Comparing Small Language Models and Large Language Models' ability to Extract Triples

Muhammad Qasim Ali

SPARC Machine Learning Workshop

ABSTRACT

We score [1] gpt-4 and [2] spacy, a large and small language model respectively, in their ability to extract triples from a [3] dataset (webnlg-dataset).

We created a program in python to iterate through the dataset and prompt the models to perform:
- **Named Entity Recognition (NER)**
- **Relation Extraction (RE)**

We plotted our results on histograms and Knowledge Graphs to check them.

Through testing we found the gpt-4 was superior at Few Shot Data Extraction than spacy.
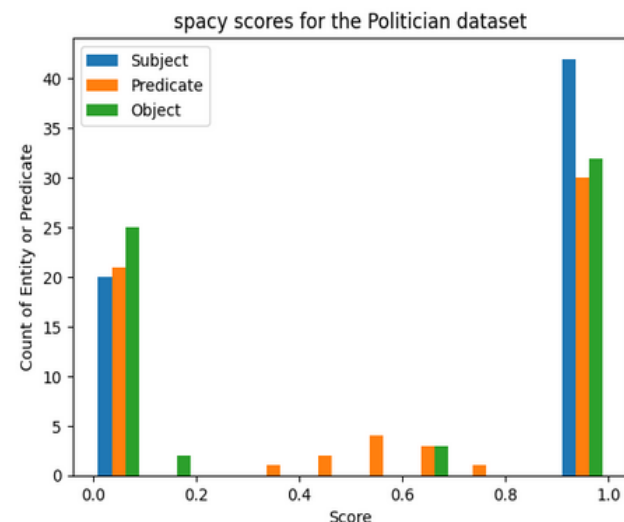
RESULTS

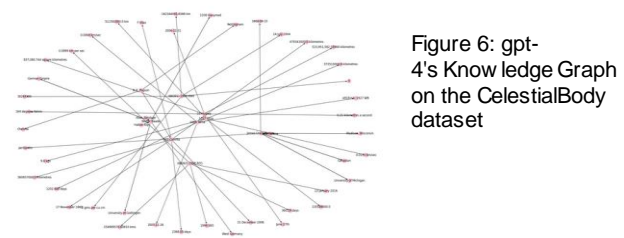Figure 4: gpt-4's performance on the Politician dataset

Figure 3: spacy's performance on the Politician dataset

Figure 6: gpt-4's Knowledge Graph on the CelestialBody dataset

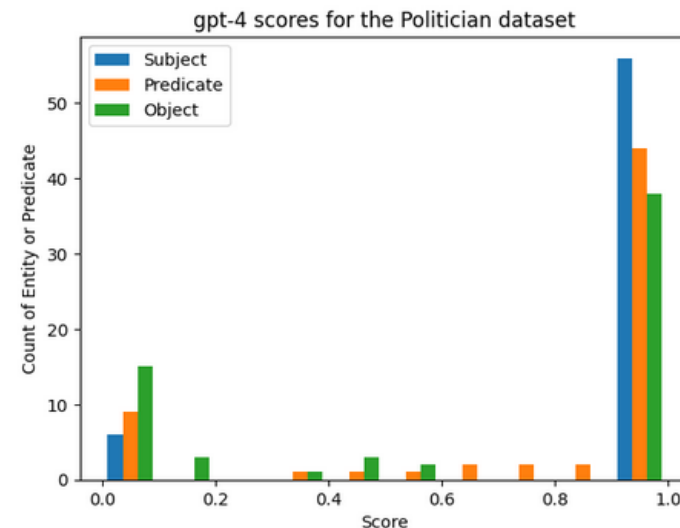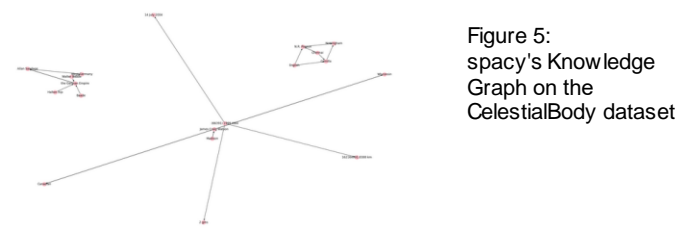Figure 5: spacy's Knowledge Graph on the CelestialBody dataset

INTRODUCTION

SLM's and LLM's differ in their ability to perform actions that they have been untrained for. LLM's are good due to the large dataset they were trained on, in addition to its deep neural network, while SML's are trained on simple sets and are structured more simply.

We compare their ability to extract triples [Figure 1], on a dataset they aren't trained on or familiar with. triples are 2 'entities' or objects joined via a 'predicate' or conjunctive.

We can plot these results on knowledge graphs and histograms to show how well the models can extract triples and their accuracy at extracting them.
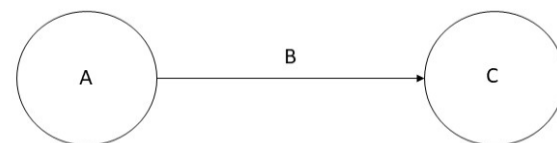

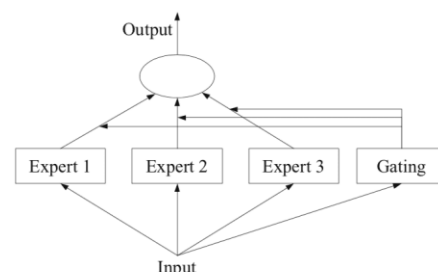Figure 1: [4] *An introduction to knowledge graphs* (2021)

Figure 2: [5] *A gentle introduction to mixture of experts ensembles (2021)*

DISCUSSION

On knowledge graphs gpt-4 extracted more entities and intern made more links on the knowledge graph.

Based on my results, gpt-4 performed better at triple extraction that spacy.
This could be because:
- Spacy is an outdated SLM and gpt-4 is the newest and most advanced LLM currently

- Spacy's limited ontology

- We could give gpt-4 examples of extracted triples through prompt engineering

While gpt-4 did perform better, both did fail to extract triples on multiple occasions, so while they can extract triples, they cannot do it consistently.

AIMS

- To build a program to automatedly iterate through a dataset and prompting gpt-4 and spacy for an output

- Score the output of the models against the truth triples

- Graph the accuracy and the links between entities

FURTHER WORK

- Test models against more complex or larger sets of data

- Try to get spacy to extract without classifiers

- Continue prompt engineering for gpt-4 so it runs better

CONCLUSIONS

- Gpt-4 performed better than spacy when it came to the extraction of triples and labeling of predicates

- Both did fail to extract entities a multitude of times, so they are not consistent in their ability to extract triples

REFERENCES

[1] OpenAI, R., 2023. GPT-4 technical report. *arXiv*, pp.2303-08774.
[2] English · spacy models documentation (no date) English. Available at: https://spacy.io/models/en (Accessed: 14 August 2023).
[3] Gardent, C., Shimorina, A., Narayan, S. and Perez-Beltrachini, L., 2017, July. Creating training corpora for nlg micro-planning. In *55th annual meeting of the Association for Computational Linguistics (ACL).*
[4] Vinay K. Chaudhri, V., Chittar, N. and Genesereth, M. (2021) *An introduction to knowledge graphs, SAIL Blog.* Available at: http://ai.stanford.edu/blog/introduction-to-knowledge-graphs/ (Accessed: 07 August 2023).
[5] Brownlee, J. (2021) *A gentle introduction to mixture of experts ensembles, MachineLearningMastery.com.* Available at: https://machinelearningmastery.com/mixture-of-experts/ (Accessed: 07 August 2023).