The Research of Paper Influence Based on Citation Context -- A Case Study of the Nobel Prize Winner's Paper

Shengbo Liu[1], Kun Ding[1], Bo Wang[2], Delong Tang[1], Zhao Qu[1]

[1] *liushengbo1121@gmail.com, dingk@dlut.edu.cn, tangdl@mail.dlut.edu.cn, qz_031@mail.dlut.edu.cn*
WISElab, Dalian University of Technology, No. 2, Linggong Road, Ganjingzi district, Dalian, 116024, China

[2] *bowang1121@gmail.com*
School of Management, Dalian Ploytechnic University, #1 Qinggongyuan, Dalian, 116000,China

Abstract

Citation context was used to measure the influence of highly cited papers. The themes of citation context were analyzed with bibliometrics methods. The citation context was classified into three categories as positive, negative and neutral. And the neutral citations were also classified into three sub categories, related work in background or introduction, theoretical foundation, and experimental foundation. The citation contexts of a highly cited paper of O'Keefe were extracted as the experiment data set. The results showed that the co-occurrence method was very useful for describing the themes of the citation contexts. The citation contexts of the selected paper were divided into five themes. The classification of citation contexts could provide more information about how and why a paper was highly cited. There was no negative citation in this experiment, and more than 10% citation contexts were positive citation. About 50% of the neutral citations were belonging to related work in background or introduction. The detailed influence of the target paper was also illustrated in our research.

Conference Topic

Citation and co-citation analysis

Introduction

Citation frequency is a commonly used indicator to measure the importance of a paper. Recently, *Nature* asked Thomson Reuters, which now owns the SCI, to list the 100 most highly cited papers published from 1900 to 2014. The results revealed some surprises, many of the world's most famous papers do not rank in the top 100 (Van Noorden, Maher, & Nuzzo, 2014). John P. A. Ioannidis and colleagues surveyed the most-cited authors of biomedical research for their views on their own influential published work. The results showed that the most important paper was indeed one of author's most-cited ones. But they described most of their chart-topping work as evolutionary, not revolutionary (Ioannidis, Boyack, Small, Sorensen, & Klavans, 2014). Although the citation frequency is an important indicator to measure the influence of a paper, it is hard to reveal why others always cited this paper and what influence it makes. Citation context refers to the text surrounding the references (Henry Small, 1982). It could provide more detailed information about citation.

In this paper, we take John O'Keefe's (Nobel Prize winner in Physiology or Medicine 2014) most influence paper as instance. The influence of this paper will be analyzed based on citation context. Our analysis will provide a richer understanding of which knowledge claims made by O'Keefe have had the greatest impact on later work.

Related work

Citation context analysis

Citation context can be defined as the sentences that contain the citation of a particular reference. For example, the sentence "This comparison is made using BLASTX [18]" is the citation context of reference [18].

Citation content can be used to identify the nature of a citation. The attributions and functions of a cited paper can be identified from the semantics of the contextual sentences (A. Siddharthan, Teufel, S., 2007). Nanba and Okumura (Nanba, 1999, 2005) collected citation context information from multiple papers cited by the same paper and generated a summary of the paper based on this citation context information. They also extracted citing sentences from citation contexts and generated a review. Elkiss et al. (Elkiss, 2008) generated the citation summarization based on citation context to describe the topic of cited paper. Mei (Mei, 2008) and Mohammad (Mohammad, 2009) found that the summarization of citation contexts is very different from the abstract of the cited reference. Liu and Chen(Liu & Chen, 2013) studied the differences between latent topics in abstracts and citation contexts. The results showed that topics from citing sentences tend to include more specific terms than topics from abstracts of citing papers. Nakov (Nakov, 2004) referred to citation contexts as citances – a set of sentences that surrounding a particular citation. Citances can be used in abstract summarization and other Natural Language Processing (NLP) tasks such as corpora comparison, entity recognition, and relation extraction. Small (H. Small, 1979) studied the context of co-citation and analyzed the context in which the co-citation paper mentioned. In addition, he analyzed the sentiment of the co-citation context (H. Small, 2011).

Anderson (Anderson, 2010) analyzed the citation contexts of a classic paper in organizational learning which was published by Walsh and Ungson in the Academy of Management Review. The results provided a richer understanding of which knowledge claims made by Walsh and Ungson have been retrieved and have had the greatest impact on later work in the area of organizational memory, and also what criticisms have been leveled against their claims. Chang(Chang, 2013) compared the citing topics of *Little Science*, *Big Science* in natural sciences and humanities and social sciences through citation context. He found that the citing topics in natural sciences and humanities and social sciences were very similar, but the cited motivation had some differences.

The classification and function of citation context

Citation context contains the direct related information between cited paper and citing paper. It could be used to reveal the nature of a citation. The cited motivation of each citation is different, so the value of each citation will be different. For example, some of the citation contexts support the claims in the cited paper, and some of them may take the opposite opinion about the views or methods in the cited paper. Spiegel-Rösing (Spiegel-Rösing, 1977) studied the citation context of Science Study in 1977 and classify the citation context into 13 categories, including use the data of cited paper, use the method of cited paper, compare the work of cited paper and citing paper and so on. In order to provide more information for literature management, Teufel reclassified the above 13 categories into four categories, (1) Explicit statement of weakness, (2) Contrast or comparison with other work, (3) Agreement /usage /compatibility with other work, (4) A

neutral category(Teufel, Siddharthan, & Tidhar, 2006) . Cue phrases were used to identify the category of each citation context. The similar method was also employed in Liu's (Liu et al.) work in which the citation context was classified as positive citation, negative citation, and neutral citation. Other people like Small (Henry Small, 1982), McCain (McCain & Turner, 1989), Siddharthan (A. Siddharthan & Teufel, 2007), Swales (Swales, 1990) also did some work about citation context classification.

Data and Method

Our procedure consists of three major components, 1. Data collection and preprocessing, 2. Theme analysis of citation context, and 3. The classification of citation context. Details are explained in corresponding sections.

Data collection and preprocessing

The 2014 Nobel Prize in Physiology or Medicine is awarded to Dr. John M. O'Keefe, Dr. May-Britt Moser and Dr. Edvard I. Moser for their discoveries of nerve cells in the brain that enable a sense of place and navigation. The scientific background was introduced in the document "The Brain's Navigational Place and Grid Cell System" . The keywords this document were selected manually and used to retrieve the award field in Web of Science. The search query was shown as follows :

TI=(hippocamp* AND (place OR Position* OR spatial)) OR (("grid cell*" OR Position* OR Navigation* OR spatial OR place) And ("entorhinal cortex" OR brain OR cerebral))

The time period was from 1945 to 2014, and 4441 papers were collected.

The citation context collection was built through three steps. First, the paper with the first author O'Keefe and the highest citation frequency was selected. Second, the papers which cited the chosen paper were downloaded with full text. Actually, we could just find less than 20% full text papers. Last, the citation contexts of the chosen papers were extracted from the full text for further analysis. The extraction method has been introduced in our previous work (Liu & Chen, 2013).

The theme analysis of citation context

The theme analysis includes two tasks. One is counting the frequency of noun phrases appeared in citation contexts. Another is mapping the co- occurrence network of noun phrases.

Part-of-speech is needed before extract noun phrases. There are many tools to label part-of-speech, such as PosTagger, CLAWS POS tagger. Stanford Log-linear Part-Of-Speech Tagger (Toutanova & Manning, 2000) was employed in this work, which was developed by NLP group of Stanford University. The noun phrase formation rules was designed with the same method described in Wang's paper (Wang, Liu, Ding, Liu, & Xu, 2014). When counting the frequency of noun phrases. If one citation context contains two same noun phrases, it will count once.

In bibliometrics analysis, co-occurrence method was often used to detect subjects/themes (Hofer, Smejkal, Bilgin, & Wuehrer, 2010; Lee, 2008; Zhang et al., 2012). But few of the related works use this method to detect the theme of citation context. Pajek software was employed to mapping the noun phrases co-occurrence network of citation context. We expect to identify the citing themes through drawing the co-occurrence map.

The classification of citation context

Following the work of Spiegel-Rösing (Spiegel-Rösing, 1977) and Teufel (Teufel et al., 2006), citation contexts will be classified into three categories as positive, negative and neutral. Table 1 shows the description of each category. We divided the positive category into three sub categories and the negative category into two sub categories.

Table 1. The description of each category

Category		Description
Positive	(1)	Affirm or praise the cited work
	(2)	Apply the methods, tools or databases of the cited paper
	(3)	Comparison of methods and results
Negative	(1)	Point out the weakness of the citation
	(2)	Contain negative cue words
Neutral	(1)	Contain no cue words

To our knowledge, the proportion of neutral citations occupy more than others. So we will classify the neutral citation into three sub categories based on the citation motivation.
  (1) Related work in background or introduction. Introduce the related work with no comments.
  (2) Theoretical foundation. Concepts, principles, methods, or results which will be used in citing paper.
  (3) Experimental foundation. Including experimental conditions, processes, environment, and results.

Results and discussion

Target paper detecting

Table 2 shows top ten highly cited papers in Nobel Prize award field. The highest cited paper was "PLACE NAVIGATION IMPAIRED IN RATS WITH HIPPOCAMPAL - LESIONS" which published in *Nature* in 1982. It has been cited 3589 times. Although this paper got highly cited in Nobel Prize award field, it did not appear in "Scientific background" document, which was the instruction of why the winner got this prize. The author Morris R.G.M did not get Nobel Prize. The Nobel Prize was given to the author of the second highest cited paper "HIPPOCAMPUS AS A SPATIAL MAP - PRELIMINARY EVIDENCE FROM UNIT ACTIVITY IN FREELY-MOVING RAT". The result is similar to the work of Van Noorden (Van Noorden et al., 2014) that the Nobel Prize winner's paper did not get the highest citation frequency.

O'Keefe who is the Nobel Prize winner had three papers ranked in top ten high cited papers in Nobel Prize award field. The highest cited paper had been cited 1812 times. This paper was selected as the target paper. The seminal work of this paper was the discovery of "place cell".

It is hard to download all the 1812 citing papers. So 200 citing papers with full text were selected in our experiment. There were 228 citing sentences. The target paper was average cited 1.14 times in each citing paper.

Table 2. Top ten high-cited papers in Nobel Prize award field.

Author	Title	Journal	Year	Cited frequency
Morris, R. G. M., P. Garrud, et al	PLACE NAVIGATION IMPAIRED IN RATS WITH HIPPOCAMPAL-LESIONS	Nature	1982	3589
Okeefe, J. and Dostrovs.J	HIPPOCAMPUS AS A SPATIAL MAP - PRELIMINARY EVIDENCE FROM UNIT ACTIVITY IN FREELY-MOVING RAT	Brain Research	1971	1812
Okeefe, J. and M. L. Recce	PHASE RELATIONSHIP BETWEEN HIPPOCAMPAL PLACE UNITS AND THE EEG THETA-RHYTHM	Hippocampus	1993	1033
Tsien, J. Z., P. T. Huerta, et al	The essential role of hippocampal CA1 NMDA receptor-dependent synaptic plasticity in spatial memory	Cell	1996	919
Grant, S. G. N., T. J. Odell, et al	IMPAIRED LONG-TERM POTENTIATION, SPATIAL-LEARNING, AND HIPPOCAMPAL DEVELOPMENT IN FYN MUTANT MICE	Science	1992	827
Hafting, T., M. Fyhn, et al	Microstructure of a spatial map in the entorhinal cortex	Nature	2005	773
Cohen, L., S. Dehaene, et al	The visual word form area - Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients	Brain	2000	755
Burgess, N., E. A. Maguire, et al	The human hippocampus and spatial and episodic memory	Neuron	2002	669
Packard, M. G. and J. L. McGaugh	Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning	Neurobiology of Learning and Memory	1996	666
Okeefe, J	PLACE UNITS IN HIPPOCAMPUS OF FREELY MOVING RAT	Experimental Neurology	1976	657

The themes of citation context

299 noun phrases were extracted from the citation contexts. Table 3 listed twenty high frequency noun phrases. The term "place cell" got the highest frequency of 76, because the most contributing work of the target paper was the discovery of place cell. Hippocampus, environment, rat, fire, neuron were all the important terms in target paper. Some of the terms were not mentioned in the target paper, such as cognitive map and ca3.

Table 3. Top twenty high cited papers in Nobel Prize award field.

No.	Noun phrase	Frequency	No.	Noun phrase	Frequency
1	place cell	76	11	discovery	17
2	hippocampus	74	12	place field	15
3	environment	55	13	rodent	13
4	rat	44	14	ca3	13
5	animal	40	15	space	12
6	cell	31	16	ca1	12
7	location	29	17	position	11
8	fire	25	18	pyramidal cell	9
9	cognitive map	19	19	region	9
10	neuron	18	20	navigation	9

Figure 1 showed the co-occurrence map of the noun phrases. Each node represents a noun phrase. The size of the node was proportional to the number of terms co-occurred with it. We set the co-occurrence threshold as more than once and got 71 nodes in the map.

The map could divide into five parts manually based on the relationship of terms. Part A was mainly involving navigation, which was not mention too much in cited paper. It was the following research of place cell. Part B was related to neuron region, including CA1 and CA3. CA1 was discussed in the cited paper, but CA3 was found in the later work. Part C was related to experimental process about firing pattern of rat. Part D was the experimental environment. The definition of place field was widely cited. Part E was about the concept of place cell.
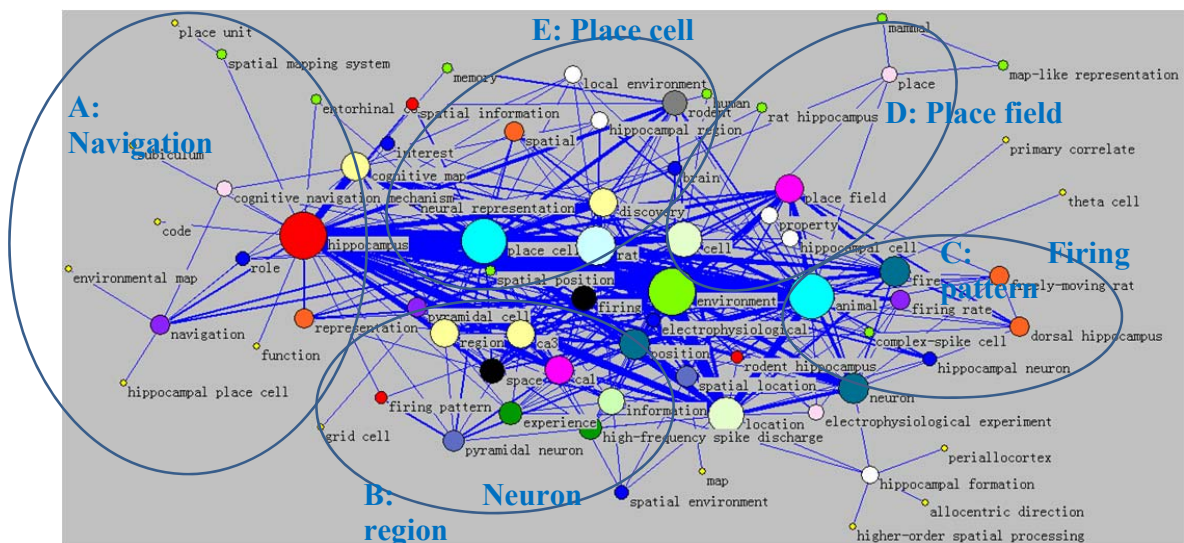

Figure 1. Co-occurrence map of the noun phrases.

Table 4. Example of positive citations

No.	Positive citation
1	The discovery of place cells [1]-[5] in the hippocampal regions of rats **consolidated** the idea that hippocampus probably represents a cognitive map of the local environment of an animal……
2	The concept of cognitive map for navigation, carried out mainly by Tolman [10], was **fuelled** by the discovery of the so-called place cells in the hippocampus of the rat and has widely increased our understanding of cognitive navigation mechanisms [11]
3	The **breakthrough** came in 1971 with the discovery of the rat s cognitive map in the cells of the hippocampus [16]……
4	The idea of the formation of a cognitive map was first proposed by Tolman [45] in the late 40s and was later **supported by the discovery** of place cells by o keefe and dostrovsky [35]
5	The **striking discovery** of place cells in the rat hippocampus [51] has triggered a wave of interest on spatial learning that holds until today

Table 5. Sub categories distribution of neutral citations

Category	Related work	Theoretical foundation	Experimental foundation
Counts	114	49	41

The classification results

The classification results showed that most of the citations were neutral citation. There was no negative citation in our datasets. 24 of 228 citation contexts were positive citations and 204 citations were neutral citations. Table 4 listed some examples of positive citations.

The sub categories distribution of neutral citations was shown in table 5. Nearly half of the citations were cited as related work. Theoretical foundation had 49 citations, and most of them were related to place cell or place field. 41 of 204 neutral citations were classified into experimental foundation, including ca1 neuron fire experiment, rodent studies and so on.

Conclusion and discussion

Citation context was used to measure the influence of paper in this research. The influence was identified from two aspects, the theme of the citation context and the classification of the citation context. The results showed that the traditional bibliometrics methods could be utilized in identify the themes of citation context. The citation contexts were divided into five themes in our experiment. The classification results showed that there were no negative citations of O'Keefe's most influential paper. More than 10% citation contexts were positive citations.

Through the citation context analysis of the influence paper, the detailed influence of the high influence paper could be revealed. The influence themes are more wide than the abstract of the target paper and the proportion of the positive citations takes more account than it appears in some journals (Liu et al., 2014).

There is only one case study in this paper. Although we could get some insightful results from this case study, comparative experiments are still needed in our future work.

Acknowledgments

References

Anderson, M. H. & Sun, P.Y.T. (2010). What have scholars retrieved from Walsh and Ungson (1991)? A citation context study. *Management Learning, 41*(2), 131-145.

Chang, Y.-W. (2013). A comparison of citation contexts between natural sciences and social sciences and humanities. *Scientometrics, 96*(2), 535-553.

Elkiss, A., Shen, S., Fader, A., Erkan, G., States, D. & Radev, D. (2008). Blind men and elephants: What do citation summaries tell us about a research article? *Journal of the American society for information science and technology, 59*(1), 51-62.

Hofer, K. M., Smejkal, A. E., Bilgin, F. Z., & Wuehrer, G. A. (2010). Conference proceedings as a matter of bibliometric studies: the Academy of International Business 2006–2008. *Scientometrics*, *84*(3), 845-862.

Ioannidis, J. P. A., Boyack, K. W., Small, H., Sorensen, A. A., & Klavans, R. (2014). Bibliometrics: Is your most cited work your best? *Nature*, *514*(7524), 561-562.

Lee, W. H. (2008). How to identify emerging research fields using scientometrics: An example in the field of Information Security. *Scientometrics 76*(3), 503.

Liu, S., & Chen, C. (2013). The differences between latent topics in abstracts and citation contexts of citing papers. *Journal of the American Society for Information Science and Technology*, *64*(3), 627-639.

Liu, S., Chen, C., Ding, K., Wang, B., Xu, K., & Lin, Y. (2014). Literature retrieval based on citation context. *Scientometrics*, *101*(2), 1293-1307.

McCain, K. W., & Turner, K. (1989). Citation context analysis and aging patterns of journal articles in molecular genetics. *Scientometrics*, *17*(1), 127-163.

Mei, Q. & Zhai, C. (2008). Generating impact-based summaries for scientific literature. *Proceedings of ACL '08*, Columbus.

Mohammad, S., Dorr, B., Egan, M., Hassan, A., Muthukrishan, P., Qazvinian, V., Radev, D. & Zajic, D. (2009). Using citations to generate surveys of scientific paradigms. *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, Boulder.

Nakov, P. I., Schwartz, A.S. & Hearst, M.A. (2004). Citances: Citation sentences for semantic analysis of bioscience text. *SIGIR 2004 Workshop on Search and Discovery in Bioinformatics*, Sheffield.

Nanba, H. & Okumura, M. (1999). Towards multi-paper summarization using reference information. *16th International Joint Conference on Artificial Intelligence*, Stockholm.

Nanba, H., & Okumura, M. (2005). Automatic detection of survey articles. *The Research and Advanced Technology for Digital Libraries*, Berlin.

Siddharthan, A. & Teufel, S. (2007). Whose idea was this, and why does it matter? Attributing scientific work to citations. *Proceedings of NAACL/HLT-07*, Rochester.

Small, H. (1979). Co-citation context analysis: The relationship between bibliometric structure and knowledge. *Proceedings of the ASIS Annual Meeting*, Medford.

Small, H. (1982). Citation context analysis. *Progress in communication sciences*, *3*, 287-310.

Small, H. (2011). Interpreting maps of science using citation context sentiments: a preliminary investgation. *Scientometrics*, *87*(2), 373-388.

Spiegel-Rösing, I. (1977). Science studies: Bibliometric and content analysis. *Social Studies of Science*, 97-113.

Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press (Cambridge England and New York).

Teufel, S., Siddharthan, A., & Tidhar, D. (2006). Automatic classification of citation function. *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*.

Toutanova, K., & Manning, C. D. (2000). Enriching the knowledge sources used in a maximum entropy part-of-speech tagger. *Proceedings of the 2000 Joint SIGDAT conference on Empirical methods in natural language processing and very large corpora: held in conjunction with the 38th Annual Meeting of the Association for Computational Linguistics-Volume 13*.

Van Noorden, R., Maher, B., & Nuzzo, R. (2014). The top 100 papers. *Nature*, *514*(7524), 550-553.

Wang, B., Liu, S., Ding, K., Liu, Z., & Xu, J. (2014). Identifying technological topics and institution-topic distribution probability for patent competitive intelligence analysis: a case study in LTE technology. *Scientometrics*, *101*(1), 685-704.

Zhang, J., Xie, J., Hou, W., Tu, X., Xu, J., Song, F., Wang, Z. & Lu, Z. (2012). Mapping the knowledge structure of research on patient adherence: Knowledge domain visualization based co-word analysis and social network analysis. *PloS One*, *7*(4), e34497.