Learning Perceptual Inference by Contrasting

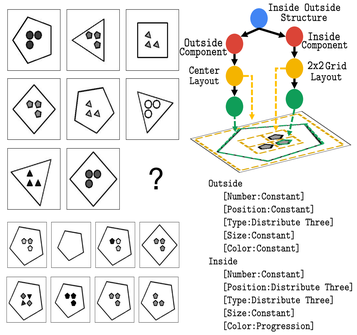
Chi Zhang[★,1,4] Baoxiong Jia[★,1] Feng Gao[3,4] Yixin Zhu[3,4] Hongjing Lu[2] Song-Chun Zhu[1,3,4]

[1] Department of Computer Science, UCLA [2] Department of Psychology, UCLA [3] Department of Statistics, UCLA

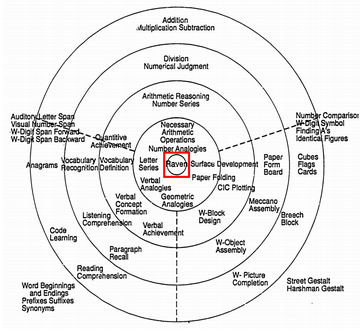
[4] International Center for AI and Robot Autonomy

{chi.zhang, baoxiongjia, f.gao, yixin.zhu, hongjing, sczhu}@ucla.edu

Motivation



Raven's Progressive Matrices



Cognitive Ability Test

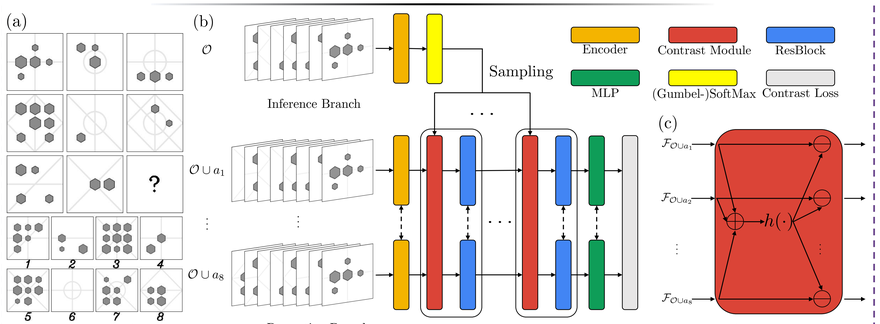
What is ***not so right*** about previous methods:

- Permutation sensitivity. Raven's Progressive Matrices (RPM) should be permutation-invariant with respect to swapped rows or columns and answer set permutation.
- Classification? We argue that answers should be ranked according to its *appropriateness*. A "wrong" answer doesn't mean it is wrong in every way.

We are inspired by the following perspectives:

- Study on contrast effects on cognitive science, biology, and computer science.
- Interplay between perception and inference detailed in Carpenter et al. for humans to solve RPM.
- Permutation-invariance is required.
- Treating it as ranking rather than classification.

CoPINet



Formulation

- A ranking perspective $p(a_\star|\mathcal{O}) \geq p(a'|\mathcal{O}), \quad \forall a' \in \mathcal{A}, a' \neq a_\star$

Contrast

- Model-level contrast $\text{Contrast}(\mathcal{F}_{\mathcal{O}\cup a}) = \mathcal{F}_{\mathcal{O}\cup a} - h\left(\sum_{a'\in\mathcal{A}} \mathcal{F}_{\mathcal{O}\cup a'}\right)$
- Objective-level contrast
  - Model $p(a|\mathcal{O}) = \frac{1}{Z}\exp(f(\mathcal{O}\cup a))$ and take log

    $\log p(a_\star|\mathcal{O}) - \log p(a'|\mathcal{O}) = f(\mathcal{O}\cup a_\star) - f(\mathcal{O}\cup a') \geq 0$
  - Push the difference to ***infinity***

    $f(\mathcal{O}\cup a_\star) - f(\mathcal{O}\cup a') \to \infty \iff \sigma(f(\mathcal{O}\cup a_\star) - f(\mathcal{O}\cup a')) \to 1$
  - Transform it into sufficient conditions

    $f(\mathcal{O}\cup a_\star) - b(\mathcal{O}\cup a_\star) \to \infty \iff \sigma(f(\mathcal{O}\cup a_\star) - b(\mathcal{O}\cup a_\star)) \to 1$

    $f(\mathcal{O}\cup a') - b(\mathcal{O}\cup a') \to -\infty \iff \sigma(f(\mathcal{O}\cup a') - b(\mathcal{O}\cup a')) \to 0$
  - Loss $\ell = \log(\sigma(f(\mathcal{O}\cup a_\star) - b(\mathcal{O}\cup a_\star))) + \sum \log(1 - \sigma(f(\mathcal{O}\cup a') - b(\mathcal{O}\cup a')))$

Perceptual Inference

- Take hidden rules into consideration

  $\log p(a|\mathcal{O}) = \log \sum_{\mathcal{T}} p(a|\mathcal{T},\mathcal{O})p(\mathcal{T}|\mathcal{O}) = \log \mathbb{E}_{\mathcal{T}\sim p(\mathcal{T}|\mathcal{O})}[p(a|\mathcal{T},\mathcal{O})]$
- Loss $\ell = \log(\sigma(f(\mathcal{O}\cup a_\star, \hat{\mathcal{T}}) - b(\mathcal{O}\cup a_\star))) + \sum \log(1 - \sigma(f(\mathcal{O}\cup a', \hat{\mathcal{T}}) - b(\mathcal{O}\cup a')))$

Performance

General performance and ablation on RAVEN

Method	Acc	Center	2x2Grid	3x3Grid	L-R	U-D	O-IC	O-IG
LSTM	13.07%	13.19%	14.13%	13.69%	12.84%	12.35%	12.15%	12.99%
WReN-NoTag-Aux	17.62%	17.66%	29.02%	34.67%	7.69%	7.89%	12.30%	13.94%
CNN	36.97%	33.58%	30.30%	33.53%	39.43%	41.26%	43.20%	37.54%
ResNet	53.43%	52.82%	41.86%	44.29%	58.77%	60.16%	63.19%	53.12%
ResNet+DRT	59.56%	58.08%	46.53%	50.40%	65.82%	67.11%	69.09%	60.11%
CoPINet	**91.42%**	**95.05%**	**77.45%**	**78.85%**	**99.10%**	**99.65%**	**98.50%**	**91.35%**
WReN-NoTag-NoAux	15.07%	12.30%	28.62%	29.22%	7.20%	6.55%	8.33%	13.10%
WReN-Tag-NoAux	17.94%	15.38%	29.81%	32.94%	11.06%	10.96%	11.06%	14.54%
WReN-Tag-Aux	33.97%	58.38%	38.89%	37.70%	21.58%	19.74%	38.84%	22.57%
CoPINet-Backbone-XE	20.75%	24.00%	23.25%	23.05%	15.00%	13.90%	21.25%	24.80%
CoPINet-Contrast-XE	86.16%	87.25%	71.05%	74.45%	97.25%	97.05%	93.20%	82.90%
CoPINet-Contrast-CL	90.04%	94.30%	74.00%	76.85%	99.05%	99.35%	98.00%	88.70%
Human	84.41%	95.45%	81.82%	79.55%	86.36%	81.81%	86.36%	81.81%
Solver	100%	100%	100%	100%	100%	100%	100%	100%

General performance and ablation on PGM

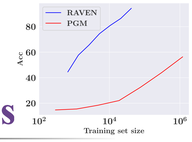
Method	CNN	LSTM	ResNet	Wild-ResNet	WReN-NoTag-Aux	CoPINet
Acc	33.00%	35.80%	42.00%	48.00%	49.10%	**56.37%**

Method	WReN-NoTag-NoAux	WReN-NoTag-Aux	WReN-Tag-NoAux	WReN-Tag-Aux
Acc	39.25%	49.10%	62.45%	77.94%

Method	CoPINet-Backbone-XE	CoPINet-Contrast-XE	CoPINet-Contrast-CL	CoPINet
Acc	42.10%	51.04%	54.19%	56.37%

Small data learning of CoPINet on RAVEN and PGM

- Log-linear on RAVEN
- Log-quadratic on PGM



Remaining Questions

- Generalization: Generalize to other configurations
- Generability: Answer generation
- Transferability: Apply knowledge learned elsewhere