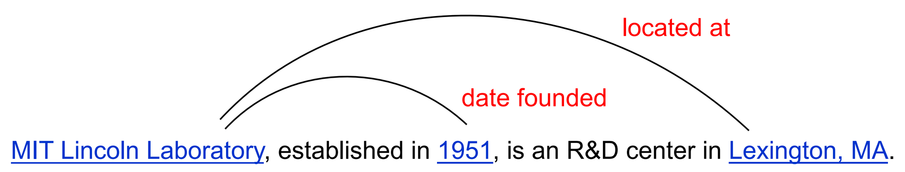
Zero-Shot Relation Extraction using Large Language Models (LLMs)

Motivation & Problem

- **Motivation:** The Department of Defense (DoD) conducts study and analysis for human networks of interest to the U.S. government (such as terrorists, gangs, cartels, and traffickers) and provides recommendations of potential courses of action to decision makers
- MIT LL develops technologies to aid in the construction, characterization, and modeling of human networks
- Automatic extraction of named entities of interest and their relationships from intelligence reports are integral components
- **Problem:** Identifying relations between entities in documents from different sources and sentences structures need to be executed with <u>high precision</u>
- Our goal is to minimize the number of annotations that analysts must manually complete while maintaining, or improving, precision for relation extraction (RE) tasks

RE Task

- **Closed RE Problem Definition:** Given a sentence s_i, the set of entities E in s_i, and a set of relation types of interest R; extract the set of all existing relations between pairs of entities present in s_i.
- **Open RE Problem Definition:** Given a sentence s_i, extract the set of all existing entities E and their corresponding relationships R.



- **RE Output:** Set of sentence-level sets triplets (e_i, r_j, e_k) where $e_i \in E$, $e_k \in E$, $r_j \in R$ and $i \neq k$

Datasets

Re-TACRED (Revised-TACRED)[1]:

- Sentence-level RE dataset with 38 total relations, ~58.5k training, ~13.5k testing, and ~19.5k validation samples
- Included sentences are sourced from a newswire and webtext corpus from 2009-2004
- 23% of the original TACRED dataset was improved to make Re-TACRED

Synthetically generated dataset:

- Sentence-level RE dataset with 24 total relations, ~4.7k training, ~1.2k testing, and ~1.4k validation samples
- Included sentences were generated following the sentence structures found in intelligence reports
- The gold standard relations <u>contain only a subset of the relationships</u> present in the sentences

Dataset downsampling:

- To evaluate our baseline's performance across zero- and few-shot instances, Re-TACRED and the synthetically generated dataset were downsampled such that each relation under evaluation had at most 100 and 50 sentence examples in the test set, respectively
- The subset of relations that did not have enough sentence examples were discarded

Targeted Zero-Shot Relation Extraction Approach

Worked Example

s_1: MIT Lincoln Laboratory (MIT LL), established in 1951, is an R&D center in Lexington, MA.

E: {MIT LL, 1951, Lexington MA} R: {date founded, located at, CEO of, subsidiary}

Ground truth relations in s_1: {<MIT LL, date founded, 1951>, <MIT LL, located at, Lexington MA >}

Method 1: Closed Relation Extraction with T/F Prompting

For every possible entity pair and relation combination; prompt an LLM to determine if the relation is "True" based on the provided context in s_i.

Benefits:

- Existing evaluation approaches are sufficient

Disadvantages:

- The approach is too inefficient, requiring queries over all pairs of entities, over all relation types ($O(|R|*|E|^2)$)

Raw Text & Relations of Interest → Extract Entities → Preprocessing (Entity pairs → Relation QA) → Relation Extraction Model (QA LLM)

Step 1) Extract entities in the input text

- Our initial T/F prompting experiments use all entities listed in the ground truth datasets.
- Future work will incorporate an NER model for entity extraction to reduce reliance on entity ground truth data.

Step 2) Identify all for the set of possible (e_i, r_j, e_k) pairings for the given entity and relation sets

MIT LL, 1951, date founded	MIT LL, Lexington MA, date founded	1951, Lexington MA, date founded
MIT LL, 1951, located at	MIT LL, Lexington MA, located at	1951, Lexington MA, located at
MIT LL, 1951, CEO of	MIT LL, Lexington MA, CEO of	1951, Lexington MA, CEO of
MIT LL, 1951, subsidiary	MIT LL, Lexington MA, subsidiary	1951, Lexington MA, subsidiary

Step 3) Prompt the LLM to determine if the relationship exists in s_1 for each (e_i, r_j, e_k) pairing

Prompt:
context: "MIT Lincoln Laboratory, established in 1951, is a research and development center in Lexington, MA."

Using only the context text provided above, label all of the provided statements 'True' or 'False'. Answer 'True' only if all the statements are correct based off the provided context. Begin your response with 'True' or 'False' followed by a detailed explanation of your reasoning.

statements: "MIT Lincoln Laboratory is an organization. 1951 is a date. MIT Lincoln Laboratory was founded in 1951."

LLM Response:
True. All the statements are supported in the provided context because…

Method 2: Open Relation Extraction

For every sentence, prompt LLM to output all relation triplets (e_i, r_j, e_k). Provide a couple of general relation examples as part of the prompt.

Benefits of Open RE:

- Single model for joint entity and relation extraction
- Enables discovery of new relations

Disadvantages:

- Output needs to be filtered
- Challenging to evaluate
- Potential evaluation approaches:
  1. Use LLM
  2. Use embeddings and cosine similarity between canonical form of the relation and extracted triplet
  3. Alternative evaluation approaches and metrics, e.g., *topic similarity, uniqueness, factualness, granularity, completeness*[3]

Raw Text → LLM Joint Entity and Relation Extraction

Step 1) Prompt the LLM to extract the entities and relations for in s_1

Prompt:
Identify all combinations of subjective, action, objective, and possible adverbials for any given sentences, and present them in the form of triplets: <subjective, action, objective>.

Here are some examples:
Sentence: A different judge then ordered the case reviewed by a higher court.
Relations:
<A different judge, ordered, the case>, <the case, reviewed by, a higher court>
Sentence: Amy, a wife of Stan, supports his work.
Relations: <Amy, married to, Stan>, <Stan, married to, Amy>, <Amy, supports, Stan's work>

Make the triplets as atomic as possible. Make sure that each triplet cannot be broken down further. Make sure to resolve coreferences. Make sure to output a complete list of triplets of the form <subjective, action, objective> for the given sentence. Do not output anything besides the triplets.

Now here is the sentence: "MIT Lincoln Laboratory (MIT LL), established in 1951, is an R&D center in Lexington, MA."

LLM Response:
<MIT Lincoln Laboratory, established, in 1951>, <MIT Lincoln Laboratory, is, an R&D center>, <MIT Lincoln Laboratory, located in, Lexington>, <MIT Lincoln Laboratory, located in, MA>

Step 2) Filter and process the output of Step 1)

- Work in progress

Final Output

Output: Sentence-level sets of relations classified as 'True' or identified by the open RE prompt;
{<MIT LL, date founded, 1951>, <MIT LL, located at, Lexington MA>}

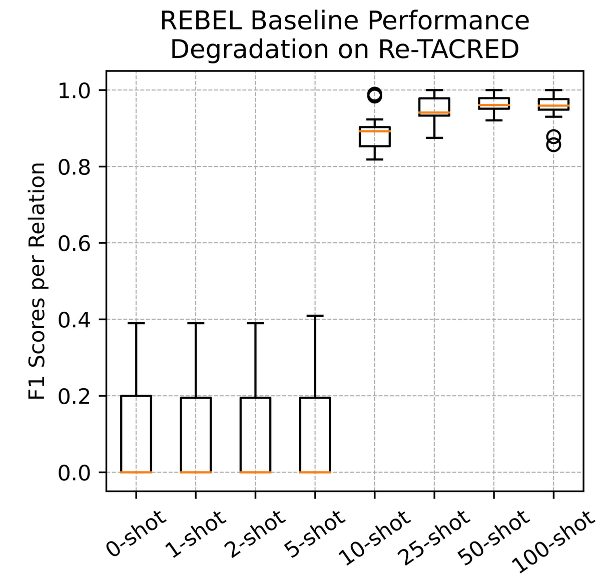
Experimental Results

High-performing RE Baseline: REBEL[2]

- Input data is formatted using triplet linearization (e.g., <triplet> example subject <subj> related entity object <obj>) and <u>assumes entities are provided</u>
- Relation extraction training and testing is done with the autoregressive model, BART-large and on Wikipedia hyperlinks with Wikidata entities
- REBEL requires fine-tuning for each dataset to prioritize contextual information provided in the test sets (reliant on lots of data for fine-tuning)

LLMs used for RE prompting:

- Mixtral 8x7B Instruct, SOTA instruction fine-tuned LLM using a "Mixture of Experts" model architecture



Zero-shot performance results on the Re-TACRED dataset

		alternate name	father	employer	top members	country of citizenship	country of headquarters	headquarters location	country of residence	identity	charge	city of residence	position held	age
REBEL BASELINE	precision	0.00	0.57	0.21	0.00	0.31	0.00	0.46	0.00	0.00	0.00	0.00	0.17	0.50
	recall	0.00	0.12	0.31	0.00	0.23	0.00	0.34	0.00	0.00	0.00	0.00	0.08	0.02
	F1	0.00	0.20	0.25	0.00	0.26	0.00	0.39	0.00	0.00	0.00	0.00	0.11	0.04
METHOD 1 APPROACH	precision	0.25	0.11	0.44	0.56	0.15	0.88	0.43	0.84	0.06	0.96	0.84	0.76	0.75
	recall	0.70	0.08	0.64	0.92	0.13	0.83	0.54	0.75	0.25	0.69	0.64	0.87	0.85
	F1	0.37	0.09	0.52	0.69	0.14	0.85	0.47	0.79	0.10	0.81	0.72	0.81	0.80

Average F1 Scores – Method 1 Approach: 0.5145; REBEL Baseline: 0.0970

Zero-shot evaluation on 3 selected relations from the synthetically generated dataset

		spouse	meet with	member of
REBEL BASELINE	precision	0.03	0.00	0.12
	recall	0.50	0.00	0.80
	F1	0.56	0.00	0.20
METHOD 1 APPROACH	precision	0.23	0.18	0.43
	recall	1.00	1.00	0.94
	F1	0.38	0.30	0.59
METHOD 2 APPROACH*	precision	1.00	0.97	1.00
	recall	0.92	0.62	0.76
	F1	0.96	0.76	0.81

*Upper bound on performance with manual evaluation since filtering algorithm will introduce additional errors

Our LLM approaches show promise for zero-shot relation extraction but are challenging to evaluate, and require significant computing resources and prompt engineering

Challenges and Future Work

Promising directions for low-resource relation extraction

Generative LLM approaches: Possible directions include few-shot prompting, classic fine-tuning, and fine-tuning with chain-of-though explanations

Main Challenges:

- With open RE, how do we evaluate if extracted relations are correct, given language variability and the fact that a model could hallucinate any relation types and that it tends to generate verbose responses?
- For both closed and open RE, generative LLMs perform quite slowly

Improve existing approaches: Modifying techniques that do not necessarily use generative LLMs

Main Challenge:

- The approaches tend to operate in a fully supervised setting and require large amounts of training data to perform well (e.g., REBEL)

1. Stoica, George, et al. "Re-tacred: Addressing shortcomings of the tacred dataset." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 35. No. 15. 2021.
2. Cabot, Pere-Lluís H., et al. "REBEL: Relation extraction by end-to-end language generation." Findings of the Association for Computational Linguistics: EMNLP 2021.
3. Jiang, Pengcheng, et al. "GenRES: Rethinking Evaluation for Generative Relation Extraction in the Era of Large Language Models." *arXiv preprint arXiv:2402.10744* (2024).

LINCOLN LABORATORY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Sheila Alemany Blanco, Danielle Sullivan, Olga Simek, Erica Cai, Benjamin Miller, Christopher Smith
MIT Lincoln Laboratory, Lexington, MA