Study on the Difference between Summary Peer Reviews and Abstracts of Scientific Papers

Chong Chen[†]
School of Government
Beijing Normal University
Beijing China
chenchong@bnu.edu.cn

Jingying Zhang, Xiaoyu Chu
School of Government
Beijing Normal University
Beijing China
zh-jy, chuxiaoyu@mail.bnu.edu.cn

Jinglin Zheng
School of Government
Beijing Normal University
Beijing China
738987984@qq.com

ABSTRACT

Readers feel impeded when they are unfamiliar with terms or knowledge entities in scientific papers. If the value of certain practice, method, or theory mentioned in a paper could be revealed by experts, the difficulties on paper reading will be significantly reduced. As scientific papers are generally composed of components with different argument functions. The experts may also comment on these function components when they review the papers. In this study, the opinion of authors and reviewers on same papers are put together for comparison. The comparison is conducted from two aspects. one is the difference of function types they have, and the other is the readability of the opinion given by two groups of people, i.e., the abstracts given by authors and the summary peer reviews given by reviewers. The results show the distinction of summary peer reviews in highlighting the function components of scientific papers. The contribution of this study lies in put forward measurement on the difference comparison between summary peer reviews and abstracts. The conclusion can be used to demonstrate the value of summary peer reviews in helping readers understanding papers.

CCS CONCEPTS

• Insert CCS text here • Insert CCS text here • Insert CCS text here

KEYWORDS

Summary peer review, paper comment, function type, term density, knowledge entity

1 Introduction

Scientific papers are the records of scientific research activities. Authors claim their new findings with various knowledge entities in their papers. Readers often spend lots of time to search among numerous papers before they find valuable ones to get reference or new insights for their study. During this process, a main impediment to their paper reading lies in the readers' unfamiliar with the knowledge entities. If the value delivering by the entities to a paper could be revealed by experts, the difficulties of paper reading will be significantly reduced. In a long scientific paper, critical information to readers often lies in different function components of discourse, such as research purpose, problems, methods, experiments, contributions and so on. Quality comments on these components could help readers understanding why and how about certain study. It definitely cannot be replaced by the widely-used paper selection indices, such as citation number, recommendation rating or review scores. In short, the readers need to know how novel a research problem is, what contribution a study do to certain application, why a model do not work in specific condition and so on.

Compared with the massive customer reviews on products in online shopping, comments on papers are lacking. Yet there are indeed three sources of comment on papers, i.e., the social media reviews, the citation context and the peer reviews. The social media reviews can be disregarded since the amount is few and the quality is not assured. The citation context represents opinions of other researchers who have cited a paper with their own intention. And the peer reviews are given by expert reviewers who have to state the highlights and inadequate points of the paper to editors. Since the peer reviewers are selected authorities with broad scholarly vision and strict taste, the peer reviews are undoubtedly the ideal source to reveal the value of a study.

The peer reviews are usually composed by two parts — the summary comments and the item scoring. The reviewers would sum up a study from several aspects before they criticize the disadvantages item by item. The summary comments include the significance of a study, the novelty of the proposed model, the solidness of the methods or experiments, etc. In this paper, we consider the summary peer reviews, and focus on two problems about it, i.e., (1) the difference between summary peer reviews and abstracts on the function components, including the readability; and (2) from which aspects the summary peer reviews highlight the value of a paper when compared with its abstract.

The contribution of this paper lies in measuring the difference between summary peer reviews and abstracts from their focused function components. The results show the distinction of summary peer reviews in highlighting scientific papers. That is to say, the important part of a paper, including knowledge entities such as models, methods and theories, may be commented in a credible and easy-to-understand expression. It is believed to be able to speed up the readers understanding on scientific papers.

2 Related Work

In order to demonstrate a study, authors usually organize the content of their scientific papers in compliance with well-established norms. For example, the rhetorical structure of abstracts has long been taken to be the Introduction-Method-Results-Discussion (IMRD) [1] in many research articles. There are also some variations to the IMRD, such as Problem-Method-Results-Conclusion, Goals-Method-Results, Introduction-Method-Results-Conclusions [1-4]. Important information of a study is contained in sentences of such rhetorical function. Thus both reviewers and readers prefer to pick out pros and cons from these sentences to evaluate a study. Following this notion, researchers try to extract concerned information by dividing the scientific papers to functional structure [5] or according to rhetoric parts [6].

With the trend of structuralizing knowledge in scientific papers [15-17], researchers also put forward ontologies for science document according to discourse element, argument model and rhetorical structure. Wang et.al analyzed the existing ontologies [7] [18] [19] and proposed Functional Units Ontology (FUO) [8] based on functional unit theory [9-11]. Functional units are considered the smallest content units for academic communication and applications. They distribute among the functional sections, such as in IMRD of papers. These units are considered to ultimately meet the readers’ goal of information use [9-11]. The FUO includes 12 classes and 28 subclasses [8]. The classes are composed of *background, theme*, *method, experiment* etc. Taking class *theme* as an example, the subclasses of it include research *scope, goal* and *definition* on the theme. Wang et. al. has mapped the classes or subclasses of FUO with the IMRD structure, and stated the relation among FUO, IMRD and the goal of information use [8].

One inspiration we can get from the aforementioned work is that content of different semantic function help to reveal the critical information of a paper. They are called functional components in this paper. In fact, they can be obtained not only inside scientific papers but also from other sources, such as summary peer reviews.

Peer review is meant to ensure the quality of publications. It is important to scientific research. Peer experts strictly evaluate papers from different perspectives related with studies and writings. The reviews are written to editors and authors including affirmation of positive and negative on some important perspectives. The reviewers judge the innovation and contribution of a paper according to the research purpose, the research problem, the proposed methods, etc. They may also comment on the experiments and the results. In a word, there are rich information in the content of peer reviews.

Much research attention has been paid to peer review in recent years. They have been used to judge whether a paper will be accepted and why. Kang Dongyeop et al. created the first open dataset of review comments for academic research, PeerRead [12]. Based on this dataset, they predicted the acceptance and rejection of a paper as well as the scores of it. Philippe Vincent-lamarre and Vincent Lariviere compared the characteristic of accepted and rejected papers from the vocabulary usage and psychology linguistics based on the PeerRead dataset and full paper texts. One of the conclusion they have drawn is the rejected papers often have a higher percentage of terminology and lower readability [13]. Ke Wang et al. predicted papers’ acceptance according to the sentimental of review texts. The results showed that there is a good agreement between the sentimental polarity of the reviews and the acceptance results [14].

Till the January of 2020, the accessible open reviews have been collected in Table 1.

Table 1. Typical Open Review Sources

Venues	Name	Peer review accessibility	Start time
Journals	Acta Psychologica Sinica	webpage	2014
	Archives of Public Health	webpage	2013
Conferences	International Conference on Learning Representations(ICLR)	API	2013
	The Conference on Neural Information Processing Systems(NIPS) [12]	webpage	2013
Academic Websites	Open Review[1]	API	2013
	Sciencepaper Online[2]	webpage	2003
	Publons[3]	webpage	2012

3 Research Framework

In this study, an observation is that some function types are in common in both summary peer reviews and abstracts. But their contents are given by people with different motivation, and thus their focuses and views are rather different. The comments by reviewers would be more critical and with broad vision，which will probably benefit to readers and help them reduce the time on paper understanding.

Two tasks are designed to conduct the comparison between the reviews and abstracts. The first one is to induce the function types appeared in reviews and abstracts, and annotate the sentences to these types. The second is to analyze how the summary peer reviews highlight a paper’s value from the function they focused, and the readability of reviews and abstracts.

The first task led to a rough classification. Namely, sentences in reviews and abstracts were identified by frequent patterns, then they were manually checked to make sure that their types were correct. Some example patterns have been shown in Table 2.

In the second task, the functional sentences in reviews and abstracts were compared by terms density and the type proportion. Term density reflects the readability of a text. General readers may feel tough to understand if a text has lots of terminology. They have to spend lots of time to determine if a paper is what they want when its abstract is of term-dense. Type proportion shows the focused function of reviews and abstracts by the proportion of sentences of each function type.

[1] https://openreview.net/
[2] http://www.paper.edu.cn/
[3] https://publons.com/

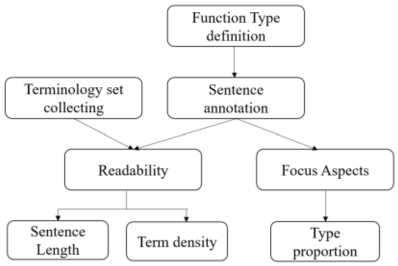

Fig.1 Research Framework.

3.1 Functional Component Types

Based on the FUO [8], six function types in scientific papers are defined in this study, i.e. *background, theme, process, result, contribution* and *strength*. The meaning of each type is listed in Table 2. Some types are borrowed from FUO classes and their subclasses, such as *background* and *theme*. Some are combined from the FUO classes and subclasses with a little change, such as *process, result* and *contribution*. Some FUO classes are dropped out according to the feature of dataset, for example the class *data*, since details of data description and data analysis are unlikely to appear in abstract and summary reviews. An example of new-added is *strength* which is not a class or subclasses in FUO. It refers to something needs to be specially highlighted about a study or a paper, such as the novelty on idea, the clarity on research design, the solidness on experiment, the ingenuity on proposed model, etc. Generally, the strength type is more common in summary peer reviews than in abstracts.

Table 2. The Definition of Function Types

Function Type	Meaning	Pattern examples
Background	Background of theories and applications; existing studies; unsettled gaps; necessity and significance of the current study;	in order to (solve the problem) …; …remain unsolved; …not (completely) studied yet;
Theme	Research scope; research goal; definition to the concerned problem;	(this study/article/paper) propose/ investigate/discuss/ demonstrate …;
Process	Hypothesis; methods; experiments; theories and research perspectives;	Base on…proposed; First…Secondly…Last; (model/approach/method) be used/implemented…;
Result	Description and evaluation on the result, Hypothesis and methods;	(experiment/result/ simulation/) show/demonstrate…; …(provide/give) a reference to …;
Contribution	Contribution to the related theories or methods; Comparison with previous studies; insight obtained; future work;	The contribution (of this study/paper)…; (This study) improve…; The improvement (of this study/research) is…;
Strength	Claim the strength or highlights of the whole study and the current paper with summary description.	(idea/ concerned problem) new/novel/critical...; (experiment/data processing/research design) is firm/well; … is significant to sth;

3.2 Comparison Approach

Besides abstracts, reviews are another source to emphasis the critical information of a study. Yet, the functional components they contained are believed to be different from those in abstracts. The difference provides new observation on a study, which may help readers discover something has not clearly state by authors in their paper. The question is how to compare them. In this study, term density and type proportion are defined as measurement.

Term density measures the density of two types of terms, i.e., the general term and the terminology of specific domain. The general term refers to concept that does not necessarily belong to specific domain. In this study, the former is calculated by average term number per sentence after removing stop words, and the latter is the average terminology number per sentence. Since the corpus is Chinese, the sentences have been segmented by segmentation tools. The word bank is composed by default word bank and the terminology bank.

The type proportion reflects which types of functional components reviewers prefer to comment. It is calculated by average percentage of sentences in each function type.

4 Experiment

4.1 Dataset

1. Reviews and Abstracts. The dataset was collected from the website of journal *Acta Psychologica Sinica*, including a total 774 papers published from 2014 to 2019. The sentence number of summary peer reviews and abstracts is respectively 2777 and 4397. The sentences have been annotated according to the function definition as shown in section 3.1. Let symbol T, R, A, n, *avgLen* and T% respectively denote the function types, the summary peer review data, the abstract data, the number of sentences in R and A, the average sentence length per type, and the type proportion. The details are shown in Table 3.

Table 3. Sentences of Different Types in Summary Peer Reviews and Abstracts

Function types T	# sentences n		*avgLen*		T%	
	R	***A***	***R***	***A***	***R***	***A***
Background	124	532	39	44	4.5%	12.1%
Theme	638	504	30	37	23.0%	11.5%
Process	519	1011	30	49	18.7%	23.0%
Result	354	1724	47	68	12.7%	39.2%
Contribution	347	575	35	61	12.5%	13.1%
Strength	795	44	23	34	28.6%	1.0%

2. Terminology. The terminology of psychology has been collected from three sources, i.e. the keywords of the 774 papers, the *Academic Hotspots of Psychology* in CNKI[4], and the *Chinese Terms in Psychology* which is published by China National Committee for Terms in Sciences and Technologies[5]. The terms from above sources were merged to a Chinese psychological academic terminology set with 8,354 terms in total. The terminology set is used for identifying the term density of sentences. In our next stage of work, it can also be used in functional term classification for obtaining knowledge entities.

4.2 Results and analysis

Readability It can be reflected from the sentence length and the term density. The shorter a text is, the better readable it is. As shown in Table 3, the average sentences length of *R* is shorter than that of *A*. The term density is shown in Table 4. In each type，the average number of terminology per sentence in *R* is less than those in *A*. Especially in the sentences of type *strength*, reviewers seldom use terminology to express their comments. As to the general terms, things are similar.

Table 4. Term Density of Summary Peer Reviews and Abstracts

Function types *T*	General terms		Terminology	
	R	***A***	***R***	***A***
Background	10.4	12.1	4.4	4.9
Theme	9.0	10.3	4.6	5.0
Process	9.1	14.3	3.4	5.4
Result	13.8	19.3	5.5	7.5
Contribution	9.7	17.3	4.1	7.0
Strength	6.4	10.7	2.1	4.8

Focus Aspect The type proportion is shown in column *T*% in Table 3. The proportion of type *background*, *process* and *result* in *A* is 12.1%, 23.0% and 39.2%, much higher than those in *R*, i.e. 4.5%, 18.7% and 12.7%. While on the other side, the proportion of type *theme* and *strength* is 23.0% and 28.6%, higher than 11.5% and 1.0% in abstracts. The result shows that reviewers usually emphasis the *theme* of a paper since it is the source of innovation. At the same time, authors try to demonstrate the research process clearly and show the results in detail. The type *strength* provides a way to highlight a paper for the reviewers. It will help readers selecting papers of new exploration and with high quality.

5 Conclusion

In this paper, the function and application of summary peer reviews have been discussed. The components of common function types in summary peer reviews and abstracts have been defined. And basic measurement for comparing the two texts have been proposed. An observation on summary peer reviews is that they provide rather different information with that in abstracts in an easy-to-read expression. It is believed that the summary peer review can be used to speed up the readers understanding on scientific papers.

ACKNOWLEDGMENTS

The authors sincerely thank for Dr. Chuanqing Wang for his kindly suggestions on the review data sources. We also thank Jiahang Sui and Yuetong Zhang for their careful annotation.

REFERENCES

[1] Graetz, N. 1985. Teaching EFL students to extract structural information from abstracts. In J. M. Ulijn & A. K. Pugh (Eds.), Reading for professional purposes. Methods and materials in teaching language (pp. 123–135). Amersfoot: Leuven.

[2] Swales, J. 1981. Aspects of article introductions. Birmingham: The University of Aston.

[3] Swales, J. 1990. Genre analysis. English in academic and research settings. Cambridge: Cambridge University Press.

[4] Trawinski, Bogdan. A methodology for writing problem-structured abstracts[J]. Information Processing and Management, 25(6):693–702. 1989.

[5] Lu W, Huang Y, Bu Y, et al. 2018. Functional structure identification of scientific documents in computer science[J]. Scientometrics, 115(1): 463-486.

[6] Liu X, Guo C, Zhang L, et al. 2014. Scholar metadata and knowledge generation with human and artificial intelligence[J]. Journal of the Association for Information Science and Technology,65(6): 1187-1201.

[7] Xiaoguang WANG and Ningyuan SONG. Review on the scientific paper component ontologies[J] . Digital Library Forum , 2017(8): 2-7. Conference Name:ACM Woodstock conference

[8] Xiaoguang WANG , Menglin LI and Ningyuan SONG. Design and Application of Scientific Paper Functional Units Ontology[J], Journal of Library Science in China, 2018,44(04):73-88.

[9] Zhang L，Kopak R，Freund L，et al． 2010. A taxonomy of functional units for information use of scholarly journal article[J]. In Proceedings of the American Society for Information Science and Technology, 47(1): 1-10.

[10] Zhang L． 2012. Grasping the structure of journal articles: utilizing the functions of information units［J］. Journal of the American Society for Information Science and Technology， 63(3) : 469-480.

[11] Zhang L． 2014. Linking information through function[J]． Journal of the Association for Information Science and Technology， 65(11) : 2293-2305.

[12] Kang Dongyeop, Ammar Waleed, Dalvi Bhavana, van Zuylen Madeleine, Kohlmeier Sebastian, Hovy Eduard, Schwartz Roy. (2018). A Dataset of Peer Reviews (Peer Read): Collection, Insights and NLP Applications. NAACL, 2018, arXiv: https://arxiv.org/abs/1804.09635

[13] Philippe Vincent-lamarre, Vincent Lariviere. (2019). Content and linguistic biases in the peer review process of artificial intelligence conferences. arXiv: https://arxiv.org/abs/1911.02648

[14] Ke Wang and Xiaojun Wan. (2018). Sentiment Analysis of Peer Review Texts for Scholarly Papers. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). ACM, New York, NY, USA, 175-184. DOI: https://doi.org/10.1145/3209978.3210056

[15] Gupta S, Manning C D. Analyzing the Dynamics of Research by Extracting Key Aspects of Scientific Papers[C]. international joint conference on natural language processing, 2011: 1-9.

[16] Augenstein I, Das M, Riedel S, et al. SemEval 2017 Task 10: ScienceIE - Extracting Keyphrases and Relations from Scientific Publications[C]. meeting of the association for computational linguistics, 2017: 546-555.

[17] Kata Gábor, Davide Buscaldi, et al. SemEval 2018 Task 7：Semantic Relation Extraction and Classification in Scientific Papers[C]. meeting of the association for computational linguistics, 2018: 679–688. DOI: 10.18653/v1/S18-1111。

[18] Hélène de Ribaupierre, Gilles Falquet. User-centric design and evaluation of a semantic annotation model for scientific documents. In Proceedings of the 14th International Conference on Knowledge Technologies and Data-driven Business（i-KNOW '14）, September 16 -19, 2014, Graz, Austria. Publisher ACM New York, NY, USA.

[4] China Academic Journal Network Publishing Database
[5] http://shuyu.cnki.net/

[19] Jung Y. A semantic annotation framework for scientific publications[J]. Quality & Quantity, 2017, 51(3): 1009-1025.