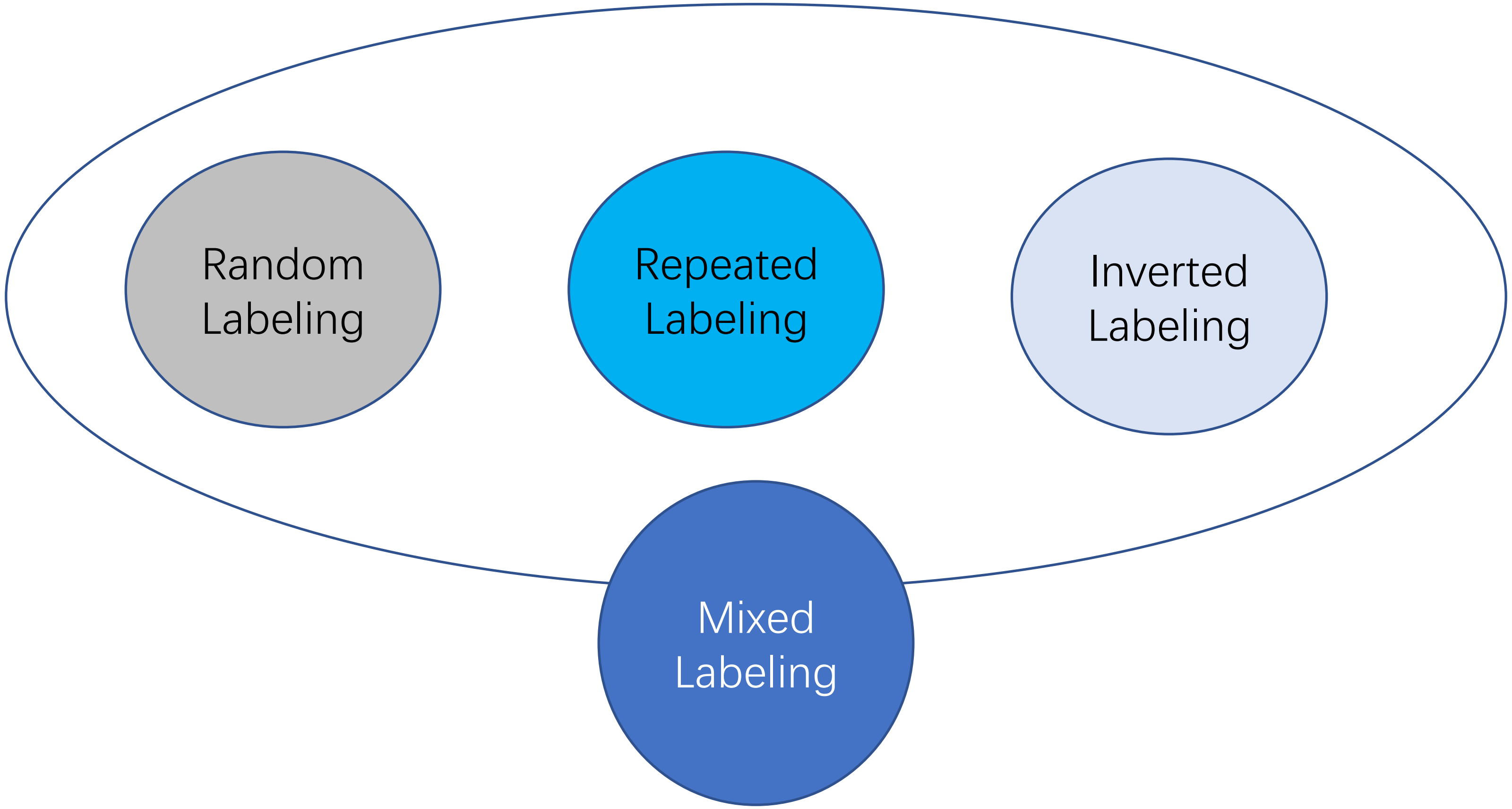
$$\sum_{n=1}^{N} \theta_{e,n} = 1$$

$[y = n]$ equals to 1 if $y = n$



The Proposed Annotation Model: GPM

- Annotators behavior classification

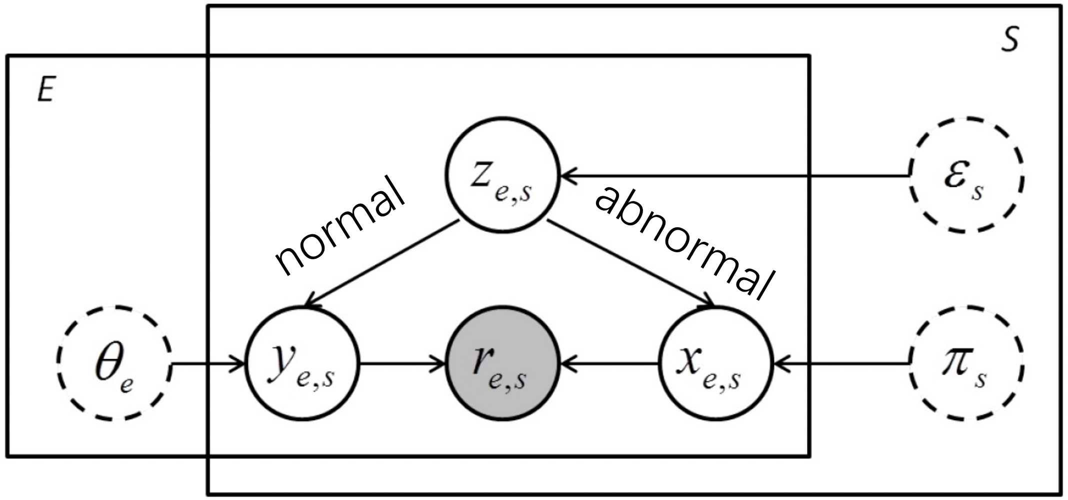


Yunus Emre Kara, Gaye Genc, Oya Aran, and Lale Akarun. 2015. Modeling annotator behaviors for crowd labeling. Neurocomputing 160 (2015), 141–156.

The Proposed Annotation Model: GPM

- For an annotator *s*
- Given an quality assessment task on object (image/video) *e*
- Using 1-5 Likert scale

$r_{e,s}$ *is the provided label for e by annotator s*



Using latent variable $z_{e,s}$ to control whether or not the annotator is in abnormal behavior

θ_e, ϵ_s and π_s are parameters, $y_{e,s}$, $x_{e,s}$ and $z_{e,s}$ are latent variables, $r_{e,s}$ is the provided label by annotator s.

$$p(Z|\epsilon) = \prod_{e,s \in A} B(z_{e,s}|\epsilon_s)$$

$$p(X|\pi) = \prod_{e,s \in A} D(x_{e,s}|\pi_s)$$

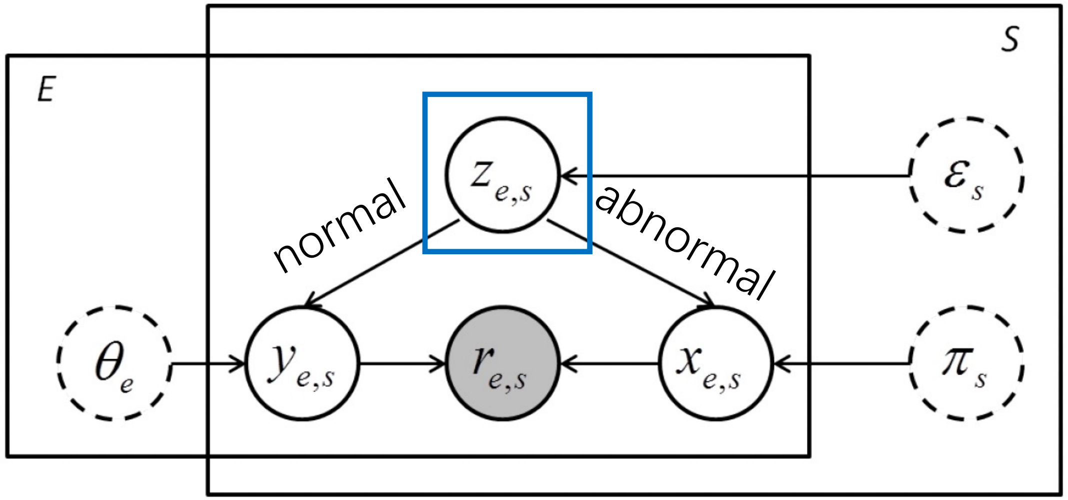
$$p(Y|\theta) = \prod_{e,s \in A} Cat(y_{e,s}|\theta_e)$$

$$p(R|X,Y,Z) = \prod_{e,s \in A} p(x_{e,s})^{[z_{e,s}=0]} p(y_{e,s})^{[z_{e,s}=1]},$$

A represents the set of all possible combinations of labelled objects and annotators

The Proposed Annotation Model: GPM

- For an annotator *s*
- Given an quality assessment task on object (image/video) *e*
- Using 1-5 Likert scale



Using latent variable $z_{e,s}$ to control whether or not the annotator is in abnormal behavior

θ_e , ϵ_e , and π_e are parameters, $y_{e,s}$, $x_{e,s}$ and $z_{e,s}$ are latent variables

$$\boxed{p(Z|\epsilon) = \prod_{e,s \in A} B(z_{e,s}|\epsilon_s)}$$

$$p(X|\pi) = \prod_{e,s \in A} D(x_{e,s}|\pi_s)$$

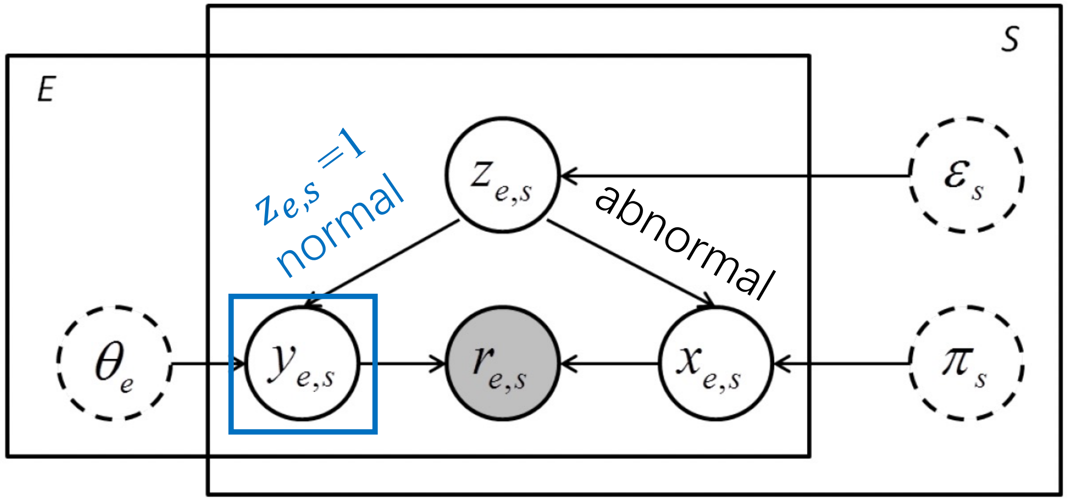
$$p(Y|\theta) = \prod_{e,s \in A} Cat(y_{e,s}|\theta_e)$$

$$p(R|X,Y,Z) = \prod_{e,s \in A} p(x_{e,s})^{[z_{e,s}=0]} p(y_{e,s})^{[z_{e,s}=1]},$$

A represents the set of all possible combinations of labelled objects and annotators

The Proposed Annotation Model: GPM

- For an annotator *s*
- Given an quality assessment task on object (image/video) *e*
- Using 1-5 Likert scale



Using latent variable $Z_{e,s}$ to control whether or not the annotator is in abnormal behavior

θ_e , ϵ_e , *and* π_e *are parameters*, $y_{e,s}$, $x_{e,s}$ *and* $z_{e,s}$ *are latent variables*

$$p(Z|\epsilon) = \prod_{e,s\in A} B(z_{e,s}|\epsilon_s)$$

$$p(X|\pi) = \prod_{e,s\in A} D(x_{e,s}|\pi_s)$$

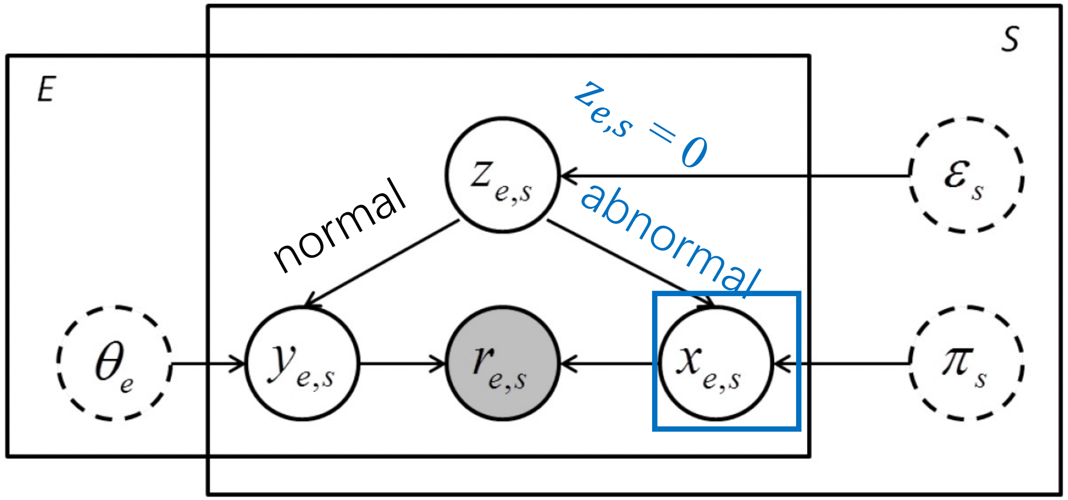
$$\boxed{p(Y|\theta) = \prod_{e,s\in A} Cat(y_{e,s}|\theta_e)}$$

$$p(R|X,Y,Z) = \prod_{e,s\in A} p(x_{e,s})^{[z_{e,s}=0]} p(y_{e,s})^{[z_{e,s}=1]},$$

A represents the set of all possible combinations of labelled objects and annotators

The Proposed Annotation Model: GPM

- For an annotator *s*
- Given an quality assessment task on object (image/video) *e*
- Using 1-5 Likert scale



Using latent variable $Z_{e,s}$ to control whether or not the annotator is in abnormal behavior

θ_e , ϵ_e , *and* π_e *are parameters*, $y_{e,s}$, $x_{e,s}$ *and* $z_{e,s}$ *are latent variables*

$$p(Z|\epsilon) = \prod_{e,s \in A} B(z_{e,s}|\epsilon_s)$$

$$\boxed{p(X|\pi) = \prod_{e,s \in A} D(x_{e,s}|\pi_s)}$$

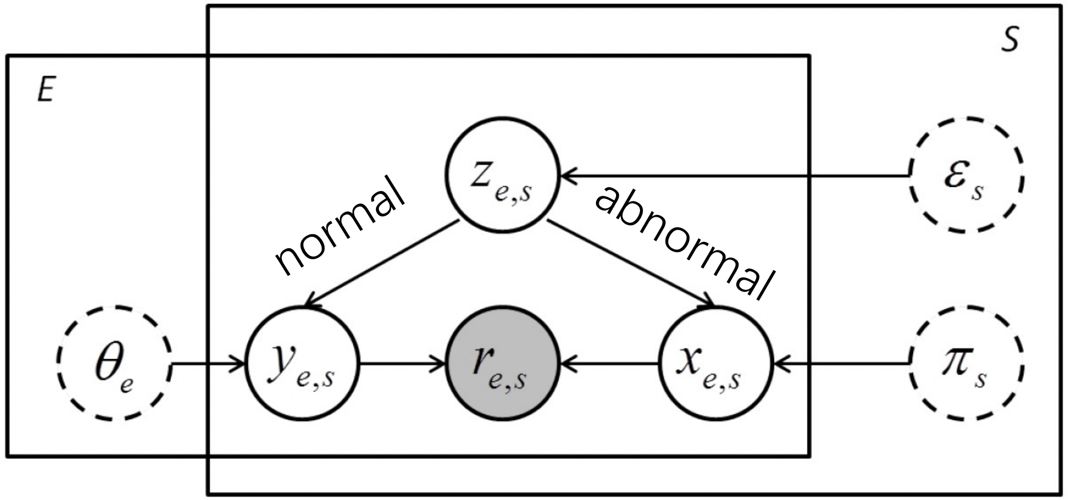
$$p(Y|\theta) = \prod_{e,s \in A} Cat(y_{e,s}|\theta_e)$$

$$p(R|X,Y,Z) = \prod_{e,s \in A} p(x_{e,s})^{[z_{e,s}=0]} p(y_{e,s})^{[z_{e,s}=1]},$$

A represents the set of all possible combinations of labelled objects and annotators

The Proposed Annotation Model: GPM

- For an annotator *s*
- Given an quality assessment task on object (image/video) *e*
- Using 1-5 Likert scale



Using latent variable $z_{e,s}$ to control whether or not the annotator is in abnormal behavior

θ_e, ϵ_s and π_s are parameters, $y_{e,s}$, $x_{e,s}$ and $z_{e,s}$ are latent variables, $r_{e,s}$ is the provided label by annotator s.

the complete conditional density

$$p(Z|\epsilon) = \prod_{e,s \in A} B(z_{e,s}|\epsilon_s)$$

$$p(X|\pi) = \prod_{e,s \in A} D(x_{e,s}|\pi_s)$$

$$p(Y|\theta) = \prod_{e,s \in A} Cat(y_{e,s}|\theta_e)$$

$$\boxed{p(R|X,Y,Z) = \prod_{e,s \in A} p(x_{e,s})^{[z_{e,s}=0]} p(y_{e,s})^{[z_{e,s}=1]},}$$

A represents the set of all possible combinations of labelled objects and annotators

The Proposed Annotation Model: GPM

$$
\begin{aligned}
p(R|Y, X, Z, \pi, \epsilon, \theta) &= \sum_{Z} p(R|X, Y, Z) p(Z|\epsilon) \\
&= p(Z = 1|\epsilon) p(Y|\theta) \\
&+ p(Z = 0|\epsilon) p(X|\pi) \\
&= \prod_{e,s \in A} [\epsilon_s (\prod_{n=1}^{N} \theta_{e,n}^{[r_{e,s}=n]}) \\
&+ (1 - \epsilon_s)(\prod_{n=1}^{N} \pi_{s,n}^{[r_{e,s}=n]})]
\end{aligned}
$$

$r_{e,s}$ is the provided label by annotator s

Parameter Estimation Using EM algorithm

Prediction of Ground Truth

$$\hat{v}_e = \sum_{n=1}^{N} n \cdot \theta_{e,n}$$

Subject to:

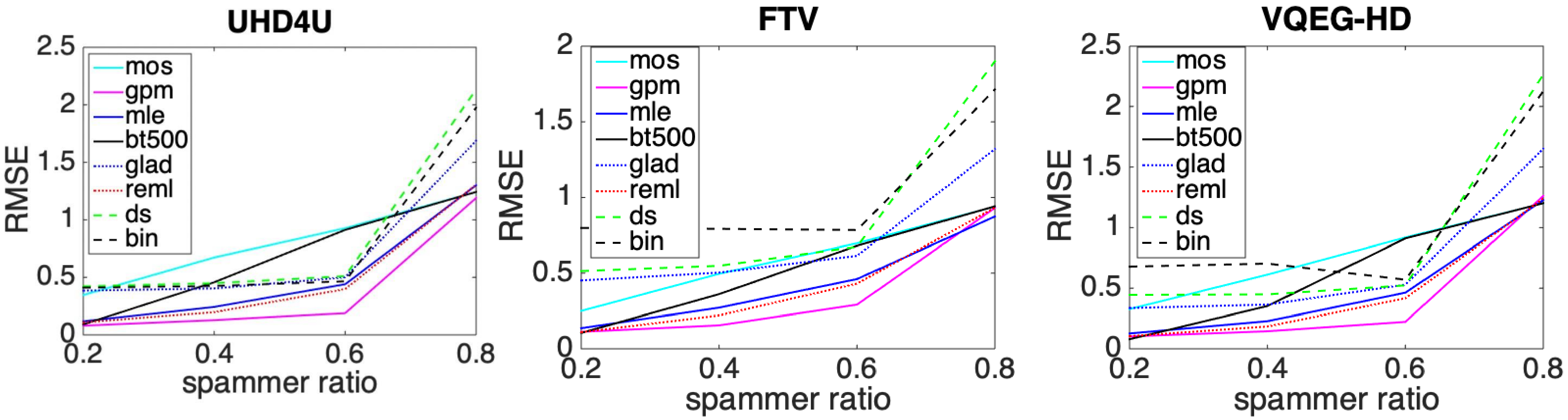
$$1 \geq \theta_{e,n} \geq 0, \sum_{n=1}^{N} \theta_{e,n} = 1$$

$$1 \geq \pi_{s,n} \geq 0, \sum_{n=1}^{N} \pi_{s,n} = 1$$

Experimental results

- Random selection of the spammer
- Simulate mixed abnormal behavior
- Replace the real data by error data
- Simulate 100 times

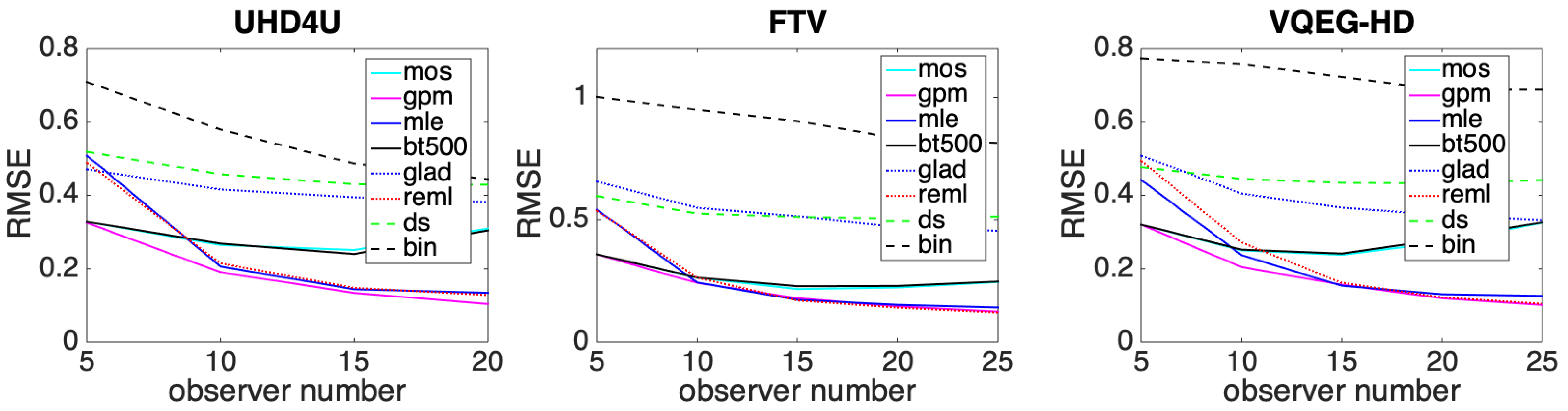
1) The influence of spammer ratio on recovered ground truth



Experimental Results

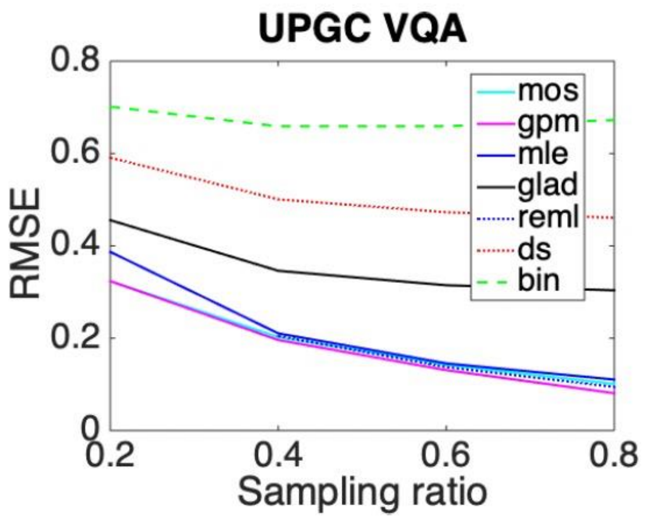
- Random sampling annotators
- Simulate mixed abnormal behavior
- Fix ‘mixed’ behavior = 20%
- Replace the real data by error data

2) Influence of annotator number on inferring the ground truth



Experiment on real crowdsourcing data

- **UPGC crowdsourcing data**
  - -1074 UPGC video sequences
  - -181 annotators
  - -23962 collected labels
  - -22 annotators/video



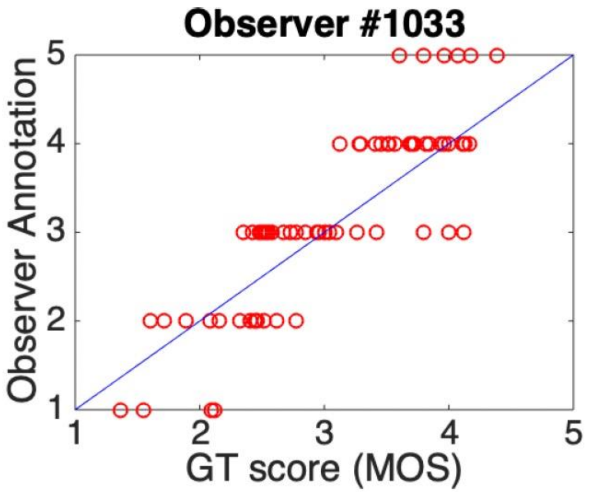
- **MovieLens 20M review data**
  - - 174 movies
  - - 69 annotators
  - - 2833 ratings
  - - 16 annotators/movie

GT: 5662 movies labeled by 15147 annotators

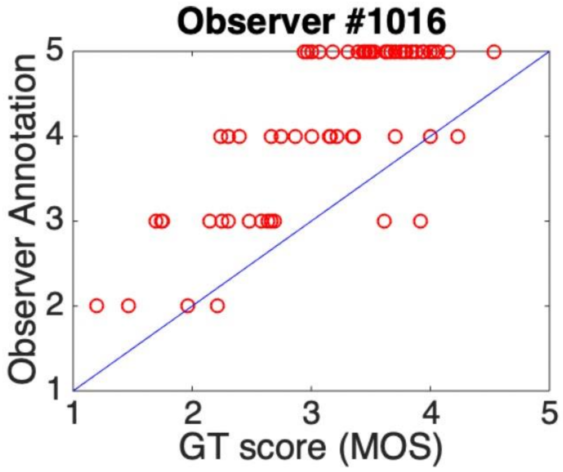
Model	PLCC ↑	ROCC ↑	RMSE ↓
D&S [6]	0.4073	0.4957	1.2287
GLAD [33]	0.5347	0.6029	0.6361
Bin [25]	0.7122	0.7166	0.5849
REML[26]	0.7066	0.7282	0.4292
MLE [18]	0.8369	0.8219	0.2483
Proposed	**0.8620**	**0.8420**	**0.2228**

Detected abnormal behavior

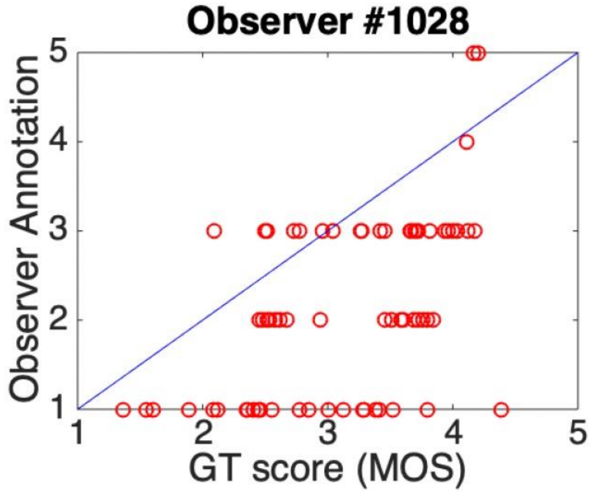
- In UPGC crowdsourcing video database



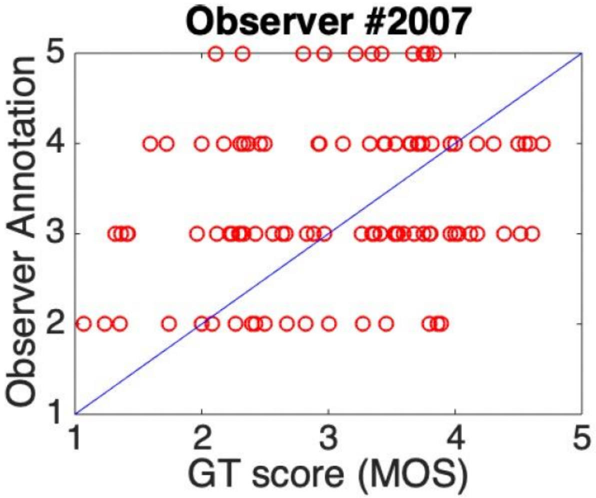
Normal annotation



Optimistic annotation



Picky annotation



Random annotation

Conclusion

- A probabilistic graphic model is proposed to recover ground truth and detect abnormal behavior
- The ground truth of the visual quality is a distribution
  - Not Gaussian
  - But an ordinal categorical distribution → more general
- Each annotator has a probability to make a mistake
  - If this probability smaller than 0.5 → spammer
  - Data is expensive, using model to denoise
- The proposed model outperforms the other SOTA methods.