Weekly Study Report

Wang Ma

2025-03-25

Electrical, Computer, and Systems Engineering Department
Rensselaer Polytechnic Institute

Outline

1. Continuing Experiments

1.1 The Results on Corrupted CIFAR-10

Previously, we trained model on Clean/high-quality data and test on the corrupted data, and we found that

- Aleatoric uncertainty and Epistemic uncertainty increase simultaneously, and they have a strong linear relationship, this phenomenon is also found by some published paper.
- Increasing AU and EU are reasonable, because AU measures the randomness in the data and EU measures whether the test data align withe the training data distribution, but the linear relationship seems problematic.

Last week, we

- Trained model on clean CIFAR-10 and Corrupted CIFAR-10 with Corruption Types: Gaussian Noise, Impulse Noise, Gaussian Blur and Contrast.
- Then we test the new model on Corrupted CIFAR-10, to see what's the ranking changes of the selected corruption types and what about others.

1.1 The Results on Corrupted CIFAR-10

My expectation is

- The EU should decrease, because the training data contains corrupted data, the model will have the knowledge of the corrupted version of data.
- The AU whether keeps or decreases, but I expect that the relative ranking should keep. Like if *Impulse Noise* data has the highest AU in the original model (only trained on clean data), then it should also have highest AU in the new model (trained on both clean and corrupted data).
- This way, even though the AU and EU are dependent, but at least AU can be robust.

1.1 The Results on Corrupted CIFAR-10

1.1.1 Preliminary Results

	Epistemic Uncertainty	Aleatoric Uncertainty	Total Uncertainty
Rank Correlation	0.6234	0.833	0.7662

- AU has high rank correlation, which means AU is robust and the rank mostly keeps.
- EU has lower rank correlation, this means after training on corrupted images, the model is can perform well on some corruption types, so the EU of them will decrease and change the ranking.

1.1.2 Unexpected Results

The model trained on different corruption types obatin a good generalization ability, the performance increases also on other corruption types.

1.1 The Results on Corrupted CIFAR-10

Corruption Type	Noise Level	Original Rank	Current Rank	Rank Change
contrast	4	25	63	-38
speckle_noise	2	31	61	-30
gaussian_blur	4	22	51	-29
shot_noise	3	14	40	-26
speckle_noise	3	23	49	-26
shot_noise	2	43	68	-25
gaussian_blur	3	41	64	-23
gaussian_blur	5	7	30	-23
contrast	3	57	80	-23
defocus_blur	5	11	34	-23
impulse_noise	2	24	46	-22
zoom_blur	4	29	50	-21
gaussian_noise	2	21	42	-21
impulse_noise	1	53	72	-19
impulse_noise	3	13	32	-19
gaussian_noise	1	50	69	-19
speckle_noise	4	12	31	-19
shot_noise	4	10	28	-18

1.2 Continuing Experiments

- Previously, we trained the model with 4 corruption types, which enables the model to generalize very well on unselected corruption types. This is good to the model performance, but not good for our analysis.
- So this time we need to **limited the selected corruption types** in the training. We only train the model on clean data and corrupted data with gaussian noise, then we analysis the results.

The expectation:

- Only the Epistemic Uncertainty on Gaussian Noise data decreases, and others' should be relevant consistent.
- The rank correlation of aleatoric uncertainty should be consistent.

Only Train on Gaussian Noise	Epistemic Uncertainty	Aleatoric Uncertainty	Total Uncertainty
Rank Correlation	0.533	0.644	0.602

1.2 Continuing Experiments

Unexpected results: different corruption types have complex relationships, the model only trained on gaussian noise can generalize to other noise types.

Epistemic Uncertainty - Noise Type-Wise Ranking Changes

Corruption Type	Avg Rank Change	Avg Absolute Change
gaussian_noise	-54.60	54.60
shot_noise	-49.60	49.60
speckle_noise	-48.80	48.80
impulse_noise	-28.40	28.40
glass_blur	+24.20	24.20
spatter	+23.60	23.60
elastic_transform	+21.60	21.60
fog	+21.00	21.00
contrast	+17.60	17.60
snow	+16.40	16.40

Aleatoric Uncertainty - Noise Type-Wise Ranking Changes

Corruption Type	Avg Rank Change	Avg Absolute Change
shot_noise	-43.40	43.40
gaussian_noise	-43.20	43.20
speckle_noise	-43.00	43.00
spatter	+24.40	24.40
fog	+24.00	24.00
contrast	+22.60	22.60
elastic_transform	+18.40	18.40
jpeg_compression	-18.00	18.00
impulse_noise	-16.80	16.80
glass_blur	+15.40	15.40

1.2 Continuing Experiments

Uncertainty Values Comparison

Aleatoric Uncertainty

Corruption Type	Original Value	Gaussian-only Value	Absolute Change	Percentage Change	Accuracy Change
contrast	0.1872	0.3060	+0.1187	+63.42%	-6.55%
fog	0.1295	0.2026	+0.0731	+56.45%	-6.80%
gaussian_noise	0.3021	0.1320	-0.1701	-56.29%	+42.68%
shot_noise	0.2493	0.1193	-0.1300	-52.16%	+27.13%
speckle_noise	0.2399	0.1198	-0.1201	-50.06%	+23.73%
spatter	0.1196	0.1663	+0.0467	+39.00%	-1.74%
impulse_noise	0.2977	0.1862	-0.1114	-37.43%	+31.09%
elastic_transform	0.1374	0.1772	+0.0399	+29.02%	-1.87%
brightness	0.0834	0.1066	+0.0232	+27.83%	-0.51%
glass_blur	0.2040	0.2550	+0.0510	+24.99%	-2.00%
jpeg_compression	0.1673	0.1337	-0.0336	-20.09%	+4.99%
saturate	0.1148	0.1356	+0.0208	+18.12%	-0.40%
snow	0.1460	0.1689	+0.0229	+15.69%	-0.36%
motion_blur	0.1889	0.2096	+0.0207	+10.95%	+1.05%

Epistemic Uncertainty - Noise Type-Wise Ranking Changes

Corruption Type	Avg Rank Change	Avg Absolute Change
gaussian_noise	-54.60	54.60
shot_noise	-49.60	49.60
speckle_noise	-48.80	48.80
impulse_noise	-28.40	28.40
glass_blur	+24.20	24.20
spatter	+23.60	23.60
elastic_transform	+21.60	21.60
fog	+21.00	21.00
contrast	+17.60	17.60
snow	+16.40	16.40

1.2 Continuing Experiments

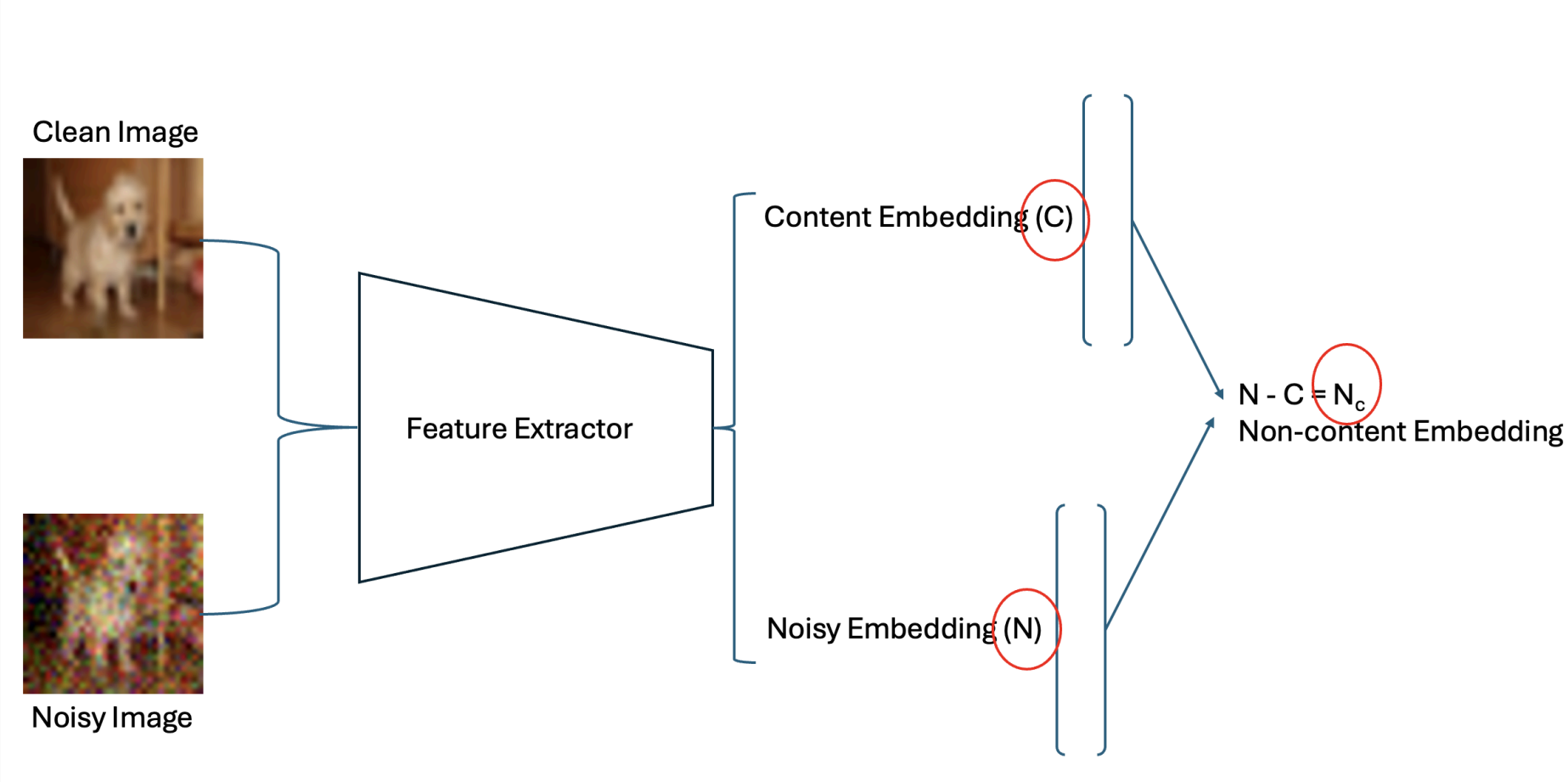
Uncertainty Values Comparison Epistemic Uncertainty

Corruption Type	Original Value	Gaussian-only Value	Absolute Change	Percentage Change	Accuracy Change
gaussian_noise	0.2715	0.0873	-0.1841	-67.82%	+42.68%
shot_noise	0.2181	0.0815	-0.1366	-62.62%	+27.13%
speckle_noise	0.2091	0.0829	-0.1262	-60.36%	+23.73%
impulse_noise	0.2925	0.1235	-0.1689	-57.76%	+31.09%
fog	0.1079	0.1528	+0.0449	+41.67%	-6.80%
spatter	0.0877	0.1211	+0.0334	+38.04%	-1.74%
brightness	0.0618	0.0847	+0.0229	+37.02%	-0.51%
glass_blur	0.1366	0.1772	+0.0406	+29.69%	-2.00%
elastic_transform	0.1087	0.1393	+0.0306	+28.14%	-1.87%
saturate	0.0910	0.1131	+0.0221	+24.24%	-0.40%
snow	0.1068	0.1302	+0.0234	+21.91%	-0.36%
contrast	0.1717	0.2085	+0.0368	+21.45%	-6.55%
jpeg_compression	0.1277	0.1014	-0.0263	-20.62%	+4.99%

2. An Experiment Design to Disentangle Uncertainty

2. An Experiment Design to Disentangle Uncertainty

Non-VAE approach



2. An Experiment Design to Disentangle Uncertainty

After getting the Content Embedding C and Non-content Embedding N_c, we can do

- regression analysis like
  - AU = $f_1(N_c)$
  - EU = $f_2(C, N_c)$
- correlation analysis about $\{C, N_c, EU\}$ and $\{N_c, AU\}$

Here, C and N_c are vectors of the size 512×1, and AU, EU are single numbers.

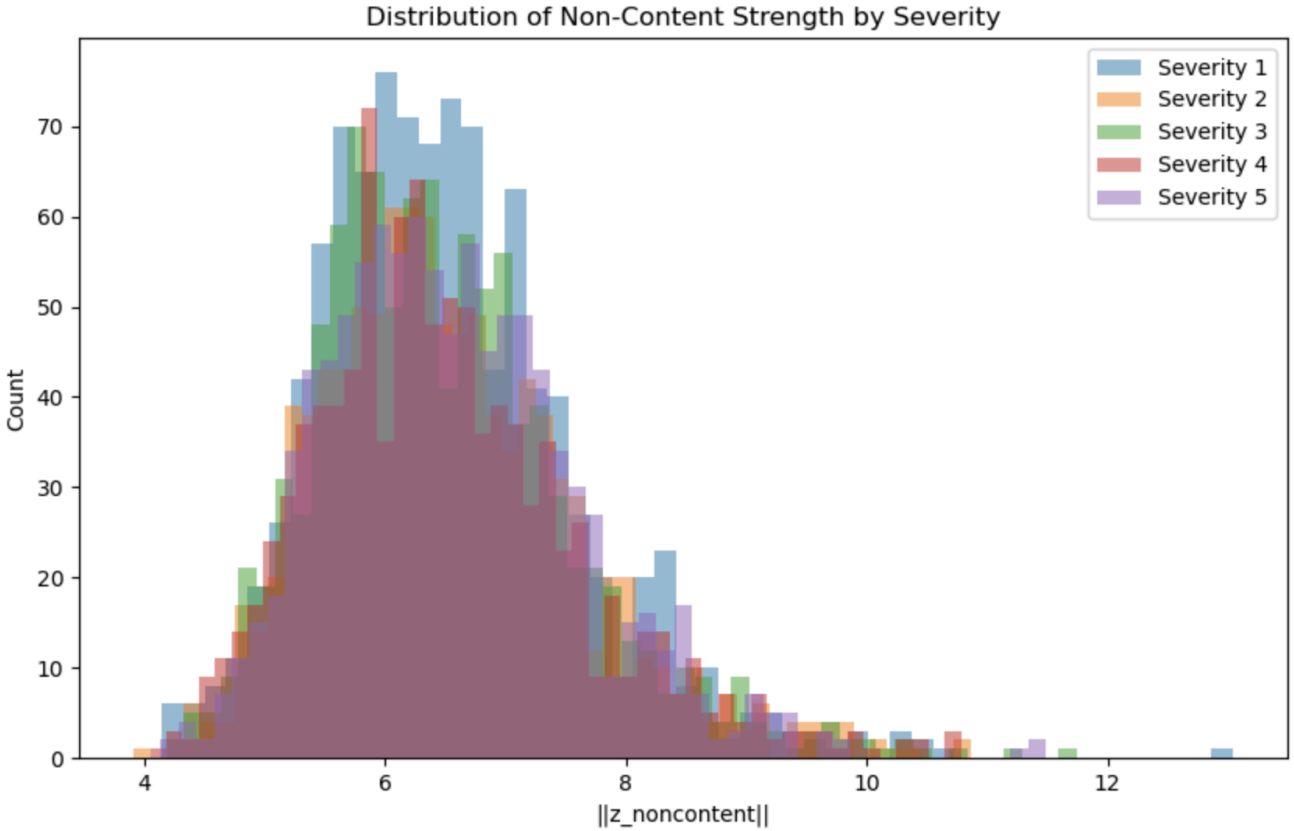
2.1 The results are not good

In the correlation analysis and regression analysis (linear regression and MLP are used), we found there is nearly no relation in $\{C, N_c \rightarrow EU\}$ and $\{N_c \rightarrow AU\}$.

In the analysis, the correlation coefficient are close to 0 and the regression R^2 are very small.

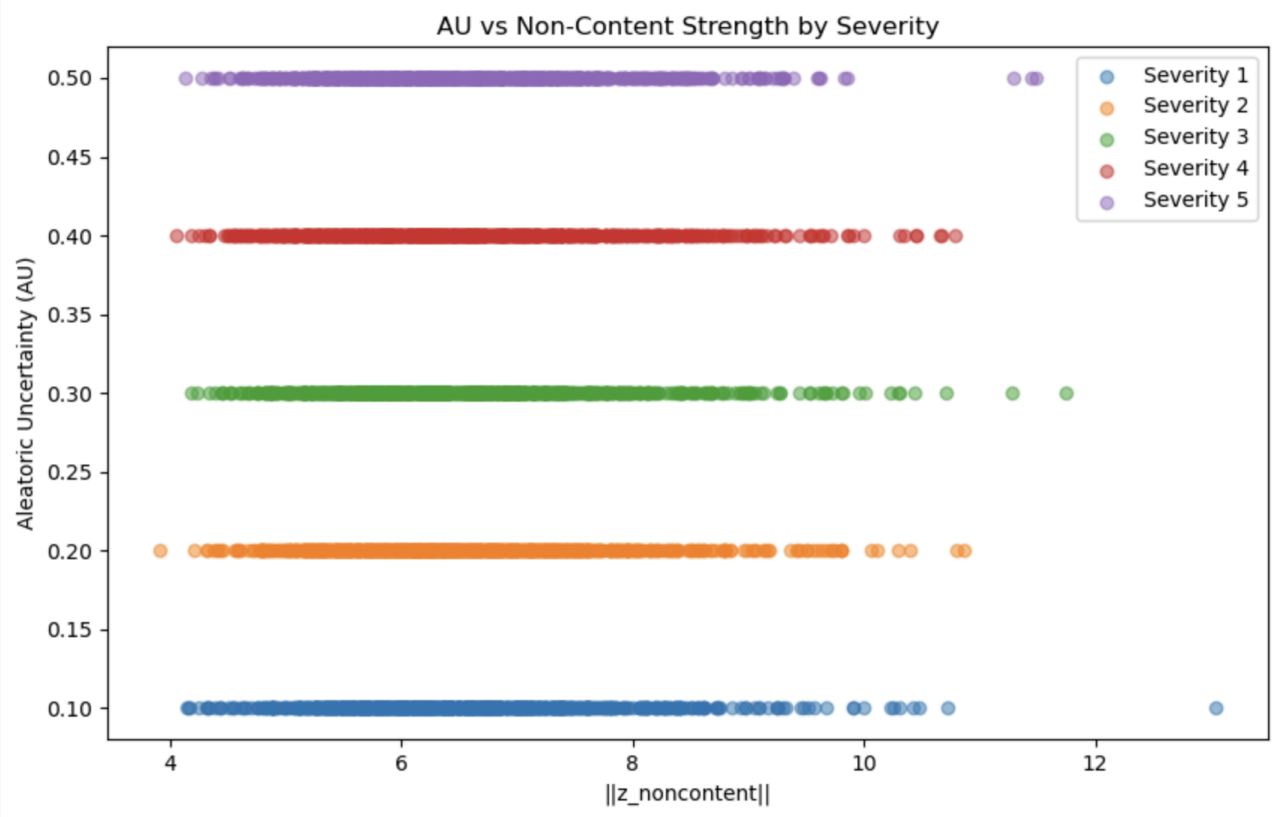
And we find that for different severities, the norm of the Non-content N_c vectors are similar. There's no such relationship that high severity has higher values in the non-content vector, and low severity has lower values.

2.1 The results are not good



2.1 The results are not good

The Aleatoric uncertainty does not rely on the norm of the non-content vector.



2.1 The results are not good

Different image, different severity, the uncertainty scores may have different sclaes.

2.2 Reflection

1. Check the implementation of the pipeline.
2. Non-linear regression methods should be considered.
3. The deterministic feature extractor may not sensitive to noise.
4. The modeling $\{N_c \rightarrow AU\}$ and $\{C, N_c \rightarrow EU\}$ may be not correct.
5. We have found that saliency map is robust to noise, which may be a good supervisor to (correct) Epistemic Uncertainty, we can think about how to use a saliency map to guide to calibrate epistemic uncertainty or combining them in downstream tasks to recognize high AU data.

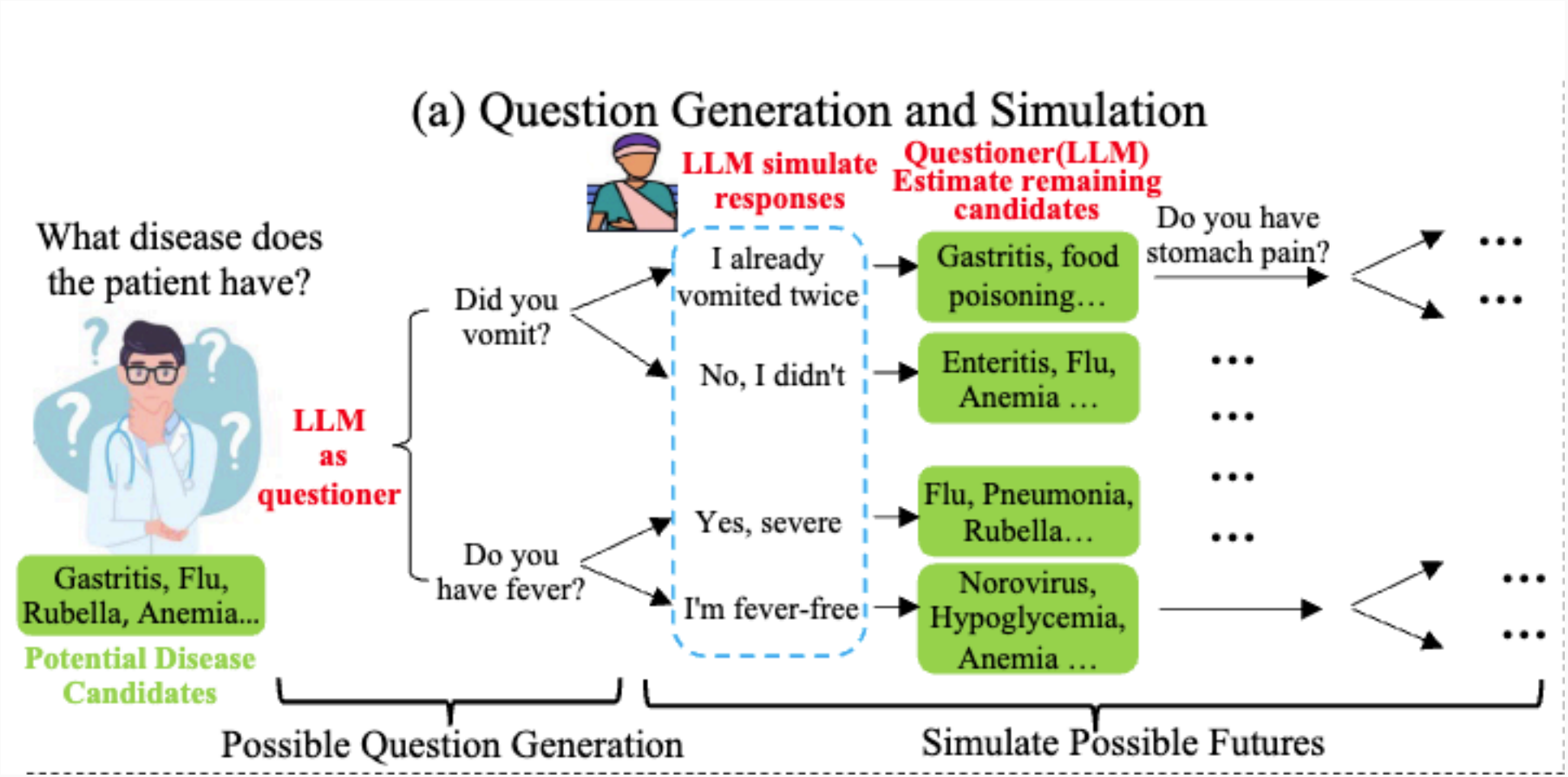
3. [NeurIPS 2024] Uncertainty of Thoughts Enhance Information Seeking in LLMs

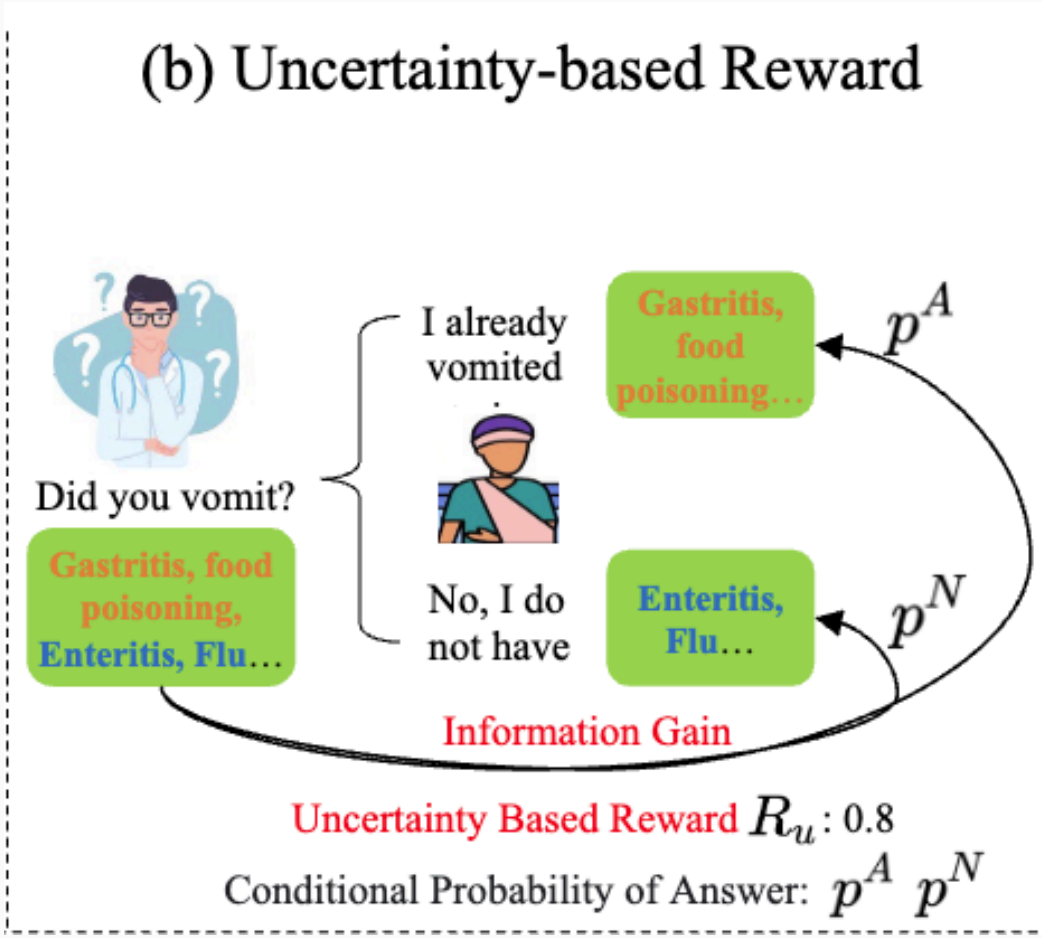
Uncertainty of Thoughts: Uncertainty-Aware Planning Enhances Information Seeking in Large Language Models

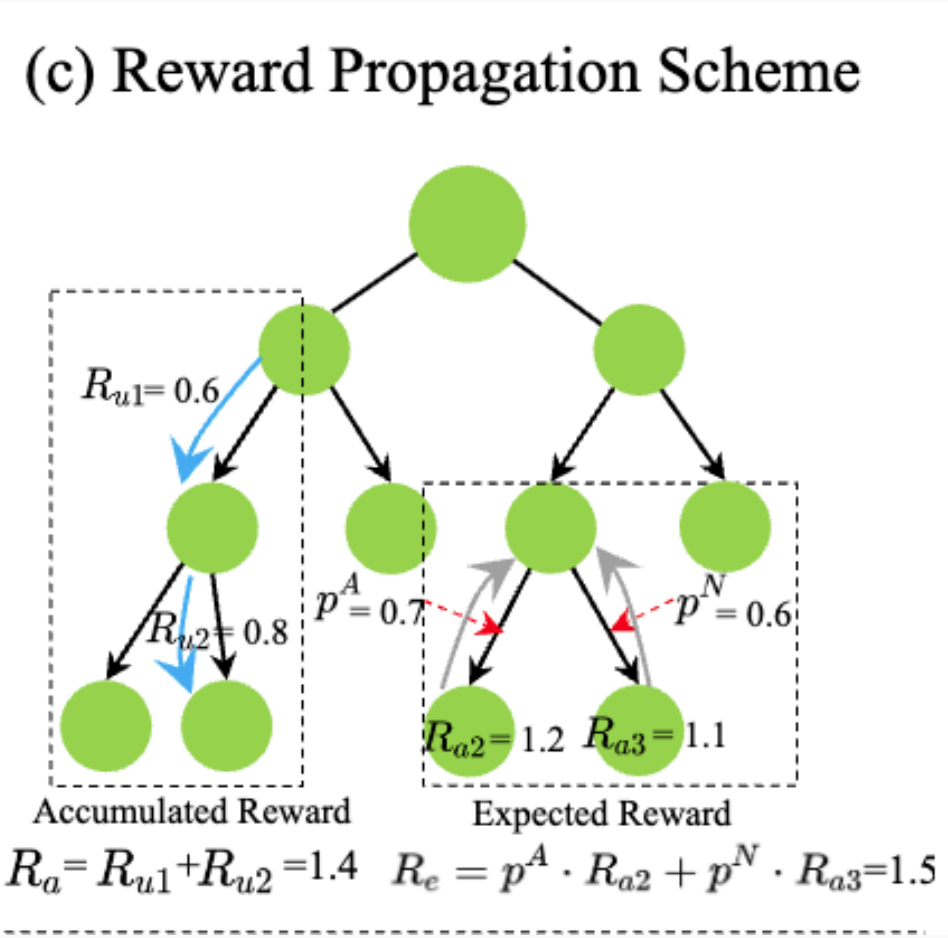
Zhiyuan Hu[1*] **Chumin Liu**[2] **Xidong Feng**[3] **Yilun Zhao**[4]
See-Kiong Ng[1] **Anh Tuan Luu**[2] **Junxian He**[5] **Pang Wei Koh**[6] **Bryan Hooi**[1]
[1] National University of Singapore [2] Nanyang Technological University
[3] University College London [4] Yale University
[5] The Hong Kong University of Science and Technology [6] University of Washington

3.1 The Proposed Method

- Question Answering settings (Medical Diagnosis, the doctors need to know the patient's disease by asking questions.)
- The LLM is the Questioner, the goal is in every round, proposing a question that can reduce the most uncertainty of the Questioner.


(a) Question Generation and Simulation
LLM simulate responses
Questioner(LLM) Estimate remaining candidates
What disease does the patient have?
LLM as questioner
Did you vomit?
Do you have fever?
I already vomited twice
No, I didn't
Yes, severe
I'm fever-free
Gastritis, food poisoning…
Enteritis, Flu, Anemia …
Flu, Pneumonia, Rubella…
Norovirus, Hypoglycemia, Anemia …
Do you have stomach pain?
Gastritis, Flu, Rubella, Anemia...
Potential Disease Candidates
Possible Question Generation
Simulate Possible Futures


(b) Uncertainty-based Reward
Did you vomit?
I already vomited
No, I do not have
Gastritis, food poisoning…
Enteritis, Flu…
Gastritis, food poisoning, Enteritis, Flu…
p^A
p^N
Information Gain
Uncertainty Based Reward R_u: 0.8
Conditional Probability of Answer: p^A p^N


(c) Reward Propagation Scheme
R_{u1}= 0.6
R_{u2}= 0.8
p^A= 0.7
p^N= 0.6
R_{a2}= 1.2
R_{a3}= 1.1
Accumulated Reward
Expected Reward
$R_a = R_{u1} + R_{u2} = 1.4$
$R_e = p^A \cdot R_{a2} + p^N \cdot R_{a3} = 1.5$

3.2 Where is the Uncertainty?

- First Round.

  Candidates: $\Omega = \{$Gastritis (0.25), food Poisoning(0.25), Enteritis(0.25), Flu (0.25)$\}$

  Uncertainty = $H(\Omega) = \sum_{i}^{4} p_i \log p_i$

- After one Question-Answer conversation, in the second round:

  Candidates: $\Omega_1 = \{$Gastritis (0.5), Enteritis(0.25), Flu (0.25)$\}$

  Uncertainty = $H(\Omega_1) = \sum_{i}^{3} p_i \log p_i$

Then the uncertainty reduction is $H(\Omega) - H(\Omega_1)$. The model tends to ask the questions and can reduce the most uncertainty in every round to make the conversation efficient and accurate.

3.2 Where is the Uncertainty?

So this is a paper specifically for QA task, from the entropy reduction point of view, they want to make the conversation efficient and accurate.