

# Comparison Study on Summary Peer Reviews and Abstracts of Scientific Papers <sup>□</sup>

Chong Chen<sup>†</sup>  
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 delivering by the entities to a paper could be revealed by experts, the difficulties of paper reading will be significantly reduced. The things that reflect the value of a study, including knowledge entities, have been contained in different types of function components in the discourse of scientific papers. Summary peer reviews are taken as expert comments on papers. In this study, they are compared with the abstracts, namely the author comment on papers. The comparison includes two aspects, one is the difference in functional components they have and the other is the readability.

## 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 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 and find valuable ones to get reference or new insights for their study. During this process, a main cause to impede 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. The value delivering by knowledge entities is contained in the discourse of papers, more specifically, in functional components, such as research purpose, problems, methods, experiments, contributions and so on. Thus comments on these parts are critical to help readers understanding scientific papers, not just the citation number, the recommendation rating or the 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.

Although comments on papers are lacking compared with the massive customer reviews on products in online shopping, there are indeed four sources of comment on papers, i.e., the social media reviews, the abstract, the citation context and the summary peer reviews. The social media reviews can be disregarded since the amount is few and the quality is not assured. The abstract represents authors' viewpoint of certain paper; the citation context represents opinions of other researchers who have cited a paper with their own intention; and the summary 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 summary 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 scores. The reviewers would sum up a study from several aspects before they start to score it, such as the significance of the research purpose, the novelty of the proposed problems, the solidness of the methods or experiments, etc. In this paper, we consider the summary part and focus on two problems about it, (1) The difference between summary peer reviews with abstracts on these aspects; and (2) how do the summary peer reviews highlight the value of a paper comparing with its abstract.

The contribution of this paper lies in revealing the roles of summary peer reviews in highlighting the knowledge entities of scientific papers. The knowledge entities, such as the research questions, models, methods and the contributions, may be commented in a credible and easy-to-understand expression which can be used to speed up the readers understanding on the 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) (Graetz, 1985) in many research articles. Some different variations to the IMRD, such as Problem-

Method-Results-Conclusion, Goals-Method-Results, Introduction-Method-Results-Conclusions, have also been used (Swales, 1982; Swales, 1990; Graetz, 1985; Bhatia, 1993; Trawinski, 1989). As a whole, important information of a study is usually contained in sentences of such rhetorical function. Thus both reviewers and readers prefer to pick out pros and cons from these sentences for evaluating a study. Following this notion, researchers try to extract concerned information by dividing the scientific papers to functional discourse (Lu 2018) or rhetoric parts (Liu X, 2014).

With the trend of structuralizing knowledge in scientific papers, researchers also put forward ontologies for science document according to discourse element, argument model and rhetorical structure. Wang et.al analyzed the existing ontologies (2017) and proposed Functional Units Ontology (FUO) (wang et.al 2018) according to functional unit theory (Zhang 2010). Functional units were considered the smallest content units for academic communication and applications. They distribute among the IMRD of papers and are considered to ultimately meet the readers' goal of information use. The FUO includes 12 classes and 28 subclasses. 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 (2018) 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.

One inspiration we can get from the aforementioned work is that fine-grained components with different semantic function help to reveal the critical information of a paper. In fact, the functional components can be defined not only in the content of scientific papers but also in corpus of other sources, such as summary peer reviews. We can find some functional components in common from summary peer reviews and abstracts of scientific papers. But they are given by people with different motivation, and thus their focuses and views are rather different. The comment by reviewers would be more critical and with broad vision, which will probably benefit to readers and help them reduce the time on paper selection.

### About the peer review

Peer review is in itself meant to ensure the quality of publications and is important to scientific research (Ziman, 2002). Peer experts strictly evaluate papers from different perspectives related with studies and writings. The reviews are written to the editors and the authors including affirmation of positive and negative on some important perspectives. The reviewers judge the innovation and contribution according to the research purposes, the research problems, the proposed methods. They may also comment on the experiments and the results. In a word, there are rich and dense information in the content of peer reviews.

Much research attention has been paid to peer review in recent years. The peer reviews 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 (Kang Dongyeop et al., 2018). 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 (Philippe Vincent-lamarre, Vincent Lariviere, 2019). 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 (Ke Wang, 2018).

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 (心理学报)	webpage	2014
	Archives of Public Health	webpage	2013
Conferences	International Conference on Learning Representations(ICLR) (David Seorgel et al., 2013)	API	2013
	The Conference on Neural Information Processing Systems(NIPS) (Dongyeop Kang et al., 2018)	webpage	2013
	Open Review <sup>1</sup> (David Seorgel et al., 2013)	API	2013
Academic Websites	Sciencepaper Online <sup>2</sup> (中国科技论文在线)	webpage	2003
	Publons <sup>3</sup>	web	2012

## 3 Research Framework

In this study, summary peer reviews were supposed to have information of similar types but with different observation if comparing with those in abstracts. Two tasks were designed to conduct the comparison. The first one was to induce the function types appeared in reviews and abstracts; and the second was to analyze how the summary peer reviews highlight a paper's value from the function they focused.

The first task led to a rough classification. That is, sentences in reviews and abstracts were identified by frequent patterns, then they were manually checked to make sure that their types were correct. The patterns were previously annotated in each functional component, and some examples are shown in Table 2.

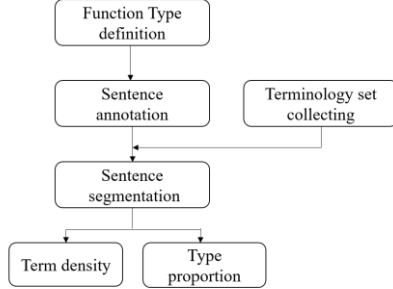
In the second task, the functional sentences in reviews and abstracts were compared by terms density and the type proportion. Term density reflects the readability of a text. General readers

<sup>1</sup> <https://openreview.net/>

<sup>2</sup> <http://www.paper.edu.cn/>

<sup>3</sup> <https://publons.com/>

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-density. Type proportion shows the focused function of reviews and abstracts by the proportion of sentences of each functional type.



**Fig.1 Research Framework.**

### 3.1 Functional Component Types

Based on the FUO [Wang, 2018], 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	(experiment/result/ simulation/)

	and methods;	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 emphasize 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 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. 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 summery 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 sentence, and the type proportion as 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 was collected from three sources, i.e. the keywords of the 774 papers, the *Academic Hotspots of Psychology* in CNKI<sup>4</sup>, and the *Chinese Terms in Psychology* published by China National Committee for Terms in Sciences and Technologies<sup>5</sup>. The terms from above sources were merged to a Chinese psychological academic terminology set with 8,354 terms in total. The terminology set was 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

**Table 4. Term Density of Summary Peer Reviews and Abstracts**

Function types <i>T</i>	General term density		Terminology density	
	<i>R</i>	<i>A</i>	<i>R</i>	<i>A</i>
Background	0.275	0.120	0.274	0.114
Theme	0.301	0.155	0.286	0.142
Process	0.318	0.122	0.297	0.115
Result	0.289	0.120	0.286	0.114
Contribution	0.282	0.118	0.286	0.120
Strength	0.279	0.092	0.297	0.132

The type proportion has shown in column *T* in Table 3. The proportion of type *background*, *process* and *result* is 12.1%, 23.0% and 39.2%, which is much higher in abstracts than that in summary peer reviews, namely 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 ease readers who want to select paper of new exploration and with high quality.

## ACKNOWLEDGMENTS

The authors heartedly thank for Dr. Chuanqing Wang for introduce the review data source. And we also thank for the members in annotation team, Jiahang Sui, Yuetong Zhang.

## REFERENCES

- [1] Lu W, Huang Y, Bu Y, et al. Functional structure identification of scientific documents in computer science[J]. *Scientometrics*, 2018, 115(1): 463-486
- [2] Liu X, Guo C, Zhang L, et al. Scholar metadata and knowledge generation with human and artificial intelligence[J]. *Journal of the Association for Information Science and Technology*, 2014, 65(6): 1187-1201.
- [3] 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). Amersfoort: Leuven.
- [4] Swales, J. (1981). *Aspects of article introductions*. Birmingham: The University of Aston.
- [5] Swales, J. (1990). *Genre analysis. English in academic and research settings*. Cambridge: Cambridge University Press.
- [6] Trawinski, Bogdan. A methodology for writing problem-structured abstracts[J]. *Information Processing and Management*, 25(6):693-702. 1989.
- [7] 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.
- [8] Swales J.1990. *Genre analysis: English in academic and research settings[M]*. Cambridge: Cambridge University Press.
- [9] Zhang L, Kopak R, Freund L, et al. 2011. Making functional units functional: the role of rhetorical structure in use of scholarly journal articles[J]. *International Journal of Information Management*, 31(1) : 21-29.
- [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] de Waard A, Buitelaar P, Eigner T. Identifying the epistemic value of discourse segments in biology texts [C] // *Proceedings of the Eighth International Conference on Computational Semantics*. Association for Computational Linguistics, 2009: 351 - 354.
- [13] Hounbo H, Mercer R E. Method Mention Extraction from Scientific Research Papers[C]. *international conference on computational linguistics*, 2012: 1211-1222.
- [14] Soldatova L N, King R D. An ontology of scientific experiments.[J]. *Journal of the Royal Society Interface*, 2006, 3(11): 795-803.
- [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] Tsai C T, Kundu G, Roth D. Concept-based analysis of scientific literature[C]// *ACM International Conference on Conference on Information & Knowledge Management*. 2013:1733-1738.
- [17] 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.
- [18] Kata Gabor, 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.
- [19] Teufel, Simone, Moens, Marc. Summarizing Scientific Articles: Experiments with Relevance and Rhetorical Status[J]. *Computational Linguistics*, 2002, 28(4):409-445.
- [20] H     de Ribaupierre and Gilles Falquet. New trends for reading scientific documents. In *Proceedings of the 4th ACM workshop on Online books, complementary social media and crowdsourcing (BooksOnline '11)*. ACM, New York, NY, USA, 19-24, 2011, DOI: <https://doi.org/10.1145/2064058.2064064>
- [21] H     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.
- [22] Jung Y. A semantic annotation framework for scientific publications[J]. *Quality & Quantity*, 2017, 51(3): 1009-1025.
- [23] 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. Xiaoguang WANG and Ningyuan SONG. Review on the scientific paper component ontologies[J]. *Digital Library Forum*, 2017(8): 2-7.

<sup>4</sup> China Academic Journal Network Publishing Database

<sup>5</sup> <http://shuyi.cnki.net/>