Localizing Text Across Domains Using RARR Attribution Technique

Soumen Kumar Mondal
23m2157@iitb.ac.in

Guide: Prof. Preethi Jyothi

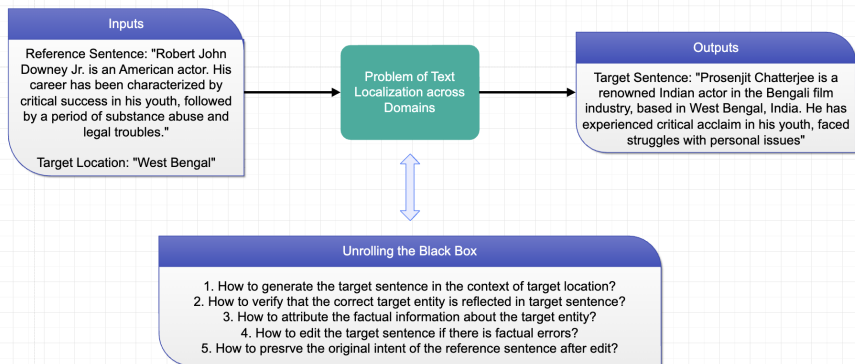Indian Institute of Technology Bombay

May 8, 2024
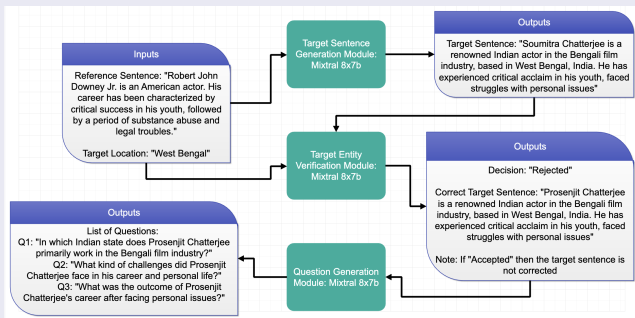
Problem Statement

What is Text Localization?

Text localization and adaptation across domains involve modifying textual content to suit different cultural and geographical contexts while maintaining the original message's context, style and intent.

Problem of Text Localization

Inputs

Reference Sentence: "Robert John Downey Jr. is an American actor. His career has been characterized by critical success in his youth, followed by a period of substance abuse and legal troubles."

Target Location: "West Bengal"

Problem of Text Localization across Domains

Outputs

Target Sentence: "Prosenjit Chatterjee is a renowned Indian actor in the Bengali film industry, based in West Bengal, India. He has experienced critical acclaim in his youth, faced struggles with personal issues"

Unrolling the Black Box

1. How to generate the target sentence in the context of target location?
2. How to verify that the correct target entity is reflected in target sentence?
3. How to attribute the factual information about the target entity?
4. How to edit the target sentence if there is factual errors?
5. How to presrve the original intent of the reference sentence after edit?

RARR System

What is RARR?

The Retrofit Attribution using Research and Revision (RARR) model is an approach designed to use for text attribution, ensuring that the content does not contain any factual misinformation while preserving the intent of the original text.

Overview of Original RARR System[1]

[1] Luyu Gao et al. "RARR: Researching and Revising What Language Models Say, Using Language Models". In: (2023). arXiv: 2210.08726 [cs.CL].

Stage 1: Question Generation

Workflow of Stage 1

Critical Analysis of Stage 1

- Mixtral 8x7b model[2] output is not deterministic when the reference sentence contains facts such as date, time, place of incidents.
- Target Entity Verification Module can be checked by GPT 3.5 or 4 for effective results.

[2] Albert Q. Jiang and Alexandre Sablayrolles. "Mixtral of Experts". In: (2024). arXiv: 2401.04088 [cs.LG].
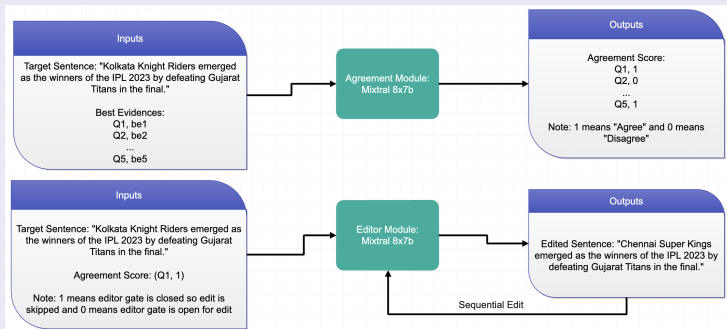
Stage 2: Evidence Retrieval

Workflow of Stage 2
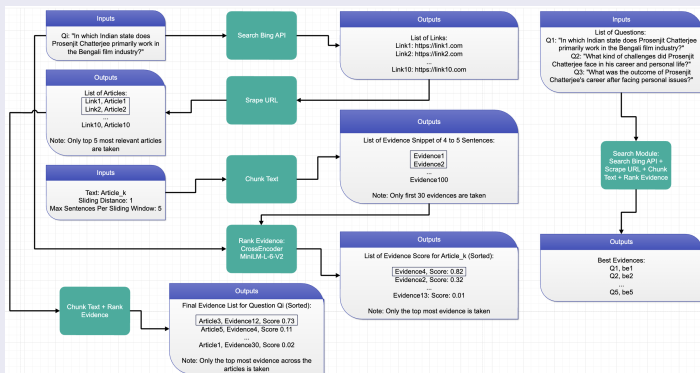
Critical Analysis of Stage 2

- Bing is unable to retrieve the relevant links if the target location is absent in the question.
- The first 30 passages from each article may not contain the relevant information.
- Instead of CrossEncoder, different embedding model can be used.

Stage 3: Attribution

Workflow of Stage 3

Critical Analysis of Stage 3

- Sequential editing increases the length of the output target sentence which fails to preserve the style and intent of original target sentence.
- RARR helps in attribution but fails to outperform Mixtral 8x7b localize text based on target location.

Evaluation Dataset

Sample Example of Dataset

Category, Reference and Target Location	Reference Sentence	Target Sentence	Common Questions
Actor/Actress, US, Kerala	Dwayne Douglas Johnson, also known by his ring name the Rock, is an American actor and professional wrestler currently signed to WWE.	Mohanlal, also known by his pet name "Lalettan", is an Indian actor and producer, currently working in the Malayalam film industry.	(i) Can you give an example of an actor who professionally holds another title as well? (ii) Can you provide the name of an actor who is well-known in the entertainment industry by another name?

Critical Analysis of Dataset

- There could be multiple answer to the common question in the context of target location.
- Human annotation is needed to improve the quality of common questions.

Common Question Evaluation Metric

Definition

- We use few shot prompting with GPT 3.5 to check if the answer to the common question in the context of target location is contained in the target sentence.
- S^i: A target sentence (RARR edited target sentence output or Mixtral 8x7b output.
- m_i: The number of generated questions.
- $\{q_j^i\}_{j=1}^{m_i}$: The list of generated questions.
- $\{A_j^i\}_{j=1}^{m_i}$: The binary score provided by GPT 3.5.
- N: The total number of sentences.
- $Q = \sum_{i=1}^{N} m_i$: Total number of questions over this N sentences.

$$C_{CQ} = \frac{1}{Q} \sum_{i=1}^{N} \sum_{j=1}^{m_i} A_j^i \tag{1}$$

Interpretation

- A higher score C_{CQ} indicates that the model is more effective at editing target sentences that contain accurate and relevant answers to the posed questions.
- This measure directly reflects the model's capability to understand and adapt the context of the reference location to the target location.

Results and Discussion

Attribution Results by RARR

Reference Sentence, Target Sentence and Entity Verification

"claim_ref": "Robert John Downey Jr. is an American actor. His career has been characterized by critical success in his youth, followed by a period of substance abuse and legal troubles, and a surge in popular and commercial success later in his career.",
"claim_target": "Prosenjit Chatterjee is a renowned Indian actor in the Bengali film industry, based in West Bengal, India. He has experienced critical acclaim in his youth, faced struggles with personal issues, and made a comeback with successful films later in his career.",
"is_target_claim_ok": "Accepted",

Questions, Best Evidences, Agreement and Edit Revisions - 1

"claim": "Prosenjit Chatterjee is a renowned Indian actor in the Bengali film industry, based in West Bengal, India. He has experienced critical acclaim in his youth, faced struggles with personal issues, and made a comeback with successful films later in his career.",
"query": "In which Indian state does Prosenjit Chatterjee primarily work in the Bengali film industry?",
"evidence": "Prosenjit Chatterjee (born 30 September 1962) is an Indian actor and producer. He is widely regarded as one of the leading actors in modern Bengali cinema. He predominantly works in Bengali cinema . He is the son of veteran Bollywood actor Biswajit Chatterjee .",
"is_open": true,
"decision": "This disagrees with what you said.",
"edited_claim": "Prosenjit Chatterjee is a renowned Indian actor based in Kolkata, West Bengal, India. He has experienced critical acclaim in his youth, faced struggles with personal issues, and made a comeback with successful films later in his career. He predominantly works in Bengali cinema. He is the son of veteran Bollywood actor Biswajit Chatterjee."

Questions, Best Evidences, Agreement and Edit Revisions - 3

"claim": "Prosenjit Chatterjee is a well-known Indian actor based in Kolkata, West Bengal, India. He has gained critical acclaim in his youth, faced struggles with personal issues, and made a comeback with successful films later in his career. He predominantly works in Bengali cinema. He is the son of veteran Bollywood actor Biswajit Chatterjee.",
"query": "What was the outcome of Prosenjit Chatterjee's career after facing personal issues?",
"evidence": "With a career spanning more than 40 years, Chatterjee continues to mesmerize and reinvent himself as an actor, defying the limits of age and artistic boundaries. His latest OTT stint is a testament to how Prosenjit is reinventing himself with each of his works. The celebrated actor has embraced the spirit of innovation, constantly pushing his creative boundaries to remain relevant and enthralling in an ever-changing industry",
"is_open": false,
"decision": "This agrees with what you said.",
"edited_claim": "Prosenjit Chatterjee is a well-known Indian actor based in Kolkata, West Bengal, India. He has gained critical acclaim in his youth, faced struggles with personal issues, and made a comeback with successful films later in his career. He predominantly works in Bengali cinema. He is the son of veteran Bollywood actor Biswajit Chatterjee."

Original Sentence and Final Edit Revisions

"claim_original": "Prosenjit Chatterjee is a renowned Indian actor in the Bengali film industry, based in West Bengal, India. He has experienced critical acclaim in his youth, faced struggles with personal issues, and made a comeback with successful films later in his career.",
"claim_attributed": "Prosenjit Chatterjee is a well-known Indian actor based in Kolkata, West Bengal, India. He has gained critical acclaim in his youth, faced struggles with personal issues, and made a comeback with successful films later in his career. He predominantly works in Bengali cinema. He is the son of veteran Bollywood actor Biswajit Chatterjee."

Key Findings

- Sequential edit increases the length of the target sentence. Furthermore, it changes the style and intent of original sentence.

Results and Discussion [Contd.]

Localization Results

Model	Common Question Evaluation Metric C_{CQ}
RARR Baseline	0.5836
RARR Single Edit	0.6411
Mixtral 8x7b	0.6439

Discussion

- The results are averaged over 200 samples with 3 instances of target sentence generation.
- The results show that the Mixtral 8x7b outperforms RARR model in the common question evaluation metric indicating the improvement required in the RARR baseline model.
- However, this metric does not capture the attribution quality of the RARR.
- It has also been observed that the RARR model is better than Mixtral 8x7b in terms of attribution but cannot outperform Mixtral 8x7b in text localization.
- Ablation study shows that the single edit is effective for the common question metric than the sequential edit.
- Changing the number of passages and embedding type do not affect the common question metric.

Conclusion and Future Work

Conclusion

- This study has explored the effectiveness of the RARR model in the field of text localization and adaptation across diverse geographical domains.
- The evaluation of the model using a comprehensive dataset and a set of common questions has revealed that the RARR model performs better in attribution of factual information but cannot beat Mixtral 8x7b for text localization.

Future Works

- **Improving Model Robustness:** Future iterations of the RARR model will focus on enhancing the search module of RARR model indicating the use of Retrieval Augmented Models.
- **Enhanced Evaluation Metrics:** The metric currently is solely based on GPT 3.5 model. However, we can use GPT 4 model[3] for the evaluation purpose.

---

[3] OpenAI. "GPT-4 Technical Report". In: (2024). arXiv: 2303.08774 [cs.CL].

References

📄 Gao, Luyu et al. "RARR: Researching and Revising What Language Models Say, Using Language Models". In: (2023). arXiv: 2210.08726 [cs.CL].

📄 Jiang, Albert Q. and Alexandre Sablayrolles. "Mixtral of Experts". In: (2024). arXiv: 2401.04088 [cs.LG].

📄 OpenAI. "GPT-4 Technical Report". In: (2024). arXiv: 2303.08774 [cs.CL].

Thank You!