

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

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

## ABSTRACT

While citation count and H-index is the most important metrics for measuring impact in Science, the intention of the given citation is usually overlooked. To examine the affects of citation sentiments on academic success, this paper conducts a large scale empirical study on University of Waterloo Computer Science faculty members' publication data. We show a negative correlation between H-index and the percentage of negative citations given by that researcher. We also show a positive correlation between number of papers published by a researcher and percentage of neutral citations given by that researcher. In addition we discuss the changes in the distribution of citation sentiments over time and the differences between same institution and different institution citation sentiments. We also provide a dataset of 9,438 papers' metadata and 697,609 citations with their citation sentiments for future research.

## CCS CONCEPTS

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

## KEYWORDS

Citation Sentiment, Citation Analysis, Citation Trends, LLM classification

### ACM Reference Format:

Onur Eren Arpacı. 2018. CiteQ: Analysis on Citation Sentiment: University of Waterloo Computer Science Faculty Case Study. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 5 pages. <https://doi.org/XXXXXXX.XXXXXXX>

## 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.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).  
*Conference acronym 'XX, June 03–05, 2018, Woodstock, NY*

© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM.  
ACM ISBN 978-1-4503-XXXX-X/18/06...\$15.00  
<https://doi.org/XXXXXXX.XXXXXXX>

The purpose of this paper is to study the citation sentiments and its affects on citation counts. We aim to answer the following research questions:

- RQ1 Do the citation sentiments 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 sentiments change over the duration of a researcher's career?
- RQ3 How did the distribution of citation sentiments change over time?
- RQ4 Is there any difference in the distribution of citation sentiments between same institution researchers and different institution researchers?

We collected the publication and citation data of the University of Waterloo Computer Science faculty members. We then used the large language model (LLM) Solar 10.7B [9] to classify the citation sentiments of each paper. We then analyzed this data to answer the research questions.

## 2 RELATED WORK

There is a large body of research on citation analysis. Most of these studies use the ACL Anthology [5] as their dataset. ACL Anthology is a dataset of NLP papers and their citations. Using the ACL Anthology creates two main differences from our study. First, it provides more contextual metadata about the citations than our dataset, such as the position of the citation in the paper, the section of the paper that the citation is in, etc. Second, it provides a large enough training dataset to train or finetune any kind of model, which our dataset also lacks. [2–4, 7] are some of the studies that use the ACL Anthology dataset. They are generally focusing on the task of citation sentiment classification itself and they do not analyze the output of this classification. None of the studies we have found use LLMs for classification. Most use sentence or word embeddings.

[13], [12] are some of the studies that analyze the citation sentiment classification output. [13] analyze the relationship between journal impact and the citation sentiment. They use PubMed Central dataset [1] for this study. [12] analyze the relationship between Nobel prize winners and their citation sentiments pre and post Nobel prize. They manually gather a 100 paper dataset for this study.

## 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

type	Positive		Negative		Neutral		Bad Context		Average	
	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.
Solar 10.7B	0.68	0.41	0.67	<b>0.67</b>	<b>0.47</b>	0.87	0.90	0.44	0.68	<b>0.60</b>
GPT-4.5	<b>0.87</b>	0.19	0.71	0.62	0.43	<b>0.96</b>	0.92	<b>0.52</b>	<b>0.73</b>	0.57
GPT-3.5	0.50	<b>0.95</b>	<b>0.80</b>	0.50	0.36	0.21	<b>1.00</b>	0.26	0.66	0.47
mistral 7x8B	0.70	0.38	0.71	0.56	0.41	0.84	0.60	0.26	0.60	0.51
llama2 7B	0.39	0.75	0.60	0.375	0.34	0.25	<b>1.00</b>	0.04	0.58	0.35
mistral	0.48	0.27	0.36	0.50	0.43	0.87	0.75	0.13	0.50	0.44

Table 1: Evaluation results of the LLMs

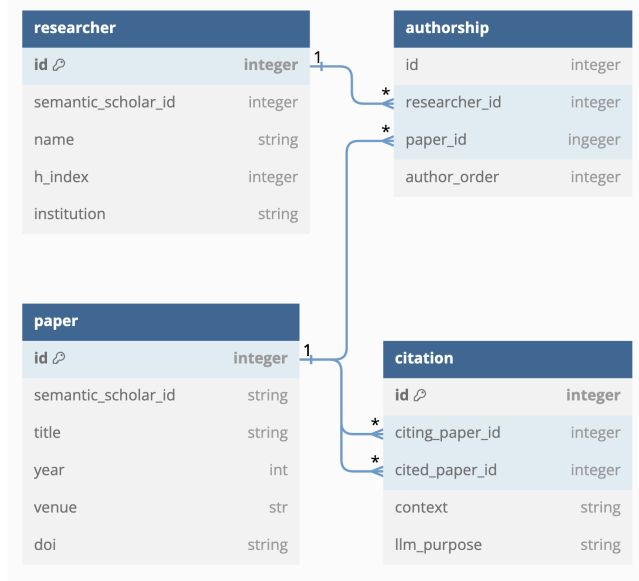


Figure 1: CiteQ database scheme

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. This database is publicly available at <https://github.com/onurerenarpaci/citeQ.git>.

### 3.2 Citation Sentiment Classification

We experimented with a number of LLMs including chatGPT-3.5 [6], chatGPT-4.5 [10], mistral and mistral 7x8B [8], llama2 [11] and Solar 10.7B [9]. Among these, only promising results were obtained from Solar 10.7B and [10]. Since with the size of the dataset we had, we could not afford to run the classification on GPT-4.5, so we chose Solar 10.7B as our LLM.

We used the following citation purpose classes: positive, neutral, negative, bad context. The positive class includes citations that are talking about the strengths of the paper, important it is. The negative class includes citations that are talking about the weaknesses of the paper, or talks about their solution being better than the paper’s solution. The neutral class includes citations that are

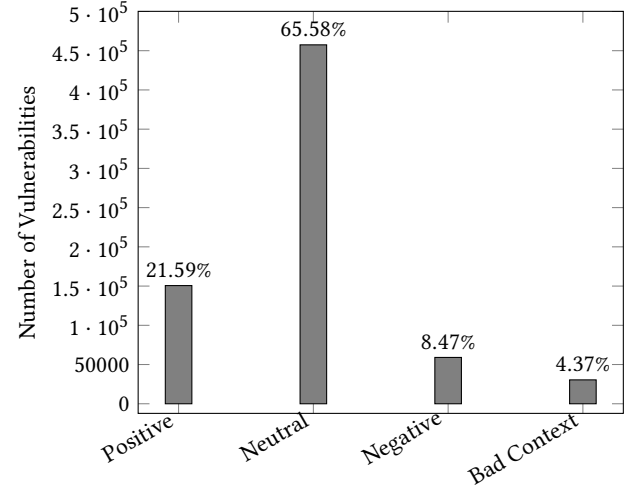


Figure 1: Distribution of citation purposes

talking about the paper without any positive or negative sentiment. The bad context class is for badly captured citation contexts that do not contain any information about the citation sentiment. With the satisfactory precision on the bad context class, we removed all the citations labeled as bad context from our dataset for all the analysis we have conducted. You can see the prompt template we used in Appendix A.

We used 6 Nvidia A100 80GB GPUs to run the classification. The total inference time was 1 day and 4 hours.

**3.2.1 Evaluation.** We hand labeled 100 citations to evaluate the accuracy of the classification. The results are shown in Table 1. We can see that the Solar 10.7B has the highest average recall, and second highest average precision. We can also see that the Solar 10.7B has the most consistent scores among classes, which is important for our analysis. Therefore, we chose Solar 10.7B as our LLM.

## 4 RESULTS

We see a distribution of 21.59% positive, 65.58% neutral, 8.47% negative, and 4.37% bad context in Figure 1. This data might not represent the actual distribution of citation sentiments in computer science because all the papers being cited belongs to University of Waterloo professors. Given that University of Waterloo has a relatively high

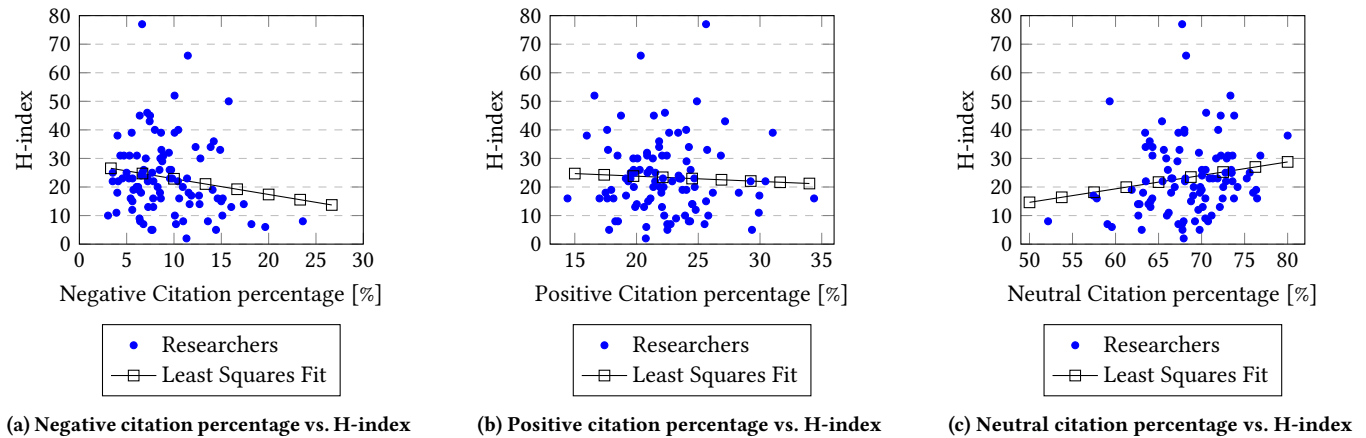


Figure 2: Citation purpose percentages vs. H-index of researchers

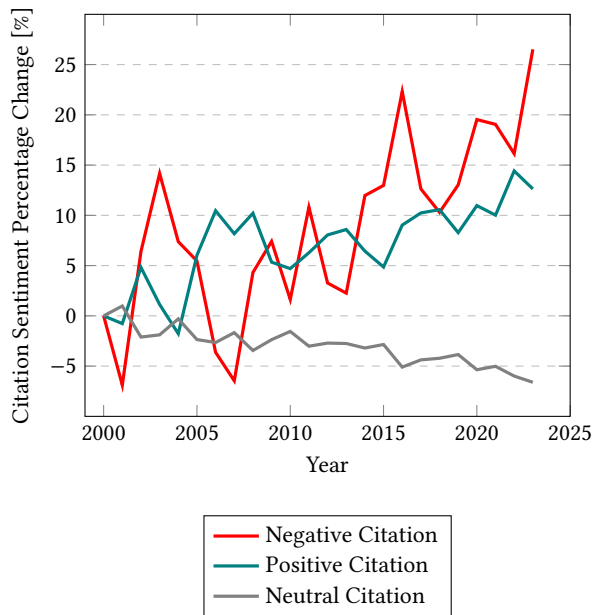


Figure 3: Citation sentiment percentage change throughout years

ranking in computer science it is possible that the distribution of citation sentiments is skewed towards positive and neutral.

#### 4.1 RQ1: Does Citation Karma Exists?

When we graph the citation sentiment percentages against the H-index of researchers, we see a negative correlation between negative citation percentage and H-index in Figure 2a. We also see a very slight negative correlation between positive citation percentage and H-index in Figure 2b. However, we see a positive correlation between neutral citation percentage and H-index in Figure 2c. While it looks easy to jump into the conclusion that there is a karma effect meaning, when you give too much negative references you will get

lower number of citations. However, this could be related to the career stage of the researcher as we will see in the next section.

#### 4.2 RQ2: How does the citation sentiment percentages change throughout the career of a researcher?

We decided to use the number of papers published by a researcher as an indication of the career stage of the researcher. When we graph the researchers citation sentiment percentages and their number of published papers we get the graphs in Figure 4, we can see that as the researchers publish more papers the negative and positive citation percentages decrease and neutral citation percentages increase. We can hypothesize that as the researchers get more experienced they tend to give more neutral references. This can also explain the negative correlation between the negative and positive citation percentage and H-index and positive correlation between neutral citation percentage and H-index. This is because H-index is also an indirect measure of the career stage of a researcher, as more experienced researchers tend to have higher H-indices.

#### 4.3 RQ3: How did the distribution of citation sentiments changed over time?

We calculated the average citation sentiment percentages for each year from 2000 to 2023, these citations include both the citations the Waterloo professors received and the citations they gave. We then graphed the percentage changes of the three classes in Figure 3. As you can see there is a consistent increase in positive and negative citation percentages and a consistent decrease in neutral citation percentage. This could potentially mean that the computer science community is getting more polarized over time.

#### 4.4 RQ4: Does being in the same institution affect the citation sentiment distribution?

We calculated the average citation sentiment percentages for citations given to University of Waterloo professors by other University of Waterloo professors and by professors from other institutions.

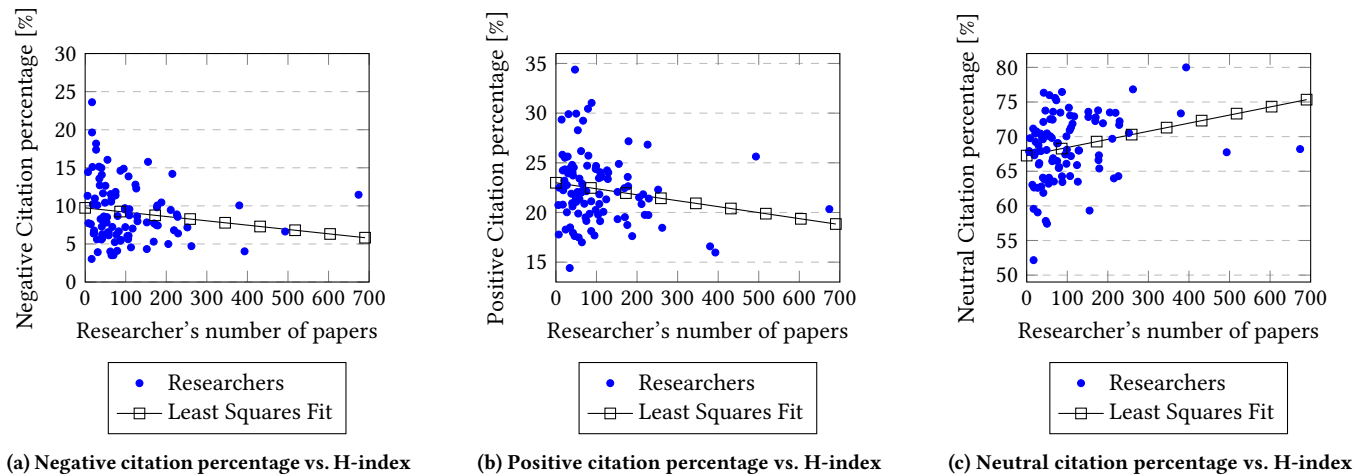


Figure 4: Citation purpose percentages vs. researchers' number of papers

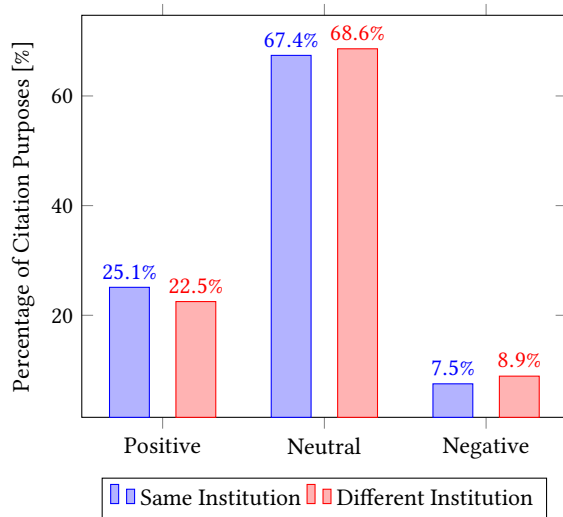


Figure 5: Citation purpose percentages within and outside of the institution

We then graphed the results in Figure 5. We can see that the distribution of citation sentiments is very similar for both cases. However, we can see a 2.6% increase in positive citation percentage and a 1.4% decrease in negative citation percentage when the citation is given by a professor from the same institution. This could potentially point to a slight positive bias towards professors from the same institution.

## 5 CONCLUSION

In conclusion, our analysis of citation sentiments among University of Waterloo professors in computer science reveals intriguing patterns. While the overall distribution suggests a predominantly positive and neutral sentiment, it's essential to acknowledge the

potential skew introduced by the university's high ranking. Examining the existence of "Citation Karma," we observed a nuanced relationship between citation sentiments and researchers' H-index, hinting at a possible correlation with career stage.

Exploring the evolution of sentiments throughout researchers' careers, we found a tendency for more experienced researchers to provide increasingly neutral references, aligning with the observed correlations. The temporal analysis indicates a growing polarization in the computer science community, reflected in rising positive and negative citation percentages alongside a decline in neutral citations over time.

Furthermore, our investigation into the influence of institutional affiliation on citation sentiment distribution suggests a subtle positive bias when citations originate from the same institution. These findings contribute to a deeper understanding of the dynamics shaping scholarly communication within a specific academic community.

In summary, our study sheds light on the complex interplay of factors influencing citation sentiments, ranging from individual career stages to institutional affiliations and temporal shifts. These insights encourage further exploration into the evolving landscape of scholarly interactions and their implications for the dynamics of academic discourse.

## REFERENCES

- [1] [n. d.]. PMC PubMed Central. <https://www.ncbi.nlm.nih.gov/pmc/>
- [2] 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.
- [3] Awais Athar. 2011. Sentiment analysis of citations using sentence structure-based features. In *Proceedings of the ACL 2011 student session*. 81–87.
- [4] Awais Athar and Simone Teufel. 2012. Context-enhanced citation sentiment detection. In *Proceedings of the 2012 conference of the North American chapter of the Association for Computational Linguistics: Human language technologies*. 597–601.
- [5] Steven Bird, Robert Dale, Bonnie J Dorr, Bryan R Gibson, Mark Thomas Joseph, Min-Yen Kan, Dongwon Lee, Brett Powley, Dragomir R Radev, Yee Fan Tan, et al. 2008. The ACL Anthology Reference Corpus: A Reference Dataset for Bibliographic Research in Computational Linguistics. In *LREC*.

- [6] 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]
- [7] Myriam Hernández-Alvarez and José M Gómez. 2015. Citation impact categorization: for scientific literature. In *2015 IEEE 18th International Conference on Computational Science and Engineering*. IEEE, 307–313.
- [8] Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Léo 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]
- [9] Dahyun Kim, Chanjun Park, Sanghoon Kim, Wonsung Lee, Wonho Song, Yunsu Kim, Hyeonwoo Kim, Yungi Kim, Hyeonju Lee, Jihoo Kim, Changbae Ahn, Seonghoon Yang, Sukyung Lee, Hyunbyung Park, Gyoungjin Gim, Mikyoung Cha, Hwalsuk Lee, and Sunghun Kim. 2023. SOLAR 10.7B: Scaling Large Language Models with Simple yet Effective Depth Up-Scaling. arXiv:Coming Soon [cs.CL]
- [10] OpenAI. 2023. GPT-4 Technical Report. arXiv:2303.08774 [cs.CL]
- [11] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajwal Bhargava, Shriti 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, 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]
- [12] Erjia Yan, Zheng Chen, and Kai Li. 2020. Authors' status and the perceived quality of their work: Measuring citation sentiment change in nobel articles. *Journal of the Association for Information Science and Technology* 71, 3 (2020), 314–324.
- [13] Erjia Yan, Zheng Chen, and Kai Li. 2020. The relationship between journal citation impact and citation sentiment: A study of 32 million citances in PubMed Central. *Quantitative Science Studies* 1, 2 (2020), 664–674.

## A CLASSIFICATION PROMPT TEMPLATE

The following is a set of citation sentiment categories, each category name is followed by a description of the category and some examples of in-text citation contexts that belongs to this category.

- **Name:** Positive

**Description:** A citing sentence is classified as "positive" when it mentions the strength of the cited approach, positively criticizes the cited approach, positively evaluates the cited source, uses the cited source as a starting point or motivation and extends on the cited work, or when the results, claims of the citing work substantiate, verify the cited paper and support each other.

**Example 1:** Researchers [13] have presented a Secure and Efficient Topology Discovery Protocol (sOFTDP) that shifts a part of the link discovery to the SDN switch.

**Example 2:** To obtain more precise and descriptive topics, we further conducted GuidedLDA [24] using some of the most salient keywords selected from our initial LDA results.

- **Name:** Negative

**Description:** A citing sentence is classified as "negative" when it mentions the weakness of the cited approach, negatively criticizes the cited approach, negatively evaluates the cited source.

**Example 1:** Most of the existing literature on the execution cost problem focus on markets where only one investor trades (for instance see [3, 4, 5, 6, 8, 9]).

**Example 2:** With nonlinear CI test, LPCMCI is computationally too expensive to be compared when the number of nodes is large.

- **Name:** Neutral

**Description:** A citing sentence is classified as "neutral" when it is a neutral description of the cited work. Use this category when there is no strong criticism negatively or positively.

**Example 1:** Since the inception of spam, many companies and research teams have combined their efforts to fight against spam deliveries using different approaches and methods [1].

**Example 2:** Conventional approaches adopt fine-tuned generative models (Zhong et al., 2020b; Guo et al., 2021; Wang et al., 2021a, inter alia ) as input generators, with a semantic parser (e.g., PCFG grammar) for sampling symbolic outputs.

- **Name:** Bad Context

**Description:** A citing sentence is classified as "bad context" when the given context is not enough to classify the citation or it does not include any citation.

**Example 1:** But then, the sum of welfare lost + retained is the optimum welfare and is bounded.

**Example 2:** Furthermore, since step (4) did not return FAIL, we must have  $g$ ; a divisor of all entries in  $V$  and  $W$  hence  $JV$  contains only polynomial entries.

Classify the following in text citation into one of these categories. First, type 'THINKING:' and write your reasoning step by step. Then type 'ANSWER:' and give your answer in a single word.  
{Citation Context}

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009