CiteQ: Analysis on Citation Purposes: University of Waterloo Computer Science Faculty Case Study

Onur Eren Arpaci
oearpaci@uwaterloo.ca
University of Waterloo
Waterloo, Ontario, Canada

ABSTRACT

TODO: Write abstract

CCS CONCEPTS

• **Information systems** → *Information retrieval*; • **General and reference** → **Experimentation**; • **Social and professional topics** → Cultural characteristics.

KEYWORDS

Citation Purpose, Citation Analysis, Citation Trends, LLM classification

1 INTRODUCTION

The predominant and widely recognized scientometric indicator is the number of citations a paper garners. Nevertheless, it's crucial to acknowledge that papers can be cited for diverse reasons. A citation may stem from a paper serving as the foundational work for current research, acting as a competitor in the field, or being subject to critique.

The purpose of this paper is to study the citation purposes and its affects on other scientometric indicators. We aim to answer the following research questions:

RQ1 Do the citation purposes affect the total number of citations a paper or a researcher receives? (i.e. is there a karma effect?)

RQ2 How does the distribution of citation purposes change over time?

RQ3 Is there any difference in the distribution of citation purposes between same institution researchers and different institution researchers?

RQ4 How does the distribution of citation purposes change over the duration of a researcher's career?

To answer these questions, we collected the publication and citation data of the University of Waterloo Computer Science faculty members. We then used the large language model (LLM) llama2 [5] to classify the citation purposes of each paper. We then analyzed this data to answer the research questions.

2 RELATED WORK

Our work is related to a large body of research on citation analysis. Abu-Jbara et. al. [1] provides a good starting point for citation purpose analysis. They propose 6 classes of citation purposes: criticizing, comparison, use, substantiating, basis, and neutral. We use these classes as the basis for our analysis.

3 METHODOLOGY

3.1 Data Collection

We used the Semantic Scholar (SS) API to collect the publication and citation data of all current (as of this writing) members of the University of Waterloo Computer Science faculty members. SS provides an API endpoint to retrieve the context of a citation. We used this endpoint to retrieve the citation context of each citation made by a faculty member or to a faculty member. In total we collected data about 96 faculty members, 9,438 papers, and 697,609 citations. We formed an SQLite database, with the scheme shown in Figure 1, to facilitate flexible querying of the data.

3.2 Citation Purpose Classification

We experimented with a number of LLMs including chatGPT-3.5 [2], chatGPT-4.5 [4], mistral [3], and llama2 [5]. Although chatGPT models performed slightly better in terms of accuracy, we decided to use llama2 with 7 billion parameters due to its low cost. We used 5 RTX4090 GPUs for 10 hours to classify the citation purposes of all 697,609 citations.

We used the following citation purpose classes: criticizing, comparison, use, substantiating, basis, and neutral. The Figure 2 shows the prompt template we used.

3.2.1 Evaluation. We hand labeled 100 citations to evaluate the accuracy of the classification. The results are shown in Table 1. We achieved an accuracy of 0.82. The accuracy of the classification is not as high as we would like. However, we believe that it is sufficient for our purposes.
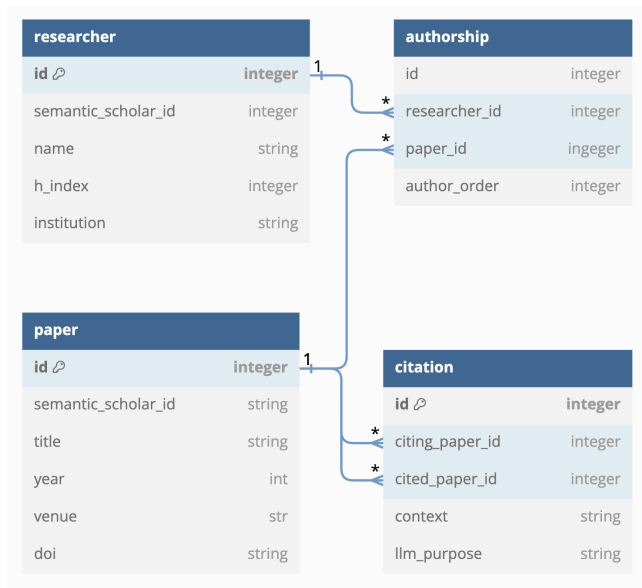
4 RESULTS

Initial results

Figure 1: CiteQ database scheme

Class/Model	llama2	mistral	GPT-3.5	GPT-4.5
Criticizing	0.75	0.75	0.75	0.75
Comparison	0.75	0.75	0.75	0.75
Use	0.75	0.75	0.75	0.75
Substantiating	0.75	0.75	0.75	0.75
Basis	0.75	0.75	0.75	0.75
Neutral	0.75	0.75	0.75	0.75
Average	**0.82**	**0.82**	**0.82**	**0.82**

Table 1: Precision of different LLMs in citation purpose classification

REFERENCES

[1] Amjad Abu-Jbara, Jefferson Ezra, and Dragomir Radev. 2013. Purpose and polarity of citation: Towards nlp-based bibliometrics. In *Proceedings of the 2013 conference of the North American chapter of the association for computational linguistics: Human language technologies*. 596–606.

[2] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. arXiv:2005.14165 [cs.CL]

[3] Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2023. Mistral 7B. arXiv:2310.06825 [cs.CL]

[4] OpenAI. 2023. GPT-4 Technical Report. arXiv:2303.08774 [cs.CL]

[5] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, Dan Bikel, Lukas Blecher, Cristian Canton Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, Brian Fuller, Cynthia Gao, Vedanuj Goswami, Naman Goyal, Anthony Hartshorn, Saghar Hosseini, Rui Hou, Hakan Inan, Marcin Kardas, Viktor Kerkez, Madian Khabsa, Isabel Kloumann, Artem Korenev, Punit Singh Koura, Marie-Anne Lachaux, Thibaut Lavril, Jenya Lee, Diana Liskovich, Yinghai Lu, Yuning Mao, Xavier Martinet, Todor Mihaylov, Pushkar Mishra, Igor Molybog, Yixin Nie, Andrew Poulton, Jeremy Reizenstein,

```
"""
The following is a set of citation purpose categories, each
category name is followed by a description of the category
and an example of a sentence that belongs to this category.

Name: Criticizing
Description: Criticism can be positive or negative. A
citing sentence is classified as "criticizing" when it
mentions the weakness/strengths of the cited approach,
negatively/positively criticizes the cited approach,
negatively/positively evaluates the cited source.
Example: Chiang (2005) introduced a constituent feature to
reward phrases that match a syntactic tree but did not
yield significant improvement.

Name: Comparison
Description: A citing sentence is classified as
"comparison" when it compares or contrasts the work in the
cited paper to the author's work. It overlaps with the
first category when the citing sentence says one approach
is not as good as the other approach. In this case we use
the first category.
Example: Our approach permits an alternative to minimum
error-rate training (MERT; Och, 2003);

Name: Use
Description: A citing sentence is classified as "use" when
the citing paper uses the method, idea or tool of the cited
paper.
Example: We perform the MERT training (Och, 2003) to tune
the optimal feature weights on the development set.

Name: Substantiating
Description: A citing sentence is classified as
"substantiating" when the results, claims of the citing
work substantiate, verify the cited paper and support each
other.
Example: It was found to produce automated scores, which
strongly correlate with human judgements about translation
fluency (Papineni et al. , 2002).

Name: Basis
Description: A citing sentence is classified as "basis"
when the author uses the cited work as starting point or
motivation and extends on the cited work.
Example: Our model is derived from the hidden-markov model
for word alignment (Vogel et al., 1996; Och and Ney, 2000).

Name: Neutral (Other)
Description: A citing sentence is classified as "neutral"
when it is a neutral description of the cited work or if it
doesn't come under any of the above categories.
Example: The solutions of these problems depend heavily on
the quality of the word alignment (Och and Ney, 2000).

Classify the following in text citation into one of these
categories. Only use a single category for each citation.
Respond with only a single word, the name of the category.

{Citation Context}
"""
```
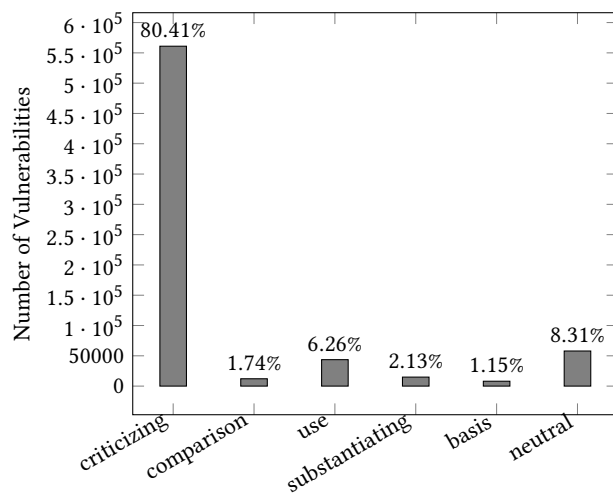
Figure 2: Prompt template used for citation purpose classification

Figure 1: Distribution of Vulnerabilities

Rashi Rungta, Kalyan Saladi, Alan Schelten, Ruan Silva, Eric Michael Smith, Ranjan Subramanian, Xiaoqing Ellen Tan, Binh Tang, Ross Taylor, Adina Williams, Jian Xiang Kuan, Puxin Xu, Zheng Yan, Iliyan Zarov, Yuchen Zhang, Angela Fan, Melanie Kambadur, Sharan Narang, Aurelien Rodriguez, Robert Stojnic, Sergey Edunov, and Thomas Scialom. 2023. Llama 2: Open Foundation and Fine-Tuned Chat Models. arXiv:2307.09288 [cs.CL]