2018 ASEE Annual Conference & Exposition
June 24-27, 2018
Salt Palace Convention Center
Salt Lake City, UT
Paper ID #22046

Developing Reviewer Profiles Using Analysis of Prior Authorship

Dr. Matthew A. Verleger, Embry-Riddle Aeronautical University

Matthew Verleger is an Associate Professor of Engineering Fundamentals at Embry-Riddle Aeronautical University in Daytona Beach, Florida. His research interests are focused on using action research methodologies to develop immediate, measurable improvements in classroom instruction and the use of Model-Eliciting Activities (MEAs) in teaching students about engineering problem solving. Dr. Verleger is an active member of ASEE. He also serves as the developer and site manager for the Model-Eliciting Activities Learning System (MEALearning.com), a site designed for implementing, managing, and researching MEAs in large classes.

Developing Reviewer Profiles Using Analysis of Prior Authorship

Abstract

Background

Peer review is a cornerstone of academic research dissemination. It is a fundamental prerequisite for "good" research, even though the process of selecting reviewers is largely shrouded in mystery. Under ideal conditions, reviewers are experts on a paper's topic, but the process of identifying that expertise varies widely from publication venue to venue.

Purpose (Hypothesis)

The purpose of this paper is to (1) identify some of the methods currently used by conference and journal editors, and (2) present an alternative methodology used by the author to identify and assign reviewers for a division of ASEE during the summer of 2017.

Design/Method

The author, as program chair of a large division, developed software that data-mined the prior conference papers of any potential reviewer to identify the words frequently used in their prior work. Using this data, a Reviewer Profile was developed that identified areas of expertise. Using a similar process, papers being reviewed were profiled to the types of expertise necessary – a Paper Need Profile. These profiles were then used to assign peer reviewers for the ASEE 2017 conference.

Results

This paper will present details of the Reviewer Profile algorithm, the Paper Need Profile algorithm, the matching algorithm, and an analysis of the effectiveness of the overall assignment methodology.

Conclusion

The conclusion will explore how this work may translate into classroom peer review as well as a discussion of how this methodology could be applied across ASEE and in other professional publication venues.

Introduction

Peer review is a cornerstone of the modern scientific process. It is meant to act as a gateway, allowing good research through, while filtering out junk science; to separate the wheat from the proverbial chaff. Yet many scientists, academics, and even the US Supreme Court agree that peer review, while essential to the scientific process, is far from a perfect system. In the 1993 Supreme Court case "Daubert v. Merrell Dow Pharmaceuticals", Justice Blackmun wrote that it was the opinion of the court that "Publication (which is but one element of peer review) is not a sine qua non of admissibility; it does not necessarily correlate with reliability… But submission to the scrutiny of the scientific community is a component of "good science," in part because it increases the likelihood that substantive flaws in methodology will be detected.” [1]. Effectively, the court recognized that, while peer review is good for science as a whole, it does not necessarily work correctly all the time. The problem with peer review is that it is a theoretically sound process that can easily fail apart on implementation. It is a methodology whose success is heavily dependent on having the most appropriate reviewer for the situation providing the right review.

Consider the case of Dr. Felisa Wolfe-Simon, a NASA Astrobiology Research Fellow. When her research found a bacterium capable of growing using arsenic instead of phosphorous, a process previously thought to be impossible, the results were understandably considered ground-breaking. NASA held a well-publicized press conference to announce the findings and claimed that this would drastically change the search for extra-terrestrial life. Dr. Wolfe-Simon and her team of researchers were published in the prestigious journal *Science* [2], an act which set off a firestorm of controversy. The day after the article was published online, Rosie Redfield, a microbiology researcher at the University of British Columbia, wrote a blog post [3] skewering Dr. Wolfe-Simon’s research, stating “If this data was presented by a PhD student at their committee meeting, I'd send them back to the bench to do more cleanup and controls.” Dr. Redfield, while later claiming that she doesn’t blame the reviewers, was effectively saying that peer review failed to keep Dr. Wolfe-Simon’s research from being published. In a December 16th tweet, Dr. Wolfe-Simon said, “We went through a solid peer-review and made responses and revision in response to them.” [4] Fundamentally, peer review did exactly what it was supposed to do – it provided the authors with additional “expert” opinions regarding the perceived quality of the research that were then incorporated into the article to improve the overall quality of the research reporting. Where peer review failed is that, based on Dr. Redfield’s and many other scientist’s objections [5], it selected the *wrong* peer reviewers for *this* research. Dr. Wolfe-Simon, through no fault of her own, was assigned reviewers who failed to adequately help her and her team address the perceived holes in their research.

Dr. Wolfe-Simon’s work is but one example of a case where peer review failed to identify flawed research that was conducted in good faith using apparently appropriate methods. The selection of qualified peer reviewers is essential to supporting scientific research validity. This paper explores this issue and demonstrates the approach used for the Educational Research and Methods division of the ASEE 2017 Annual Conference.

Background

Existing Peer Review Assignment Methods

In academic publishing, guidance on selecting peer reviewers is scarce. The World Health Organization published guidelines for medical research [6] that rely heavily on editors to identify and select reviewers:

> The editor should establish a reviewer database that includes information about the expertise of each reviewer as well as addresses and other contact information. The editor may identify potential reviewers from:
>
> - personal knowledge of the topic
> - authors referenced in the manuscript
> - membership of the society that publishes the journal
> - computer searches of databases such as PubMed

Journal management platform ScholarOne includes their "Reviewer Locator" [7] to offer up names, but the underlying process for how those names are identified is unknown. This system is used by ASEE's Journal of Engineering Education (JEE). Lisa Benson [8], current editor of JEE, offers the following guidance for Associate Editors (AEs):

> In general, reviewers should be chosen to reflect the broad, growing, global field of engineering education, including content domain experts and experts in education and the learning sciences. Also, well-qualified graduate students may be invited to review manuscripts. Consider choosing reviewers for their expertise in the methodology used by the authors. Please avoid choosing other AEs and people who work at the same institution as one of the authors.

The focus of this work is to describe an automated methodology used to assign reviewers for conference papers for the ASEE National Conference.

Methodology/Results

Context

The author of this paper was the Program Chair for the Educational Research and Methods Division (ERM) of the American Society for Engineering Education (ASEE) for the 2017 Annual Conference held in Columbus, OH. The Program Chair is responsible for managing the paper process for the division, including selecting reviewers, making accept/revise/reject decisions, and organizing the overall program. The ERM Division received 157 draft papers needing at least three (3) reviews each. 447 individuals agreed to review at least one (1) paper, with many agreeing to review 2 or more.

Reviewer Expertise Profiles

The approach used by the program chair was to develop reviewer expertise profiles based on text analysis of prior authorship. The assumption was that, if a reviewer has previously (co-)authored a paper about a topic, they are more likely to have expertise on that topic. To obtain prior authorship, the peer.asee.org domain was recursively downloaded using the GNU wget command (https://www.gnu.org/software/wget/). In addition to PDFs of each paper, the site offers index pages by author, allowing for more accurate identification of prior authorship. For each of the 447 individuals that agreed to be a reviewer, their prior work was moved to a separate directory and any work from individuals not agreeing to review was deleted. This resulted in 3681 papers from 380 authors. 67 individuals agreed to be reviewers but had not submitted a paper to ASEE recently enough for it to be included in the peer.asee.org database. Note that authorship order was not considered in this analysis, meaning that even papers in which a potential reviewer is the last author contributed toward their reviewer expertise profile.

Using the pdftotext tool in the Poppler suite - http://poppler.freedesktop.org – each of the paper PDFs was converted to plain text. A database was then populated with the following:

1. Author Name
2. Paper Title
3. Paper Text

For each word-author pairing, a Term-Frequency-Inverse-Document-Frequency (TFIDF – described in the next section) was calculated for all words used by that author. By ranking each author's TFIDF scores, the areas of expertise for that author can be revealed.

Calculating TFIDF

For every unique word that a particular author used in one of their papers, a Term-Frequency-Inverse-Document-Frequency (TFIDF) calculation was made. While there are numerous variants of the TFIDF algorithm, they all fundamentally calculate a weighting of how frequently a word appears in one work versus how frequently it appears across a larger corpus of work. For this implementation, the term frequency (TF) is calculated as the raw number of times a paper contains a word. Other variants calculate this as a percentage to account for length variability, but as conference papers are largely similar in length, this was deemed unnecessary. The inverse document frequency (IDF) is calculated as $\log_{10}$ of the inverse percentage of authors who have ever used that word. The logarithmic scaling was done to amplify the unique words in a paper, as words appearing in 15 or fewer papers had a multiplicative effect on the TFIDF.

$$TFIDF = \text{Frequency of Uses of Word by an Author} * \log_{10} \frac{\text{Total Number of Authors}}{\text{Authors who have used Word}}$$

For every word-author pairing, the algorithm calculated how many times that author used the word in any of their papers. It then calculated the number of other authors who also used that word. For example, the word "motivation" appeared 7,248 times in 1,563 papers produced by 317 unique authors. For a particular author, Alfred Kissinger[1] – a motivation researcher, the word "motivation" was used 203 times. Their TFIDF for "motivation" was:

[1] Name changed to protect anonymity

$$TFIDF_{AK,motivation} = 203 * \log_{10} \frac{380}{317} = 15.981$$

In analyzing the TFIDF equation, the first term accounts for how important a word is to a particular author, while the second term identifies how important a word is to all other authors. The more common a word is, closer the fraction is to 1 and the smaller logarithmic value is, giving words such as “the” and “participant” a very small TFIDF for this corpus of text. For a word to have a high TFIDF, an author must use it often and other authors must use it rarely.

Reviewee Needs Analysis

To assess each paper needing review, code was written to download the PDFs of draft papers from the ASEE system. Each paper was converted to plain text using a similar approach to the reviewer-profile processing. For each paper needing peer review, a TFIDF was calculated for the words in the collection of draft papers. By ranking these TFIDF values, each reviewee paper’s most important keywords were revealed.

For example, a section of the abstract for the ERM Division’s 2017 Best Paper, “Measuring Students’ Subjective Task Values Related to the Post-Undergraduate Career Search” [9] reads: “The PEPS study is grounded in Expectancy-Value Theory (EVT), which conceptualizes engagement in a task as a function of four subjective task values: attainment value, intrinsic value, utility value, and cost. The focus of this research paper is on the development and validation of survey measures to capture students’ subjective task values (STV) related to their post-undergraduate career search.” The top 10 keywords from that paper, based on their TFIDF, are shown in Table 1.

Table 1. Top 10 Keywords in 2017 ERM Best Paper

Word	Term Frequency in Paper	Document Frequency (n=157)	TFIDF
STV*	44	2	83.374
Items	76	57	33.442
Factor	69	64	26.891
Position	37	32	25.558
EFA*	17	8	21.978
CFA*	14	5	20.957
Cost	28	33	18.967
Three-Factor	9	2	17.054
EVT*	10	6	14.177
Item	26	45	14.110

Note: For reasons discussed in the conclusion, acronyms are more significant discriminators. The 4 acronyms listed stand for: Subjective Task Value, Exploratory Factor Analysis, Confirmatory Factor Analysis, and Expectancy-Value Theory.

Matching Reviewers to Reviewees

To match reviewers to reviewee papers, each paper was analyzed to obtain a ranked list of potential reviewers. For each paper, the top 10 words were selected as representative of that paper's critical concepts. For each of those 10 words, the 10 potential reviewers with the highest TFIDF score for that word were identified. Using the paper shown in Table 1, an example analysis is shown in Table 2. Note that, for the word "STV", only 2 other potential reviewers used the term in their own work. In theory, this approach would yield 100 possible reviewers for each paper, but with overlaps where a particular reviewer is in the top 10 for multiple words, such as Amy and Leonard below, the average paper identified 57 potential reviewers.

Table 2. Example Reviewer Identification

Word	**STV**	**Items**	**Factor**	**Position**
Paper TFIDF	**83.374**	**33.442**	**26.891**	**25.558**
Reviewer's TFIDF by Word	Sheldon – 32.30	Leonard – 86.11	Bernadette – 48.88	Raj – 39.54
	Stuart – 32.30	Amy – 73.19	Amy – 55.18	Howard – 37.60
		Penny – 71.47	Leonard – 43.73	Wil – 33.34

To rank each reviewer, a summation of the product of their TFIDF for each of the 10 words and the paper's TFIDF for those words was calculated and the final reviewer rating for each paper was recorded. This is shown in Table 3. This approach means that potential reviewers who have expertise for multiple words are rated higher and that reviewers who have expertise on the most important words will be weighted more highly.

Table 3. Example Reviewer Rating

	STV	**Items**	**Factor**	**Position**	**Reviewer Rating**
	83.374	**33.442**	**26.891**	**25.558**	**(Sum of Products)**
Leonard		86.11	43.73		4055.63
Amy		73.19	55.18		3931.47
Sheldon	32.30				2692.98
Stuart	32.30				2692.98
Penny		71.47			2390.10
Bernadette			48.88		1314.43
Raj				39.54	1010.56
Howard				37.60	960.98
Wil				33.34	852.10

Reviewer ratings were calculated for every paper, resulting in 8908 ratings. The next step is to assign reviews for all of the papers. This was done by first setting an assignment limit for each reviewer to prevent prolific reviewers from being assigned to an excessive number of papers.

Because each paper needed 3 reviews, a total of 471 reviewers needed to be assigned. Accounting for the limit values reviewers provided to ASEE, it was decided that no reviewer would need to review more than 2 papers, and most would only need to review 1.

The list of reviewer ratings was sorted and each reviewer was evaluated according to the following conditions:

1. Is the reviewer an author on the paper?
2. Does the paper associated with this rating already have 3 reviewers?
3. Has this reviewer already been assigned their maximum number of reviews?

If the answer to all three questions was *no*, the reviewer was assigned that paper. If any of the answers was *yes*, this review assignment was skipped and the next reviewer-paper rating was examined. This process was repeated until all papers had 3 reviewers.

The results were printed to a file and transcribed into the ASEE Monolith System by hand.

Results & Analysis

Analysis of the success of this approach is difficult to measure directly, but there are a few indirect measurements that indicate that the algorithm is capable of identifying reviewers with expertise.

The simplest measure, though least rigorous, is that 7 reviewers indicated that they had concerns about a potential conflict of interest with reviewing their assigned paper. While direct authorship of the paper was considered, the logistics of also including co-author status on other work was not considered.

ERM is not a "review to publish" division. Of the 157 papers under review, 77 papers had at least one author also agree to be a reviewer, representing a total of 120 author-reviewers (most papers had 2+ authors agree to review). Of those 120 author-reviewers, 74 (62%) of them were rated highly enough to calculate a reviewer rating scores. 37 of those 74 (50%) were in the top 5 potential reviewers for their own paper. This is not unexpected, as any author is likely to write about their area more than once. What this demonstrates is that the algorithmic approach is accurately identifying keywords and the authors that use them, so much so that it is identifying the current author based solely on the text of their prior work.

Limitations

The approach outlined above is one method for potentially identifying reviewer-reviewee pairings that results in reviewers having expertise on the work being reviewed. The approach was not without its limitations.

First is the issue that co-author relationships were not considered as conflicts of interest. The process of identifying, for each paper, if the individual being considered for assignment has ever published with one of that paper's author, is a difficult problem, given the lack of API for accessing ASEE's systems. This is compounded by the general challenges associated with tracking author/co-author relationships, such as when someone changes their last name or optionally uses a middle initial.

A more significant methodological flaw is that this approach depends on reviewers being prior authors. It does not support new and emerging authors, as there is no data to identify their areas of expertise. At the same time, it tends to overly support highly prolific authors. Because the system did not consider authorship order in weighting a potential reviewers word history, authors that are included on large numbers of papers are seen as having excessively broad expertise in areas they may only be tangentially knowledgeable about. One potential reviewer from a PhD granting institution has had a large number of papers submitted to ASEE by their various PhD students, with this reviewer typically as a late co-author in the authorship list. Because of their prominence in the ASEE database, they had a reviewer rating calculated for 139 of the 157 papers being reviewed. They were a top 10 reviewer for 73 of those 139 (52%) papers. By weighting words based on authorship order, or only including first authored papers, the algorithm may better detect experts from simply prolific authors.

Another methodological limitation is that the code only considered single-words, not phrases. Because of this, acronyms tended to be highly weighted words, but their meaning was not considered. An additional iteration of refinement would be to identify common groupings of words and consider them to be a single entity.

Next Steps

As a first attempt, the approach outlined above seemed moderately successful. Beyond addressing the identified methodological limitations, the next step is to apply this approach to other conferences or publications, building a more extensive reviewer expertise model. Additionally, refining the selection algorithm to weight different areas of expertise may be helpful. For example, if these techniques could be used to identify general frameworks, quantitative vs. qualitative methodologies, statistical techniques, etc., it may help to identify expertise with specific components of a particular paper.

Finally, an implementation problem for further study is that the approach is not yet operationalized in a way that other program chairs could easily use this technique for their own programs. Building a system to integrate with ASEE's Monolith architecture would require either internal support from ASEE or an outward facing API and building a complete stand-alone tool is viable but beyond the scope of the current research.

Acknowledgements
This material is based upon work supported by the National Science Foundation under Grant No. 1264005. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] H. Blackmun, *Blackmun, H., Daubert v. Merrell Dow Pharmaceuticals*, vol. 509 US. Supreme Court of the United States, 1993, p. 579.
[2] F. Wolfe-Simon *et al.*, “A Bacterium That Can Grow by Using Arsenic Instead of Phosphorus,” *Science (80-.).*, vol. 332, pp. 1163–1166, 2011.
[3] R. Redfield, “Arsenic-associated bacteria (NASA’s claims),” 2010. [Online]. Available: http://rrresearch.blogspot.com/2010/12/arsenic-associated-bacteria-nasas.html.
[4] F. Wolfe-Simon, “Twitter - We went through a solid peer-review and made responses and revision in response to them.,” 2010. [Online]. Available: https://twitter.com/#!/ironlisa/status/15496350829383682.
[5] Nature Editorial Staff, “Response Required,” *Nature*, 2010. [Online]. Available: http://www.nature.com/nature/journal/v468/n7326/full/468867a.html.
[6] World Health Organization, “FAME Editoral Guidelines,” Geneva, Switzerland, 2004.
[7] Clarivate Analytics, “Reviewer Locator - Clarivate,” 2018. [Online]. Available: https://clarivate.com/products/scholarone/reviewer-locator/. [Accessed: 20-Jan-2018].
[8] L. Benson, “Guidance for Associate Editors.” 2018.
[9] S. Brunhaver *et al.*, “Measuring Students’ Subjective Task Values Related to the Post-Undergraduate Career Search,” in *2017 ASEE Annual Conference & Exposition*.