A comparative investigation on citation counts and altmetrics between papers authored by top universities and companies in the research field of artificial intelligence

Luo, Feiheng; Zheng, Han; Erdt, Mojisola; Sesagiri Raamkumar, Aravind; Theng, Yin-Leng

2018

Luo, F., Zheng, H., Erdt, M., Sesagiri Raamkumar, A., & Theng, Y.-L. (2018). A comparative investigation on citation counts and altmetrics between papers authored by top universities and companies in the research field of artificial intelligence. AROSIM 2018: Altmetrics for Research Outputs Measurement and Scholarly Information Management, 105-114.

<https://hdl.handle.net/10356/79727>

https://doi.org/10.1007/978-981-13-1053-9_9

© 2018 Springer Nature Singapore Pte Ltd. This is the author created version of a work that has been peer reviewed and accepted for publication by AROSIM 2018: Altmetrics for Research Outputs Measurement and Scholarly Information Management, Springer Nature Singapore Pte Ltd. It incorporates referee's comments but changes resulting from the publishing process, such as copyediting, structural formatting, may not be reflected in this document. The published version is available at:
[http://dx.doi.org/10.1007/978-981-13-1053-9_9].

A Comparative Investigation on Citation Counts and Altmetrics between Papers Authored by Top Universities and Companies in the Research Field of Artificial Intelligence

Feiheng Luo, Han Zheng, Mojisola Erdt, Aravind Sesagiri Raamkumar, and Yin-Leng Theng

Wee Kim Wee School of Communication and Information, Nanyang Technological University, Singapore 637718

fhluo@ntu.edu.sg, zhenghan@ntu.edu.sg, mojisola.erdt@ntu.edu.sg, aravind0002@ntu.edu.sg, tyltheng@ntu.edu.sg

Abstract. Artificial Intelligence is currently a popular research field. With the development of deep learning techniques, researchers in this area have achieved impressive results in a variety of tasks. In this initial study, we explored scientific papers in Artificial Intelligence, making comparisons between papers authored by the top universities and companies from the dual perspectives of bibliometrics and altmetrics. We selected publication venues according to the venue rankings provided by Google Scholar and Scopus, and retrieved related papers along with their citation counts from Scopus. Altmetrics such as Altmetric Attention Scores and Mendeley reader counts were collected from Altmetric.com and PlumX. Top universities and companies were identified, and the retrieved papers were classified into three groups accordingly: university-authored papers, company-authored papers, and co-authored papers. Comparative results showed that university-authored papers received slightly higher citation counts than company-authored papers, while company-authored papers gained considerably more attention online. In addition, when we focused on the most impactful papers, i.e., the papers with the highest numbers of citation counts, and the papers with the largest amount of online attention, companies seemed to make a larger contribution by publishing more impactful papers than universities.

Keywords: Citation Analysis, Altmetrics, Industrial Research, Academic Research.

1 Introduction

Artificial Intelligence is currently a widely discussed topic, not only in scientific research community but also in daily life, especially after a significant milestone was accomplished when a computer program called AlphaGo defeated a human professional player in a five-game Go match in 2016 [1]. Both academia and industry have made a lot of efforts in this research area, and achieved impressive results in a number

of tasks including computer Go program, machine translation, speech recognition, object detection, to name a few [1, 2, 3, 4, 5, 6, 7]. Typically, deep learning is one of the most popular and the most impactful techniques employed in these studies. However, deep learning requires a huge amount of training data and powerful computing resources [1, 8, 9]. Compared to universities, influential companies such as Google could have more advantages on these two requirements. Hence, in the current study, we are interested in evaluating the research outputs produced by academia and industry in the research field of Artificial Intelligence.

In this paper, we conducted a comparative investigation on citation counts and altmetrics between scientific papers published by the top academic universities and industrial companies in the research fields of Artificial Intelligence. From Scopus, we extracted papers published in the relevant top publication venues that were determined through Google Scholar and Scopus, for the years 2015 and 2016. Amongst these papers, we identified the top universities and companies. Comparisons of bibliometrics and altmetrics were made among three groups of papers: university-authored papers, company-authored papers, and co-authored papers. Comparative results showed that only a small difference in citation counts among the three groups of papers was observed, while company-authored papers received much more attention online than university-authored papers.

In the following sections, we will describe data collection and data processing procedures, followed by results of a variety of comparisons. Finally, we will conclude this paper with a discussion.

2 Data

We considered the data sources Scopus, Google Scholar, Altmetric.com and PlumX. Details of data collection and data processing are presented in this section. Data collection was completed in August 2017.

2.1 Publication Venue Selection

Artificial Intelligence research field is the main focus of this study. First we identified related publication venues. Different from other research fields, Artificial Intelligence has given conference papers an equal or even higher value than journal papers, and thus both conference papers and journal papers were considered in current study. As one of the most prevailing academic search engines, Google Scholar provides top publication venue lists for different research fields. For Artificial Intelligence, Google Scholar lists the top 20 publication venues consisting of journals and conference proceedings based on the h5-index, the h-index for articles published in the last five complete years [10]. In the top 20 publication venues, we excluded arXiv venues since papers uploaded on arXiv are mostly preprints. We also removed venues in which the paper DOI was not available since paper DOIs are required for fetching altmetric data. Finally, we obtained a list of top 10 venues from Google Scholar.

We also selected publication venues from Scopus. However, the research fields in Scopus have been classified into broad categories, and Artificial Intelligence is classified under Computer Science. Thus, we selected the top 10 journals and the top 10 conference proceedings based on the Scopus CiteScore which measures average citations received per document published in the venue [11]. We also removed venues that did not contain the paper DOI.

Since the h5-index is used in ranking the venues on Google Scholar, the publication venues which publish more papers are benefitted by it. But on the other hand, the Scopus CiteScore played a similar role as journal impact factor. As a result, by combining both Google Scholar and Scopus, a total of 30 publication venues without overlaps were identified as the venues related to Artificial Intelligence in the current study.

Subsequently, we retrieved all research papers published in these 30 venues in 2015 and 2016 from Scopus. There were 14,647 retrieved papers in total, and the exported information of each paper included title, DOI and affiliations. Table 1 gives a summary of the selected venues.

Table 1. Summary of selected venues.

S/N	Venue	Source	# of papers
1	Expert Systems with Applications	GS	1,393
2	IEEE Transactions on Neural Networks and Learning Systems	GS	542
3	IEEE Transactions on Fuzzy Systems	GS	330
4	Knowledge-Based Systems	GS	68
5	Applied Soft Computing	GS	1,198
6	Neurocomputing	GS	2,901
7	Neural Networks	GS	93
8	Engineering Applications of Artificial Intelligence	GS	385
9	Neural Computing and Applications	GS	321
10	Robotics and Autonomous Systems	GS	249
11	IEEE Transactions on Pattern Analysis and Machine Intelligence	Scopus	386
12	IEEE Transactions on Evolutionary Computation	Scopus	124
13	Wiley Interdisciplinary Reviews: Computational Molecular Science	Scopus	32
14	ACM Computing Surveys	Scopus	111
15	International Journal of Computer Vision	Scopus	150
16	IEEE Communications Magazine	Scopus	526
17	IEEE Internet of Things Journal	Scopus	162
18	IEEE Journal on Selected Areas in Communications	Scopus	477
19	IEEE Wireless Communications	Scopus	181
20	IEEE Signal Processing Magazine	Scopus	125
21	Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition	Scopus	1,245
22	Proceedings of the IEEE International Conference on Computer Vision	Scopus	600
23	International Conference on Architectural Support for Programming Languages and Operating Systems	Scopus	102
24	Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining	Scopus	455

25	Proceedings of the Annual ACM Symposium on Theory of Computing	Scopus	185
26	Proceedings of the ACM SIGMOD International Conference on Management of Data	Scopus	390
27	Proceedings - International Conference on Software Engineering	Scopus	582
28	Proceedings of the Annual International Conference on Mobile Computing and Networking, MOBICOM	Scopus	223
29	Proceedings - IEEE INFOCOM	Scopus	1,004
30	Proceedings of the ACM SIGPLAN Conference on Programming Language Design and Implementation	Scopus	107

2.2 Prestigious Affiliation Identification

This study intends to explore the comparisons between papers authored by the academic institutes and papers authored by industry in terms of bibliometrics and altmetrics. Since the top publication venues were selected in the last subsection, we considered the prestigious affiliations, due to our assumption that less prestigious affiliations might publish fewer papers in those top venues.

In the 14,647 retrieved papers, an affiliation could be recorded in different names. For example, “Google” also had another recorded name as “Google Inc.”. We used string match to consolidate the affiliation names, so that every affiliation entry has a unique name. We sorted the affiliations by the number of affiliated papers, and identified the top 10 universities and the top 10 companies. These identified affiliations published the most papers in the selected venues. In general, academic universities are more engaged in the research community, and thus we additionally included the top 10 universities on the QS World University Rankings 2015 for Computer Science [12]. Since there was an overlap between the two lists of top 10 universities, we finally obtained 19 unique universities and 10 companies, which are summarized in Table 2.

Table 2. Summary of identified universities and companies.

Affiliation	Type	# of papers
MIT	University	175
Stanford University	University	181
University of Oxford	University	47
Carnegie Mellon University	University	210
Harvard University	University	58
University of California, Berkeley	University	211
University of Cambridge	University	39
The Hong Kong University of Science and Technology	University	108
ETH Zurich - Swiss Federal Institute of Technology	University	119
Princeton University	University	75
Tsinghua University	University	450
Harbin Institute of Technology	University	229
Southeast University	University	198

Nanyang Technological University	University	203
Huazhong University of Science and Technology	University	205
Xidian University	University	194
Shanghai Jiao Tong University	University	210
Zhejiang University	University	180
National University of Singapore	University	173
Microsoft Research	Company	318
IBM Research	Company	136
Google	Company	121
Huawei Technologies	Company	88
NEC Laboratories America	Company	61
Ericsson Research	Company	59
Nokia	Company	59
Samsung Electronics	Company	57
Yahoo Labs	Company	51
Adobe Research	Company	49

2.3 Retrieval of Bibliometric and Altmetric Data

We collected citation counts of the retrieved papers using the Scopus API. Out of the total 14,647 papers, citation counts could be extracted for 14,474 papers (98.8%), while access errors were encountered when fetching citation counts of the remaining 173 papers. Please note that an available citation count simply meant that the citation count of the particular paper could be retrieved from Scopus via the API, but it did not mean the number was necessarily greater than zero.

We harvested altmetric data from two predominant sources: PlumX and Altmetric.com. Most of the papers had been indexed in PlumX, since PlumX and Scopus both belong to Elsevier. When we used Altmetric.com API, the number of papers with retrieved data was much lower. However, we found that if the preprint version of a paper was uploaded to arXiv, then the preprint version could be also indexed in Altmetric.com, which meant that a particular paper could have two different records on Altmetric.com: one for the preprint version and the other for the print version. It was also found that for some of the papers, only the preprint versions were indexed in Altmetric.com. Thus, we used full title match to obtain arXiv ids for all the papers if they were uploaded to arXiv, and fed their arXiv ids to the Altmetric.com API to retrieve altmetric data.

From PlumX, we collected four types of metrics: usage data (e.g., views count and downloads count), captures data (e.g., Mendeley readers count and bookmarks count), mentions data (e.g., blog mentions count), and social media metrics data (e.g., Facebook posts count and tweets count), all of which were the main indicators in PlumX. From Altmetric.com, we collected three metrics: Altmetric Attention Scores, Mendeley readers, and tweets. The Altmetric Attention Score is an indicator of the amount of attention the paper has received online [13]. It is a weighted count of the online attention a paper has received, such as references in Wikipedia and mentions in social networks.

In addition, we also collected Mendeley readers count and tweets count, since both metrics were available for relatively more papers. For a certain paper, if it had two records in Altmetric.com, we selected the higher value for each metric.

3 Results

Comparisons of collected citation counts and altmetrics were made between papers published by the 19 identified universities and papers published by the 10 companies. In the comparisons, if there was an author affiliated with the identified university or the identified company, then the corresponding paper was defined as a university-authored paper or a company-authored paper. Additionally, a paper was defined as a co-authored paper, if its associated affiliations involved both the identified university and company. Comparison results are presented in this section.

3.1 Comparison of Citation Counts

Figure 1 compares citation counts among the three paper groups: university-authored papers, company-authored papers, and co-authored papers. The left subfigure of Figure 1 shows citation counts in a boxplot in which the line in the box indicates the median of citation counts. According to the left subfigure, papers published by universities had slightly higher citation counts, but overall there was not much difference among the three groups. The right subfigure of Figure 1 compares means of citation counts among the three paper groups, where the company-authored paper group obtains the highest mean of citation counts. From Figure 1, the results revealed that in general, whether the paper was published by the university or by the company, it had a limited impact on the citation count. However, when considering highly cited papers, companies may have a larger contribution. By counting the number of highly cited papers which were defined as papers with at least 100 citations, it was validated that there were in total 20 highly cited papers, consisting of 10 company-authored papers, 9 university-authored papers, and 1 co-authored paper.

Fig. 1. Comparison of citation counts among three paper groups.

3.2 Comparison of Altmetrics

Comparisons of altmetric data collected from PlumX are first presented. The four types of metrics collected from PlumX were sparse. While capture data was available for most of the papers, usage data was available for around half of the papers, and for mentions data and social media metrics data, the percentages of papers with non-null records were lower than 10%. According to Figure 2, it is observed that co-authored papers achieved the best performance for both usage and captures data. In addition, there is no major difference between the university-authored paper group and the company-authored paper group in the left subfigure of usage data, but company-authored papers indeed outperform university-authored papers in the right subfigure of captures data. Regarding mentions and social media metrics data, the comparison results were insufficiently representative, since only a very small number of papers had non-null records, but we found that the papers with the highest values of mentions and social media metrics data were published by companies.

Fig. 2. Comparisons of usage and captures collected from PlumX.

Figure 3 illustrates comparisons of the three metrics collected from Altmetric.com. We can see in Figure 3 that company-authored papers consistently achieved more attention online than university-authored papers. Specifically, the differences of median values between company-authored papers and university-authored papers are relatively small, while there are considerable differences in terms of upper range of percentiles. Similar to citation counts, the distribution of altmetrics of individual papers is also skewed, meaning that only a few papers received high attention online while a large number of papers did not. We also counted the number of most popular papers whose Altmetric Attention Scores were higher than 100. Out of the 12 most discussed-online papers, eight were published by companies, three by universities, and the last one was co-authored.

In the center subfigure of Mendeley readers, co-authored papers attracted the most readers, followed by company-authored papers and university-authored papers. However, the Altmetric Attention Score does not take Mendeley readers count into account, which could be a reason why the comparative pattern in the center subfigure is different from the patterns in the other two subfigures.

Fig. 3. Comparisons of the three metrics collected from Altmetric.com.

4 Conclusion

In this study, we explored the performance difference between scientific papers published by the top academic universities and industrial companies in the research fields of Artificial Intelligence, from both the perspectives of bibliometrics and altmetrics. A total of 30 publication venues were selected from Google Scholar and Scopus, and the papers published in these venues in 2015 and 2016 were retrieved from Scopus. Based on the retrieved papers, we identified 19 top universities and 10 top companies in Artificial Intelligence. While most of the papers were published by universities, around 20% by companies, and the remaining were co-authored by universities and companies. Citation counts and various types of altmetrics of the papers were extracted from Scopus, PlumX and Altmetric.com.

Comparisons of citation counts among the three groups of papers showed that university-authored papers received slightly higher citation counts but the difference was relatively small. On the contrary, comparisons of altmetrics revealed that company-authored papers and co-authored papers obtained considerably more attention online than university-authored papers. However, the comparison results could be affected by the unbalanced numbers of the top universities and companies. In addition, some important venues in Artificial Intelligence such as the International Conference on Machine Learning (ICML) were not included in the current study due to the lack of DOIs of the papers published in these venues.

Most Internet users might not be scientific researchers, and thanks to the diverse products and services provided by companies, people would be more familiar with the large, well-known companies and would pay more attention to them online. In addition, research conducted by companies could be more product-oriented and closer to daily

life. These factors may motivate people to discuss more about papers authored by companies online. However, regarding paper citation count, it depends essentially on the academic quality, rigor and novelty of the paper. Hence, there was not much difference in citation counts between university-authored papers and company-authored papers. This is simply our intuitive explanation, while a deep understanding should be acquired through a series of well-designed surveys and interviews.

Although there is a consistent growth of studies on altmetrics, the relationship between bibliometrics and altmetrics has not been fully revealed [14]. In our study, we focus on the papers published by industries, since this typical category of papers would probably have advantages in receiving more attention online. In future work, we plan a thorough investigation of the relationship between citation counts and altmetrics for company-authored papers. A larger set of company-authored papers would be collected and investigated on the content of their altmetrics. We would also look into the qualitative differences between altmetrics for company-authored papers and university-authored papers.

Acknowledgement

This research is supported by the National Research Foundation, Prime Minister's Office, Singapore under its Science of Research, Innovation and Enterprise programme (SRIE Award No. NRF2014-NRF-SRIE001-019). We also thank Altmetric.com for providing access to Fetch API.

References

1. Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., ... & Dieleman, S.: Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587), 484-489 (2016).
2. Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y.: Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078* (2014).
3. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L. and Polosukhin, I.: Attention Is All You Need. *arXiv preprint arXiv:1706.03762* (2017).
4. He, K., Zhang, X., Ren, S., & Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778 (2016).
5. Ren, S., He, K., Girshick, R., & Sun, J.: Faster R-CNN: Towards real-time object detection with region proposal networks. In: *Advances in neural information processing systems*, pp. 91-99 (2015).
6. Graves, A., Mohamed, A. R., & Hinton, G.: Speech recognition with deep recurrent neural networks. In: *Proceedings of the IEEE International Conference on Acoustics, speech and signal processing*, pp. 6645-6649 (2013).
7. Chorowski, J. K., Bahdanau, D., Serdyuk, D., Cho, K., & Bengio, Y.: Attention-based models for speech recognition. In *Advances in Neural Information Processing Systems*, pp. 577-585 (2015).

8. LeCun, Y., Bengio, Y., & Hinton, G.: Deep learning. *Nature*, 521(7553), 436-444 (2015).
9. Schmidhuber, J.: Deep learning in neural networks: An overview. *Neural networks*, 61, 85-117 (2015).
10. Google Scholar top publications in Artificial Intelligence, https://scholar.google.com.sg/citations?view_op=top_venues&hl=en&vq=eng_artificialintelligence, last accessed 2017/10/06.
11. Scopus list of sources, <https://www.scopus.com/source/browse.url>, last accessed 2017/10/06.
12. QS World University Rankings by Subject 2015 - Computer Science & Information Systems, <https://www.topuniversities.com/university-rankings/university-subject-rankings/2015/computer-science-information-systems>, last accessed 2017/10/06.
13. Altmetric Attention Score, <https://www.altmetric.com/blog/the-altmetric-score-is-now-the-altmetric-attention-score/>, last accessed 2017/10/06.
14. Erdt, M., Nagarajan, A., Sin, S. C. J., & Theng, Y. L.: Altmetrics: an analysis of the state-of-the-art in measuring research impact on social media. *Scientometrics*, 109(2), 1117-1166 (2016).