ShortScience.org - Reproducing Intuition

Joseph Paul Cohen
Institute for Reproducible Research
and
Montreal Institute for Learning Algorithms
Université of Montréal
cohenjos@iro.umontreal.ca

Henry Z. Lo
Institute for Reproducible Research
henryzlo@cs.umb.edu

Abstract

We present ShortScience.org, a platform for post-publication discussion of research papers. On ShortScience.org, the research community can read and write summaries of papers in order to increase accessible and reproducibility. Summaries contain the perspective and insight of other readers, why they liked or disliked it, and their attempt to demystify complicated sections. ShortScience.org has over 600 paper summaries, all of which are searchable and organized by paper, conference, and year. Many regular contributors are expert machine learning researchers. We present statistics from the last year of operation, user demographics, and responses from a usage survey. Results indicate that ShortScience benefits students most, by providing short, understandable summaries reflecting expert opinions.

1 Overview

ShortScience.org is a platform for post-publication discussion of research papers. Users can write summaries for research papers on the site. Interested readers can read these summaries to get multiple perspectives on the given paper, in addition to the author's, thus gaining better understanding. Many regular contributors are expert machine learning researchers, whose descriptions make papers, and by extension the field of research, more accessible for all.

Papers can be hard to understand, for a variety of reasons:

- Different communities have different nomenclature to describe the same concepts
- There is a lot of jargon in papers, often making vanilla ideas sound novel
- Some ideas are very complex and could use multiple perspectives to get a more complete understanding
- Some parts of ideas may be obscure so that flaws in papers cannot be found
- Authors are encouraged to make the work seem as significant and important as possible for it to be accepted
- Some readers do not have access to papers directly and rely on second hand knowledge

Asking multiple domain experts to explain is an excellent way to understand a piece of research. However, not everyone has access to an expert, let alone multiple. ShortScience.org provides a platform for experts and non-experts alike to share notes on papers. These notes are available to all, providing a variety of explanations to help everyone better understand.

Figure 1: Example summary available on `ShortScience.org`. Each paper has information such as venue, abstract, and useful links followed by a summary box. The summary contains author, votes, view-source and the summary text, which can be formatted in Markdown and contain LATEX math, images, and videos.

2   Approach

The `ShortScience.org` platform provides three main features:

- Post summaries/notes on papers (public, private, or anonymous)
- Comment on summaries/notes
- Search, browse by venues, and follow users

Summaries can be written for any paper in three main databases, which includes anything with a DOI, on arχiv, or on Bibsonomy [3]. These summaries can be voted on by each user using a simple up or down metric. Each summary can be set as private which is useful for personal organization of papers.

`ShortScience.org` is run and managed by the Institute for Reproducible Research (IRR), a U.S. Non-Profit organization. The IRR also manages the project `academictorrents.com` which is a system that facilitates the movement of large datasets for research [2, 1].

3   Community Impact

Over the last year of the site's operation, `ShortScience.org` has received 34,938 unique users to the 626 public and 83 private summaries. These users visited the site 118,874 times and spent an average duration of 1.41 minutes per visit. These users come from all over the world, are mainly
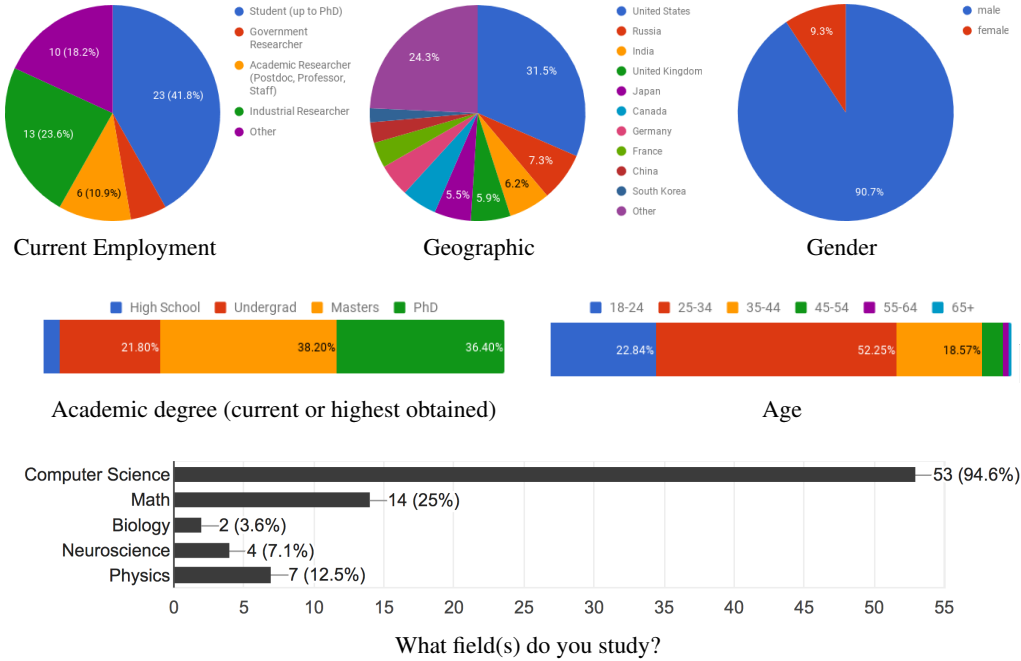
Figure 2: User demographics, collected using a survey and website statistics

focused in Computer Science, typically enrolled in Masters or PhD programs, and younger than 30. More detailed demographics are shown in Figure 2. Based on a sample of 55 users, we found:

- 60% of users read 5 or more summaries
- 87% of users found reading these summaries useful in understanding papers
- **82% of users read summaries for papers that they would not have otherwise read**

These usage statistics suggest that summaries are helpful for both readers, in terms of understanding, and for authors in terms of readers reached.

3.1 Gender

Users were only 9.3% are female. Because the primary content on the site is Machine Learning related, this may reflect a trend in Machine Learning that differs from Computer Science as a whole. The National Science Board's Science and Engineering Indicators report [5] states 25.3% (671,000/2,647,000) are employed as computer and mathematical scientists in 2016. Supporting this number, the Survey of Earned Doctorates [6] reports 24% (943/3,825) earned a PhD in mathematics and computer sciences in 2015. These numbers indicate a bias in Machine Learning.

3.2 Reproducibility

We define reproducibility as recreating the intuition the author tried to describe in their paper and as recreating the experiments in order to verify results. Recreating an experiment alone will not guarantee the intuition can be passed on to the reader, however recreating the intuition directly can enable a research to implement their own solution to verify results.

We assess intuition reproducibility explicitly with user reported success in Figure 4. In our survey we found 87% of users were able to use the platform to understand a research paper. While the majority of users did not try to directly reproduce research using the site, 10.9% (6/55 users surveyed) did and were successful while 5.5% (3/55) reported the platform not helping them and 83.6% (46/55) did not try to reproduce results.
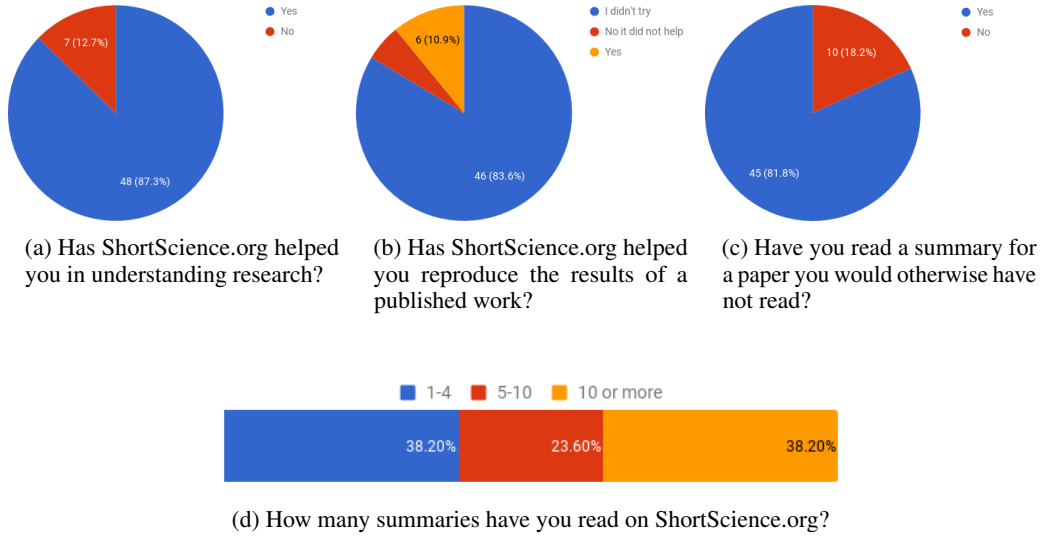
(a) Has ShortScience.org helped you in understanding research?

(b) Has ShortScience.org helped you reproduce the results of a published work?

(c) Have you read a summary for a paper you would otherwise have not read?

(d) How many summaries have you read on ShortScience.org?

Figure 3: Questions on usage

3.3 Usefulness

Responses from the survey (4a) indicate that the project is perceived to be useful. A more detailed version of this poll is shown in Figure 4b which allows us to use the Net Promoter Score (NPS) evaluation [4]. NPS asks the question "How likely are you to recommend `ShortScience.org` to a friend or colleague?" and present 11 choices between 0 and 10. From the responses, the NPS is calculated as $\frac{\# \text{ promoters} - \# \text{ detractors}}{\# \text{ total respondents}}$ where promoters are those who responded $9 - 10$ and detractors responded between $0 - 6$. The European variant accounts for respondents giving lower scores even though they are satisfied and alters these numbers to $8 - 10$ and $0 - 5$. We observe a score of 31 using the U.S. scale and 60 using the European variant. The range of possible scores are between -100 and $+100$, so the observed scores are fairly good.
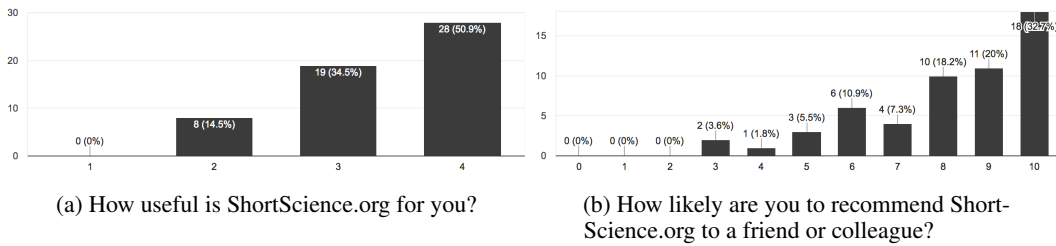

(a) How useful is ShortScience.org for you?

(b) How likely are you to recommend Short-Science.org to a friend or colleague?

Figure 4: General sentiment towards the site

4 Conclusion

Here we presented `ShortScience.org`, which aims to make research more accessible by making the ideas more understandable. After one year of operation the site has made impact, as measured by survey results. 82% of users read summaries for papers that they would not have otherwise read. The project has also helped 87% of users understand the research papers they are reading and 10.9% directly reproduce results of a paper. The project has impact on the machine learning community and is expected to have more in the future.

References

[1] J. P. Cohen and H. Z. Lo. Academic Torrents: A Community-Maintained Distributed Repository. In *Annual Conference of the Extreme Science and Engineering Discovery Environment*, 2014.

[2] J. P. Cohen and H. Z. Lo. Academic Torrents: Scalable Data Distribution. *Neural Information Processing Systems 2015 Challenges in Machine Learning (CiML) workshop*, 2016.

[3] A. Hotho, R. Jäschke, C. Schmitz, and G. Stumme. BibSonomy: A Social Bookmark and Publication Sharing System. In *Proceedings of the First Conceptual Structures Tool Interoperability Workshop at the 14th International Conference on Conceptual Structures*, 2006.

[4] T. L. Keiningham, L. Aksoy, B. Cooil, T. W. Andreassen, and L. Williams. A holistic examination of Net Promoter. *Journal of Database Marketing & Customer Strategy Management*, 2008.

[5] National Science Board. Science and Engineering Indicators, 2016.

[6] National Science Foundation. Doctorate Recipients from U.S. Universities, 2015.